USDOT Tier 1
University Transportation Center
on Improving Rail Transportation
Infrastructure Sustainability and Durability

Final Report UD-1

A STOCHASTIC RAIL WEAR FORECAST USING MIXTURE DENSITY NETWORKS AND THE LAPLACE DISTRIBUTION

By

Mr. Joseph Palese, PhD, MBA, PE
Senior Scientist
Department of Civil & Environmental Engineering, University of Delaware

Allan Zarembski PhD, PE, FASME, Hon. Mbr. AREMA
Professor and Director of Railroad Engineering and Safety Program
Department of Civil & Environmental Engineering, University of Delaware

and

Nii Attoh-Okine PhD, PE, FASCE
Professor, Department of Civil & Environmental Engineering, University of Delaware

October 31, 2019

Grant Number: 69A3551747132

DISCLAIMER

ABSTRACT

The US rail network consists of more than 232,000 miles of track (as of 2016). More than 160,000 miles are classified as Class I railway (belonging to a railway with operating revenue greater than $458M). (Association of American Railroads, 2017). In 2016, Class I railroads replaced or installed nearly 7,000 miles of rail. (Association of American Railroads, 2016), which represents 1.5% of the total amount of rail in the US. This corresponds to an average (steady state) rail life of approximately 33 years.

Rail wear research has been conducted since the 1860's. Much research has been performed in the area of mechanics of rail wear and the parameters that most contribute to the rate of wear. Early research by Archard (Archard, 1953) focused on applied load and material properties (rail metallurgy/hardness) and developed wear coefficients for understanding the relative wear performance of various materials under different loading conditions.

With the advent of automated inspection cars, railways had a method to measure and monitor rail wear over time (and accumulated traffic) more accurately and efficiently then measuring by hand. This however resulted in large amounts of data that were only used for go/no-go (threshold) analyses. Some work was performed to automate the process and determine wear rates from the measured wear data, but was limited in scope.

This report formulates, solves and implements a stochastic rail wear forecasting model using mixture density networks (MDNs). MDNs take advantage of artificial intelligence, using a neural network, to define a median and shape parameter as output for a defined distribution, given causal factors (independent variables). In the case of rail wear rates, it was shown in a previous report that the rail wear rates follow a two side exponential distribution (Laplace distribution).

Specifically, a stochastic modeling approach is presented and applied to more than 270 miles of railway track to develop a stochastic forecast of future rail requirements based on rail wear.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

The US rail network consists of more than 232,000 miles of track (as of 2016). More than 160,000 miles are classified as Class I railway (belonging to a railway with operating revenue greater than $458M). (Association of American Railroads, 2017). In 2016, Class I railroads replaced or installed nearly 7,000 miles of rail. (Association of American Railroads, 2016), which represents 1.5% of the total amount of rail in the US. This corresponds to an average (steady state) rail life of approximately 33 years.

It was recognized early on that enhancements in the steel making process and reduction in forces on the rail allowed for prolonged life of the rail, due to a reduction in the rate of wear. (Adams, 1865) Since rail is one of the railways most valuable assets, the study of its life cycle costs has been deemed important since the introduction of conventional tee-rail. (Sandberg, 1869)

It is extremely important to understand how rail degrades, when it fails, and what maintenance interventions can be deployed to improve the life of the asset. Rail typically fails due to internal fatigue, surface fatigue, and wear. (Hay, 1982) Maintenance activities that can be utilized to extend the life of the rail include lubrication in curves, top of rail friction modification, and profile grinding. (Zarembski, 2005)

Rail wear research has been conducted since the 1860's. Much research has been performed in the area of mechanics of rail wear and the parameters that most contribute to the rate of wear. Early research by Archard (Archard, 1953) focused on applied load and material properties (rail metallurgy/hardness) and developed wear coefficients for understanding the relative wear performance of various materials under different loading conditions. This research concentrated on laboratory testing with special test rigs to control various parameters and measure the rate of wear.

As a consequence of increased axle loads and resulting severe wear rates, railways used laboratory testing to evaluate enhanced rail steels and lubricants to study potential extensions in rail life in track.

With this basic understanding, railway researchers developed empirical and statistical rail wear rate prediction models to be used to define appropriate track components and maintenance practices that maximize rail life as well as understand future rail requirements for planning purposes.

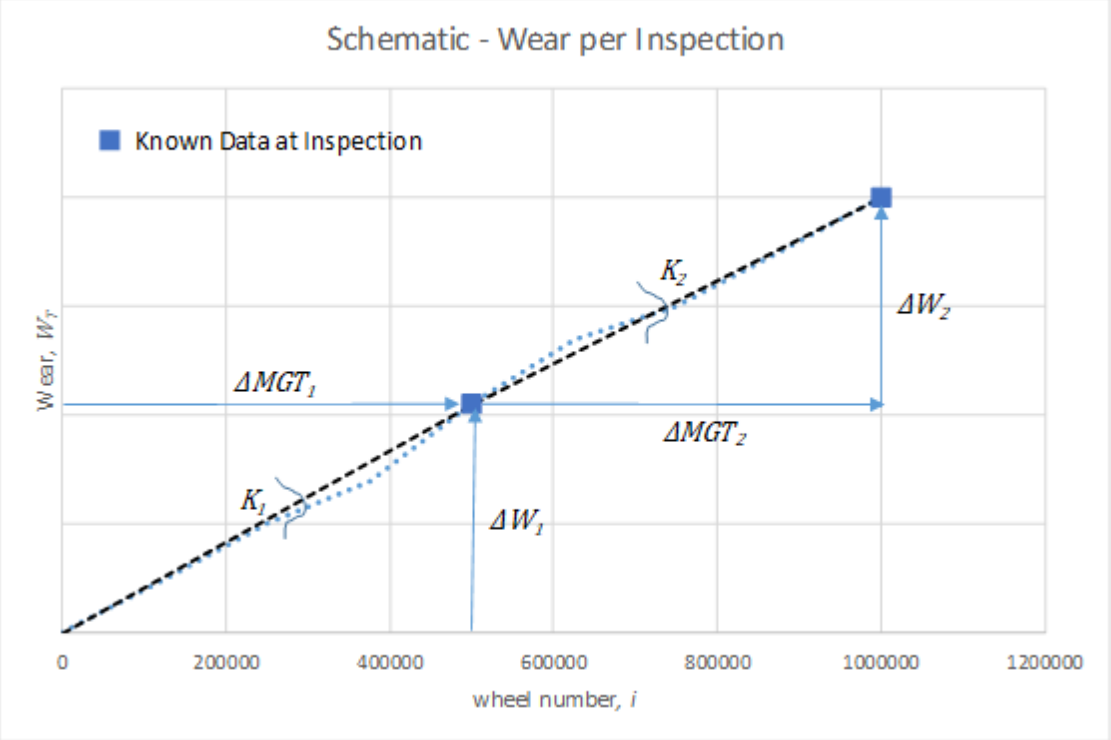
With the advent of automated inspection cars, railways had a method to measure and monitor rail wear over time (and accumulated traffic) more accurately and efficiently then measuring by hand. This however resulted in large amounts of data that were only used for go/no-go (threshold) analyses. Some work was performed to automate the process and determine wear rates from the measured wear data, but was limited in scope.

Additional increases in computing power and emerging advanced data science techniques allow for the implementation of higher order stochastic processes to be developed and applied to the

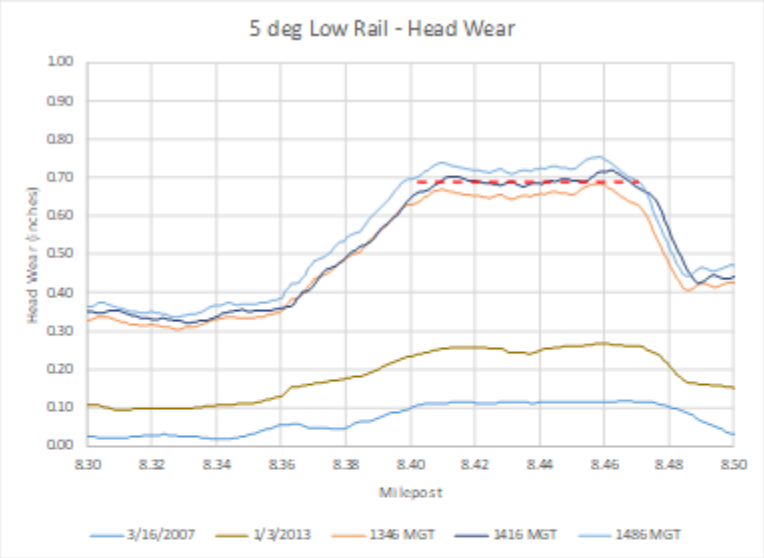
large amounts of rail wear data. This allows for a better understanding of the rate of wear in-situ and prediction of failure or maintenance interventions, which prolong the life of the rail.

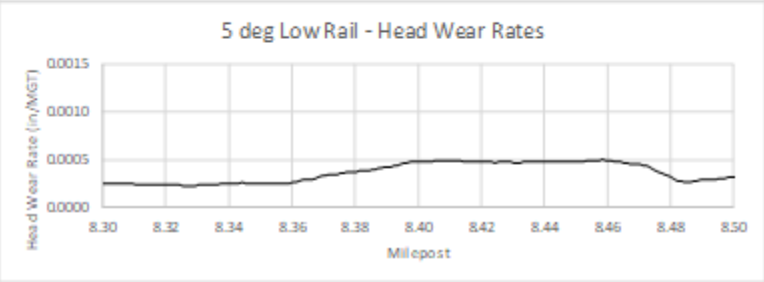
This report formulates, solves and implements a stochastic rail wear forecasting model using mixture density networks (MDNs). MDNs take advantage of artificial intelligence, using a neural network, to define a median and shape parameter as output for a defined distribution, given causal factors (independent variables). In the case of rail wear rates, it was shown in a previous report that the rail wear rates follow a two side exponential distribution (Laplace distribution). (Joseph W Palese, Zarembski, & Attoh-Okine, 2019)

Rail wear, and associated inspection points can be depicted according to the figure below. This figure shows that for a segment of rail the wear will accumulate linearly, but will have a distribution about the mean (or median in the case of a Laplace distribution).



When considering rail wear data streams as a time series, Auto Regressive Integrated Moving Average (ARIMA) time series techniques can be utilized to forecast the next wear point. This accounts for distribution in the data measurement points, and handles errors in the inspection data. An example of ARIMA forecasting for a segment is show in in the figure below. This figure shows that wear can be non-uniform throughout a given rail segment with homogeneous characteristics (known causal factors). Thus, unknown, or latent characteristics, influence the forces and resulting wear on the rail.

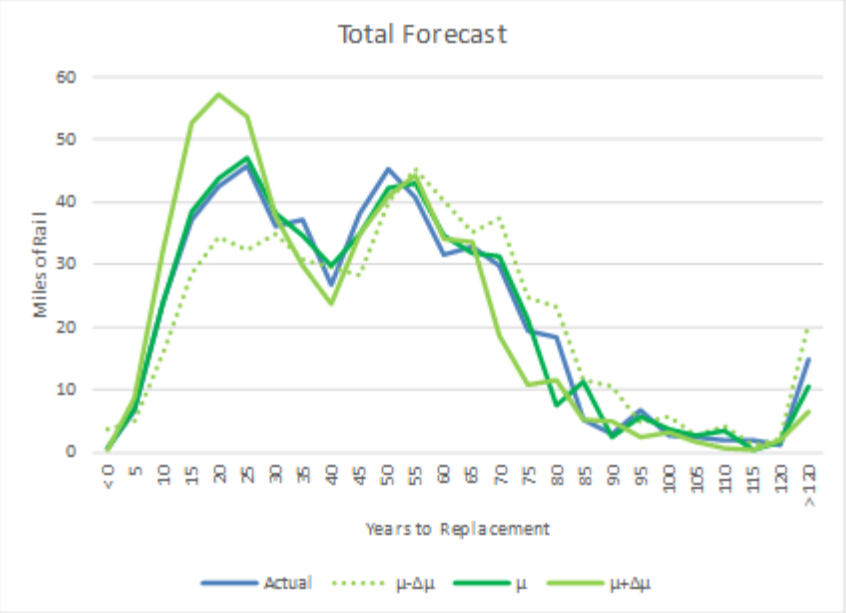




Mixture density networks leverage a neural network, and expected outcome and distribution, to determine rail wear rates. An example of this is approach is shown in the table below.

					Head Wear Rates (in/100MGT)						Gage Wear Rates (in/100MGT)				
Deg	SE (in)	Ann MGT	Rail		ARIMA	Actual	μ	b	C		ARIMA	Actual	μ	b	C
5	1.5	31.35	High		0.0515	0.0523	0.0511	0.0174	46.9%		0.0081	0.0096	0.0341	0.0058	99.3%
7.25	1.75	31.35	High		0.0932	0.1022	0.1013	0.0182	47.6%		0.0211	0.0502	0.0496	0.0060	45.2%
5	1.5	31.35	Low		0.0479	0.0476	0.0482	0.0119	52.6%						
7.25	1.75	31.35	Low		0.1273	0.1150	0.0766	0.0126	2.4%						

Combining these techniques and applying them to a route (270+ miles) of railway track results in a future forecast of rail wear requirements. Taking advantage of the stochastic output from the MDN (associated with a Laplace distribution) and considering a middle 50% probability band on the wear rates results in a stochastic forecast as shown in the figure below.



The figure above shows that the forecast has a median probability of annual rail requirements by year as well as a range associated with a +/- 25% probability departure from the median.

INTRODUCTION

The US rail network consists of more than 232,000 miles of track (as of 2016). More than 160,000 miles are classified as Class I railway (belonging to a railway with operating revenue greater than $458M). (Association of American Railroads, 2017). In 2016, Class I railroads replaced or installed nearly 7,000 miles of rail. (Association of American Railroads, 2016), which represents 1.5% of the total amount of rail in the US. This corresponds to an average (steady state) rail life of approximately 33 years.

It was recognized early on that enhancements in the steel making process and reduction in forces on the rail allowed for prolonged life of the rail, due to a reduction in the rate of wear. (Adams, 1865) Since rail is one of the railways most valuable assets, the study of its life cycle costs has been deemed important since the introduction of conventional tee-rail. (Sandberg, 1869)

It is extremely important to understand how rail degrades, when it fails, and what maintenance interventions can be deployed to improve the life of the asset. Rail typically fails due to internal fatigue, surface fatigue, and wear. (Hay, 1982) Maintenance activities that can be utilized to extend the life of the rail include lubrication in curves, top of rail friction modification, and profile grinding. (Zarembski, 2005)

Rail wear research has been conducted since the 1860's. Much research has been performed in the area of mechanics of rail wear and the parameters that most contribute to the rate of wear. Early research by Archard (Archard, 1953) focused on applied load and material properties (rail metallurgy/hardness) and developed wear coefficients for understanding the relative wear performance of various materials under different loading conditions. This research concentrated on laboratory testing with special test rigs to control various parameters and measure the rate of wear.

Research quickly moved into tribology and the effects of lubrication on the rate of wear. As researchers began to understand the modes of wear, additional laboratory research was conducted to identify rail wear regimes and to "map" wear in order to identify conditions under which rail wear would transition from adhesive, to plastic flow and fatigue, to abrasive wear.

Research subsequently transitioned to contact mechanics and the effects of the wheel/rail interface on rail wear, which included contact area as well as slip/sliding and associated creep forces. An entire symposium dedicated to contact mechanics was initiated in the 1980's, and is still held to this day[1].

As a consequence of increased axle loads and resulting severe wear rates, railways used laboratory testing to evaluate enhanced rail steels and lubricants to study potential extensions in rail life in track.

[1] Most recently, 11th International Conference on Contact Mechanics and Wear of Rail/Wheel Systems, Delft technical University, the Netherlands, 2018.

With this basic understanding, railway researchers developed empirical and statistical rail wear rate prediction models to be used to define appropriate track components and maintenance practices that maximize rail life as well as understand future rail requirements for planning purposes.

With the advent of automated inspection cars, railways had a method to measure and monitor rail wear over time (and accumulated traffic) more accurately and efficiently then measuring by hand. This however resulted in large amounts of data that were only used for go/no-go (threshold) analyses. Some work was performed to automate the process and determine wear rates from the measured wear data, but was limited in scope.

Additional increases in computing power and emerging advanced data science techniques allow for the implementation of higher order stochastic processes to be developed and applied to the large amounts of rail wear data. This allows for a better understanding of the rate of wear in-situ and prediction of failure or maintenance interventions, which prolong the life of the rail.

This report formulates, solves and implements a stochastic rail wear forecasting model using mixture density networks (MDNs). MDNs take advantage of artificial intelligence, using a neural network, to define a median and shape parameter as output for a defined distribution, given causal factors (independent variables). In the case of rail wear rates, it was shown in a previous report that the rail wear rates follow a two side exponential distribution (Laplace distribution). (Joseph W Palese et al., 2019)

Specifically, a stochastic modeling approach is presented and applied to more than 270 miles of railway track to develop a stochastic forecast of future rail requirements based on rail wear.

Description of Rail Wear Parameters

Rail wear is the cross sectional loss of material from the new or as-rolled section, as shown in Figure 1, where the dashed line is the original cross section, and the solid line is the worn or as measured cross section.

.

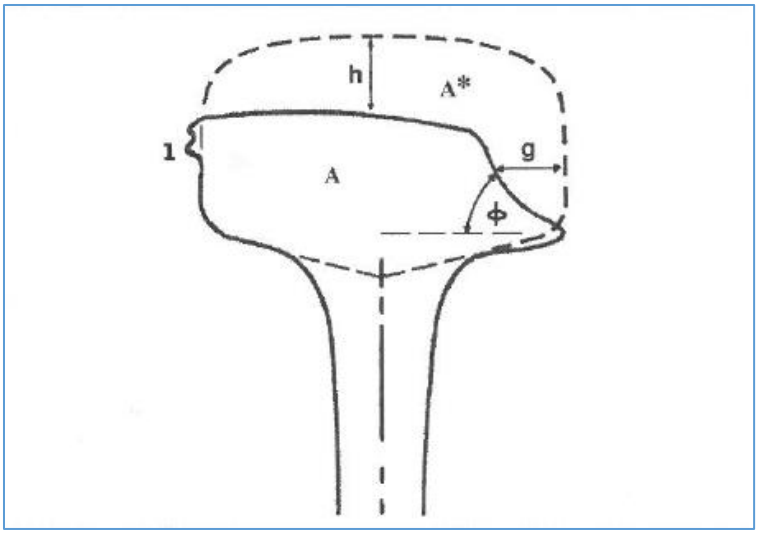


Figure 1 Rail wear defined (Zarembski, 1992).

Typical rail wear measurement parameters are as follows:

A: Area of rail head remaining
A^*: Area of rail head lost
h: Vertical wear (sometimes called head or top) taken at the center of the rail
g: Lateral wear (sometimes called gage or side), normally 5/8"[2] below current top of rail
l: Lip from plastic flow
Φ: Gage face angle: the angle at which the worn plane of the gage face of the rail makes with the horizontal, and at which the wheel flange makes contact.

Railroads in the US traditionally measure the rail wear (at defined locations transversely across the rail head) and compare against defined standards for safety and maintenance. An example of typical rail wear standards is shown below in Table 1.

[2] Note that English units are used throughout this dissertation, as this is the standard for US railways.

Table 1 Typical Rail Wear Standards (Zarembski & Bohara, 1993)

Rail Section	Max Head Wear (inches)	Max Gage Face Wear (inches)
100 ARA-B	3/8 - 7/16	3/8 - 7/16
115 RE	5/8	1/2 - 5/8
132 RE	3/4	5/8 - 3/4

* Replace if wheel flanges strike top of joint bars

* Maximum allowable angle of side wear 32 degrees from the vertical

Note that the standards are a function of the rail size, usually referred to as the rail section (defined in lbs/yard). Many rail sections exist in the US and have different cross sectional parameters, including the overall rail height, head height and head width. As rail wears, the reduction of these parameters effect the ability of the rail to withstand induced stresses, and as such, the allowable section loss varies as a function of the size of the original section. (Zarembski & Bohara, 1993)

When defining the life of the rail as it relates to rail wear, the rate of wear, along with allowable limits, and current state of wear are used to define the overall life and life left on the rail. This is achieved by determining when the allowable limit is exceeded.

Measuring Rail Wear

Railways measure the wear on the rail in track through two primary methods, manual and automated. The measurement of rail wear has been cited back to the late 19th century. (Coxe, 1879) The manual method makes uses of hand held gauges that conform to the web and base of the rail and provide a reference point for measuring the top wear and side wear independently.

Note that since there are many different rail sections available to railways, these gauges are developed individually for each rail section, or a configurable gauge is used (See Figure 2). The wear measurements are read from the gauge and recorded manually. Measurements are made at the discretion of the inspector, typically several locations in a curve or tangent stretch of track, multiple times per year.

Many railways utilize automated inspection systems that use laser-camera machine vision to measure the profile of the rail. These images are converted to Cartesian coordinates that are then compared to the original section and the rail wear parameters are computed automatically (See Figure 3).

Figure 2 Rail wear gauge (Western Safety Products, 2018)

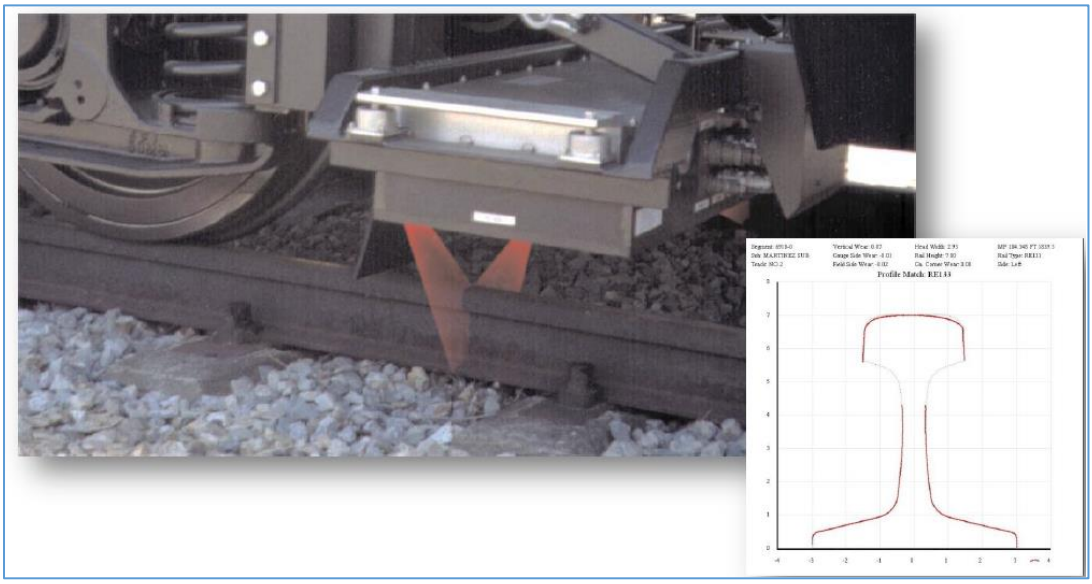


Figure 3 Automated rail profile measurement system, KLD Orian (Magnus, 2014).

These automated systems can capture in-situ rail profiles at a defined distance interval, typically every 10-15 feet, with some systems capturing rail profile every inch. (J.W. Palese, Zarembski, Hartsough, & Ozturk, 2015) The data is then post processed to; determine the original rail section, align the measured rail section to the original rail section (usually utilizing the base fillet radii), and calculate the wear parameters. The output of this system is semi continuous wear

measurements (based on the measurement interval) and results in typical output for the track shown in Figure 4.

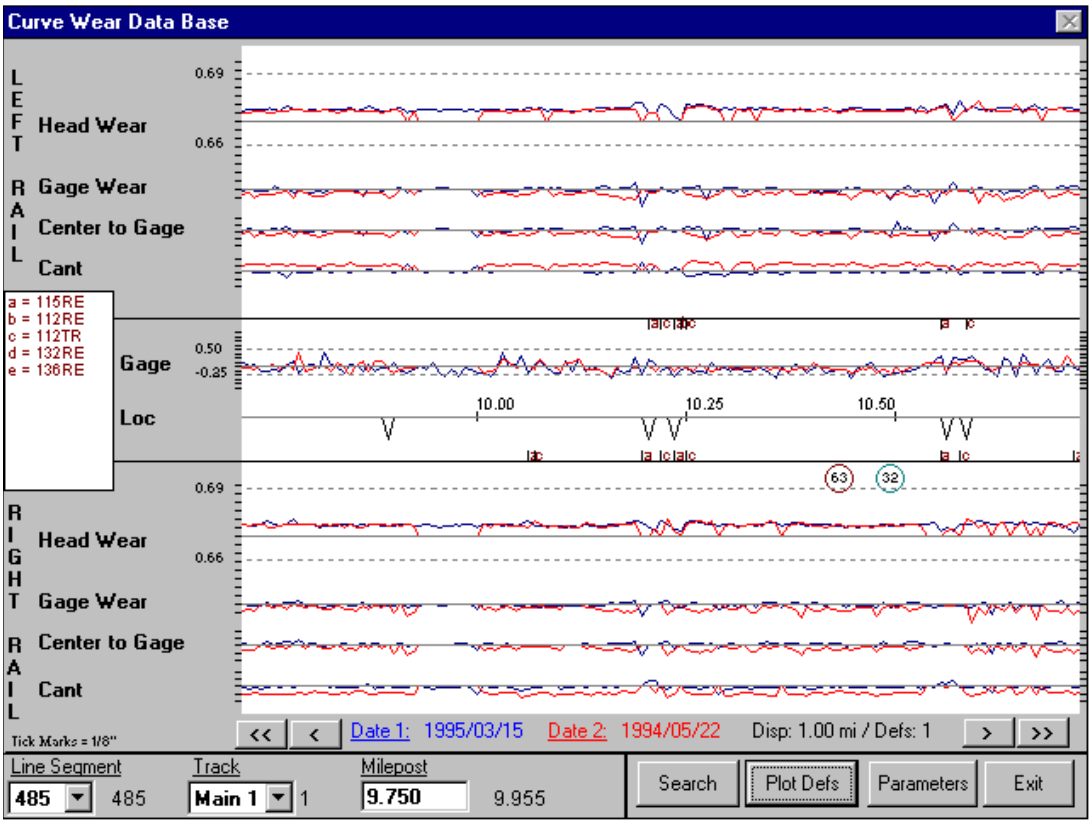


Figure 4 Typical semi continuous rail wear output (Zarembski, Thornton, Palese, & Forte, 1998).

REVIEW OF RAIL WEAR MODELING AND PROBLEM FORMULATION

Much research has been conducted to understand the rail wear phenomenon and those parameters that predominantly effect the rate of rail wear. This research has been performed to date in the laboratory and in the field, and has focused on understanding the effects of axle load, wheel/rail contact (rolling-sliding) and the stress at the wheel/rail interface, metallurgy (chemical composition and hardness) of the rail and wheel, lubrication, and track/operating parameters that effect the force at the wheel/rail interface. Typical rail wear regimes that have been modelled are shown in Figure 5.

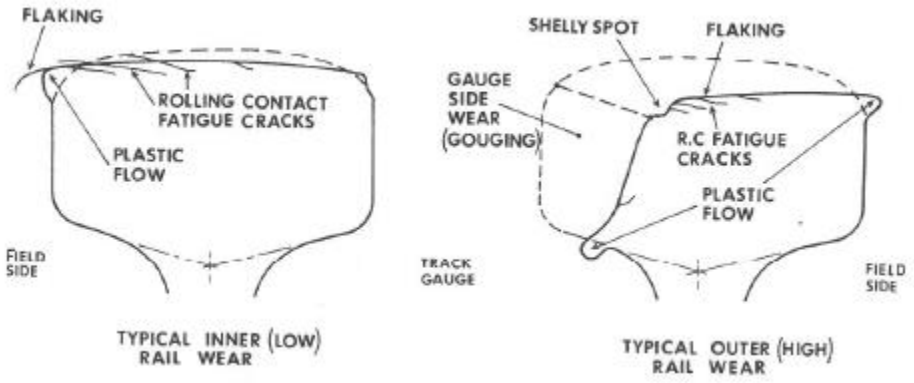


Figure 5 Typical curve rail wear on high and low rails (Ghonem, Kalousek, Stone, & Dibble, 1982).

It is understood that many modeling approaches have been developed to predict wear relationships for two bodies in rolling-sliding contact under different loading conditions. The simplest form of the model is Archard's wear equation, with the wear coefficient encompassing all influencing parameters.

This chapter reviews traditional rail wear modeling approaches (both mechanistic and statistical) utilized to date that have the goal of evaluating the life of the rail. Thus, little will be offered on the mechanistic and empirical approaches of modeling wear in the laboratory, unless it relates directly to modeling the life of the rail. In addition, the formulation of a systems approach to rail wear modeling using readily available inspection data is presented.

Traditional Rail Wear Modeling

Traditional rail wear modeling can be broken down into two primary areas, mechanistic (deterministic) and statistical. The mechanistic approaches have been primarily empirical based, while the statistical processes have evolved from general statistical approaches (linear regression of wear to determine average wear rates) to recent advances in modeling using stochastic processes.

Mechanistic Wear Modeling

Mechanistic wear modeling has focused on rail wear rate and rail life as a function of applied load, with additional focus on the sub elements effecting wear that include, wheel/rail contact, creep forces, effects of lubrication, effect of metallurgy, and operational parameters.

Considering the work of Archard, and that wear can be defined as an energy dissipation process, whereby the volume of wear is related to frictional energy by some wear constant, the following general wear equation results (Magel & Kalousek, 2002):

$$V = K_{s,p} \iint_{contact\ area} s(x,y)Q(x,y)dxdy \tag{1}$$

Where

V = volume of wear
$K_{s,p}$ = wear coefficient as a function of slip and pressure[3]
$s(x,y)$ = creepage at any point in the contact zone
$Q(x,y)$ = surface traction at any point in the contact zone

Recognizing early on that predicting actual rail wear rates in track due to the complexity, and interaction, of wear processes, the following equation was presented that included effect of lubrication and rail hardness (Kalousek & Bethune, 1978):

$$V = \frac{C\eta L}{\sin(\gamma)H^a} \tag{2}$$

Where

V = volume of wear
$L/\sin(\gamma)$ = flange force (perpendicular to wheel flange/rail interface)
η = mean value of lateral and vertical creep (increasing with angle of attack)
H = metal hardness
a = exponent defining nonlinear effect of rail hardness on wear rate
C = coefficient dependent on type and amount of lubrication at the wheel/rail interface

The above equation still requires empirical calibration through testing.

An equation for modeling rail wear as a fatigue process (combined with adhesion) was proposed, in its simplest form, as follows (Steele & Devine, 1982; Stone & Steele, 1978):

$$R = \frac{Wh}{P} f \left[\frac{\tau_a}{\tau_f'}\right]^{-1/b} \tag{3}$$

Where

[3] Note that the $K_{s,p}$ wear coefficients are determined empirically and have been the focus of much research, as described previously.

R = rate of wear (volume per cycle of loading)
$\frac{Wh}{P}$ = volume in which fatigue damage accumulates
h = depth at which fatigue damage occurs
W = function of load
P = proportionality constant related to hardness
$f\left[\frac{\tau_a}{\tau_f'}\right]^{-1/b}$ = inversely proportional to number of fatigue cycles to produce failure
$\frac{\tau_a}{\tau_f'}$ = applied stress/monotonic fracture strength
b = steel material property (usually -0.1)

Considering angle of attack as a key driver in the rate of wear, one early mechanistic approach utilized Archard's fundamental relationship and expanded it to allow for shifting of wear regime with increases in angle of attack. This mechanistic approach was calibrated with field data. The fundamental resulting relationship for rail life (γ) as a function of angle of attack (ψ) is as follows (Ghonem & Kalousek, 1984):

$$\begin{aligned} \gamma(\psi) &= l(\psi - m)^2 + n \qquad \varphi \leq 0.0058\, rad \\ \gamma(\psi) &= p\psi + q \qquad\qquad \varphi \geq 0.0058\, rad \end{aligned} \tag{4}$$

Where l, m, n, p and q are constants which can be defined for a specific set of boundary conditions. These constants are determined through testing and practically implementing the approach is limited.

Early Empirical Models

Early empirical models relied on ease of calculation, as available computing resources were limited and calculations intensive. This class of models utilized straight forward empirical relationships to calculate the predicted rail wear rate or rail life using a calculator. These models were based on mechanistic relationships developed in the laboratory and field environment, and normally calibrated with limited field data.

One study at FAST resulted in an empirical equation for wear rate (inches of wear per million gross tons of traffic, MGT) as a function of rail hardness, applied stress (for a 19 kip wheel load), and lubrication based on observations of wear data in the field for variations of these parameters as follows (Steele & Reiff, 1982):

$$\log(R) = -3.4437 + 27.41\left[\frac{\tau_{oct}}{H^{\mu}}\right] \tag{5}$$

Where
R = rate of wear (in/MGT)
τ_{oct} = surface octahedral shear stress
H = rail hardness (BHN)
μ = coefficient of friction

This equation showed very good agreement to the specific conditions for which the data was obtained.

Another study that built on early work in Europe and North America was sponsored by the Association of American Railroads. (Reiner & Staplin, 1984) The study presented the Couard-Gant model developed in Europe in the 1930's and calibrated with 50,000 field measurements as follows:

$$L_y = \frac{W}{T(a + bD^x)(1 + dg^2) + c} \tag{6}$$

Where

L_y = rail life in years
W = allowable wear area (in^2)
T = annual traffic density (MGT/year)
D = curvature (degrees)
g = gradient (percent)
a = nominal wear rate of rail in tangent track (in^2/MGT)
= 0.0007
b = nominal wear rate of the high rail in curved track (in^2/MGT/(degree^x))
= 0.000173
c = annual rate of corrosion (in^2/year)
= 0.0045 for tangent track
= 0.0052 for curved track
d = constant relating wear to grades
= 0.023
x = 2.0

This equation did not conform to North American conditions where the rail was considerably harder, the car type differed (Europe had many two axles cars), and geometry for curves (particularly elevation in curves) was substantially different. As a result, the Chessie System gathered 1,500 rail wear measurements and recalibrated the equation in the 1970's with the following changes:

a = 0.00056
b = 0.000715
x = 1.2

As this equation did not take several critical parameters into account, such as, wheel load, rail hardness and lubrication, the study resulted in a revised rail life equation as follows:

$$L_c = \frac{TW_c}{K_c\left[Ta\left(\frac{P}{23}\right)^2 + TbD^x\left(\frac{P}{23}\right)^y\right][1 + dG^2] + c_c} \tag{7}$$

$$L_t = \frac{TW_t}{K_t\left[Ta\left(\frac{P}{23}\right)^2\right][1 + dG^2] + c_t} \quad (8)$$

Where

L_t = tangent rail life in MGT
L_c = curved rail life in MGT
K_t = 1.0 for standard carbon rail and no lubrication
K_c = 1.0 for standard carbon rail and no lubrication
W_t = allowable cross sectional area loss in tangent rail (in^2)
W_c = allowable cross sectional area loss in curved rail (in^2)
T = annual traffic density (MGT/year)
D = curvature (degrees)
G = gradient (percent)
P = static wheel load (kips)
a = nominal wear rate of rail in tangent track (in^2/MGT)
= 0.00056
b = incremental wear rate of the high rail in curved track (in^2/MGT/(degree^x))
= 0.000715
c_t = annual rate of corrosion for tangent track (in^2/year)
= 0.0045
C_c = annual rate of corrosion for curved track (in^2/year)
= 0.0052
y = modified Barwell exponent
= (3.70-0.40D), D≥2°
= (2.14-0.14D), D≥5°
d = constant relating wear to grades
= 0.023
x = 1.3

For the rail at the time, this equation provided predicted results within 15% of the actual values for the test data.

Wheel/Rail Contact Models

Various wheel/rail contact equations were presented previously. This section covers the application of these equations, as well as revised equations, to the simulation of wear. Several specific wheel/rail contact models have been developed that have been used as well, to include CONTACT™ and FASTSIM™ developed by Kalker. (Kalker, 1990)

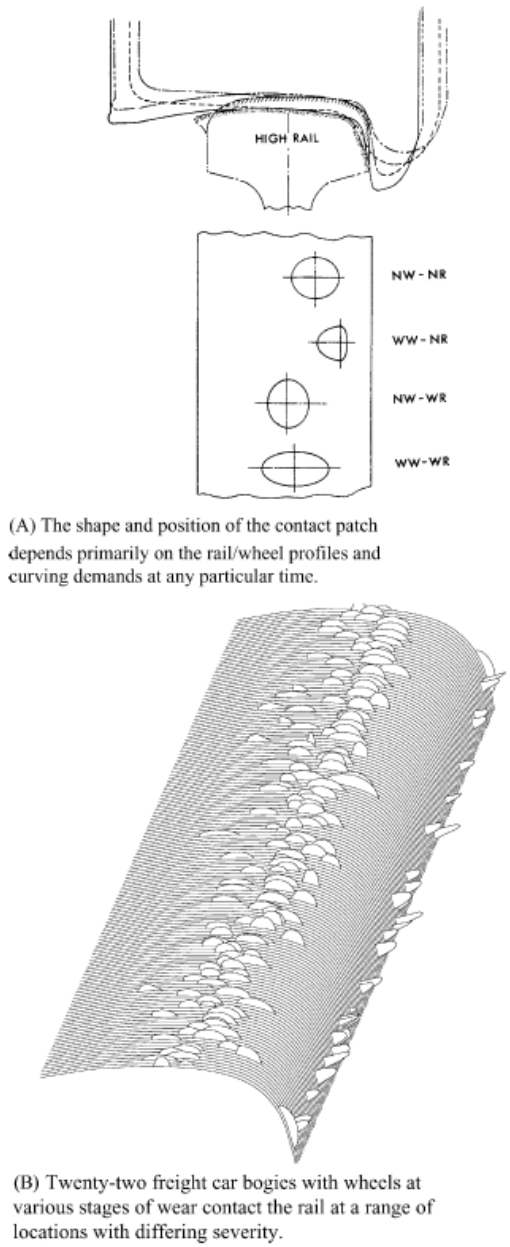
In order to investigate the combination of wear and plastic deformation, an enhanced contact simulation model was developed using a modified Winkler method, and verified using finite element modeling. The results were used to simulate wear, showing the variation in wear with various contact locations and contact stresses. (Telliskivi & Olofsson, 2004) This research was extended to a non-Hertzian contact shape and allowing for variable creep along the contact patch. A simulation model was developed and calibrated with disc-on-disc laboratory results. Good

quantitative agreement was found between the experimental and simulation results for wear, along with good qualitative agreement of the resulting surface conditions. (Telliskivi, 2004)

As indicated previously, wheel/rail wear and contact position change as a function of current wear (shape) on the wheels and rails, resulting in various applied stress levels. (Tournay & Mulder, 1996) In order to simulate wear, the stochastic nature of wheel/rail contact can be considered. This was the subject of one study on Amtrak which developed the model PUMMEL™ and a quasi-static curving model (using NUCARS™) to distribute wheel rail contact of wheels in the various stages of wear (varying profiles) on the rail (along a route profile) to predict wheel profile wear. The variation in contact is shown in Figure 6. This approach was used to design wheel profiles that would have a decreased wear rate in normal operations. The study points out that the actual friction conditions were unknown and could dramatically affect the results. (Magel, Kalousek, & Caldwell, 2005)

Considering plastic flow in the near surface layer, one study utilized a three dimensional elasto-plastic finite element model to simulate accumulated plastic strain in the rail head. The resulting simulated plastic flow was compared to observed flow in real rail head conditions of tangent track with simulated results larger than actual. (Busquet, Baillet, Bordreuil, & Berthier, 2005)

When simulating friction in the contact zone, it is important to understand the effects of the roughness associated with each surface (wheel and rail), as well as the material properties in contact. This was the subject of a study whereby simulation of various contact was formulated and solved with physical models. The result was the development of an algorithm for modeling the friction coefficient at the submicron level. It follows that a probabilistic value of roughness (and resulting friction coefficient) results in a probabilistic wear rate. (Bucher et al., 2006)



(A) The shape and position of the contact patch depends primarily on the rail/wheel profiles and curving demands at any particular time.

(B) Twenty-two freight car bogies with wheels at various stages of wear contact the rail at a range of locations with differing severity.

Figure 6 Variation in location and amplitude of contact using PUMMEL™ (Magel et al., 2005).

A disc-on-disc test, where specimens were selected from in-service rails and wheels, developed contact model relationships for sliding velocity, contact pressure, and creep ratio to develop specific wear rates for input to Archard's wear model. The resulting contact relationships resulted in predictions of wear within 8% of actual. However, the author's note that these results are only valid for interpolation, and that extrapolation to field results may not be valid. (Ramalho, 2015)

Another study evaluated various contact models to evaluate the effect of wheel/rail contact simulation on wheel wear. This study did not consider the plastic flow regime. The study showed

that a significant tradeoff between accuracy of the wheel/rail contact and computation time, thus choosing the right contact model for simulation requires careful consideration. (Tao, Wen, Zhao, & Jin, 2016)

Simulation Models

The core of many simulation modeling approaches is dynamic simulation of train/track interaction to define the dynamic forces as well as contact location. There are commercially available multi-body dynamic simulation models that have been utilized (GENSYS™), as well as specific rail vehicle dynamic simulation models (MEDYNA™, NUCARS™, VAMPIRE™, etc.) that are cited in the references (Olofsson & Telliskivi, 2003). As the wheel and rail wear together, attention has been focused on the wear of the rail, the wear of the wheel, and mutual rail/wheel wear. As such, various approaches are presented for all three.

An extension of this analytic approach evaluated wheel profile wear on tangent rail using a combined model approach, specifically, a vehicle dynamic simulation model and several wear models with assumed contact. (Fries & Dávila, 1987) The simulation provided reasonable resulting wear profile shape, however, an important finding was the actual magnitude of wear was dependent on the wear model applied (subject to laboratory test data). The authors noted that the shortcomings of the approach were no vehicle curving in the simulation and proper wear model selection. The authors also noted that the same approach could be applied to rail wear.

This approach was extended to include a more robust dynamic simulation, along with simulated wheel/rail contact, curving, and an energy dissipation wear model. (Pearce & Sherratt, 1991) The authors noted that each cycle of a train over a route was modelled and consumed large amounts of computation time, 29.5 hours to model a wheel through its 125,000 mile life.

One wheel wear prediction tool took advantage of a full vehicle multi-body simulation model, along with measured rail profiles to model a metro vehicle through a line with varying curvature. The model included track irregularities and coefficient of friction, along with material properties as input to the simulation model to calculate the tracking position and forces. These in turn were used with CONTACT™ and FASTSIM™ to determine contact stresses and predict wheel wear using an energy/work based model. The results showed very good agreement between a measured wheel and a simulated wheel after 125,000 miles of wheel travel. (Jendel, 2002)

The wheel wear simulation was extended to include braking and additional change in wheel/rail contact utilizing Archard's wear model with associated wear maps. Modeling braking through applied torque and changes in coefficient of friction provided expected and reasonable results, though no field validation was performed. (Enblom & Berg, 2005) This same approach was extended to rail wear, showing expected behavior, however, the magnitudes for changes in lubrication were not as expected. (Enblom & Berg, 2008)

A study for predicting wheel wear comparing various wear models in a combined simulation approach was performed showing similar profile shapes, with variations in magnitude of wear. An incremental approach was utilized whereby each resulting profile from a wear cycle was used as

input to the next wear cycle simulation. Selection of the appropriate contact model had significant influence on the results. (Pombo et al., 2011)

Most mathematical (deterministic) simulation models utilize a combination of vehicle dynamic simulation models to determine forces as a function of vehicle and track conditions, a wheel/rail contact model to determine applied stresses, and a local wear model (empirical based on laboratory test data). It is reasonable to expect a compounding of errors and an extensive amount of computational requirements using this approach. One study evaluated wheel wear with a combination of models, and validated the results against controlled condition in a full-scale laboratory environment. The simulated results were overstated and further testing was recommended. (Braghin, Lewis, Dwyer-Joyce, & Bruni, 2006)

Finite element modeling for the wheel/rail contact to understand contact fatigue and contact stresses, particularly how it relates to head wear, was performed in a study for the French railways. The study provided illustrative examples with the intent of extending the research to further applications. (Van & Maitournam, 2002)

A similar approach (combination of models) that evaluates wear on curves only was performed for China Railways, and evaluated the incremental wear associated with each wheel of a passing truck. Since the lead wheel on the outside rail has a larger creep associated with increased stick/slip, the incremental wear caused by the lead wheel on the outside rail is the largest contributor. The model was able to determine wear contribution for each wheel in the truck. (X. Jin, Wen, Xiao, & Zhou, 2007) Another method developed for the China Railways utilized a three-dimensional nonlinear dynamic finite element method, along with the FASTSIM™ contact model, calibrated to disc-on-disc test data. This simulation method resulted in profiles that showed reasonably good agreement between the test data and simulation results at the gage corner where large lateral creepage occurs. (Chongyi, Chengguo, & Ying, 2010)

Combining wear and plastic flow was the subject of one study that used a combined simulation approach using GENSYS™ as the dynamic simulation model, FASTSIM™ as the contact model and Archard's wear model. Plastic strain was modelled using a 2D finite element model with a plane strain assumption (idealized). The wear coefficients were calibrated using a full-scale laboratory fixture. The authors noted weakness in the contact model with respect to wear sensitivity. In addition, the model is only valid for a limited set of conditions. (Brouzoulis, Torstensson, Stock, & Ekh, 2011)

Another study for Japanese Railways used a combined simulation approach and accounted for additional variables such as angle of attack, lateral force, hardness, and lubrication. The wear model used was Archard's wear coefficient model, and wear coefficients were calculated using laboratory data. The predicted results from the model had reasonable agreement with the real-life in-track data. The author's noted the weakness of the model to account for slip. (Y. Jin, Ishida, & Namura, 2011)

A model developed and validated on the Italian railways utilized the dynamic simulation model SimpackRail and a global wheel/rail contact model (developed by the authors) to analyze the

wheel/rail system, for a particular vehicle and route. The calibrated model showed very good agreement to actual measured in-track profile data.
Figure 7 shows the evolution of the rail head profile through 200 MGT. (Ignesti, Malvezzi, Marini, Meli, & Rindi, 2012) This approach was expanded to include a matrix of track conditions (curvature) as input as opposed to a single line. (Innocenti, Marini, Meli, Pallini, & Rindi, 2014)

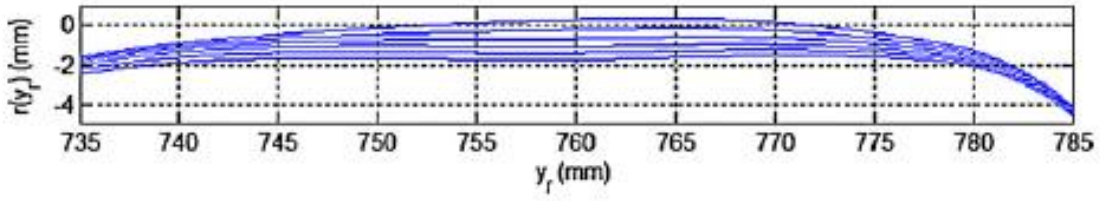


Figure 7 Rail head profile evolution through simulation (Ignesti et al., 2012).

A similar approach was utilized by China Railways to simulate rail profile evolution using a complex dynamic simulation model that models the track and the vehicle (a matrix of vehicles in this case), a contact model and Archard's wear model. Extensive field tests were used to calibrate the model for the line being analyzed. The simulated results showed good agreement in shape, but significant differences in magnitude of wear (See Figure 8), indicating a missed modeling component or calibration issue. (Li et al., 2016) This approach was extended to turnouts as well. (P. Wang, Xu, Xie, & Chen, 2016)

Considering that a ductile material subject to repeated rolling contact, very high levels of shear strain can be accumulated near the surface. When the strain levels exceed the life of the material, particles flake off in the form of wear. One simulation approach divided the rail into layers and elements and used simulation of contact to accumulate strain. When an element failed, adjacent elements would accumulate strain at a higher rate and eventually fail. The study showed wear rates comparable with other laboratory testing. (Franklin, Widiyarta, & Kapoor, 2001)

As with the genesis of a number of simulation models, the combined approach has been built upon to add in other causal factors. The effect of the roughness of the rail and wheel surfaces was investigated using two-dimensional finite element modeling and showed that asperities of a certain height and shape would plastically deform under just a few loading cycles, and rapidly approached a steady state. (Daves & Fischer, 2002)

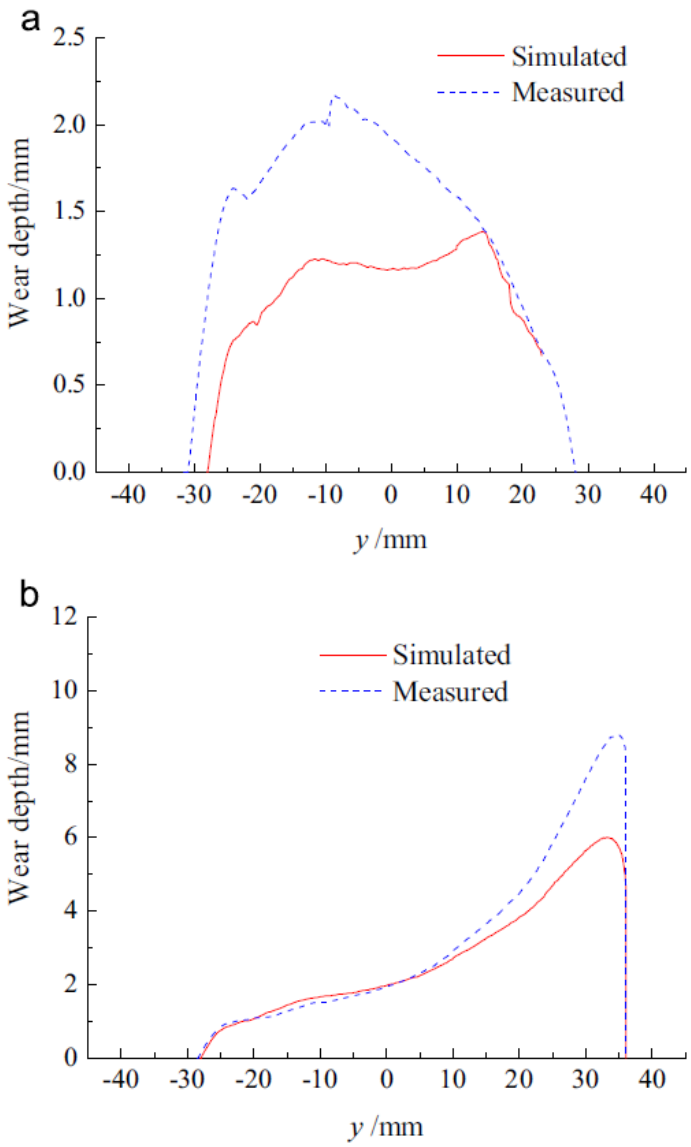


Figure 8 Simulated wear compared to actual wear, (a) low rail, (b) high rail (Li et al., 2016).

One study introduced a surface roughness algorithm (for both the wheel and the rail) into the contact model to identify high asperity contact stress location, which shifts to the plastic regime, greatly effecting the rate of wear. (Bărbînţă & Creţu, 2018) This was also the subject of an in-depth study by the Spanish railways utilizing a combined modeling approach. The major result was based on high precision measurement of the pressure at the contact surfaces (using ultrasonics), which showed a high variation to that of Hertzian theory (most widely used). This was the effect of asperities carrying the load due to the roughness of the surface. Thus, peaks in wear, normally found in experimental observations, can be attributed to the roughness and resulting high contact pressures (Rovira, Roda, Marshall, Brunskill, & Lewis, 2011).

Another study accounted for vertical stiffness of the track (combined rail and support stiffness) by introducing a nonlinear multibody system dynamic analysis model, along with a contact model and energy dissipation wear model. The study concluded that track stiffness had little effect on wear in tangent track. However, when irregularities were introduced into the track, and resulting dynamic forces were accounted for, wear was affected as expected. (Aceituno, Wang, Wang, & Shabana, 2017)

To this end, while the simulation models may not provide exact replication of field conditions, such models are very useful for understanding relative changes in operations and maintenance. One Brazilian study used the combined simulation model approach to understand the effects of a new heavy axle load car design on rail wear (specifically rail life). The approach allowed for modeling the increased vehicle weight, along with the increased dynamic forces as a function of the new car characteristics, and their effect on rail wear. This allowed for understanding the economic benefits of higher performing operations, along with their cost impact. (Urban, Barbosa, Santos, Joy, & Shu, 2015) A similar approach was used on a study performed for China Railways, focusing on curve wear. (J. Wang, Chen, Li, & Wu, 2015)

Statistical/Stochastic Wear Modeling

While mechanistic modeling approaches are deterministic in nature (using empirical data), a stochastic process allows for uncertainty of the input and output variables to provide a prediction that may not be precise, rather it is probabilistic. When considering rail wear modeling, there is a probabilistic distribution of many of the input variables, including, but not limited to (Costello & Premathilaka, 2011):

- Wheel/rail contact due to variations in the wheel and rail shape
- Coefficient of friction due to lubrication application and tracking effectiveness
- Rail hardness due to work hardening and limited information on installed rail metallurgy
- Actual applied forces on any given rail due to track geometry variations and vehicle conditions
- Variations in support condition

Thus, deterministic modeling will not readily result in a proper prediction of rail wear. Stochastic modeling allows for the uncertainty of the inputs to provide a probabilistic output for rail wear.

An early study utilized a stochastic process that evaluated a population of wheel and rail profile shapes, combined with a dynamic simulation model (MEDYNA™) and an elliptical contact model (FASTSIM™) to determine incremental wear on both the rail and wheel. Accumulated wear on the rail and wheels was modelled based on dissipated energy. This study showed the power of simulation modeling under semi-ideal conditions. (Zobory, 1997)

This study was extended to classes of wheels and rails for simulation, as well as a probability of lateral track irregularities, to develop predicted profiles for a railway line with a distribution of curves. The presented results were not validated, and a full spectrum of vehicle types was not

considered. (Szabo & Zobory, 1998) This approach was further explored utilizing updated contact models to predict wheel and rail profiles for use as input to a dynamic simulation model (Enblom & Berg, 2005)

When performing stochastic processes for rail wear, it is important to understand the compounded error that can be introduced in the measurements, including direct measurement error, as well as longitudinal positioning error. This is particularly important when performing second level analyses, whereby measured data (with inherent measurement error) is transformed into rate date (error compounds). A reliability model was proposed for use with laboratory test data to characterize the results. (Ramalho, 2010)

Recently, researchers have started to apply stochastic processes to rail wear modeling to gain a better understanding of the wear mechanisms and overall asset needs of the railway on a probabilistic basis. In essence, rail wear modeling is a continuous deterioration model with time or accumulated traffic, and can be considered at the rail profile level, a rail segment level (curve), or a system level (entire railroad or subset).

Early deterioration models evaluated average wear on a track segment (curve or tangent) for each rail (high or low) for each primary measurement (head wear or side wear) through linear regression techniques. The resulting wear rates were then used with replacement thresholds to determine the wear rate that controls replacement (head or gage). (Zarembski, 2010) This technique has been used in practice for many decades, and is in fact a large component of the empirical aspects of the mechanistic modeling approach.

Another cited study on metro track performed multivariate regression analyses on head and gage wear as a function of accumulated tonnage and curvature to show relationships specific to the traffic and rail section associated with that track with consistent results in rail wear rate relationships. (M.C. Jeong, Kim, Lee, Kang, & Kong, 2010)

The linear regression method can be used on a profile level or rail segment level and repeated as necessary. Note that this approach is considered an absolute approach as all of the historic data is used to develop a linear wear rate (sometimes non-linear if the data suggests nonlinearity) that can be used to project wear forward. An incremental model would allow for the most recent data to predict the next state, however, the data available is generally not accurate enough for this approach.

This approach can be extended to look at a group of rail segments and perform multivariate regression analysis to develop a wear equation that considers available independent variables (their correlation) and influence on wear rate.

One study utilized available data (for a defined rail section) to develop a model for determining wear. (Premathilaka & Costello, 2010) While several independent variables were known, two models for high rail gage wear (HS in inches) and low rail head wear (LT in inches) were developed based on rail age (t in years), annual tonnage (T in MGT), and angle of attack (ψ in degrees) as follows:

$$\mathrm{HS} = 0.4783tT\psi \quad (9)$$
$$\mathrm{LT} = 0.1936tT\psi \quad (10)$$

Where, angle of attack defined herein is, ψ = I/2, and I is the angle of deflection of the curve.

This approach would require multiple models for independent variables not in the equation such as, rail section, traffic type, rail metallurgy, etc.

One recent study evaluated the rail requirements due to wear deterioration using a Markov process using Markov chains. Since rail wear is discrete in time with a finite state space and satisfies the Markov property (the future state depends on the present state), this approach is relevant to stochastic modeling. This study determined the next state, i.e. what percentage of rail will move from one state to the next (up to and including failure) for several rail sections for a defined sub-network. The two primary component of the Markov chain, Initial Condition Vector (ICV) and Transition Probability Matrix (TPM) were determined using the actual historic wear data (measured with handheld gauges) and modified using engineering expertise when data was not available. This study, however, did not offer a validation by segmenting the data, but stated it was in process. (Costello & Premathilaka, 2011)

An additional study that utilized Markov theory for predicting the deterioration of rail wear studied one curve with limited (6 months) data to develop the ICV and TPM. The measured data was converted into wear condition indices (integer defining rail quality as a level of wear, based on percentage of threshold) and utilized in the Markov process. The results showed that the test case was already condemned. (Zakeri & Shahriari, 2012)

The University of Huddlefield employed artificial neural networks to predict wheel and rail specimen wear based on laboratory data. In this study, controlled tests were performed on specimens under dry, wet and lubricated conditions over many cycles. A nonlinear autoregressive model with exogenous input neural network (NARXNN) was trained using supervised training methods under different contact conditions. The test data was used to train, validate and test the model. Predictions using the NARXNN showed good agreement with a mean absolute percentage area (MAPE) of less than 8%. This approach proved valid for laboratory test data, but has yet to be employed with field data, which will have larger variation in the inputs (and effected by other parameters) that must be considered. (Shebani & Iwnicki, 2018)

A particle filter method based on Bayesian theory was utilized on a stretch of metro track for data obtained from six quarterly inspections, focusing only on curves. This method allowed for uncertainty in the measurement data, as is typically found in automated measurement systems. The first three measurement points were run through the particle filter model for 1000 particles, and predicted out to the fourth measurement. While 87% of the actual fourth inspection measurements fell within the 95% confidence interval of the predicted measurements, it was seen that the 95% confidence interval was larger than practical, +/- 0.2". No validation with respect to predicted value was provided. (Min Chul Jeong, Lee, Cha, Zi, & Kong, 2019)

Discussion

Early research focused on laboratory testing in a controlled environment to evaluate the wear performance of various materials. It is recognized that it is extremely difficult, if not impossible, to develop appropriate wear relationships of rail in the field based on first principles. Due to the difficult nature of evaluating rail steels in track subjected to many variations in causal parameters that effect the rate of wear, laboratory testing to develop wear relationships as a function of primary metallurgical parameters and mechanical variables, not just ultimate tensile strength of the material, were initially deemed the most effective method. (Morton, Cannon, Clayton, & Jones, 1978)

Initial understanding of the rail-wheel tribosystem, and its various causal factors (inputs), functional relationships (tribosystem envelope) and failure mechanisms (outputs) is depicted in the diagram of Figure 9. The interrelationship and subcategorization of the input variables is complex.

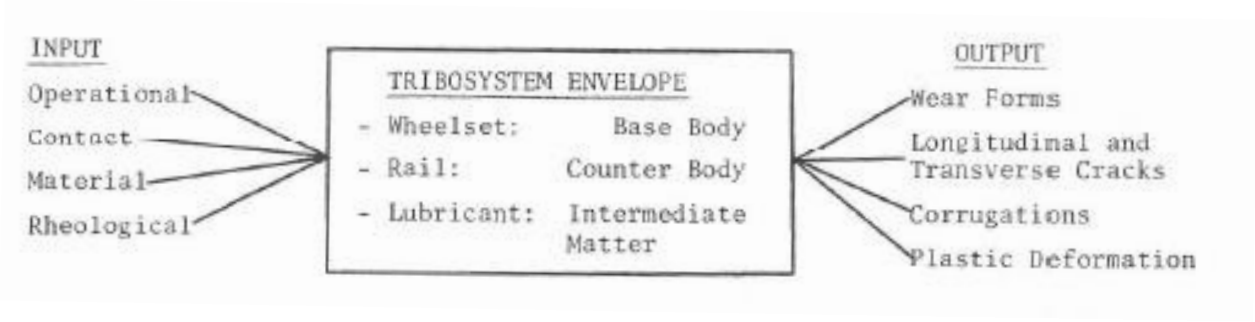


Figure 9 Rail-wheel tribosystem (Ghonem et al., 1982).

The importance of understanding the relationship between individual wear mechanisms and service environment can assist railways in understanding effective maintenance mechanisms to improve the incremental return on investment, particularly in response to shifts in traffic (and corresponding axle load). (Roney, 1982) It is also important to recognize that as wear accumulates, the shape of the rail (and wheel) will change, and impacts the overall instantaneous rate of wear. (Zakharov & Zharov, 2002)

In addition, it was noted in several studies that the contact and stress regimes change based on wheel and rail shape, particularly as these components go through wear progression, and in fact, the failure mode can shift through the wear modes as well as shift to a fatigue dominated mode. (Tournay, 2008) The development of wear and fatigue and relationship between modes was addressed with the addition of shakedown and truck stiffness (Dirks & Enblom, 2011) and was recently extended to additional data and verified. (Dirks, Enblom, & Berg, 2016) In addition, the relationship to the development of rolling contact fatigue and rail wear (both rate and shape) was evaluated with respect to increase in derailment risk. (Magel, Mutton, Ekberg, & Kapoor, 2016) Using an energy dissipation wear model in a combined simulation approach, it was shown that optimal rail profiles could reduce the energy dissipation and corresponding wear rate by a factor of five. (Alarcon, Burgelman, Meza, Toro, & Li, 2016)

Many of the simulation models assume a wear model based on Archard or energy dissipation through a frictional work-wear model. The models that rely on wear coefficients do not effectively handle transitions from one mode of wear to another. Some limited work for determining transitions of wear coefficients to be used in a wear coefficient model (Archard's model) has been performed. (Vuong & Meehan, 2009)

The results of this research have culminated in a number of computer models that utilize empirical equations to predict rail life, predominantly as a function of wear, in addition to the models previously presented. (Innotrack, 2008)

Table 2 Empirical Rail Life Models

Organization	Model
Queensland University of Technology, Australia	Integrated Track Degradation Model (ITDM)
Association of American Railroads (USA)	Railway Track Life-Cycle Model (RTLM^{TM})
British Railway Rail Safety and Standards Board	Whole Life Rail Model (WLRM)

A relatively recent literature review of the rail wear prediction and simulation approaches to date concluded that there are opportunities for integrated research in a cross functional modeling approach to account for the interactions between adhesive wear, plastic flow, rail surface damage and development of cracks. (Enblom, 2009)

In light of the previous study, it is recognized that the wear and instantaneous wear rate at contiguous locations along the track will vary. This is due to any number of factors described previously. One study evaluated the effect of track geometry anomalies on non-uniform wear through a segment of rail utilizing a dynamic simulation model, contact model and material wear model. The simulations showed increased wear as a result of increased forces due to track geometry variation. As with other models, agreement with real field conditions varied, as not all causality is captured in the simulation approach, but provides a mechanism to understand relative effects. (Sun, Guo, & Zhai, 2019)

Finally, it was recognized early on that wear mechanism transitions prohibit an effective general model of wear behavior from being developed. To date, the research supports this claim, however, as research continues, the limitations in these approaches may be overcome, particularly in light of the tremendous amounts of data being captured. As evidenced in the above studies, particular phenomena can be easily modelled and tested in the laboratory, and can result in useful simulation models to provide relative wear relationships. This class of models is effective for understanding the wear relationships between load, metallurgy, and other external factors. (Clayton, 1995).

A summary of relevant models and the factors considered in these models is presented in Table 3.

Table 3 Summary of Relevant Models and Variables Considered

Reference	Wear Coefficient	Load	Creep	Traction	Slip	Pressure	Flange Force	Hardness	Lubrication	Angle of Attack	Metallurgy	Stress	Grade	Curvature	Annual MGT	Wheel Profile	Rail Profile	Rail Age
Magel, 2002	✓	✓	✓	✓	*	*												
Kalousek, 1978	✓	✓	✓				✓	✓	*	✓								
Steele, 1982		✓						✓										
Ghonem, 1984	✓									✓								
Steele, 1982								✓	✓			✓						
Reiner, 1984	✓	✓						*	*				✓	✓	✓			
Magel, 2005		✓	✓	✓	✓	✓		✓					✓	✓		✓		
Ramalho, 2015	✓	✓	*	*	*	*												
Ignesti, 2012	✓	✓				✓		✓	✓	✓			✓		✓	✓		
Zobory, 1997		✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓		✓	✓	
Prematilaka, 2010	✓									✓					✓			✓

- = Handled through Wear Coefficient

Problem Formulation

The past research discussed above has shown that relationships for wear can be developed in the laboratory as relates to independent drivers, such as applied force/stress, metallurgy, lubrication and several other factors. When considering rail in track, these factors interact, are unknown, or vary substantially. This section discusses a generalized formulation for understanding the rail wear problem for rail in track.

Wear Model Attributes

A common conclusion of past research showed that modeling rail wear under variant field conditions is difficult, if not impossible, since many of the driving factors are unknown, cannot be measured, or vary for each wheel cycle. The research presented thus far has shown that much attention has been given to the effects of the individual elements associated with the wear process; applied load and resulting contact stress based on wheel/rail geometry, rail metallurgy, lubrication, and operational components. While this research has provided insights to the individual component effects, little success has been achieved in the area of a systems approach, and all the variability that enters through real life conditions.

Figure 10 shows a schematic of the rail wear process and its major contributing factors. Thus, in the simplest sense, and consistent with the research presented, rail wear is accumulated by each passing wheel. Each passing wheel (from a distribution of wheels) may have a different contribution, and is a function of the load and resulting stresses applied for that wheel, in the confines of the tribosystem and environment for which it is applied. In addition, artificial wear due to rail maintenance (rail grinding) can occur.

The tribosystem in which the wheel operates is defined by the material properties in contact, and level of lubrication. The applied stresses associated with each passing wheel are a function of the contact geometry and the applied forces, which are in turn determined by the track design and condition, vehicle design and condition, and kinematics of the operation.

Figure 10 shows the interconnected nodes, which drive rail wear (circles), along with the contributing factors (ellipses). The contributing factors are data elements that change over time and are often measured by the railway, with availability of data color coded as follows:

Green = Known and measured frequently
Yellow = Sometimes known but not measured frequently
Red = Unknown or not measured

Figure 10 Rail wear flow diagram.

The system defined above contributes to the overall wear of the rail (and wheel) with each passing wheel making an independent contribution.

General Formulation

Wear can be defined as the volumetric loss of material due to applied slip and pressure as a function of a wear coefficient, which is in turn, a function of the material, geometric, and tribological properties of the system. As shown in the research, the wear coefficient is linear for an applied load cycle. This can be written generally for any applied load cycle as the following equation:

$$\Delta w_i = k_i \Delta n \tag{11}$$

Where

Δw_i = incremental wear due to load cycle
k_i = wear coefficient
Δn = load cycle = 1

Considering multiple load cycles, whereby the wear coefficient is specific to each load cycle, the accumulated wear for total cycles, N, can be written as follows:

$$W = \sum_{i=1}^{N} \Delta w_i = \sum_{i=1}^{N} k_i \Delta n \tag{12}$$

This is shown schematically in Figure 11. Thus, each applied load cycle may have a different rate. While certain aspects of the wear associated with each load cycle may remain constant (rail profile/hardness, curvature, grade, etc.), others may vary (wheel profile/hardness, lubrication, force, etc.).

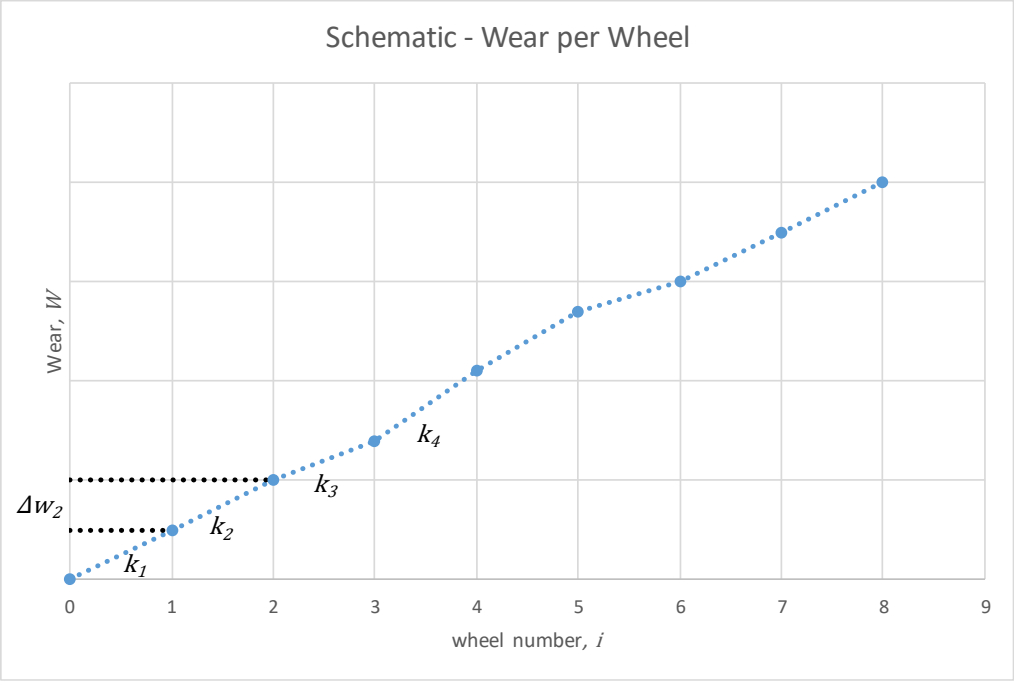


Figure 11 Schematic representation of accumulated wear of each load cycle.

For millions of cycles, with intermittently known points, the average rate of wear of all applied cycles between known points must be considered, with the average rate having some unknown distribution. This is shown schematically in Figure 12. Thus, the total wear can be written in the same form as Equation 3.12 for known data points as follows:

$$W_T = \sum_{i=1}^{N} \Delta W_i = \sum_{i=1}^{N} K_i \Delta MGT_i \tag{13}$$

Where

W_T = Total wear
K_i = Average wear coefficient (rate) between inspections
ΔMGT_i = MGT between inspections
N = number of inspection

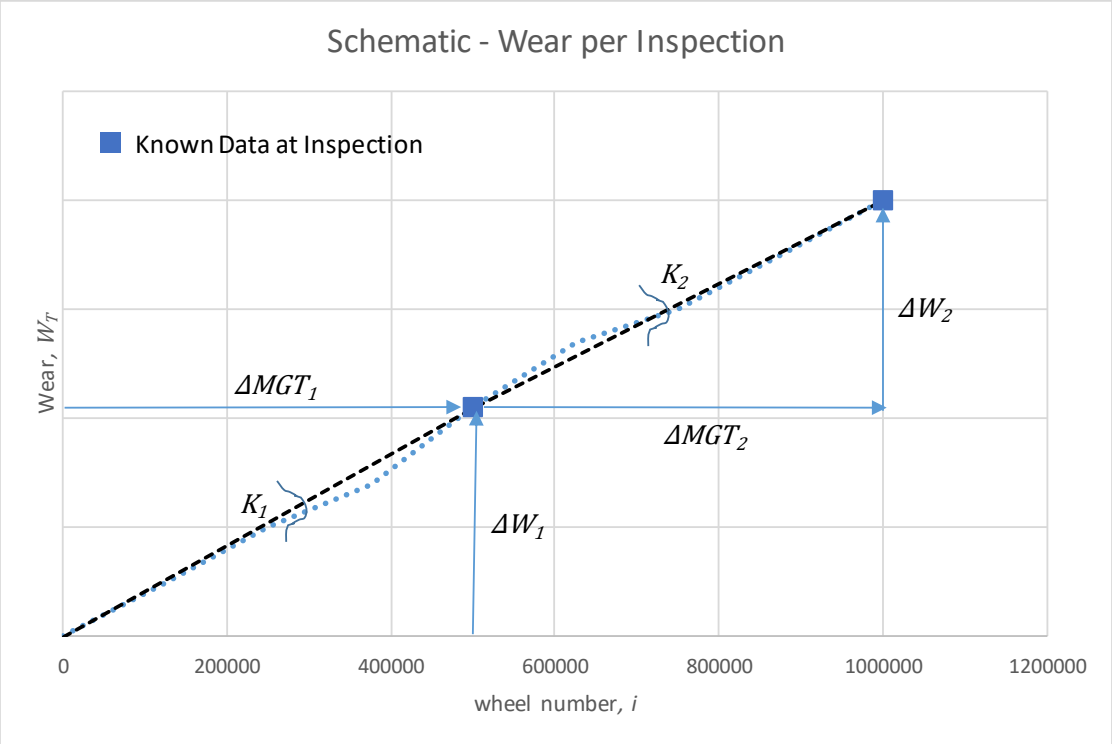


Figure 12 Schematic representation of accumulated wear per known point (inspection).

For multiple inspections over the life of the rail (or partial life), the average rate of wear for the rail can be represented as shown in Figure 13. This figure shows variation for three segments. Thus, the wear for a particular segment (i) can be defined as follows:

$$W_T = (\lambda_i \pm \varepsilon_i)\, MGT \tag{14}$$

Where

λ_i = the average rate for known parameters
ε_i = some contribution associated with unknown measurable parameters

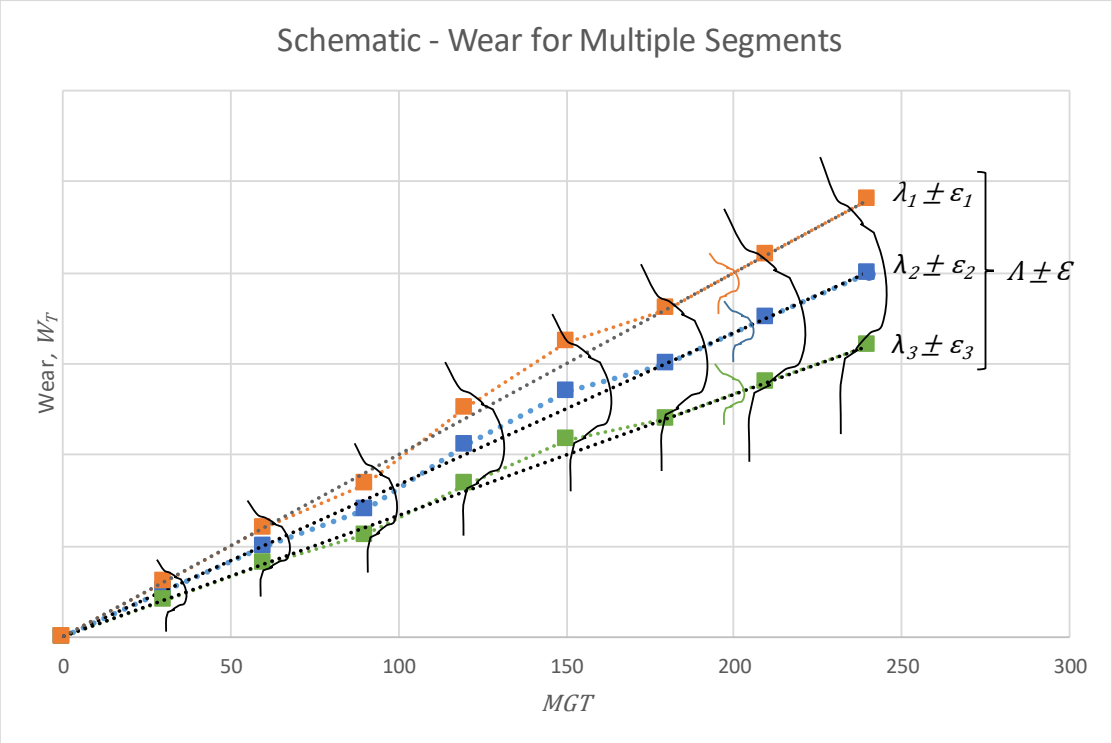


Figure 13 Schematic representation of accumulated wear for multiple segments.

For multiple segments with identical known parameters, the wear for multiple segments can be combined to provide an estimate of total wear as follows:

$$W_T = (\Lambda \pm \mathcal{E})\, MGT \tag{15}$$

Where
Λ = the average rate for known parameters
$\mathcal{E}$ = some accumulated contribution associated with unknown measurable parameters

It is often more appropriate to consider the rate of wear as the predominant result of any rail wear equation. Thus, Equation 3.15 can be rewritten as follows:

$$W'_T(\Theta, \Psi) = \Lambda(\Theta) \pm \mathcal{E}(\Psi) \tag{16}$$

Where
$W'_T(\theta, \varphi)$ = Wear rate, as a function of input parameters
$\Theta = \theta_1, \theta_2, \ldots, \theta_n$ = vector of known parameters
$\Psi = \psi_1, \psi_2, \ldots, \psi_n$ = vector of unknown parameters

Equation 3.16 provides a stochastic approach to determining the wear rate, whereby, the major contribution $\{\Lambda(\Theta)\}$ from known and measured parameters can be determined, and a contribution $\{\mathcal{E}(\Psi)\}$ from unknown/unmeasured parameters has a probability of occurring.

Determining and understanding Λ and $\mathcal{E}$ from a population of available and measured data is paramount to understanding how rail wears in real life conditions.

RAIL WEAR FORECASTING MODEL DEVELOPMENT

The research to date that has been presented thus far has shown that much attention has been given to the effects of the individual elements associated with the wear process; applied load and resulting contact stress based on wheel/rail geometry, rail metallurgy, lubrication, and operational components. While this research has provided insights to the individual component effects, limited success has been achieved in the area of a systems approach, and all the variability that enters through real life conditions.

Considering the generalized wear equations presented previously, a system wear approach can be developed for use in evaluating the wear of the rail based on readily available measurements taken during routine inspections. This must be coupled with the unknown (or latent) variables and factors that enter the wear process.

Given that the research and data has shown a large amount of uncertainty in the causal factors for wear, several stochastic processes are used to develop the proposed rail wear life model. Given its nature, the new model is termed RAIL Wear Analysis using Stochastic Processes (RAIL-WASP).

A flowchart of the RAIL-WASP model framework is shown in Figure 14, and encompasses the deterministic research to date, along with the observations made from the exploratory data analysis. This flowchart shows the conceptual framework for development of multi-stepped approach to determining the life of the rail due to wear and its performance with respect to wear.

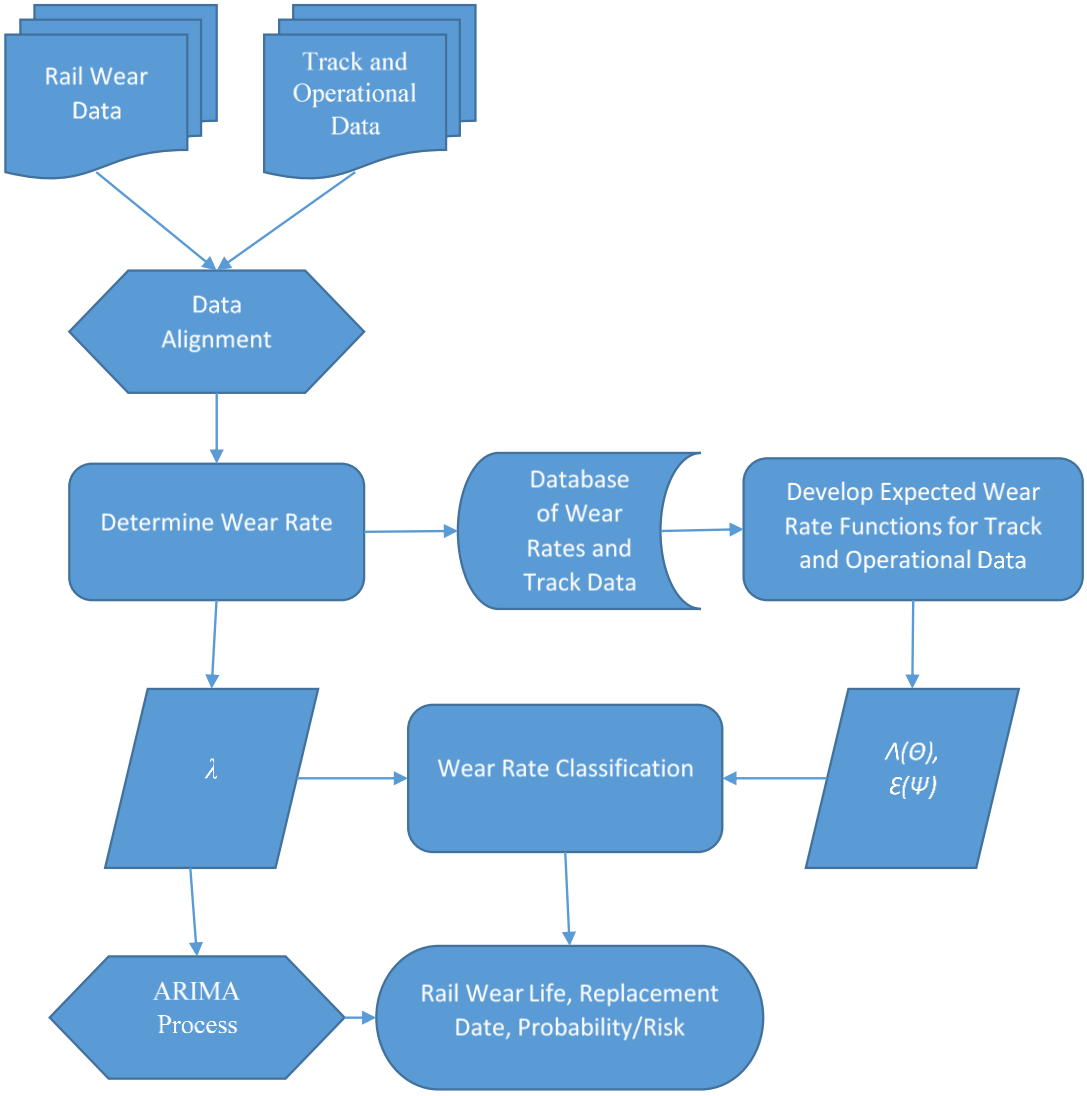


Figure 14 Flowchart of the RAIL-WASP model.

The next sections introduce the encompassing stochastic processes.

Stochastic Processes used in RAIL-WASP

Several stochastic processes used to develop RAIL-WASP were identified and/or modified, and are as follows:

- Auto Regressive Integrated Moving Average (ARIMA)
- Mixture Density Networks (MDNs)

ARIMA

Auto Regressive Integrated Moving Average (ARIMA) is a method for understanding and forecasting time series data, utilizing the patterns within the past data to predict future values. The method was first introduced in 1970 (Box & Jenkins, 1970) and can be broken down into three primary components, as follows (Brownlee, 2017):

- **AR**, Auto Regression: Utilize dependent relationship between an observation and some number of past observations with defined lag, i.e., an observation is a function of the lagged observations.
- **I**, Integrated: Use an order of differencing, subtracting consecutive observations, to make a series stationary.
- **MA**, Moving Average: Utilize dependency between an observation and the residual errors from a moving average of past (lagged) observations.

Addressing the difference component first, the **I** corresponds to an order of differencing, first, second, etc. to make the series stationary. Differencing is defined by the following equation:

$$Y'_i = Y_i - Y_{i-1} \tag{17}$$

Where

Y'_i = the differenced result for the i^{th} observation
Y_i = the i^{th} observation
Y_{i-1} = the preceding observation

Second order differencing would be to repeat the process, and so on.

The **AR** component is defined by the following equation:

$$Y_i = \alpha + \phi_1 Y_{i-1} + \phi_2 Y_{i-2} + \cdots + \phi_p Y_{i-p} + \epsilon_1 \tag{18}$$

Where

Y_i = the i^{th} observation
Y_{i-p} = p^{th} lagged observation
α = model intercept
ϕ_p = model coefficients
ϵ_1 = error component

Likewise, the **MA** component is defined as follows:

$$Y_i = \alpha + \epsilon_i + \theta_1 \epsilon_{i-1} + \theta_2 \epsilon_{i-2} + \cdots + \theta_q \epsilon_{i-q} \tag{19}$$

Where

Y_i = the i^{th} observation
α = model intercept
θ_q = model coefficients
ϵ_{i-q} = error component

Combining the **AR** and **MA** components, implementing an order of differencing using **I**, and introducing a lag operator (L), the equation for ARIMA(p,d,q) can be represented as follows (Kongcharoen & Kruangpradit, 2013):

$$\left(1-\sum_{j=1}^{p}\phi_j L^j\right)(1-L)^d Y_i = \left(1+\sum_{j=1}^{q}\theta_j L^j\right)\epsilon_i \qquad (20)$$

Where

p = lag order
d = order of differencing
p = order of moving average

When implementing an ARIMA model for forecasting time series, the data must be made at least weakly stationary by finding the order of differencing, d, that achieves this. Note that if $d = 0$ then the data is already stationary, and if $d = 1$, the data trends linearly, $d = 2$, the data trends to the squared power. Then the lag order, p and order of moving average, q must be determined.

The Box-Jenkins approach for determining the optimal values for the ARIMA parameters is an iterative three-step approach as follows (Brownlee, 2017; Croarkin & Tobias, 2013):

- Identification: Use data to identify appropriate model and corresponding parameters that best describe the data.
    - Evaluate stationarity and seasonality checking that mean and variance do not change significantly over time. Make stationary by:
        - Order of differencing, d (avoid over differencing)
        - Transformation
        - De-Trending
    - Identify ARIMA parameters (p, q)
        - Auto Correlation Function (ACF): Evaluates correlation for various lag values. Lag values with the largest ACF that fall outside the 95% confidence limit are candidate lags. The lag at which the ACF trails off is the order of **MA**, q.
        - Partial Auto Correlation Function (PACF): Evaluates correlation for various lag values, not accounting for prior lagged observations. Lag values with the largest PACF that fall outside the 95% confidence limit are candidate lags. The number of lags which cross the PACF confidence interval before it trails off is the order of **AR**, p.
- Estimation: Estimating the coefficients of the model is done through maximum likelihood estimation and is handled in many commercially available statistical software packages.
- Diagnostics/Validation: Validate the model by insuring:
    - Residuals are random, mean of zero and symmetric variance, i.e., white noise.

- No overfitting such that noise is not captured in the forecast and is no more complex than required to achieve similar results, i.e., minimum p, q.

The above approach may suggest multiple configurations of model coefficients. In this case the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) can be calculated, according to the following:

$$AIC = -2log(L) + 2(p + q + k)$$
$$BIC = AIC + (log(T) - 2)(p + q + k) \quad (21)$$

Where

L = likelihood of the data
k = 0 if intercept of model $\alpha = 0$
k = 1 if intercept of model $\alpha \neq 0$

The lower the AIC and/or BIC the better relative fit of the model. Note, the AIC and BIC do not provide a measure of the goodness of fit (accuracy) of the model overall, and are only utilized to choose the best model from a number of candidate models.

In addition, the Ljung-Box test (Ljung & Box, 1978) can be performed to insure that there is an overall randomness of the residuals of the resulting model, by testing for autocorrelation on the residuals for all lags in a range. Since this test is performed for a number of lags (not just at a particular lag) it tests the entire model and is defined as a portmanteau test (evaluating multiple autocorrelations).

Finally, the developed model can be used to forecast future events/values, as well as confidence intervals, based on the developed model. Given the structure of the model, this is done for a defined number of future time (or MGT) steps. As with any forecast, the further out the forecast, the less confident it is. The forecast can be determined using the data, the determined model coefficients and Equation 5.3.

In order to evaluate the effectiveness of the model, as with any statistical model, a training data set can be used to build a model. The model can then be used to forecast to known future data and compared using mean square error, or some other commonly used accuracy metric.

Use of ARIMA in Railroads

ARIMA time series forecasting has been used extensively in statistics and econometrics to understand past relationships and forecasting future events. It has also been used for spatial data relating ground penetrating radar data to highway thickness prediction (Attoh-Okine, 1994).

Limited ARIMA applications in the railroad industry have focused around operations and goods demand, along with other machine learning techniques.

In particular, passenger demand for railways and metros was modelled and forecast using ARIMA (Milenkovic, Svadlenka, Melichar, & Bojovic, 2016) and forecast demand of grain car loadings (Babcock, Lu, & Norton, 1999). These studies note the challenges of missing information regarding explanatory variables, and the ability of ARIMA to utilize historic data to make reasonable forecasts. However, certain volatility cannot enter the forecasts.

Several researchers have evaluated railway component degradation and forecast time to failure. These includes a real time ARIMA model for identifying railway car brake failures (Pacella & Anglani, 2008), and predicting the failure of switch machines using ARIMA analysis of the current draw of the switch-point machine (Abbasnejad & Mirabadi, 2017).

Relative to the rail wear research, one study evaluated wheel flange wear and wheel diameter and forecast replacement dates using an ARIMA model with very accurate results for short term forecasts (L. Wang, Zhao, Xu, & Na, 2015).

Sample Application of ARIMA for Rail Wear

Before implementing an ARIMA process on the rail wear data used in this research, it is important to understand that rail wear measurements are captured at varying intervals in time. In addition, the amount of traffic fluctuates over time, thus, it is more appropriate to correlate rail wear measurements with traffic density (MGT). Most inspection intervals result in up to 30 measurements over the lifetime of the rail. ARIMA processes requires constant intervals and work best with at least 50 measurements. Finally, rail wear is expected to increase linearly with accumulated tonnage. Thus, the series will be non-stationary by definition.

In order to satisfy the above requirements, the series is made stationary by removing the linear trend. This can be achieved using ordinary least squares regression techniques. This also allows for capturing the dominant wear rate for use later. Also, in order to achieve constant interval and have enough points to apply ARIMA, up sampling through piecewise linear interpolation (USPLI) can be employed. It is important to choose an interpolation interval that satisfies the requirements for employing the ARIMA process (at least 50 samples).

The above steps were implemented using R open source statistical software (R Core Team, 2019). Considering up-sampled and aligned wear data, and choosing one sequence at a milepost as an example (226.8842), the best-fit ARIMA model can be determined.

The measurement data by accumulated MGT[4] on the rail is shown in Figure 15. This data encompasses 11 measurements from 575 to 911 accumulated MGT.

[4] Accumulated MGT was determined based on the installation date of the rail (1997) and the MGT accumulated each year (obtained from annual MGT data) to the date of inspection.

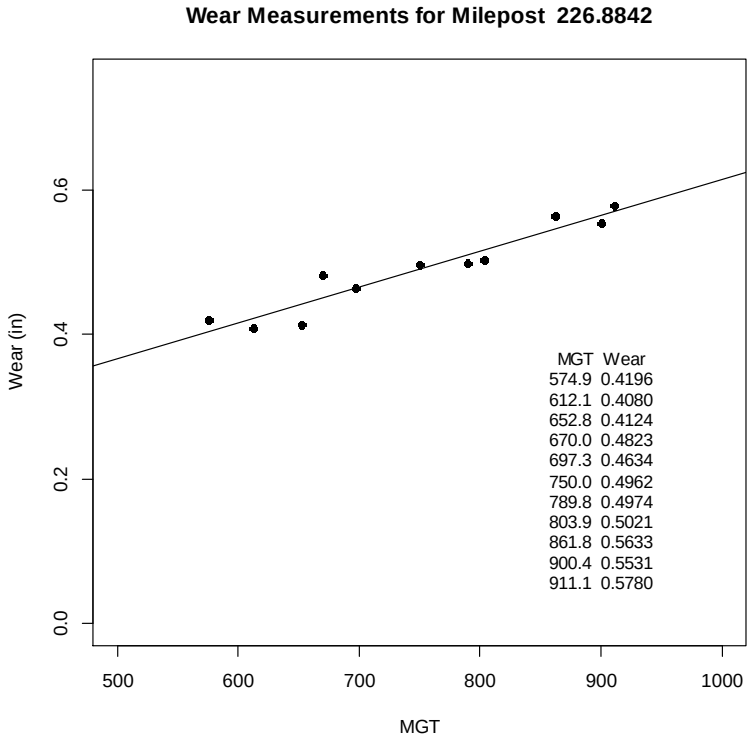


Figure 15 Wear measurement data for MP 226.8842.

The first step in developing the ARIMA model for this milepost is to make the series stationary by removing the linear trend of the series. Note that this is accomplished by subtracting out the least squares fit. The resulting stationary data is then up-sampled to provide a uniform interval between samples as well as provide a sufficient number of samples for effective ARIMA modeling. In this case a delta MGT of 5 was chosen to insure at least one interpolated point between successive inspections.

In order to determine the AR and MA components, the ACF and PACF plots are evaluated

(See
Figure 16). The PACF plot shows spikes outside the 95% confidence band of the correlation coefficient at lags 1 and 2, with further lags being insignificant. Therefore, the order of the AR(*p*) component is AR(2).

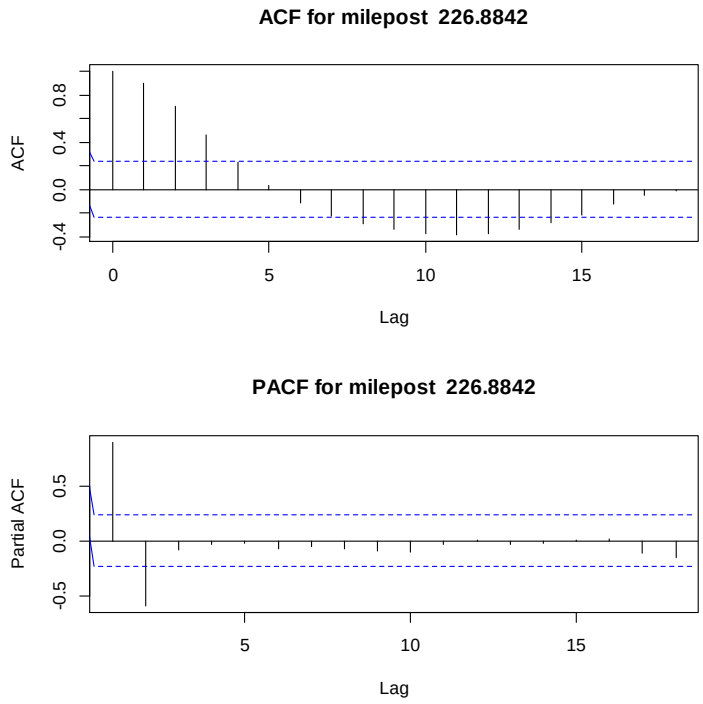


Figure 16 PACF and ACF plots for MP 226.8842.

Considering the ACF, the data appears to trend increasingly downward after lag 0, and significantly after lag 1. Possible lags for the order of MA are 0 through 3, as these lags have values outside the 95% confidence band before trailing off. Thus, following the Box-Jenkins approach, ARIMA[5] models of $p = 2$ and $q = 0$ to 3, can be considered. The most relevant model can be determined by checking the Akaike Information Criterion (AIC) and/or Bayesian Information Criterion (BIC). Using the R function `Arima`(series, order = c(p, d, q)) to determine the parameters and AIC, BIC of a prescribed model. Table 4 shows the AIC, BIC for candidate models using R. This table suggests a model of ARIMA(2,0,2) based on AIC and ARIMA(2,0,0) based on BIC, recognizing that the smallest values correspond to the best fit.

Table 4 ARIMA Model Results for MP 226.8842

p	d	q	AIC	BIC
2	0	0	-579.45	-570.57
2	0	1	-578.38	-567.28
2	0	2	-580.96	-569.86
2	0	3	-577.04	-561.50

[5] Note that the I in ARIMA is left in intentionally for discussion later. Since the trend was removed and the data is stationary, it is expected that the order of I will be zero, I(d) = I(0).

However, R also has a function `auto.arima` which iterates through candidate models (by evaluating ACF and PACF) and returns the best fit model. This function specified ARIMA(2,0,2) for this example based on AIC. The value of this approach is that it allows for checking many segments rapidly as opposed to evaluating ACF and PACF and iteratively determining AIC, BIC for each model.

Evaluating the Ljung-Box test (using the `LjungBoxTest` function available in R) at various lag values results in Figure 17.

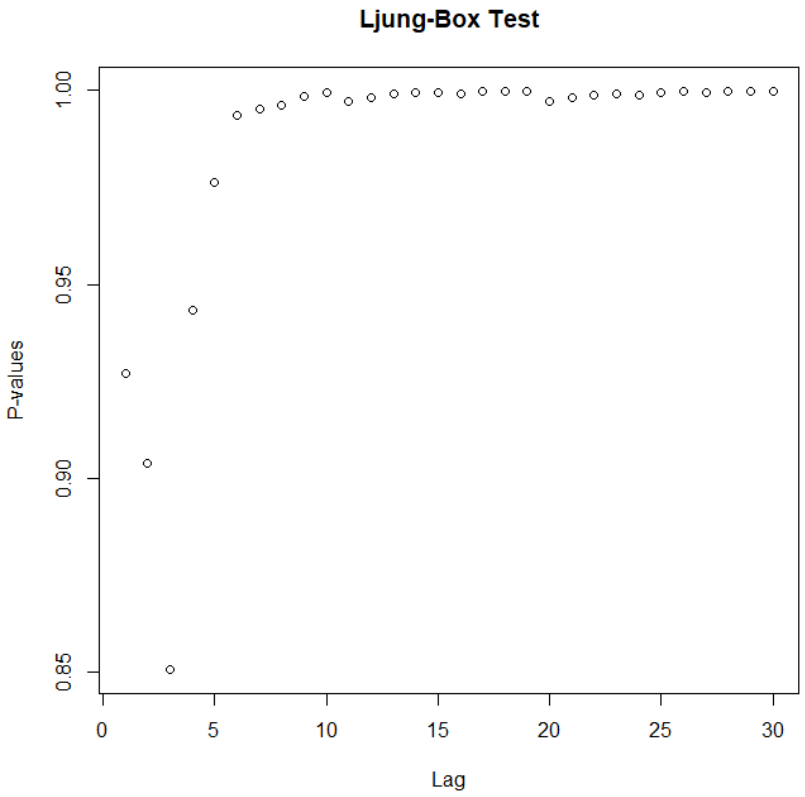


Figure 17 Ljung-Box test of ARIMA model.

Since the p-values are all well above 0.05, the model shows randomness for lag values up to 30.

The final check is that the residuals are normally distributed, that is they appear as "white noise". Figure 18 shows a plot of the residuals, as well as the ACF and distribution of residuals. This figure shows that the residual fall within the 95% confidence band and are insignificant. The distribution is skewed normal and has a zero mean.

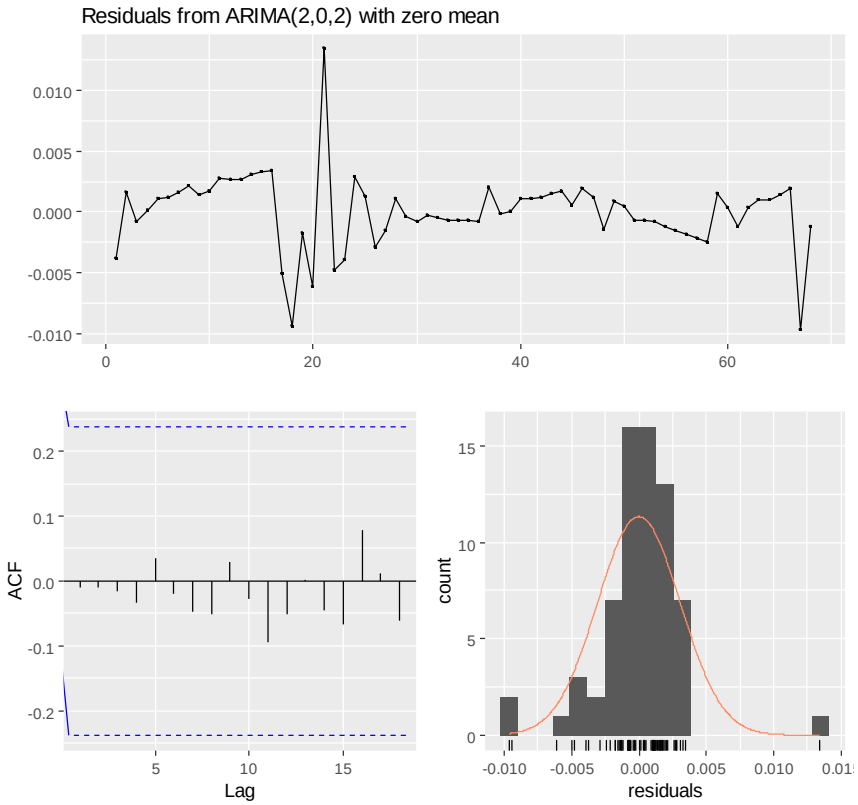


Figure 18 ARIMA(2,0,2) residual analysis for MP 226.8842.

Now that the model is established, the accuracy can be determined for its fit against the historic data. The resulting ARIMA(2,0,2) model has two coefficients each for the AR and MA components, and an intercept of zero, as follows:

- AR
  - ϕ_1 = 1.4133
  - ϕ_2 = -0.5905
- MA
  - θ_1 = 0.3414
  - θ_2 = 0.3301

These coefficients can be used along with equation 5.4 to model the original series. This is shown in Figure 19, along with the original measurements, as well as the up-sampled data in between measurements[6]. The resulting accuracy of the fit can be defined by a number of metrics. The metric chosen here is root mean square error (RMSE), as this provides the same units as the measurements themselves. The equation for RMSE is as follows:

[6] Note that this is for the data with linear trend removed which must be added back in when defining the model for forecasting.

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N}\left(\hat{Y}_i - Y_i\right)^2}{N}} \tag{22}$$

The RMSE for the ARIMA(2,0,2) model for this data is 0.00305 inches.

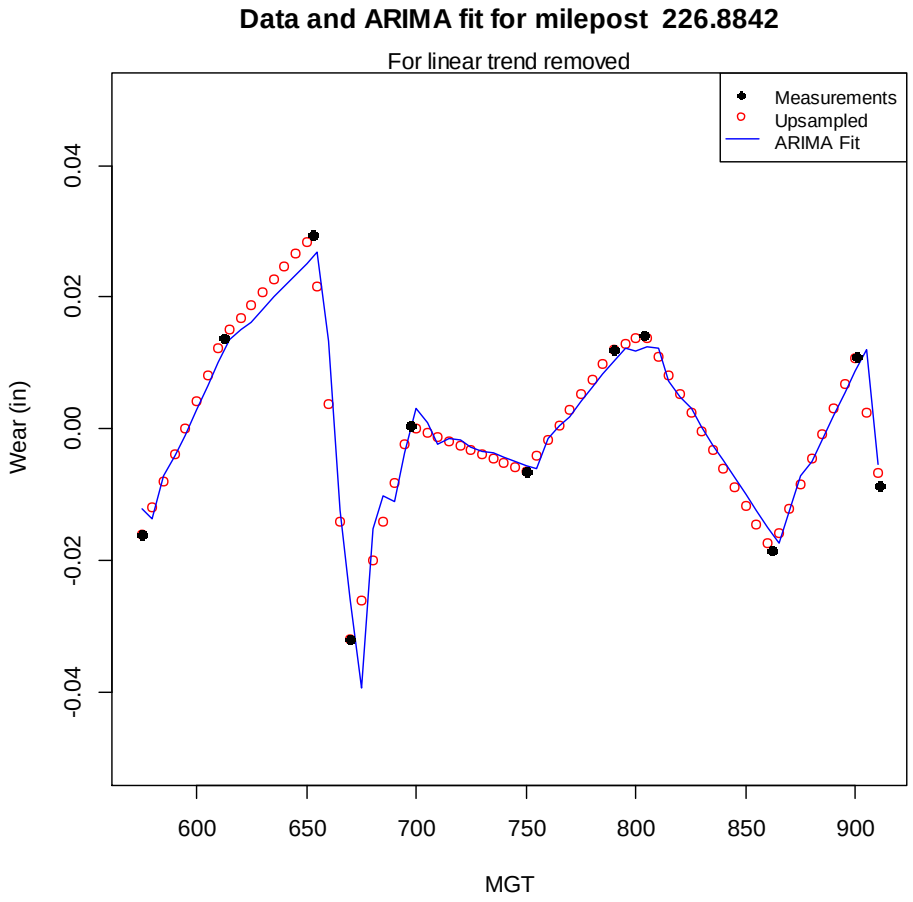


Figure 19 ARIMA(2,0,2) model with original and up-sampled data for MP 226.8842.

Equation 5.4 and the corresponding coefficients for the ARIMA(2,0,2) model can be used to forecast forward a number of defined steps. Each step corresponding to 5 MGT, as this is the interval defined in the model. This can be done simply using the `predict` function in R. Figure 20 shows the measurement data along with its linear trend and ARIMA fit (with linear trend added back in). In addition, the forecast is shown out to 1,100 MGT (along with the 95% confidence interval). It can be seen that the forecast from ARIMA converges to the linear least squares fit rapidly (50-75 MGT). This indicates a value in using ARIMA for short term forecasts, and least squares fit for longer term forecasts.

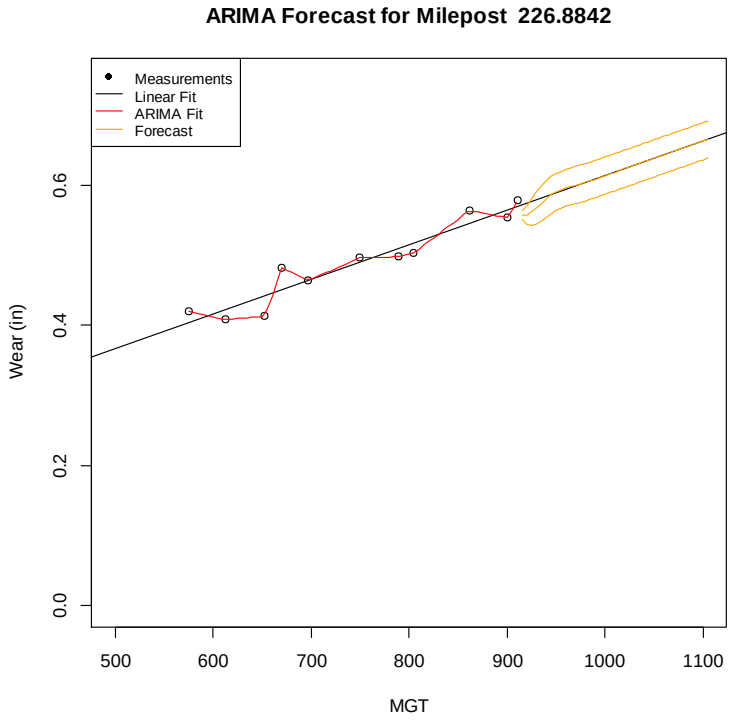


Figure 20 ARIMA Forecast for MP 226.8842.

MDNs

Traditionally, neural networks have been used in regression analyses to utilize input variables to predict target variables; that is provide a likely output given a series of inputs. This approach does not take into account the variability of the target for a particular set of inputs. In fact, the target variability is not constant and can be a function of the inputs as well.

Mixture Density Networks (MDNs), introduced in 1994, allow for the determination of conditional probability distributions associated with supervised learning of multiple target values as a function of input variables to account for the variability of the target value (Bishop, 1994). Simply put, an MDN is a combination of a feed-forward artificial neural network and a mixture of probability distributions, such that the conditional probability is output, for a vector of given inputs. This is quite important when predicting just the mean is insufficient. Many real world problems require more knowledge than just the mean to make proper decisions. Thus, determining the conditional probability distribution allows for understanding a range of answers and their associated probabilities.

The derivation of the MDN equations and approach follows (Bishop, 2006). Starting with the output, a mixture of conditional probability density would take the form:

$$p(Y = y|X = x) = \sum_{i=1}^{m} \alpha_i(x) D_i(y|\mu_i(x), \sigma_i^2(x)) \qquad (23)$$

Where,

X = a vector of inputs x
Y = some output vector of observations y conditioned on x
$\alpha_i(x)$ = i^{th} of m mixing coefficients as a function of X
$D_i(y|\mu_i(x), \sigma_i^2(x))$ = i^{th} of m conditional density kernel functions
$\mu_i(x)$ = conditional mean for a Gaussian distribution
$\sigma_i^2(x)$ = conditional variance for a Gaussian distribution

The mixing coefficients must satisfy the following constraints to ensure that the resulting conditional probability density function integrates to 1, and the variance for a Gaussian distribution must be greater than or equal to zero:

$$\sum_{i=1}^{m} \alpha_i(x) = 1$$
$$0 \leq \alpha_i(x) \leq 1$$
$$\sigma_i^2(x) \geq 0 \tag{24}$$

It should be noted that the conditional density kernel function can be any probability density function depending on the problem being solved. If the distribution has one, two, three or even four parameters, then the conditional density kernel function will be derived appropriately. However, successfully parametrizing the model with the proper number of Gaussian components (and associated mixture coefficients) will provide a good estimation for most probability density functions.

Also, note that the conditional variance is a function of the input vector X, indicating a heteroscedastic[7] model, unlike the least squares method where the variance would be constant for all X.

Considering a conditional density kernel function of Gaussian form results in the following:

$$D_i(y|\mu_i(x), \sigma_i^2(x)) = \frac{1}{\sqrt{2\pi\sigma_i^2(x)}} exp\left(-\frac{(y-\mu_i(x))^2}{2\sigma_i^2(x)}\right) \tag{25}$$

The use of a feed forward artificial neural network lends itself extremely well for determining the components of the conditional probability distribution, namely, the mixing coefficients ($\alpha_i(x)$) and the probability density function parameters ($\mu_i(x), \sigma_i^2(x)$).

A generalized representation of a mixture density network (MDN) is shown in Figure 21. This network contains the following:

[7] Heteroscedasticity is defined as sub-populations of a larger data set having differing variances.

- n inputs, X = (x_1, x_2, … x_{n-1}, x_n)
- i x j hidden layers/nodes, $H = \begin{pmatrix} h_{11} & \cdots & h_{i1} \\ \vdots & \ddots & \vdots \\ h_{1j} & \cdots & h_{ij} \end{pmatrix}$
- m x 3 outputs, $\alpha_{1\ldots m}(x)$, $\mu_{1\ldots m}(x)$, $\sigma^2_{1\ldots m}(x)$

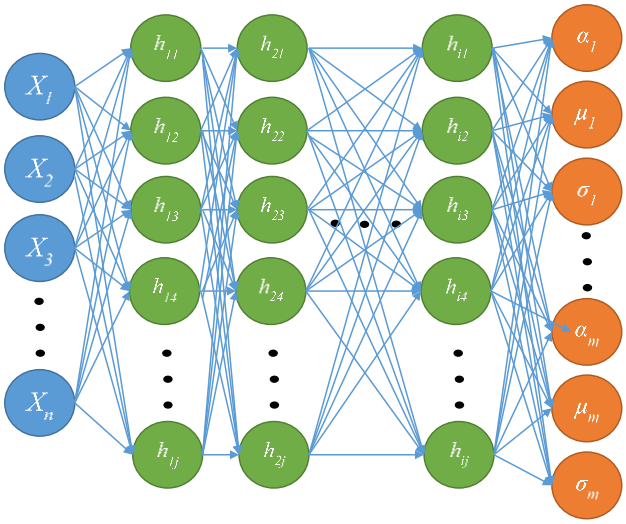


Figure 21 Generalized mixture density network (MDN).

The goal is to maximize the likelihood ($\mathcal{L}$) given the input data set, (with *c* observations) such that:

$$\mathcal{L} = \prod_{q=1}^{c} p(y_q, x_q) \tag{26}$$

A more convenient and stable mathematical (and programmatic) approach is to minimize the error, defined by taking the negative log of the likelihood as follows:

$$E = -\ln(\mathcal{L}) \tag{27}$$

For typical ANN implementation an error function (sometimes termed loss or cost function) is typically defined by sum of squares. A least squares approach is employed in the ANN to minimize the error using a standard optimization procedure (often gradient descent is employed). However, for the MDN approach, an error function equivalent to the conditional probability mixture density function is employed as follows:

$$E(w) = \sum_{q=1}^{c} -ln\left[\sum_{i}^{m} \alpha_i(x_q, w) D_i(y_q | \mu_i(x_q, w), \sigma_i^2(x_q, w))\right] \tag{28}$$

Where, *w* are the weights of the neural network and introduced here, and *D* is the Gaussian or some other distribution with associated parameters.

Minimization of the error function can be achieved by taking the partial derivatives of the error function with respect to the weights (*w*) in the neural network. As this is handled in most open source software packages for implementation, the derivation is not included here.

The MDN can be implemented by employing the following steps:

- Choose the input variables, *X*
- Choose the number of layers (*i*) and nodes (*j*)
- Specify an activation function for the nodes (that function which processes inputs to outputs at each node). The constraints showed that non-negative values are required, thus, the activation function is typically:
    - tanh
    - Exponential Linear Unit, $ELU(x) = \begin{cases} x & x \geq 0 \\ \exp(x) - 1 & x < 0 \end{cases}$
- Select a number of mixture components (*m*)
- Select an appropriate error function (Gaussian or some other distribution)
- Define the number of outputs
- Impose constraints on mixing coefficients
    - Use *softmax* function which is a normalized exponential function to force values to be between zero and 1 and sum to 1:

$$\propto_i = \frac{\exp(z_i^{\propto})}{\sum_j z_j^{\propto}}$$

- Select an optimization routine
    - Gradient descent; iterative process to determine parameters that minimize error (or cost, loss) function
- Train the MDN to determine the outputs , $\alpha_{1\ldots m}(x)$, $\mu_{1\ldots m}(x)$, $\sigma^2_{1\ldots m}(x)$

Once the network is trained predictions for the expected target values ($\mathbb{E}$, VAR) can be determined from the following:

$$\mathbb{E}(Y|X) = \int Yp(Y|X)dy = \sum_i^m \alpha_i(X)\mu_i(X)$$

$$VAR(X) = \sum_{j=1}^{m} \alpha_j(X)\left[\sigma_j^2(X) + \left\|\mu_j(X) - \sum_{i=1}^{m} \alpha_i(X)\mu_i(X)\right\|^2\right] \tag{29}$$

MDN Research and Applications

While mixture density networks have been around since the 1990's, they have only recently come to pass as a viable stochastic tool for practical implementation. Artificial neural networks have

been increasingly utilized, to include use in the railway industry. However, since the focus in this section is MDNs, the use of ANNs in railways is not included herein. To date, no utilization or application of MDNs in railways has been documented. Therefore, this section introduces a select few non-railroad applications that correlate to railway degradation analysis, and the research conducted herein.

Upon introduction in 1994, MDNs gained popularity in time series analysis, whereby it was noted that predicting only the conditional mean can result in poor, if not false, predictions, and that knowing the conditional distribution as a function of time is extremely valuable (Husmeier & Taylor, 1997). Initial focus was on improving the computational efficiencies if MDNs due to the lack of available computing power, readily available today (Husmeier & Taylor, 1998; Nikolaev, Tino, & Smirnov, 2013).

MDNs subsequently gained popularity in econometrics and financial forecasting to utilize the prediction of variance, along with the mean, to measure volatility (Schittenkopf, Dorffner, & Dockner, 1999). This is analogous to risk management in financial time series forecasting (Miazhynskaia, Dorffner, & Dockner, 2003). The ability to model uncertainty allows for more informed decision making and control of assets (Herzallah & Lowe, 2004). Taking advantage of the ability to inherently model non-linearity in inflation with additional smoothing has led to improved prediction for highly heteroscedastic data (Villani, Kohn, & Giordani, 2009).

Much attention has been paid to using MDNs for speech synthesis to account for subtle changes in inputs (Siri-Team, 2017; Zen & Senior, 2015), as well as for meteorological predictions (Cannon, 2012; Carney, Dowling, & Lee, 2005; Men, Yee, Lien, Wen, & Chen, 2016).

Finally, one research effort focused on using MDNs for developing conditional distributions when observations are missing or even entire independent variables are unknown (Brouwer & Pedrycz, 2003).

In particular, the parallel research associated with asset management, degradation forecasting and unknown independent variables is of particular import to this research effort.

Sample Application of MDNs for Head Wear Rates

Implementing an MDN can be achieved through open source software. For this sample application, Google's machine learning library, Tensorflow[8] (Martin et al., 2015) was utilized as an application programming interface along with the Python programming language, compiled as a Jupyter Notebook. The head wear rate data for tangent rail was utilized to allow for comparison. As noted, the head wear rate data shows a distinct distribution around a mean for common independent variables. This is due to the nonlinear relationship and interaction between the independent variables, as well as the presence of several unknown or unidentified independent variables. Thus, as pointed out previously, just predicting the average in such a situation leads to poor forecasting, and ultimately improperly supported decisions.

[8] TensorFlow, the TensorFlow logo and any related marks are trademarks of Google Inc.

In order to evaluate this process, three mixture density network models were developed to predict a conditional density distribution for the known independent variables. Specifically, the MDNs were trained with the tangent head rail wear rate data to produce a conditional mean and variance as outputs, using the six known independent variable as inputs (year installed, grade, rail weight, distance to lubricator, last annual MGT, and speed). The following were specified for the MDNs (See Table 5):

Table 5 MDN Specifics

	MDN – 1	MDN - 2	MDN - 3
Inputs	6	6	6
Learning Iterations	7500	7500	10000
Batch Size	50	50	50
Learning Rate	0.0001	0.0001	0.0001
Layers	15	15	15
Neurons	120	120	120
Mixture Components	1	1	3
Outputs	2	2	9
Neuron Activation Function	tanh	tanh	tanh
Standard Deviation Constraint	ELU	ELU	ELU
Distribution/Error Function	Gaussian	Laplace	Gaussian
Mixing Coefficient Normalization *	N/A	N/A	*softmax*
Optimization **	Adam	Adam	Adam

* A *softmax* function for normalizing mixing coefficients is utilized for MDNs with more than one distribution in the mixture model. This as a normalized exponential function that allows for meeting the constraint that the mixing coefficients sum to 1. However, since there is only 1 mixing coefficient for MDN -1 and MDN - 2, the scale factor is 1, and *softmax* is not required.
** Classical stochastic gradient descent optimization is sometimes computationally inefficient. AdamOptmizer (Kingma & Ba, 2015) was utilized for the determination of the network rates, with the learning rate (step size for which weights are updated during training) and error function as primary inputs. This optimization technique is efficient and robust, and takes advantage of training data entering the equations in batches (as opposed to one large matrix) for each iteration. In addition, the Laplace distribution has a differential discontinuity at the median, and thus classical gradient descent would fail. AdamOptimizer handles this discontinuity effectively.

Code was written to implement the three MDNs, each of which are discussed next.

MDN – 1: One Gaussian Distribution

A mixture density network with only one conditional density distribution, while not truly a mixture model, offers the ability to provide the conditional mean and variance, unlike artificial neural networks that give only the mean, thus resulting in important additional information regarding the data. The head wear rate data for tangent rail was used to train the network and result in a conditional mean and variance for a range of the six independent variables.
Training the model and observing the cumulative loss (error function) results in Figure 22. This figure shows the model starting to converge after 5,000 iterations. Note the spikes in the loss function which are a manifestation of the AdamOtimizer utilizing batch processing. Larger batches would result in less spikes, however more iterations would be required, and thus, more computational time.

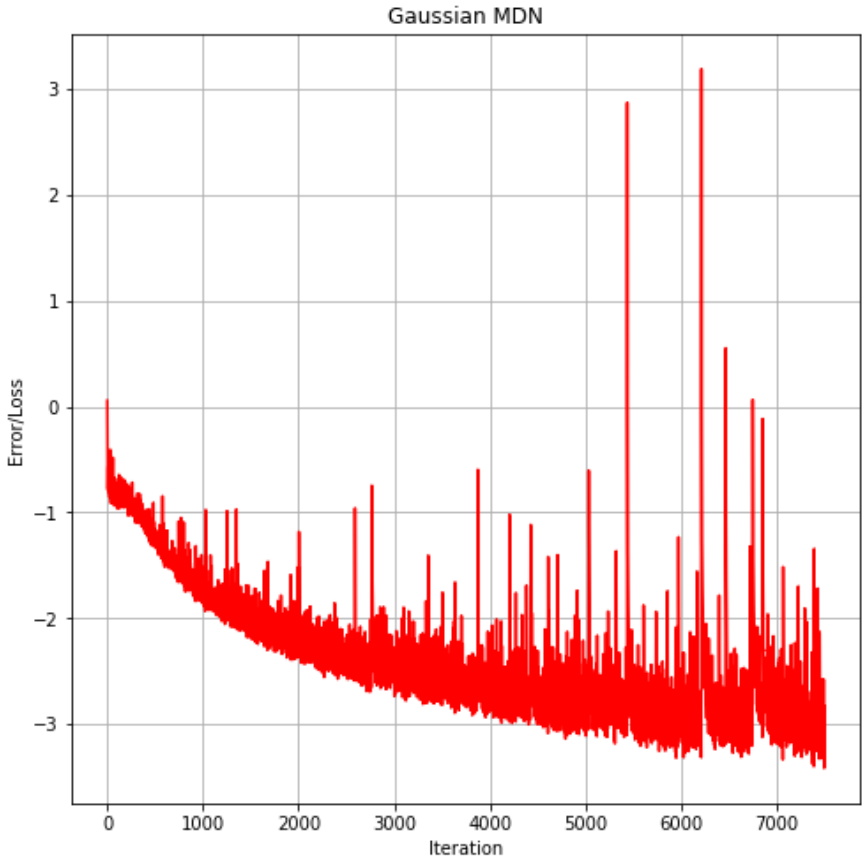


Figure 22 Error/Loss by iteration for single Gaussian MDN.

The resulting conditional mean head wear rates as compared to the actual head wear rates of the model are shown in Figure 23.

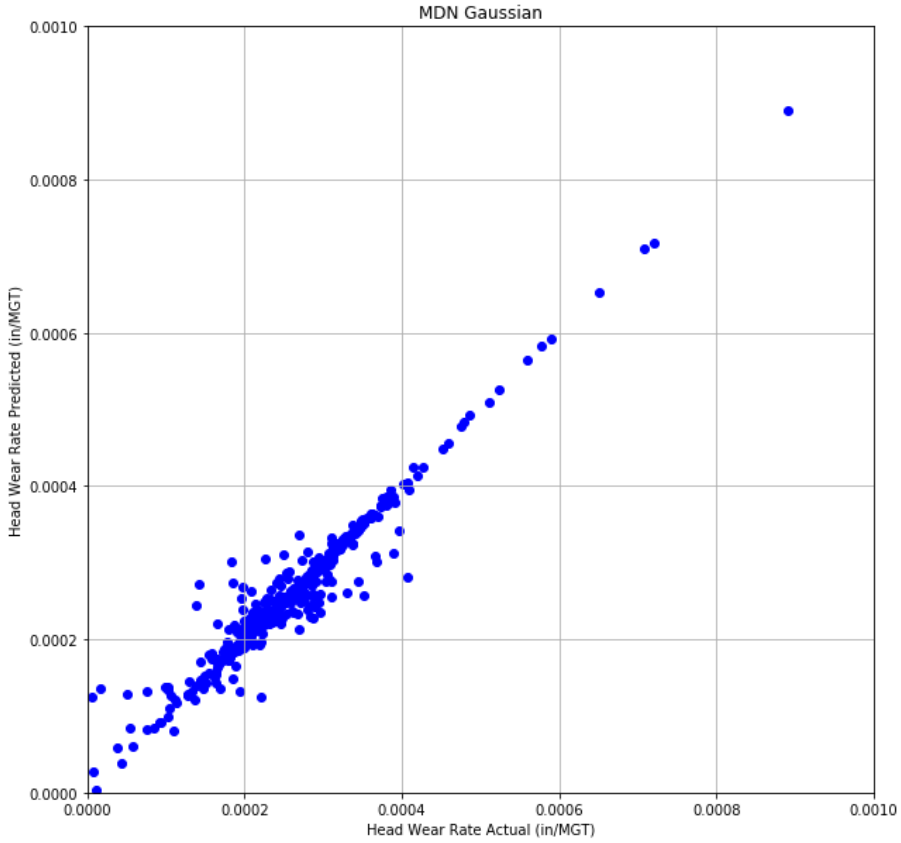


Figure 23 Predicted versus actual head wear rate for one Gaussian distribution.

The power of the MDN is providing the conditional variance. Figure 24 shows the predicted versus actual head wear rates, along with the boundaries imposed by +/- one conditional standard deviation. Note that the standard deviation is not a constant (as would be with a typical regression model), rather it is determined for each instance. This figure shows the expected variation around the expected mean. Thus, a straight line could result at "perfect" correlation of prediction/actuals, however the probability along that line would not be 50%, it would vary.

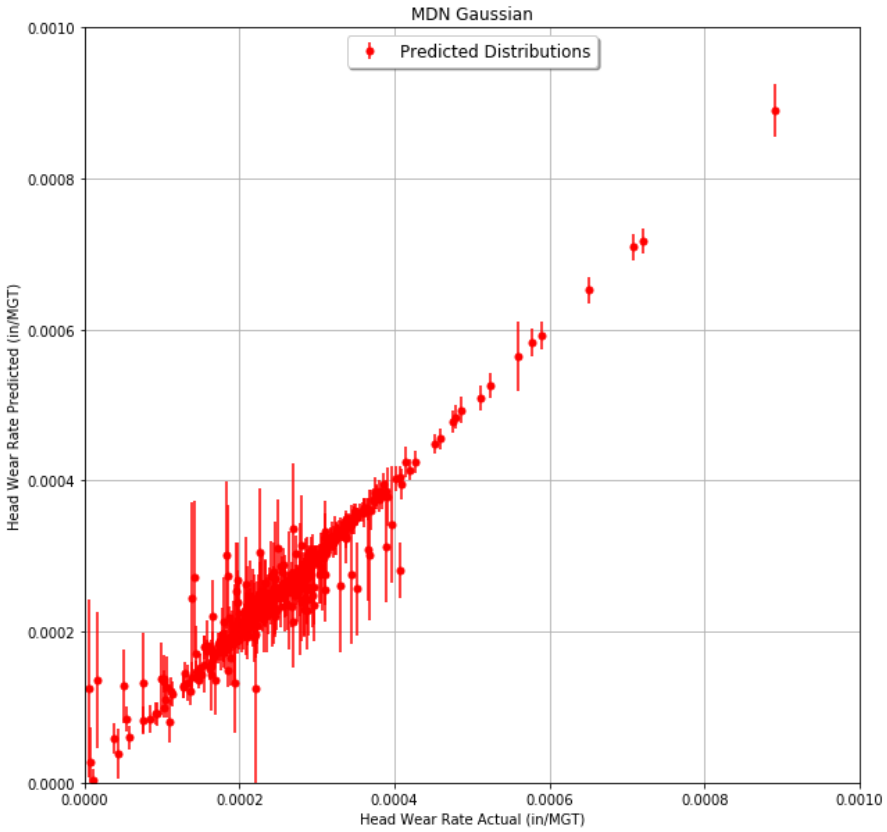


Figure 24 Predicted versus actual head wear rate for one Gaussian distribution, with one standard deviation.

MDN – 2: One Laplace Distribution

A second mixture density network was created using a Laplace distribution for the error/loss function. Thus, the outputs of the model are the location and shape parameter of the Laplace distribution, as opposed to the mean and standard deviation of the Gaussian distribution

Figure 25 shows the error by iteration. Note that the model is just starting to converge at 7,500 and shows the variation associated with the batch processing methodology, however does not result in very large divergence at any given iteration.

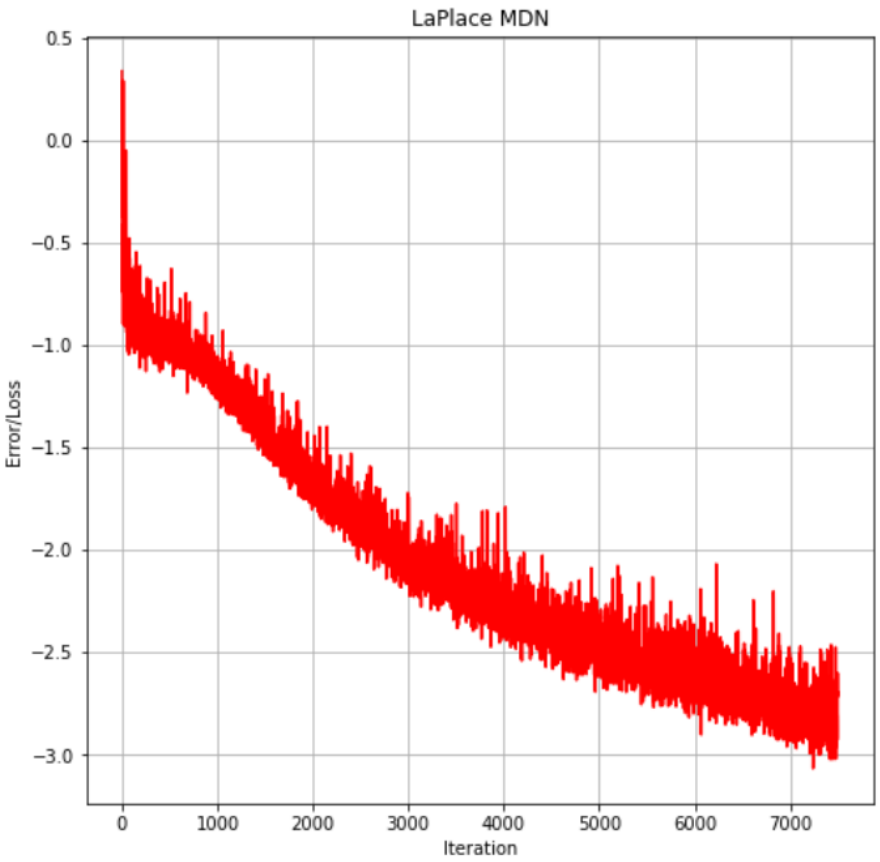


Figure 25 Error/Loss by iteration for single Laplace MDN.

The resulting comparison of predicted head wear rates to actual head wear rates, along with the extent of the conditional distribution (shown as +/- one shape factor for convenience) is provided in Figure 26.

While similar to the single Gaussian MDN, there are some significant differences. Mainly, the extreme locations (very small and very large wear rates) show larger variation (larger shape factors), indicating a wide distribution, which would result in a large range of predicted probabilities. The distributions near the median wear rate are very similar.

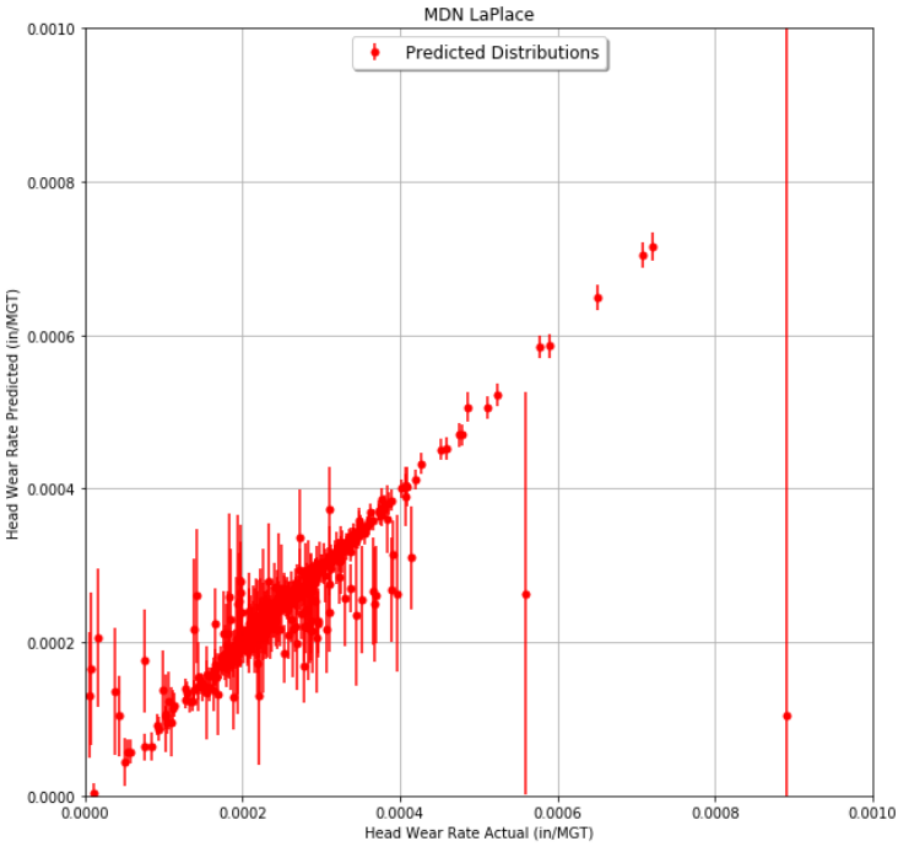


Figure 26 Predicted versus actual head wear rate for one Laplace distribution, with one shape factor.

MDN – 3: Three Gaussian Distributions

The final sample MDN takes advantage of the mixture model and is a three Gaussian distribution MDN. This results in a weighting factor, mean, and standard deviation output for each of the three mixture components, thus the MDN has nine outputs as opposed to two.

The error/loss during model training is shown in Figure 27. This figure shows convergence at around 7,500 iterations. However, there is a discontinuity at 8500 iterations, which quickly return to convergence.

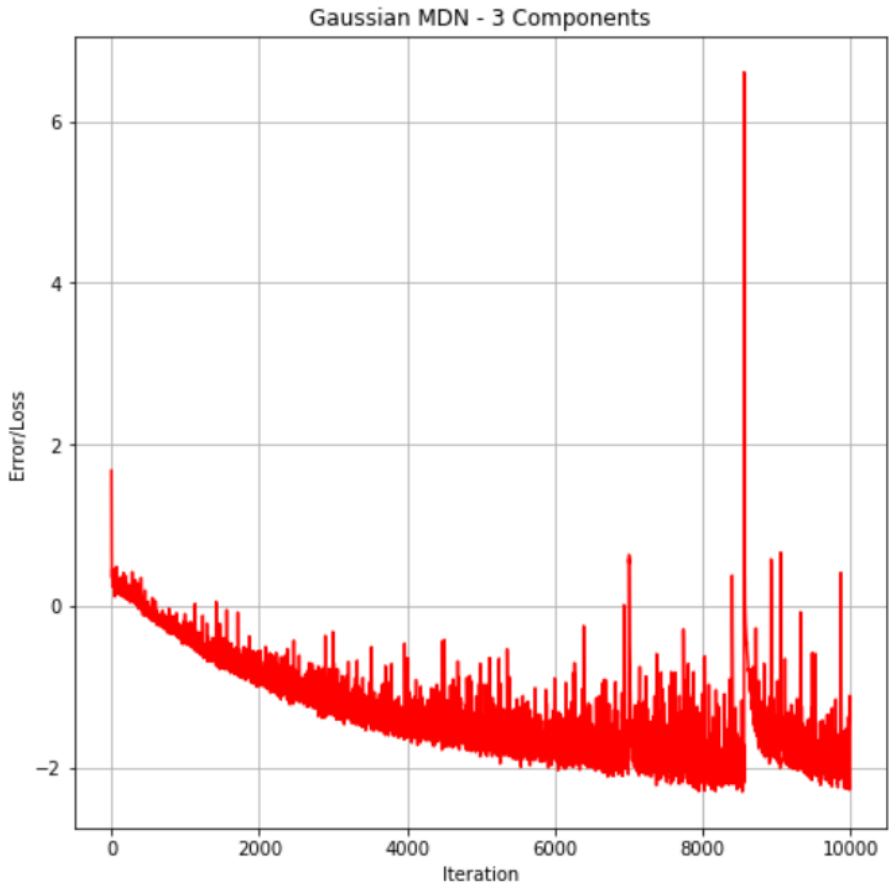


Figure 27 Error/Loss by iteration for three Gaussian MDN.

The predicted versus actual head wear rates, along with +/- one standard deviation for the three Gaussian mixture model is shown in Figure 28. This model appears to capture the salient features of the previous two models, larger distributions at extreme locations and significant variability around the mean. Note that this plot was developed using the mixture distribution as defined in Equation 5.14, which develops a weighted mean and standard deviation based on likelihood.

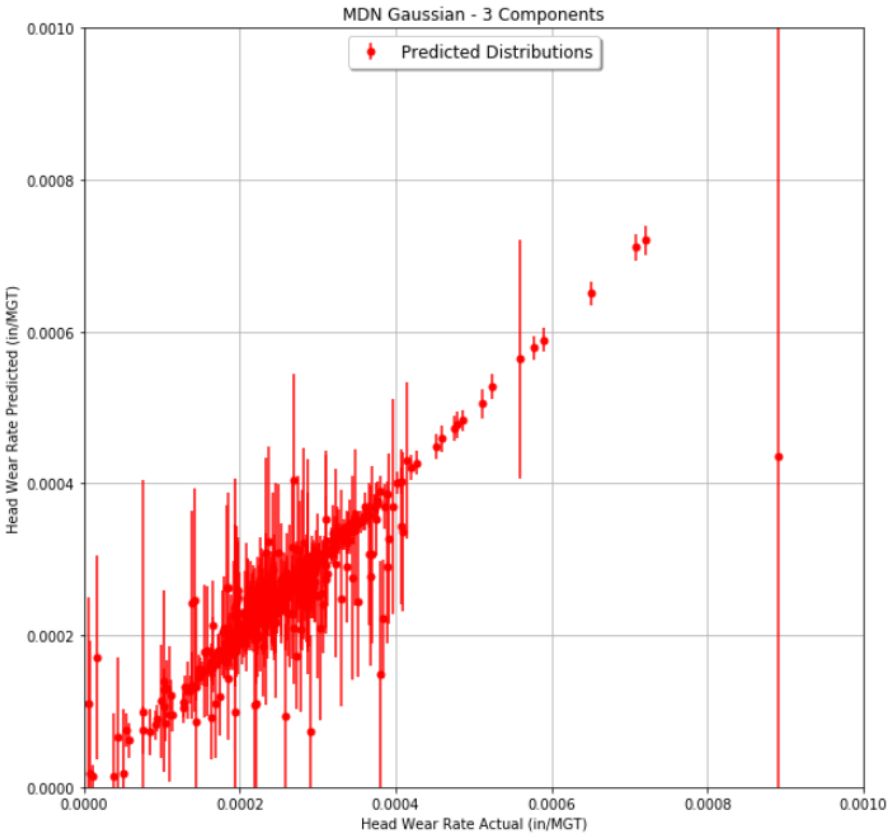


Figure 28 Predicted versus actual head wear rate for three Gaussian distributions, with one standard deviation.

Discussion

The three sample MDNs showed some overall consistent features as well as the ability to show variation in the datasets. Comparing the three MDNs, the sum square of errors (SSE) and negative log likelihood (NLL) were calculated for the predictions of each model, and are shown in Table 6. This table shows that the single Gaussian MDN provides the lowest SSE, and the three Gaussian MDN results in the lowest NLL.

Table 6 Comparison of MDNs

	SSE	NLL
MDN – 1: 1 Gaussian	2.66×10^{-7}	22.22
MDN – 2: 1 Laplace	1.12×10^{-6}	-4.76
MDN – 3: 3 Gaussians	7.15×10^{-7}	-9.47

To gain a further understanding,
Figure 29 shows the resulting conditional distributions, and their variation, for three results of input head wear rates (for varying independent variable inputs). Each plot shows the MDN single Gaussian distribution, MDN single Laplace distribution, 3 MDN Gaussian distribution components and the consolidated Gaussian distribution.
The first plot shows the single Gaussian and Laplace results are nearly identical and that the three components of the Gaussian mixture are nearly identical and equally weighted. However, the 3 Gaussian mixture model results in a broader and more varied result.

The second plot shows that the single Gaussian, Laplace and Gaussian mixture composite are very similar, and the three components of the Gaussian mixture are similar to one another.

The third plot shows that there can be distinct variation between the three MDNs. The Laplace is predicting a significantly varied distribution pushed to the right, while the single Gaussian is more narrow and centered, and the mixture Gaussian is very broad and to the left. Thus, in this instance, three distinct distributions are predicted.

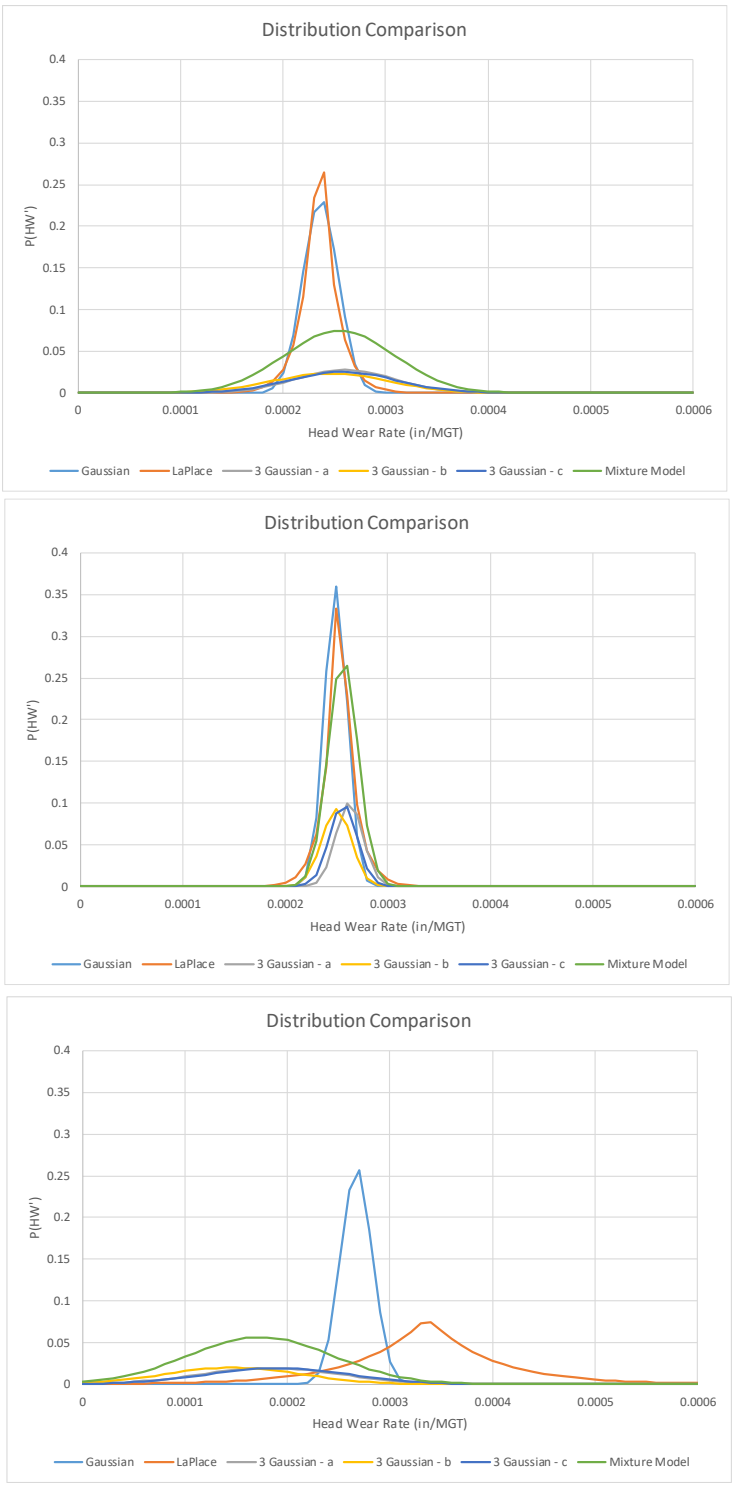


Figure 29 MDN conditional distribution comparison at three different data inputs.

For the third plot in

Figure 29, the distribution parameters for each model are shown in Table 7, along with the actual head wear rate. This table shows that the single Gaussian mean is nearest the actual value.

Table 7 Distribution Parameters for Three Models at Disparate Location

	1 Gaussian	1 LaPlace	3 Gaussian - a	3 Gaussian - b	3 Gaussian - c	3 Gaussian
Actual	0.000273	0.000273	0.000273	0.000273	0.000273	0.000273
Mu	0.000267	0.000336	0.000184	0.000146	0.000190	0.000173
Sigma	1.5E-05	6.3E-05	6.8E-05	6.6E-05	7.0E-05	7.1E-05
Alpha			0.333	0.331	0.336	1

Mixture density networks are very powerful for this type of analysis, as they allow for determining a likely range of wear rate predictions, based on any number of independent input variables.

RAIL-WASP Model

The goal of forecasting models is to develop a relationship between input variables and a target variable such that the model can evaluate changes to the input variables and determine the effect on the target output variable. Linear regression and other machine learning techniques have historically been used to develop simple and complex models, whereby, the most likely target is predicted. As has been shown, this can lead to poor decisions if the variance within the prediction is not evaluated. The model developed herein utilizes the stochastic processes presented in the previous section in a hierarchical stochastic fashion to provide detailed information regarding rail wear based on historical data.

The ARIMA process is modified and employed on an individual segment with an evaluation as to its effectiveness (validation), along with an example forecast. This is followed by a more general understanding of the rail wear rates and their variance (particularly attributed to unknown independent variables) utilizing mixture density networks (MDNs), which can be utilized to evaluate the instantaneous state of the rail and it wear rate. Finally, a classification mechanism is presented for evaluating the current condition of the rail with respect to wear.

The stochastic processes are bundled into a hierarchical model to develop rail wear forecasts, as well as classify the condition of the rail for every rail segment.

Rail Wear Determination for a Rail Segment

The concept of utilizing ARIMA modeling for forecasting rail wear was introduced in 0, along with an example result at one milepost location. As the EDA showed, the wear (and wear rate) will change longitudinally along the track based on many factors. This section presents the analysis of consecutive locations (a rail segment) to further evaluate and validate the ARIMA processing methodology for rail wear, and incorporation into a next generation rail wear model.
Consecutive rail wear measurements (in time, MGT) are aligned and up-sampled to provide contiguous, multiple measurements every 5.28 feet for 1,600 feet (306 locations). Eleven

inspections were performed over a span of six years (from 2007 to 2013). This results in eleven data points for 306 locations that can be treated as a time series and processed using ARIMA methods. Note that this rail segment is for head wear (136RE rail installed in 1997) in a low rail from the leading tangent, through an entry spiral to the full body of a 2 degree curve, and through the exit spiral.

The ARIMA modeling approach is summarized in the flowchart presented in Figure 30.

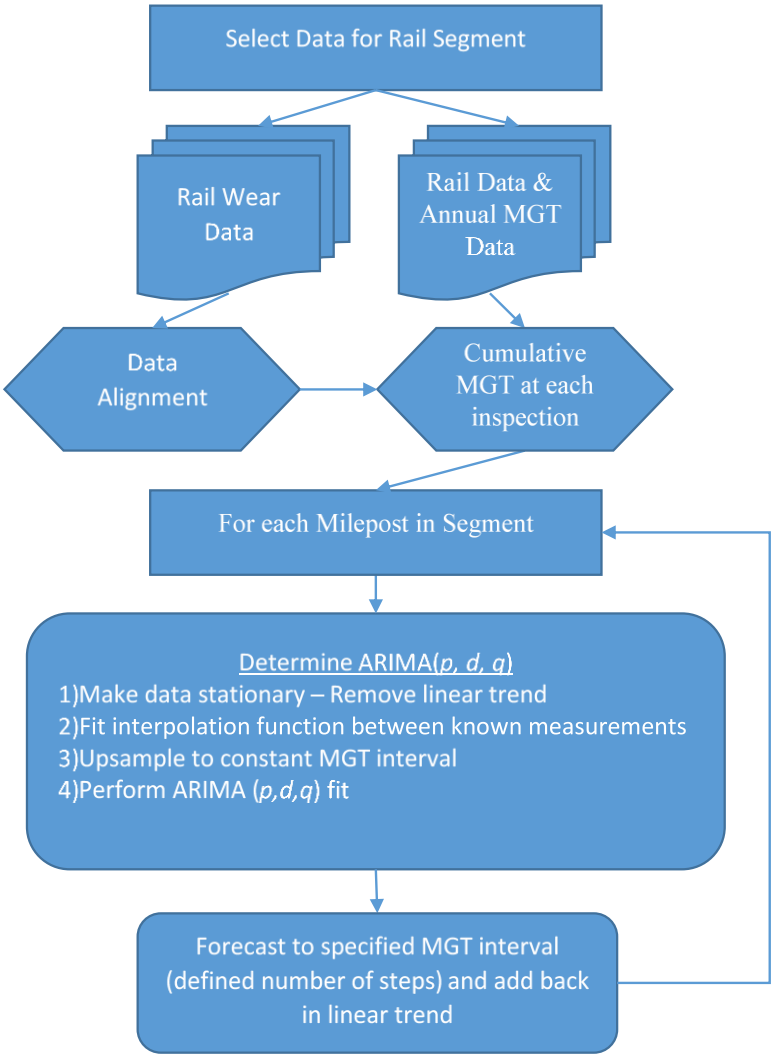


Figure 30 Flowchart of ARIMA process for a rail segment.

Validation

In order to validate the modeling approach, ARIMA fits were determined for the first eight measurement points and projected to the next measurement following the process defined in Figure 30. Thus, an ARIMA model was determined for each of the 306 locations in the curve (every 5.28 feet) using data from 3/14/2007 to 11/4/2010, or 575 to 804 MGT. this model was then used to forecast to 10/26/2011 or 862 MGT, approximately one year later.

Figure 31 shows the wear measurements for the last date that was modelled (11/4/2010; corresponding to 804 MGT) and the next measurement one year later (10/26/2011; corresponding to 862 MGT), along with the ARIMA and linear regression (LS) forecasts. In addition, the difference between the forecast and the actual measurement are shown as dotted lines at the bottom of the figure.

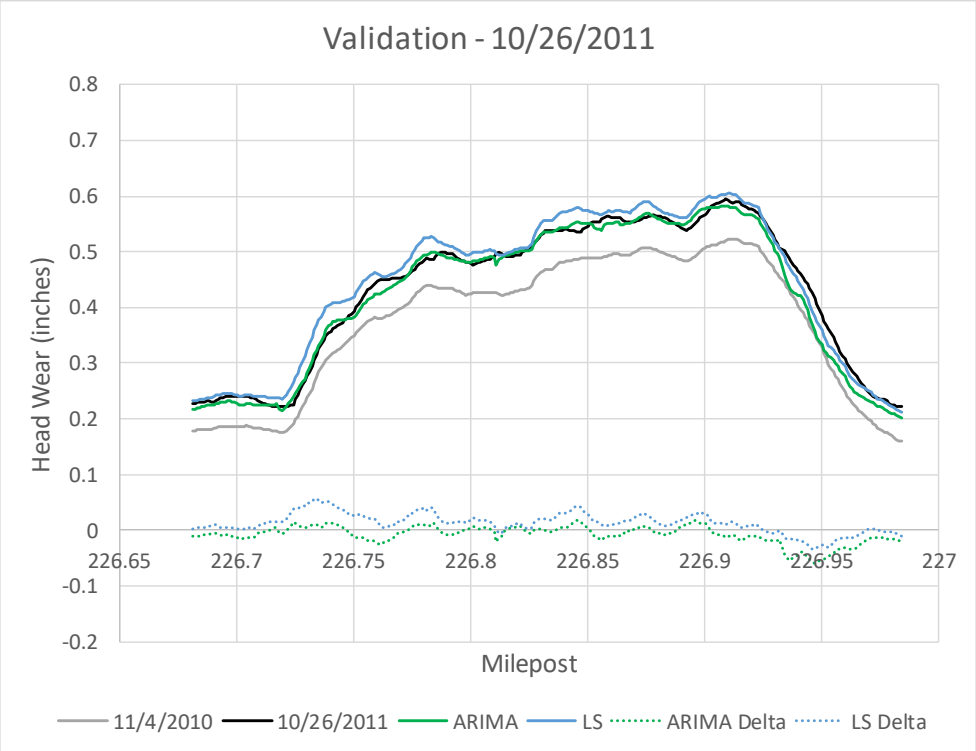


Figure 31 Segment model results.

It can be seen from Figure 31 that the ARIMA forecast more closely matches the actual measurements, i.e. has a smaller difference, while the linear regression fit tends to overstate the actual measurement. Both understate the ending spiral and provide good results in the leading tangent.

Considering the differences between the last actual measurement and corresponding predictions, the difference statistics were calculated for the wear forecasts through the entire curve and are shown in Table 8.

Table 8 Model Comparison for Rail Segment

	ARIMA	**LS**
Min (in)	-0.0586	-0.0331
Max (in)	0.0180	0.0564
Avg (In)	-0.0080	0.0131
RMSE (in)	0.0177	0.0222

These statistics show that the ARIMA model understates the forecast by 0.008" on average with a maximum understatement 0f 0.059" and maximum overstatement of 0.018", while the linear regression model overstates the forecast by 0.013" on average with a maximum understatement of 0.033" and maximum overstatement of 0,056".

Thus, in one year with an additional 58 MGT, the rail has worn between 0.035" and 0.075", with an average of 0.057".

It should be noted that while the ARIMA forecast is more accurate (RMSE = 0.0177 in), the least squares fit is quite reasonable (RMSE =. 0.0222 in)

Finally, since the ARIMA orders (and corresponding coefficients) are determined individually for each location, it is interesting to evaluate how these orders are distributed throughout the segment. Figure 32 shows the ARIMA order (p, d, q) for each milepost. This figure shows that the predominant order of AR is 2, MA is 0 or 1 and differencing is 0. It should be noted that the order of differencing, d, is expected to be 0 since the linear trend was removed. However, even after removing the linear trend some mileposts exhibited non-stationary behavior and required up to two additional orders of differencing.

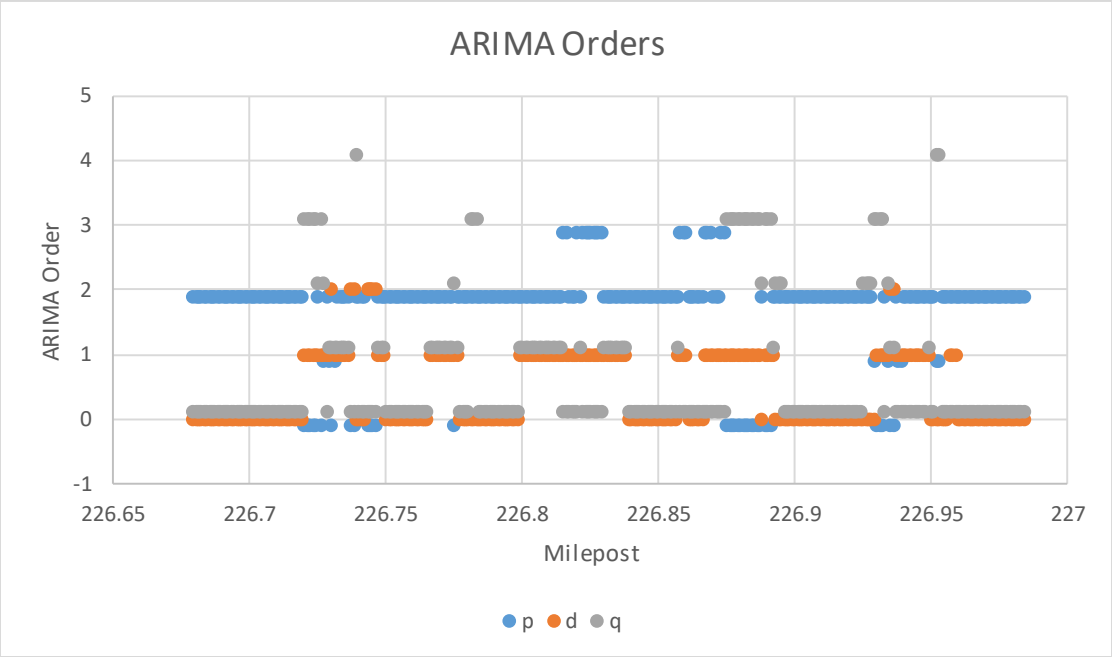


Figure 32 ARIMA orders distributed through segment.

The orders are further explained in

Table 9. The left portion of the table is an absolute count of each order from 0 to 4, confirming the observations made from the figure above. The right portion of the table shows a count of all possible derived ARIMA(p, d, q) models . Clearly ARIMA(2, 0, 0) dominates (162 occurrences), indicating that the majority of the mileposts follow only an autoregressive model of AR(2), with ARIMA(2, 1, 1) having 47 occurrences.

Table 9 ARIMA Orders for Segment

Order	*p*	*d*	*q*		ARIMA(p, d, q)			
0	35	177	210		*p*	*d*	*q*	*Count*
1	9	120	52		0	1	2	1
2	243	9	12		0	1	3	25
3	19	0	29		0	2	0	6
4	0	0	3		0	2	1	3
					1	0	3	1
					1	0	4	2
					1	1	0	2
					1	1	1	2
					1	1	2	2
					2	0	0	162
					2	0	2	8
					2	0	3	3
					2	0	4	1
					2	1	0	21
					2	1	1	47
					2	1	2	1
					3	1	0	19

Forecasting

The primary goal of the modeling approach is to forecast wear into the future to determine when the rail reaches a state at which it requires replacement due to increased risk of derailment. Using the approach outlined in Figure 30, the wear can be forecast to a defined accumulated MGT and evaluated against a defined threshold to determine its condition at any point in the future.

For this step, all eleven measurements were utilized to determine the ARIMA models at each milepost. These models were then applied and the head wear forecast from its current accumulated MGT of 912 to 1010, 1110 and 1210 MGT. The results of the forecast are shown in Figure 33. This figure shows the progression of wear over time and accumulated tonnage. Note from this figure that the linear regression forecast and ARIMA forecast are similar with some small discrepancies in the spirals and areas where the rail has non-uniform changes in wear. Also, note from this particular example that the wear is not uniform in the full body and that the wear and associated rates seem to peak at the spiral to curve points. This is likely due to increased lateral curving forces in these areas.

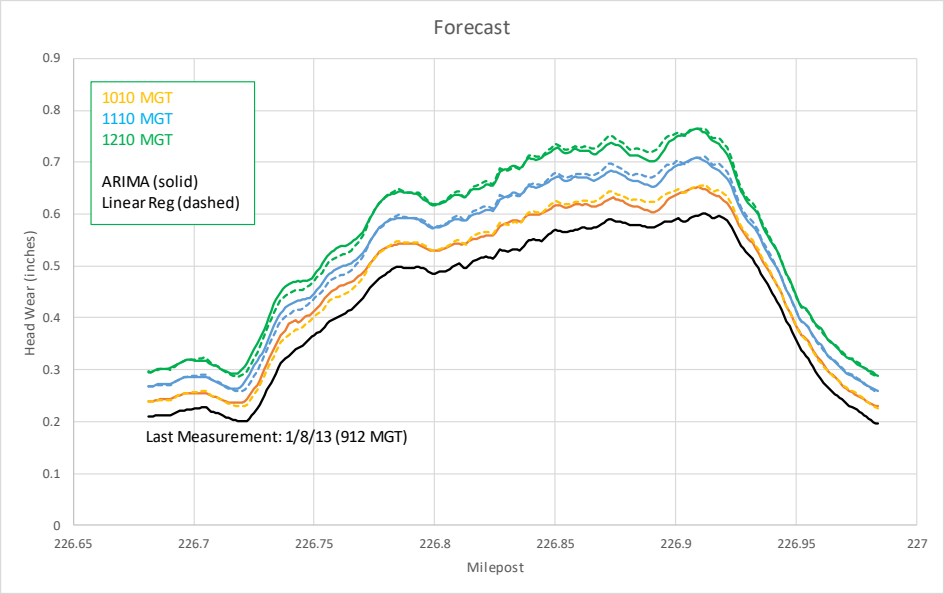


Figure 33 ARIMA forecast for entire rail segment.

Figure 33 also shows that when the curve reaches approximately 1200 MGT, a head wear of 3/4" will be reached, and that at 1100 MGT the majority of the full body will exceed 5/8". In this manner, depending on the railroads wear limits, a forecast of when the first portion of the curve will reach the limit can be forecast, or when 50% of the curve will reach the limit, or when any defined distribution of wear will be achieved. This becomes important when understanding the useful life of the asset, and future asset condition when considering deferred maintenance.

Rail Wear Rate Model

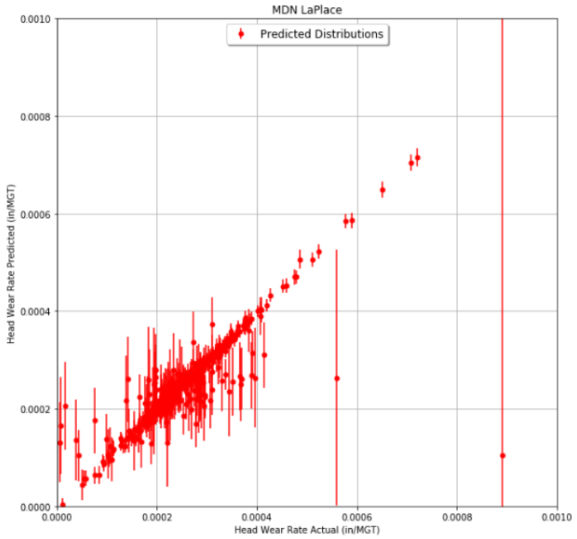
The mixture density network method described in Section 0 was utilized to develop separate models for the following:

- Tangent Rail Head Wear
- Low Rail Head Wear
- High Rail Head Wear
- High Rail Gage Wear

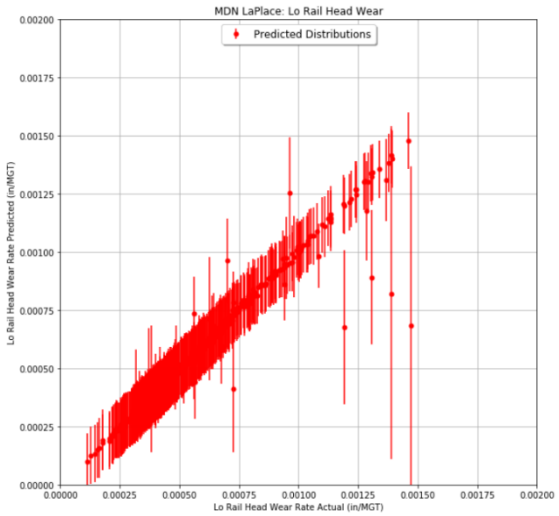
These mixture density network models were developed utilizing the actual wear rate data developed as part of the EDA, and a two-parameter Laplace distribution. The Laplace distribution was chosen because, as noted previously, the data exhibited a Laplace behavior. The conditional distribution (+/- one shape factor) compared to the actual wear rates are shown for all four models in

Figure 34[9]. Some observation can be made regarding these plots as follows:

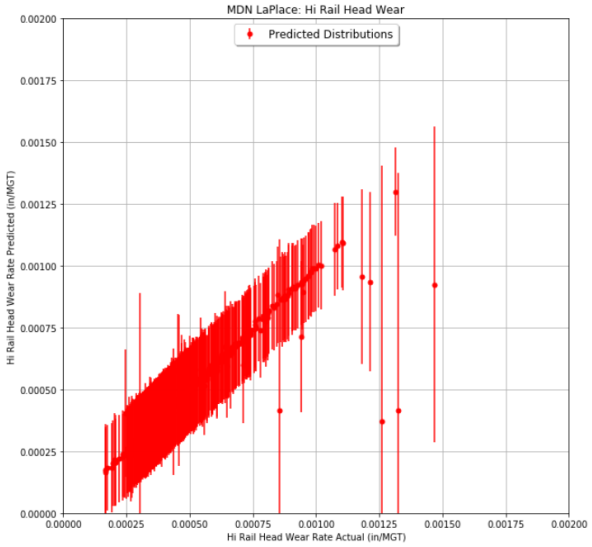
- a – Tangent Rail Head Wear: This is the same result shown in the sample application. The predictions are reasonable, and there is significant distribution around the mean values.
- b – Low Rail Head Wear: The predictions are reasonable and there is fairly uniform variance for most of the observations. Seemingly outlier points show more variance. Overall the variance is larger than that for tangent rail.
- c – High Rail Head Wear: Again, the predictions are reasonable, with a few outlier points with large variance, while most of the points have a uniform large variance (as compared to the other three modes).
- d – High Rail Gage Wear: The gage wear shows the smallest amount of variance and has the fewest number of outliers.

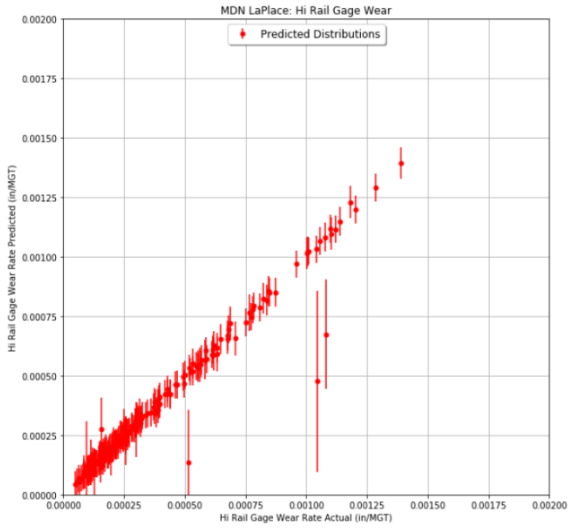


a – Tangent Head Wear



b – Low Rail Head Wear





[9] Note, the scale (in/MGT) for Tangent is 0 to 0.001, while the remainder are 0 to 0.002.

c – High Rail Head Wear d – High Rail Gage Wear

Figure 34 MDNs for all four rail wear modes.

The conditional rail wear rate determination methodology is summarized in Figure 35.

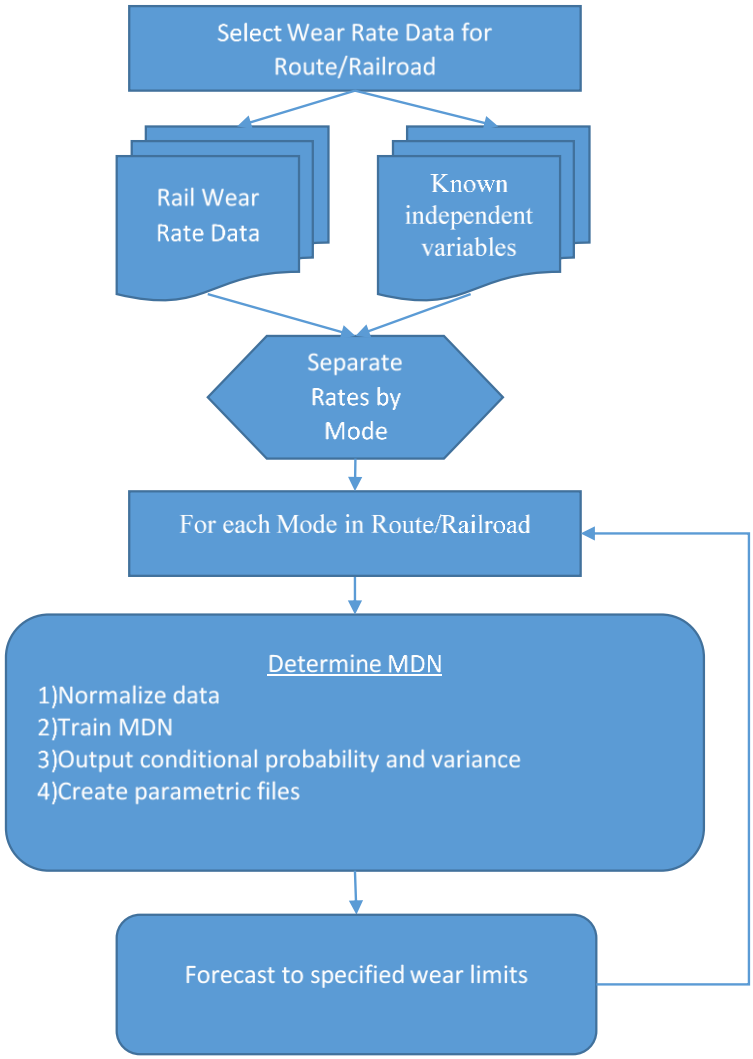


Figure 35 Flowchart of conditional wear rate determination using MDN framework.

Central Probability Determination

In order to understand the effect of the shape factor from the Laplace distribution, and to have a practical sense of the conditional probability distribution, the middle 90% probability of occurrence (+/-45% from the condition mean) is determined (See Figure 36).

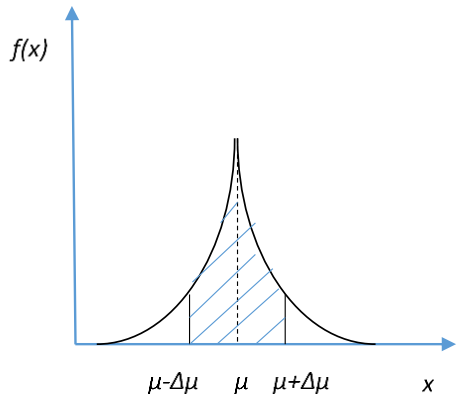


Figure 36 Central probability of Laplace distribution.

Considering Figure 36, the shaded area is the desired central probability, and $\Delta\mu$ is the offset from the mean to be determined. This is calculated from the cumulative distribution function of the Laplace distribution, which is obtained by integrating the Laplace probability density function, and is given by the following:

$$F(x) = \int_{-\infty}^{x} f(u)du = \begin{cases} \frac{1}{2}\exp(\frac{x-\mu}{b}) & if\ x < \mu \\ 1 - \frac{1}{2}\exp(-\frac{x-\mu}{b}) & if\ x \geq \mu \end{cases} \quad (30)$$

The desired value of wear rate ($\Delta\mu$) for the defined central percentage (ΔF) is determined by solving for $x = \mu \pm \Delta\mu$ from the following, given the conditional wear rate (μ) and conditional shape factor (b):

$$F(x) - F(\mu) = \Delta F/2 \quad (31)$$

Using equations (5.14) and (5.15) results in the following:

$$x = \mu \pm b\ln(1 - \Delta F)$$
$$or$$
$$\Delta\mu = b\ \ln(1 - \Delta F) \quad (32)$$

For the parametric study performed, $\Delta F = 0.9$, thus Equation (5.16) reduces to $\Delta\mu = b\ \ln(0.1)$ and the plots show the conditional mean as well as the surrounding 90% probability of occurrence (shaded area) between $\mu \pm \Delta\mu$.

Parametric Graphs

While the use of MDNs suggests a meaningful prediction, as well as includes information regarding the variance in the data, the data has shown some counterintuitive relationships, per the EDA. To better understand the predictions of the models, parametric studies were performed. This was achieved by passing suitable ranges of values for the independent variables to the MDNs and predicting the conditional distribution parameters, which can then be plotted to evaluate reasonableness.

Test files were created for tangents (6 independent variables) and curves (8 independent variables) separately, and the respective ranges shown in Table 10.

Table 10 Test File Parameter Ranges

	Tangents		Curves	
Parameter	Range	Num	Range	Num
Grade	0 to 2 by 0.25	9	0 to 1 by 0.5	3
Year Installed	1980 to 2015 by 5	8	1990 to 2015 by 5	6
Speed	20 to 50 by 10	4	20 to 40 by 10	3
Annual MGT	20 to 60 by 10	5	20 to 50 by 10	4
Lubricator Distance	0 to 2 by 0.25	9	0 to 2 by 0.25	9
Rail Weight	132, 136, 142	3	132, 136, 142	3
Curvature	N/A		0 to 14 by 2	8
Super Elevation	N/A		0 to 4 by 1	5
Total Permutations		38,880		233,280

The Tangent permutation file was input to the Tangent Rail Head Wear MDN model, while the Curve file was input to the High and Low rail curve MDN models. The resulting outputs were then used to generate parametric graphs.

Utilizing the outputs from the MDN models, several selected parametric plots are presented and discussed. Note that hundreds of plots can be created due to the number of independent variables/permutations. In order to get a representation of the models, only a select few are presented.

Tangent Rail Head Wear

Grade and Speed were chosen as the primary parametric variable to compare against head wear rate, as a function of Annual MGT, Rail Weight and Year Installed.

Figure 37 shows the tangent head wear rate for grades from 0% to 2% for rail installed in years 1985, 2000, 2015. This figure shows that as grade increases the head wear rate increases as expected. In addition, there is more variance (spread) in the middle of the grade zone (0.25% to 0.75%). While rails of differing ages behave the same for level track (0% grade), older rail has more severe wear rates at higher grades as expected, as rail has progressively gotten harder[10]. There is more variance for the older1985 rail at higher grades.

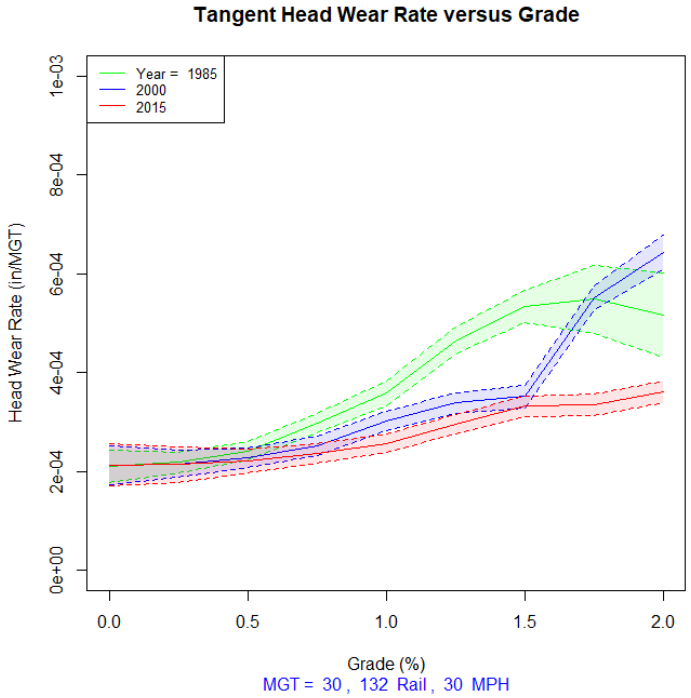


Figure 37 Tangent head wear rate as a function of grade and year installed.

Figure 38 shows the tangent head wear rate for grades from 0% to 2% for the three rail sections. The larger section (141RE) shows more severe wear rates. This is not unexpected as 141RE has an 8" head radius while 132RE and 136RE have 10" and 14" head radii respectively. The sharper the radius on the head, the more severe the contact stress. All sections show and increase in wear rate with an increase in grade.

Figure 39 shows the tangent head wear rate for grades from 0% to2% for three values of annual MGT (10, 30, 40 MGT/year). Note there are two plots, one for rail installed in 2000 and one for rail installed in 2010. These figures show an increase with head wear rate with an increase in grade and increase in annual MGT, for grades less than 1%. Greater than 1%, 20 and 30 MGT/year have

[10] Specifications for standard rail hardness have progressively increased in recent years.

significant increases in head wear rate while 40 MGT/year has a slight decrease, which is unexpected behavior.

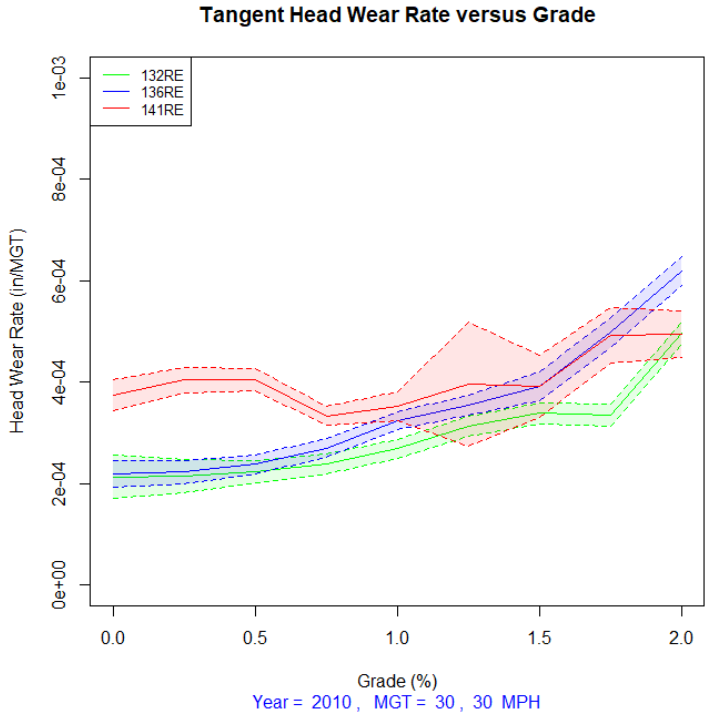


Figure 38 Tangent head wear rate as a function of grade and rail section.

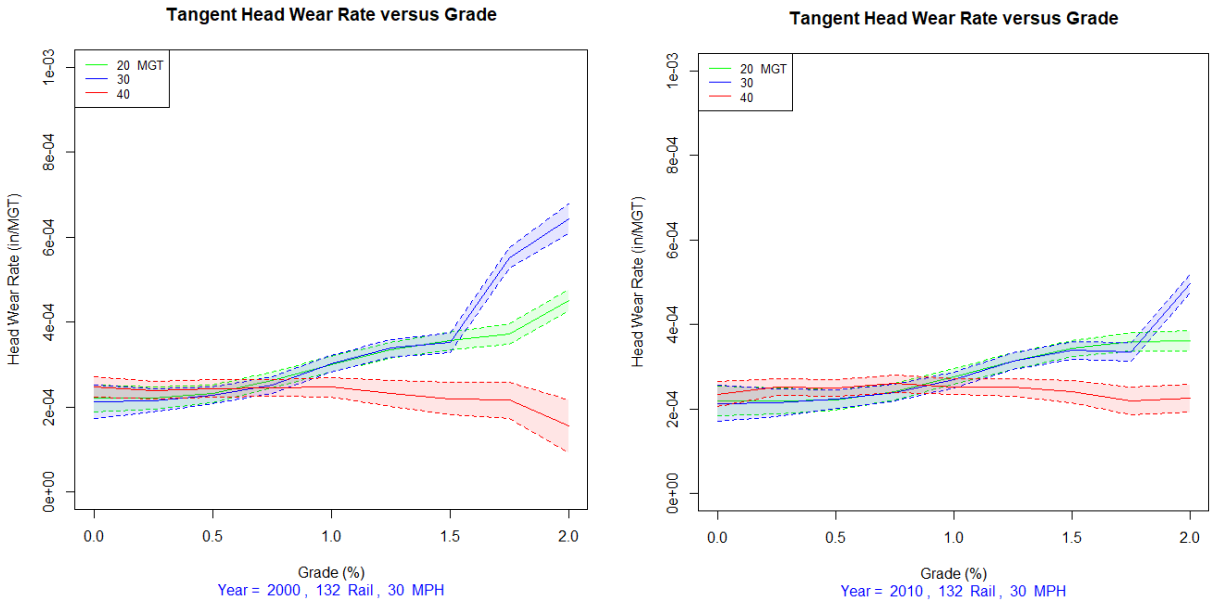


Figure 39 Tangent head wear rate as a function of grade and annual MGT.

Figure 40 shows the tangent head wear rate for grades from 0% to2% for three values of speed (20, 30, 40 mph). Again there are two plots, one for 2000 rail and one for 2010 rail. These plots shown an increase in wear rate with an increase in speed. The behavior is consistent for 20 and 30 mph. 40 mph rail behaves counterintuitively at higher grades, however, since this is a freight railroad, it is unlikely that speeds of this magnitude will be reached for significant grades, therefore the model would be invalid at higher speeds.

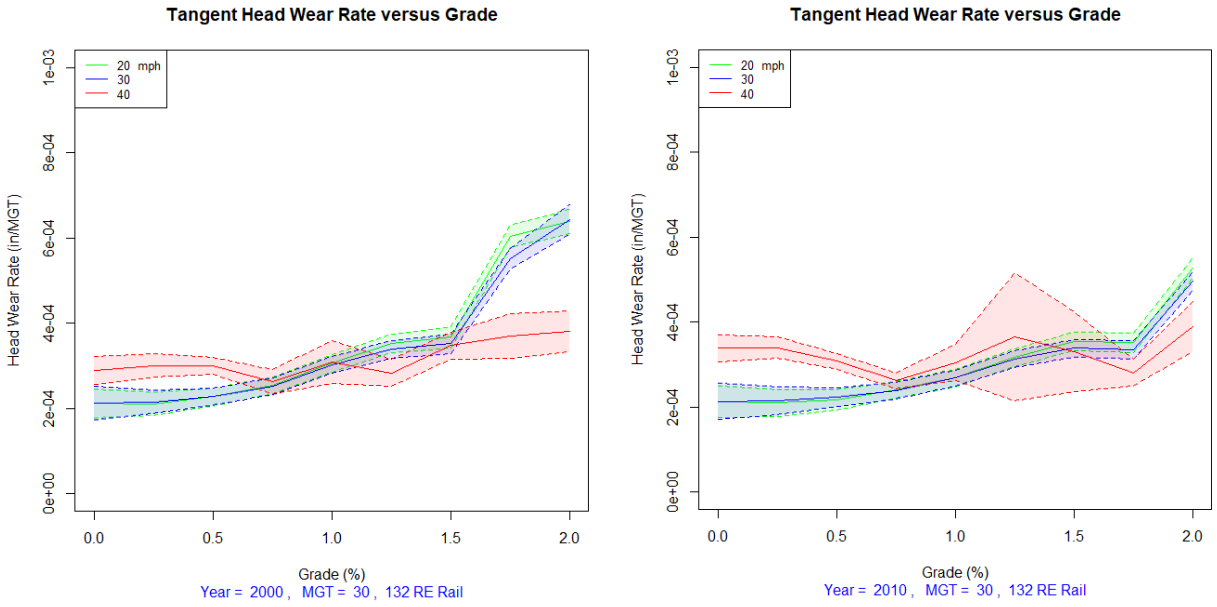


Figure 40 Tangent head wear rate as a function of grade and speed.

Figure 41 shows the tangent head wear rate as a function of speed (from 20 to 50 mph) for three installation years. This figure shows consistent behavior for all three installation years (constant), with increased variance at higher speeds.

Figure 42 shows the tangent head wear rate as a function of speed and the three rail sections. Again, the 141RE rail demonstrates a higher wear rate, while the lighter section behave similarly. The rates remain relatively constant with increases in speed (slight increase or decrease). Figure 43 shows similar behavior for three values of annual MGT.

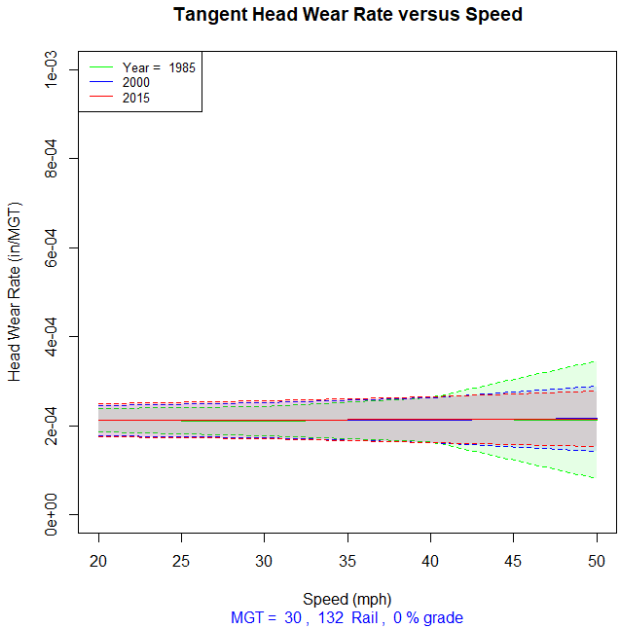


Figure 41 Tangent head wear rate as a function of speed and year installed.

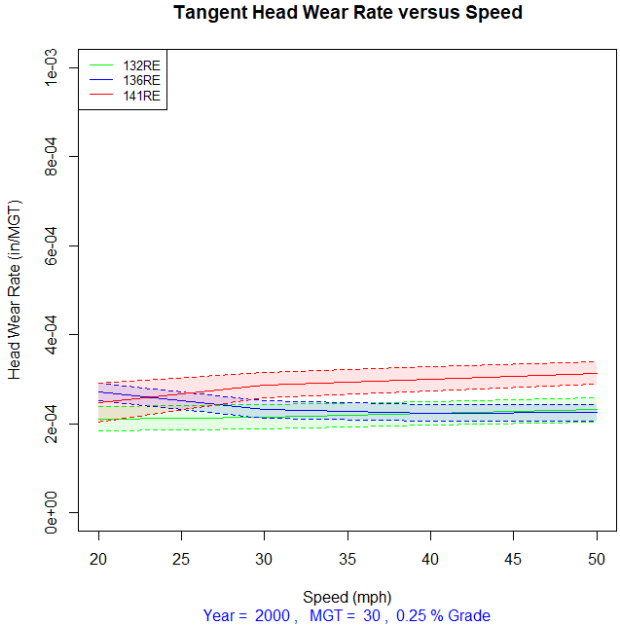


Figure 42 Tangent head wear rate as a function of speed and rail section.

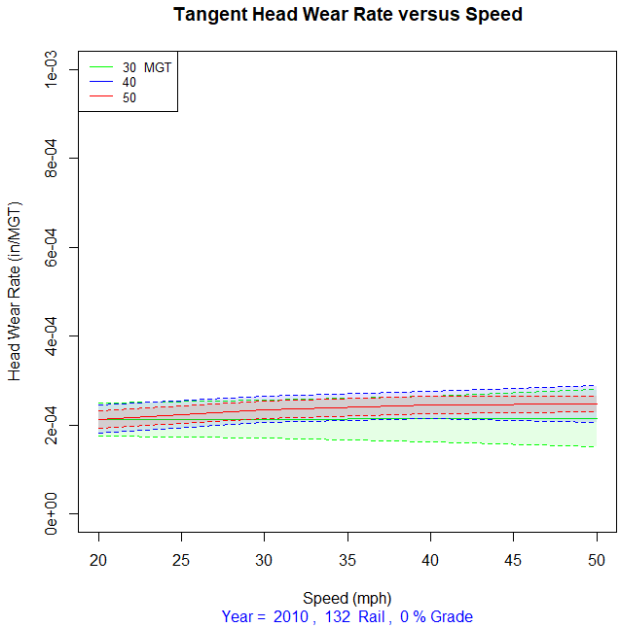


Figure 43 Tangent head wear rate as a function of speed and annual MGT.

Low Rail Head Wear

As with tangent rail, the low rail in a curve will experience head wear (at the top mating surface of the rail with the wheel tread). Due to the solid axle set of a railway vehicle, the low rail will experience more sliding (depending on the degree of curvature, amount of super elevation and speed of the vehicle), and thus is expected to have a higher rate of wear than tangent rail. The parametric study performed on the low rail head wear MDN focused on the curvature, super elevation, and speed effects.

It is important to note that speed in curves (V_{max}) is limited based on the degree of curvature (D), amount of super elevation (E) and allowable amount of unbalance (E_u). Unbalance defines the allowable over-speed in curves, and is treated as that amount of elevation (in inches) by which the curve is unbalanced. The maximum allowable speed is defined according to the following equation:

$$V_{max} = \sqrt{\frac{E + E_u}{0.0007D}} \tag{33}$$

Applying this formula for ranges of data found within the sample dataset results in Table 11. This table provide maximum allowable speed for curves from 2 to 14 degrees, with 1 to 4 inches of super elevation, for balanced operations and 3 inches of unbalance. Note that the maximum allowable speed for severe curves (> 10 degrees) is less than 30 mph. Thus, the plots from the model do not take this into account and are strictly based on the inputs to the model. Results outside the allowable range will be ignored.

Table 11 Maximum Allowable Speed by Degree Of Curvature, Super Elevations and Maximum Unbalance

	Balance Speed					Vmax for 3" Unbalance			
	Super Elevation (inches)					Super Elevation (inches)			
Degree	1	2	3	4		1	2	3	4
2	27	38	46	53		53	60	65	71
4	19	27	33	38		38	42	46	50
6	15	22	27	31		31	35	38	41
8	13	19	23	27		27	30	33	35
10	12	17	21	24		24	27	29	32
12	11	15	19	22		22	24	27	29
14	10	14	17	20		20	23	25	27

Figure 44 shows the low rail head wear by degree of curvature (for 1" of super elevation at 20 mph). The left hand plot shows the probability band for the central 90% and the right hand plot shows the probability band for the central 50%[11]. Note the wear increasing with degree of curvature and that all rail sections behave similarly, with 141RE performing slightly better. These plots are for 20 mph and 1 inch of super elevation, which fall with the maximum allowable operating speed.

Figure 45 shows the low rail head wear rate as a function of curvature and amount of super elevation (SE in inches). The left hand plot is for operations at 20mph and the right hand plot is for operations at 30 mph (note that 1" of SE is only valid to 6 degrees, 2" SE only valid 8 degrees, and 4" SE only valid to 10 degrees). These plots show generally increasing wear with increases in curvature. The data for the 20 mph plot shows some anomalous behavior as super elevation shows no real correlation between wear rate and super elevation. There is a correlation, however, for 30 mph. In addition, 20 mph shows larger rates than 30 mph. This is likely due to more vertical load transferred to the lower rail at slower speeds. At balance speed, the vertical load on each rail is theoretically equally, and below balance speed more of the weight is shifted to the low rail.

[11] The remaining plots for all curve analyses will be determined for the central 50% probability. Also, note that the wear rate scales are 0 to 0.002 in/MGT for all curve plots)

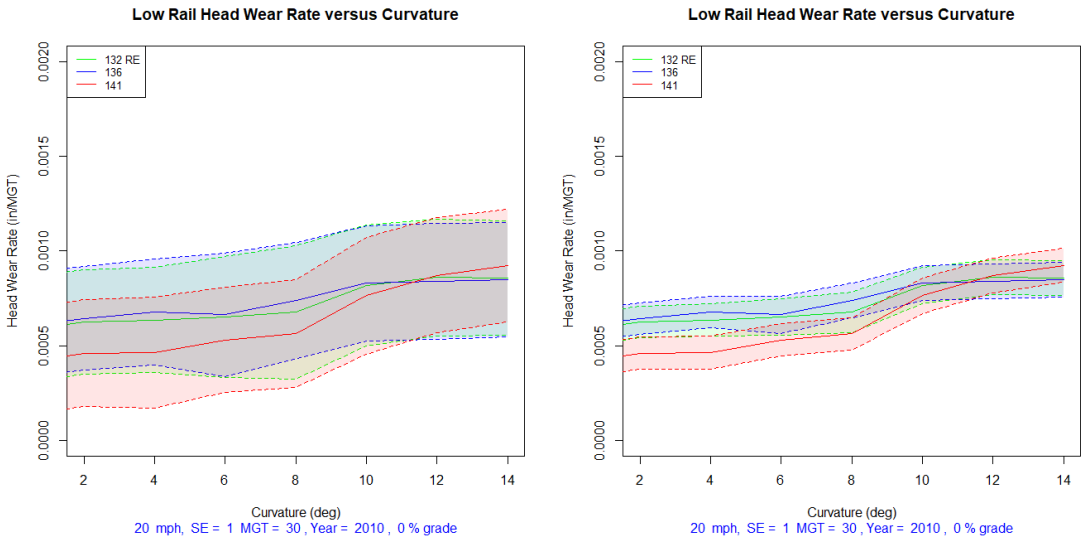


Figure 44 Low rail head wear rate as a function of curvature and rail section.

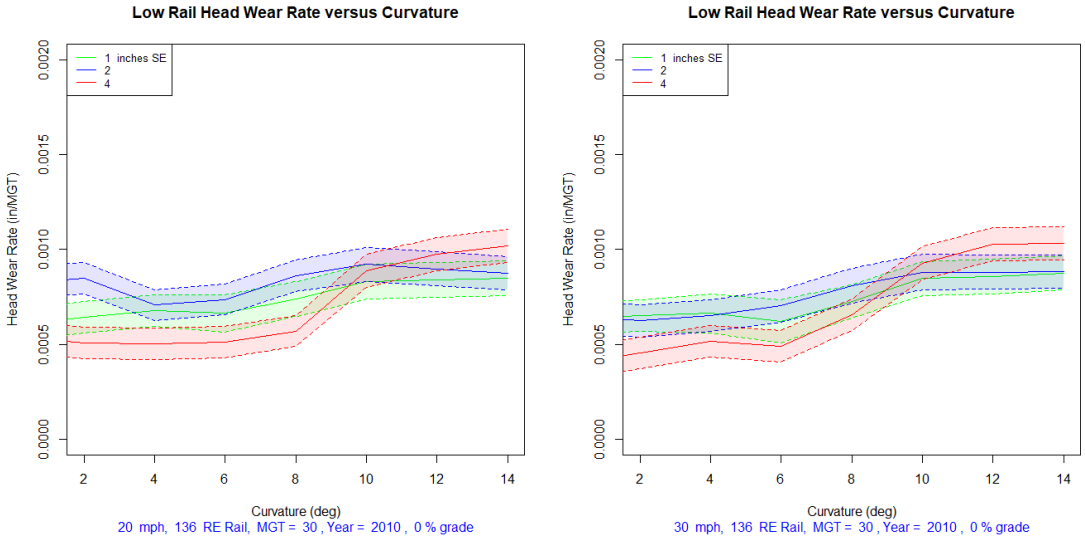


Figure 45 Low rail head wear rate as a function of curvature and super elevation.

Figure 46 shows the low rail head wear as a function of curvature for three speeds of operation (20, 30, 40 mph). The left hand plot is for 2 inches of super elevation and the right hand plot is for 4 inches of super elevation. The 2 inches of super elevation plot shows similar behavior for all speeds, however 30 mph is only valid to 8 degrees and 40 mph only valid to 4 degrees, as speeds in excess of those defined in Table 11 would not be allowed. The 4 inches of super elevation plot shows significant increases in wear rate as curvature increases (30 mph is valid to 12 degrees, 40 mph valid to 8 degrees), with mild (< 6 degrees) and shallow curves behaving similarly.

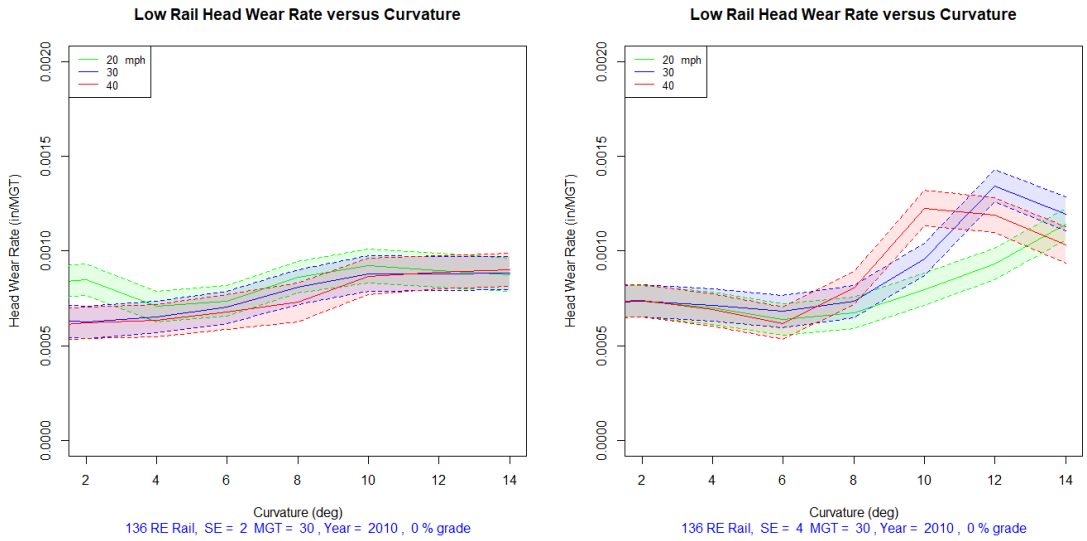


Figure 46 Low rail head wear rate as a function of curvature and speed.

High Rail Head Wear

As with the low rail in a curve, the high rail will experience head wear (in addition to gage face wear). Similar plots to those for the low rail are shown for the high rail using the high rail head wear rate MDN, and again focused on the curvature, super elevation, and speed effects.

Figure 47 shows the high rail head wear by degree of curvature (for 1" of super elevation at 20 mph). The left hand plot shows the probability band for the central 90% and the right hand plot shows the probability band for the central 50%. Note the wear is constant or decreases with increasing degree of curvature and that all rail sections behave similarly at higher degrees of curvature. A possible reason for this slight decrease in wear rate is that as the degree of curvature increases, the high rail goes into two-point contact, i.e. the wheel flange makes contact with the gage face of the rail, picking up some of the load (especially at 1" of elevation). These plots are for 20 mph and 1 inch of super elevation, which fall within the maximum allowable operating speed.

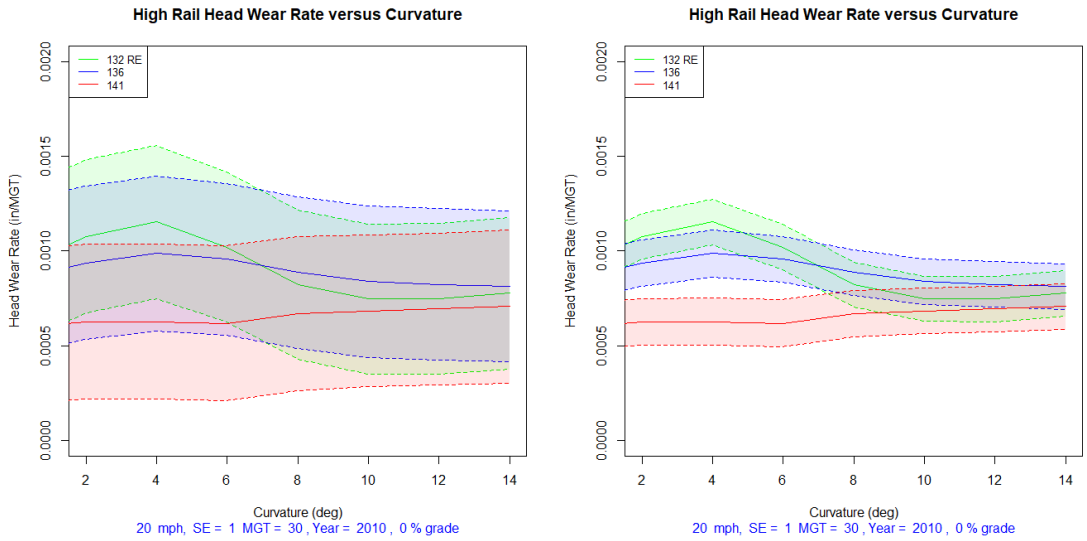


Figure 47 High rail head wear rate as a function of curvature and rail section.

Figure 48 shows the high rail head wear rate as a function of curvature and amount of super elevation (inches). The left hand plot is for operations at 20 mph and the right hand plot is for operations at 30 mph (note that 1" of SE is only valid to 6 degrees, 2" SE only valid 8 degrees, and 4" SE only valid to 10 degrees). These plots show a variation in wear with increases in curvature. Wear rates for super elevation of 4 inches are generally less than those of 1 inch of elevation. This is as expected as the vertical load would shift more to the low rail, however 2 inches of elevation is not behaving consistently.

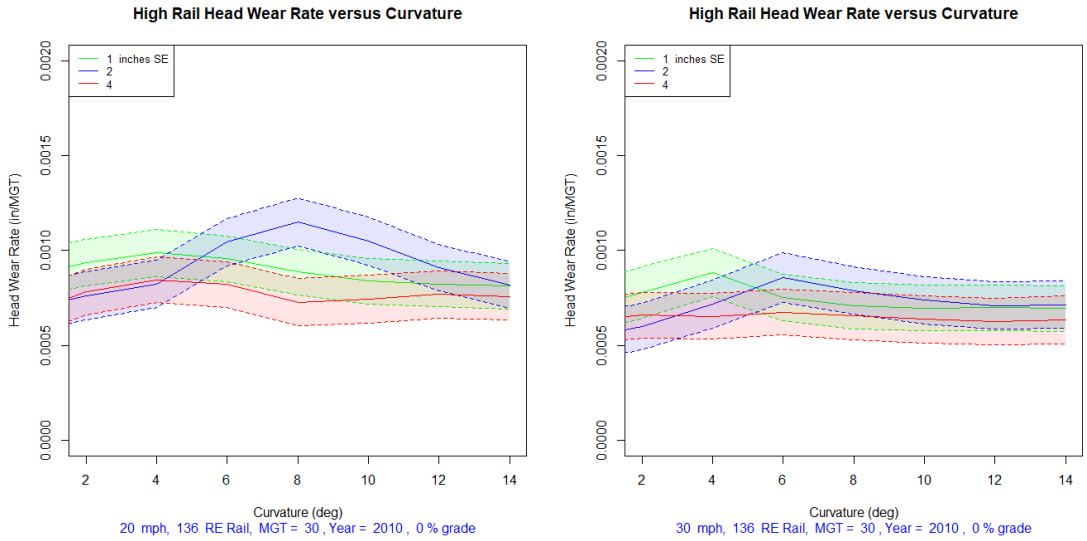


Figure 48 High rail head wear rate as a function of curvature and super elevation.

Figure 49 shows the high rail head wear as a function of curvature for three speeds of operation (20, 30, 40 mph). The left hand plot is for 2 inches of super elevation and the right hand plot is for 4 inches of super elevation. The 2 inches of super elevation plot shows lower wear rates for higher speeds, consistent with two-point contact, however 30 mph is only valid to 8 degrees and 40 mph only valid to 4 degrees. The 4 inches of elevation plot shows nearly constant wear rates as curvature increases that behave similarly with operating speed (30 mph on valid to 12 degrees, 40 mph valid to 8 degrees). This is not unexpected as 4 inches of elevation may cause consistent vertical contact forces on the high rail. This would need to be confirmed with dynamic simulation modeling.

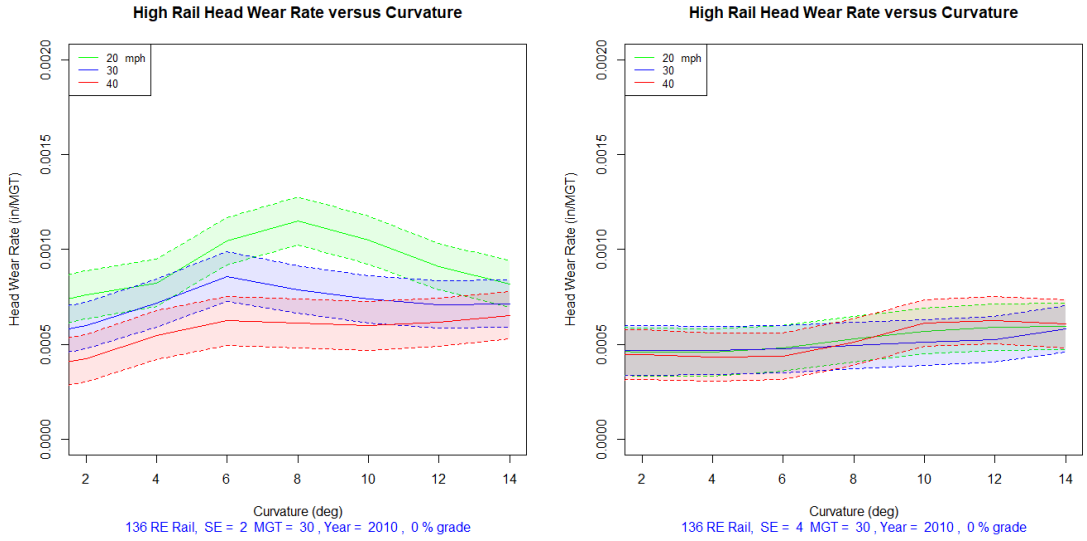


Figure 49 High rail head wear rate as a function of curvature and speed.

High Rail Gage Wear

The final MDN, high rail gage wear, was used to model the gage wear rate. The high rail experiences head wear and gage wear simultaneously, the rates of which are very much a function of the contact stresses; defined by the lateral and vertical force resultant. Comparative plots to the high rail head wear are shown for the high rail gage wear using the high rail gage wear rate MDN, and again focused on the curvature, super elevation, and speed effects.

Figure 50 shows the high rail gage wear rates by degree of curvature (for 1" of super elevation at 20 mph). The left hand plot shows the probability band for the central 90% and the right hand plot shows the probability band for the central 50%. Note that the gage wear has less variance then the head wear of the low and high rails.

Note the wear tends to decrease, then increase (at severe curves greater than 8 degrees) with increasing degree of curvature. However, the various rail sections behave differently, with the heavier sections performing better. A possible reason the heavier sections performing better is the higher lateral stiffness of the heavier rail section, but this would require further in-depth analysis. The primary reason for the "dip" in wear rate at the mild curvatures is the change over from dominant wear mode, i.e., head versus gage wear rate. Since gage is measured instantaneously for the current top of the rail, its location point of measure changes. It was noticed in the data that for mild curves, the head wear rate outpaces the gage wear rate. Thus, when the head wears faster than the gage face or side of the rail head, the point where gage wear is measured shifts (vertically) and it is possible to have a change in gage wear rate.

These plots are for 20 mph and 1 inch of super elevation, which fall with the maximum allowable operating speed.

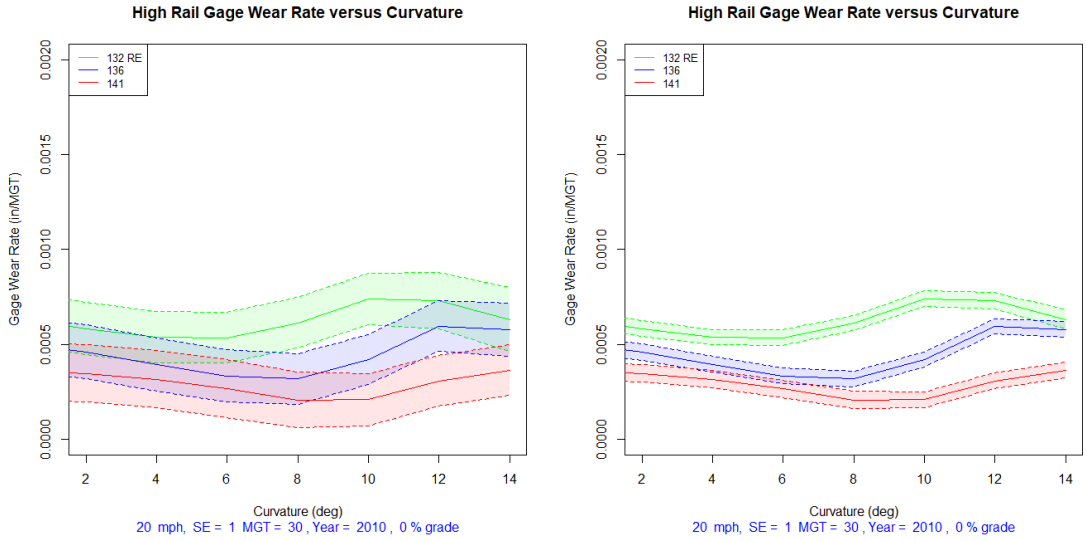


Figure 50 High rail gage wear rate as a function of curvature and rail section.

Figure 51 shows the high rail gage wear rate as a function of curvature and amount of super elevation (inches). The left hand plot is for operations at 20 mph and the right hand plot is for operations at 30 mph (note that 1" of SE is only valid to 6 degrees, 2" SE only valid 8 degrees, and 4" SE only valid to 10 degrees). The 20 mph plots show relatively constant gage wear, with an increase at severe curves, however the 2 inches of super elevation seemingly performs much worse, which is counterintuitive. Also, the effect of super elevation is opposite of that expected. Note that these plots are for the middle 50% probability band, which shows some overlap between the predictions. The 90% band would show further overlap, indicating overlapping conditional probabilities. Thus, while the means may not behave as expected, there is a probability (at the extremes) that consistent behavior can occur as expected.

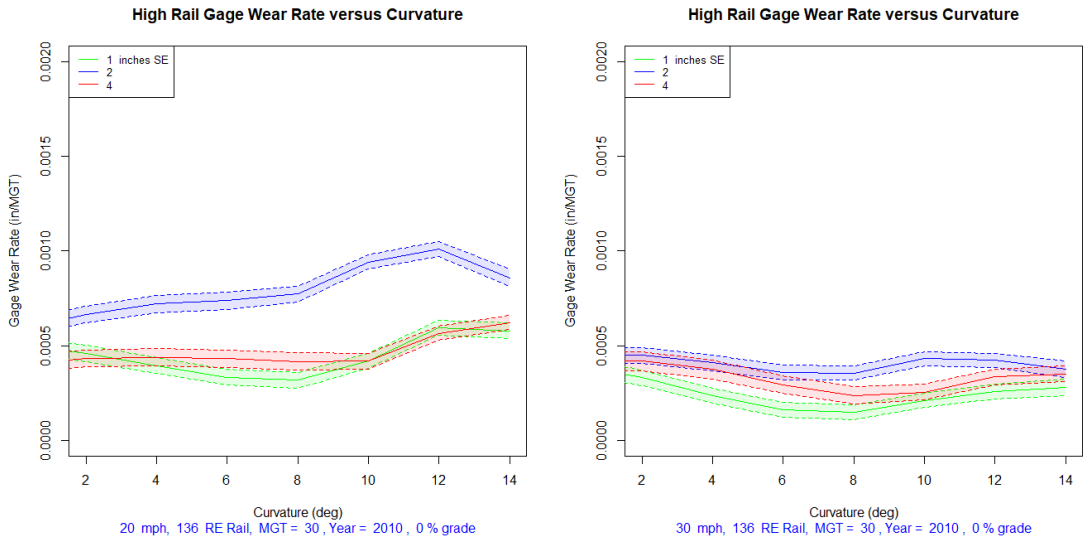


Figure 51 High rail gage wear rate as a function of curvature and super elevation.

Figure 52 shows the high rail gage wear as a function of curvature for three speeds of operation (20, 30, 40 mph). The left hand plot is for 2 inches of super elevation and the right hand plot is for 4 inches of super elevation. The 2 inches of super elevation plot shows lower wear rates for higher speeds, consistent with head wear outpacing gage wear, however 30 mph is only valid to 8 degrees and 40 mph only valid to 4 degrees. The 4 inches of elevation plot shows increasing wear rates as curvature increases. It also shows a decrease in wear rate with an increase in speed (30 mph on valid to 12 degrees, 40 mph valid to 8 degrees; where the plots tail off). This again is attributed to head wear outpacing gage wear.

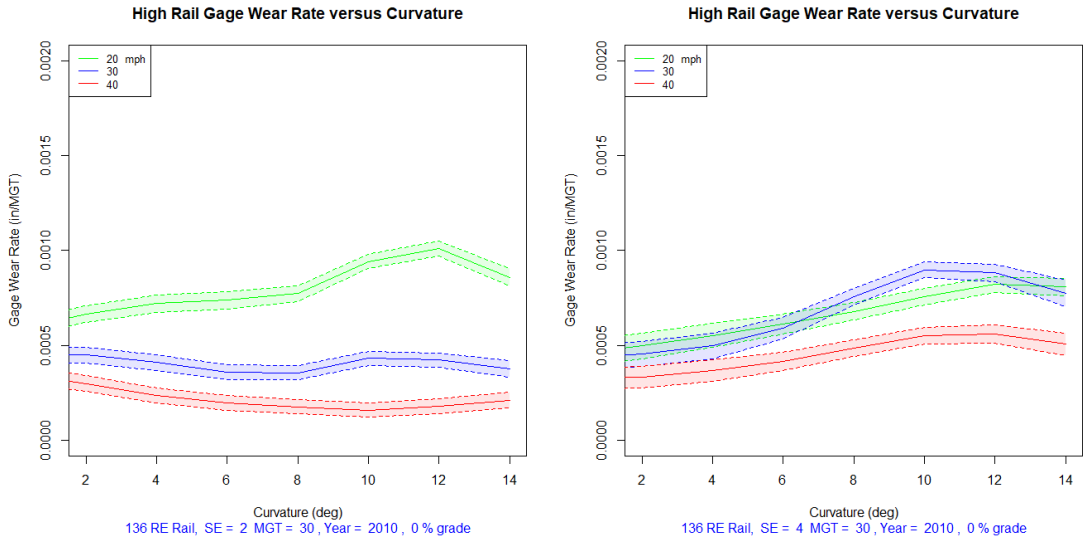


Figure 52 High rail gage wear rate as a function of curvature and speed.

The final sensitivity performed for high rail gage wear rate is with respect to lubrication. Figure 53 shows the high rail gage wear rate as a function of curvature for three levels of lubrication. The levels are defined by the proximity of a lubricator, i.e. the distance in miles to the nearest lubricator.

The left hand plot shows the results for a speed of 30 mph and 1 inch of super elevation (only valid to 6 degrees) and the right hand plot shows the results for 30 mph with 2 inches of super elevation (only valid to 8 degrees). The left hand plot clearly shows that the closer the lubricator, the lower the gage wear rate. Note that a lubricator effectiveness diminishes significantly with distance away from the curve. The right hand plot shows the same benefit of lubrication, however for the higher end of curvature (8 degrees) the "close" and "far" lubricator results converge. This is unexpected and can again be attributed to head wear outpacing gage wear in this curvature range.

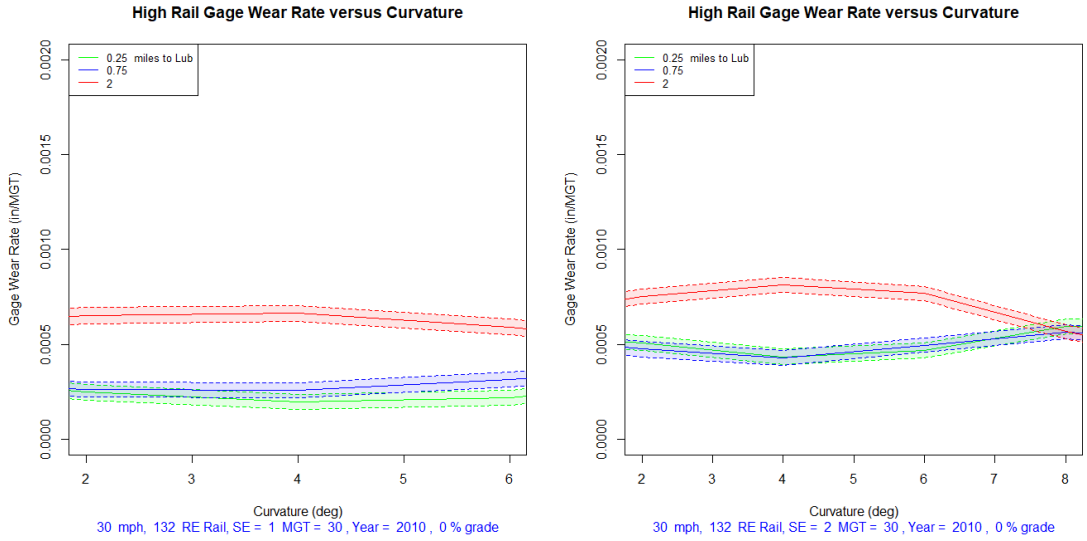


Figure 53 High rail gage wear rate as a function of curvature and distance from lubricator.

Discussion

The ability of the mixture density network model to predict a conditional distribution, as opposed to just the conditional target value, offers a significant advantage. In addition, the model is unbiased, in that the data is driving this behavior. The current results show some counterintuitive behavior that requires further research.

As a final comparison, the wear rates for each MDN were plotted by degree of curvature (See Figure 54). The left hand plot is for 20 mph with 1 inch of elevation, and the right hand plot is for 20 mph with 2 inches of elevation. These plots show distinct traits that are worth noting as follows:

- The tangent head wear is the lowest wear rate and shows very little variation.
- The low rail head wear has the highest magnitude.

- The gage wear dominates the high rail for curves greater than 4 degrees with 1 inch of elevation and curves greater than 8 degrees with 2 inch of elevation.
- The high rail head wear has the greatest variance.
- The curve MDNs do not converge to the tangent MDN using the middle 50% probability band.

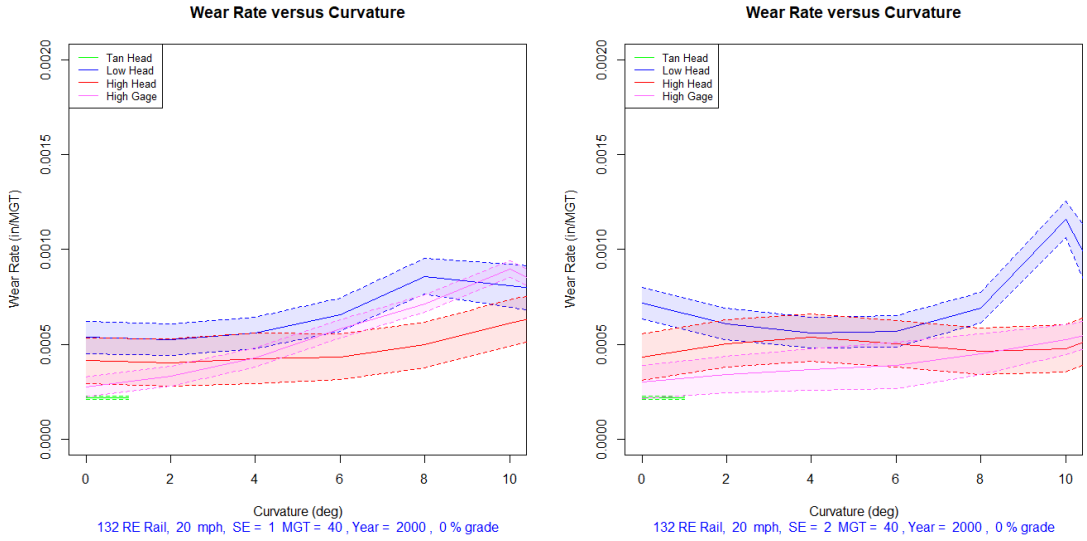


Figure 54 Wear rates by degree of curvature for each MDN (mode).

It should be noted here that for the high rail, the combined wear associated with head and gage wear rates must be accounted for. The wear mode (head or gage wear) that reaches the replacement limit the soonest (life=limit/rate) will control rail renewal.

Forecasting

The wear rate can be determined from the ARIMA process (as input to the MDN process), or for rail segments with too few wear measurements (no inspections), the MDN can be applied to determine the conditional wear rates, based on the known causal factors. Note that the MDN is used to determined conditional wear rates and variances regardless, to allow for classification based on the ARIMA wear rates. This is performed for head and gage wear separately, as appropriate.

Once the wear rate is known, forecasting to an allowable level of wear is straight-forward and follows the equation:

$$Wear(MGT) = W_0 + [W'(X) \pm \Delta W'(X)][\Delta MGT] \quad (34)$$

Where

Wear(MGT) = inches of wear at some MGT level = $MGT_0 + \Delta MGT$

W_0 = Wear in inches at current MGT, (=0 at MGT=0)
$W'(X)$ = conditional wear rate in inches/MGT
$\Delta W'(X)$ = conditional offset wear rate based on desired probability band[12]
ΔMGT = MGT from current measurement

This equation can be reconfigured to forecast ΔMGT to wear limit, as appropriate. As indicated previously, for high rails in curves, the forecast must be performed for both head and gage wear, and the shortest MGT interval to the limit controls rail replacement.

Classification of Wear Performance

Given that the approach used herein is stochastic, a classification process based on stochastic principles was developed. Recalling Equations (5.14 – 5.16), the cumulative probability of a given wear rate, based on the mean wear rate and shape factor can be determined. And recalling that the MDN approach results in a predicted conditional mean wear rate and conditional shape factor, the wear rate calculated during the ARIMA process can be measured against the predicted conditional distribution to assess performance. This is outlined in the flowchart of Figure 55.

For each rail segment, a wear rate $\{\lambda\}$ was determined using the ARIMA process and least squares regression. In addition, the MDN model determines a conditional mean wear rate $\{\mu(X)\}$ and shape factor $\{b(X)\}$ according to a Laplace distribution. It is these parameters for which the classification is determined. Taking advantage of Equation (5.14), $F(\lambda|\mu(X), b(X))$ can be determined. Noting that $F(\mu) = 0.5$ is average (based on input characteristics, X), a linear probability transformation can be utilized to classify performance (C), as follows:

$$C = 1 - F(\lambda) \tag{35}$$

Classification can be shown schematically as in Figure 56, where the vertical lines represent values of wear rate (λ) that correspond to bands of classification (C_n).

[12] Using Equation (5.16)

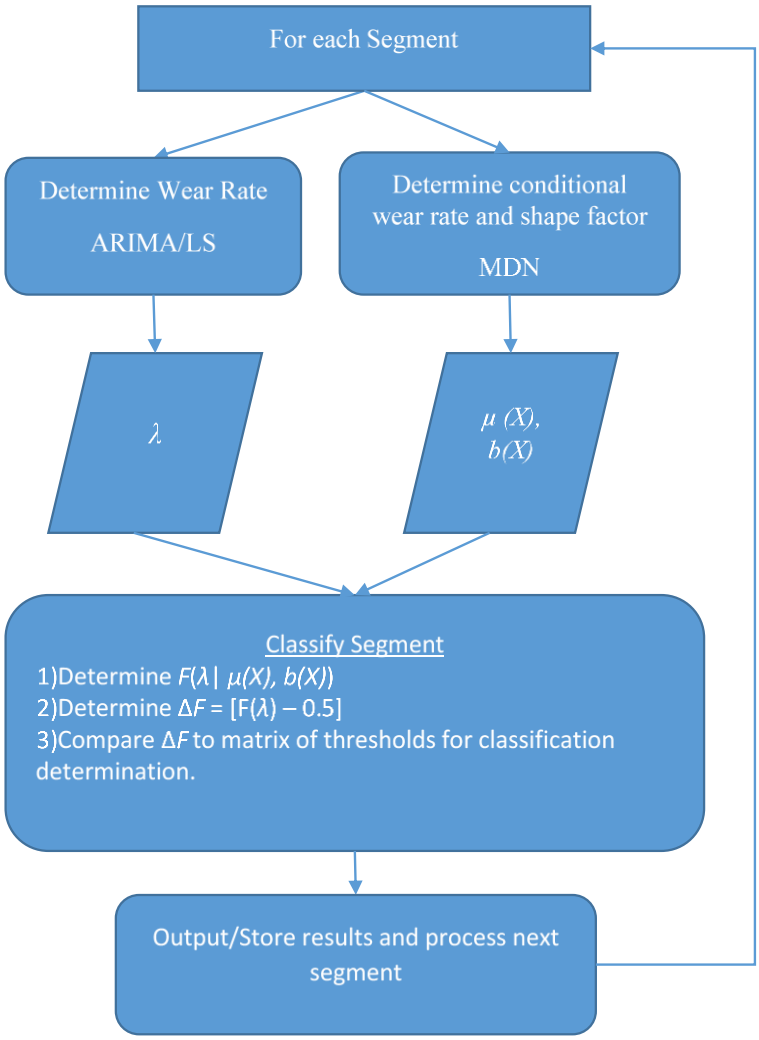


Figure 55 Flowchart of stochastic classification.

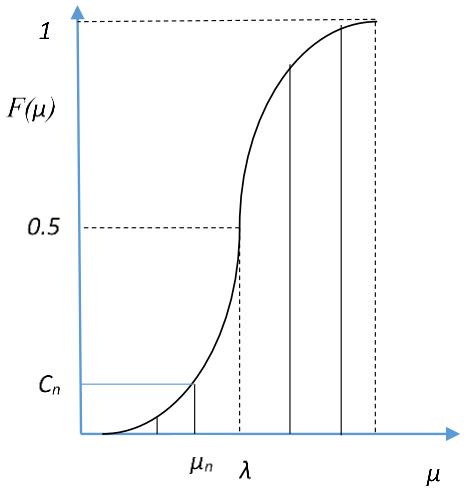


Figure 56 Schematic of classification using $F(\mu)$.

Thresholds can be set to define the performance of rail wear as follows:

$$C = \begin{cases} < 0.5 & Poor \\ = 0.5 & Average \\ > 0.5 & Good \end{cases} \tag{36}$$

Where *Poor* indicated higher than expected wear rates, and *Good* indicates lower than expected wear rates.

MODEL APPLICATION

The models and framework developed in the previous chapters are put into practice for the entire data set. Putting the flowcharts (See Figure 14, Figure 30, Figure 35 and Figure 55) into a graphical perspective shows the general implementation of the Rail-WASP model in Figure 57 below.

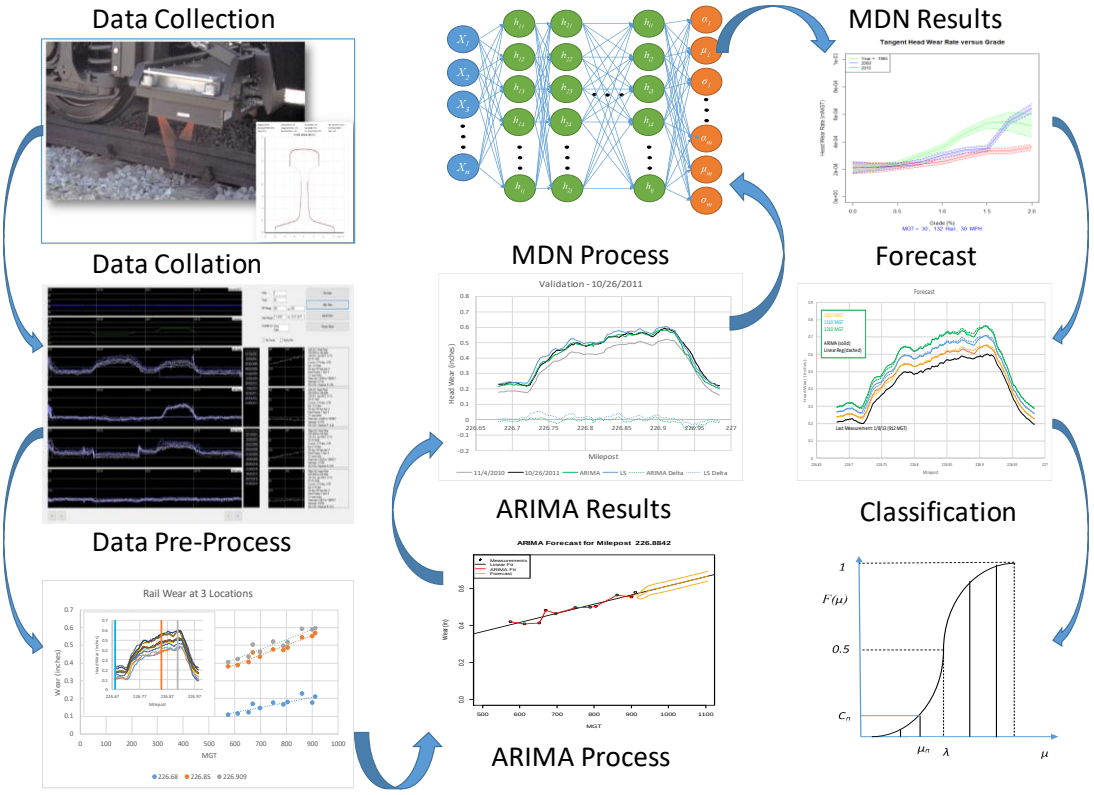


Figure 57 Rail-WASP.

The model is evaluated on a segment basis as well as a system basis.

Segment Analyses

In order to evaluate the model at the segment level, two curves are analyzed and presented here. The characteristics of the two segments are shown in

Table 12. The rails were installed between 2000 and 2002[13] and are from a 5 degree and 7.25 degree curve, with 1.5 inches and 1.75 inches of super elevation respectively.

Table 12 Segment Characteristics

					Left/High Rail		Right/Low Rail	
Rail	Length	Deg	SE	Speed	Rl Wt	Installed	Rl Wt	Installed
1	634	5.00	1.50	25	141	2002	141	2000
2	1795	7.25	1.75	25	141	2002	136	2011

The data was collected for these two segments, and was aligned using algorithms developed and coded into the model.

ARIMA Processing

The ARIMA process was performed, and several MGT steps forward were forecast using the resulting ARIMA models.

The forecast accumulated MGT values were determined for the following conditions:

- First instance of crossing wear limit[14]
- 50% of full body of curve exceeding wear limit
- 100% of full body of curve exceeding wear limit

The results of the ARIMA process and forecast are shown in Figure 58 through Figure 61. The high rail plots show the head and gage wear in the left and right panels. The top left and top right panels show the first and last inspection results (intermediate inspections values were not reproduced in these Figures, though they were used in the ARIMA modeling), and the three forecast wear sequences (first exceedance, 50% of full body, 100% of full body), for the head and gage wear respectively. The wear sequences are labeled by the accumulated MGT at which that prediction occurs. For the head wear plots the wear limit is shown for reference (since both curves reach the head wear limit much faster than the gage wear limit). In addition, the nominal wear rates through the curve are shown (by milepost location) in the bottom left and bottom right panels (head and gage wear rates respectively).

For the low rail plots, only head wear is shown, with the top panel corresponding to the prediction and the bottom panel showing how the head rates are distributed through the curve.

[13] Note that the low rail of the 7.25 degree curve was replaced in 2011. However, the data was analyzed from the inspections prior to this date, for the rail that was originally installed in 2002.

[14] The wear limit was set at 11/16" for both head and gage wear, however, any limit can be set.

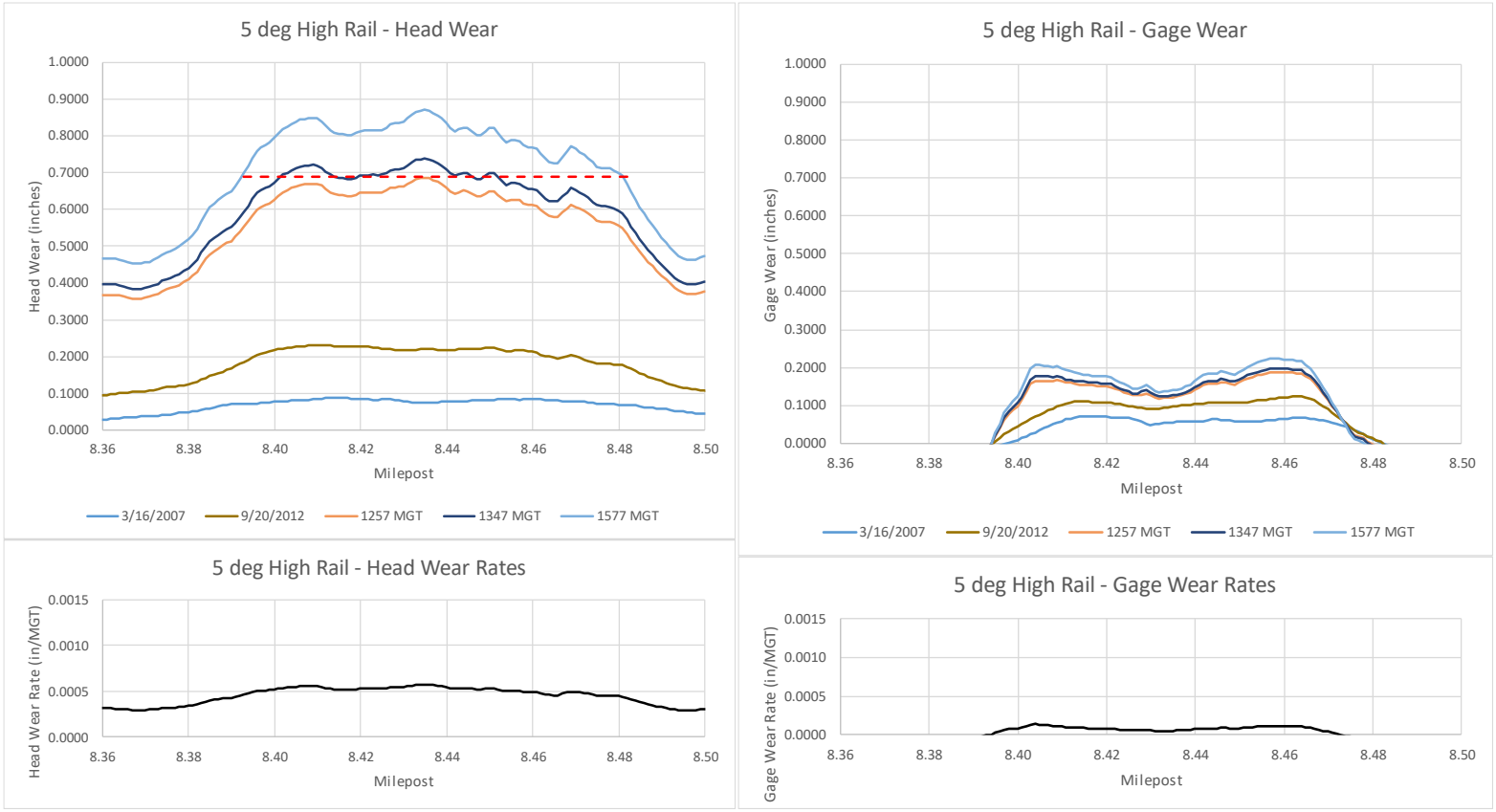


Figure 58 5.0 degree curve high rail wear results from ARIMA processing and forecasting.

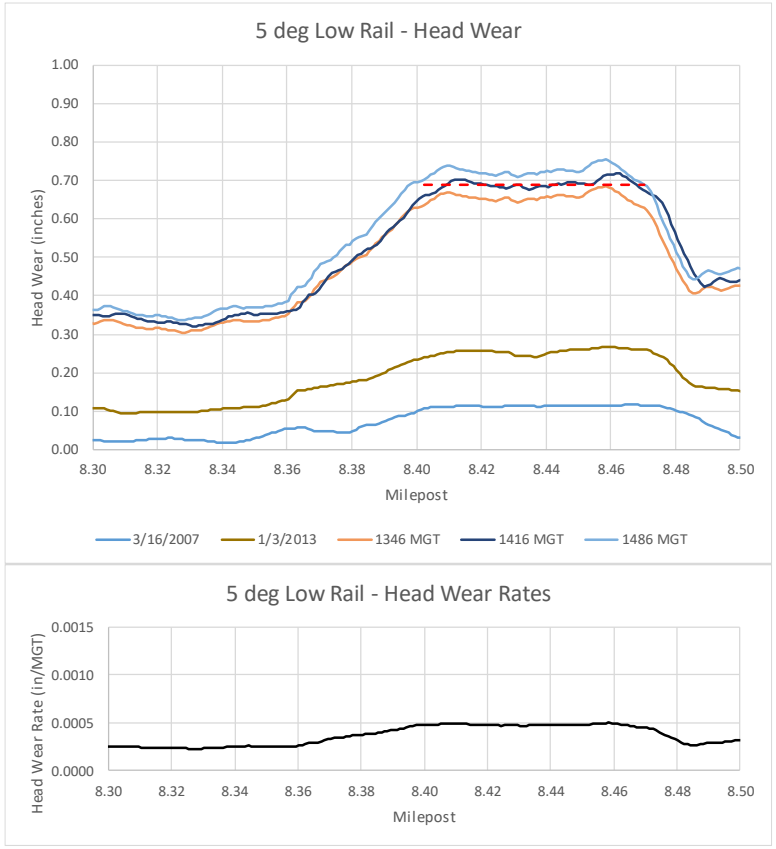


Figure 59 5.0 degree curve low rail wear results from ARIMA processing and forecasting.

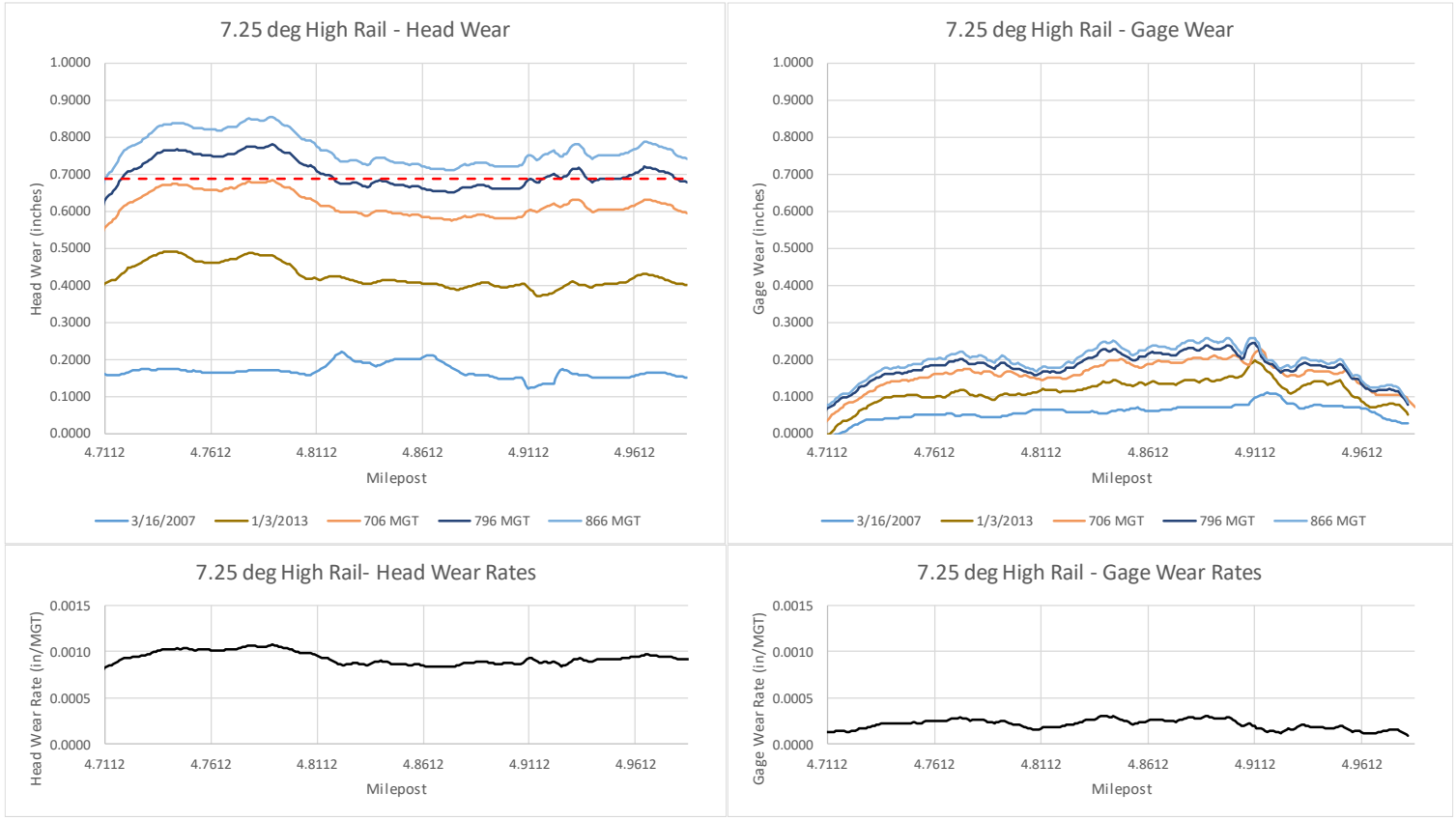


Figure 60 7.25 degree curve high rail wear results from ARIMA processing and forecasting.

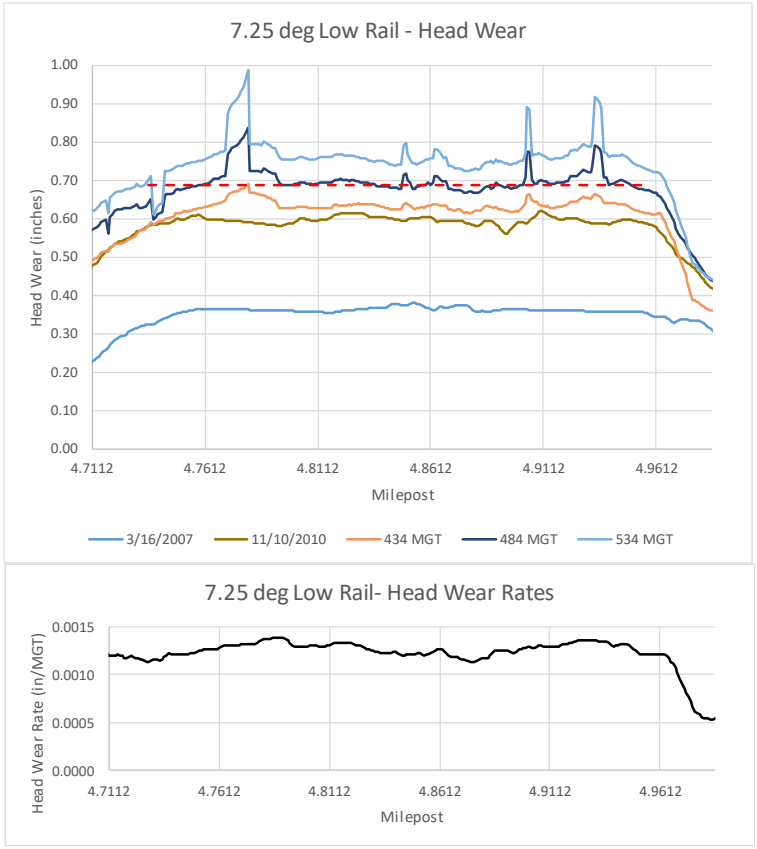


Figure 61 7.25 degree curve low rail wear results from ARIMA processing and forecasting.

Figure 58 through Figure 61 easily allow for the visualization of locations of severe wear (and corresponding severe wear rates), distribution of wear rates through the curve and time to condemnation. Note the unstable nature of the ARIMA process in Figure 61. The progression of spikes in the last wear measurement are carried through, and filtering of the distribution may be required.

Statistics for the wear rates determined using the ARIMA process are summarized in Table 13. Note that for ease of interpretations, the wear rates are shown as inches/100MGT.

Table 13 Statistics of Wear Rates Determined Through Curve Using ARIMA Process

	Wear Rates (in/100MGT)							
	5 Deg Curve				7.25 Deg Curve			
	High Rail		Low Rail		High Rail		Low Rail	
	Head	Gage	Head		Head	Gage	Head	
Minimum	0.0427	0.0003	0.0448	Minimum	0.0823	0.0092	0.1133	
Maximum	0.0574	0.0137	0.0499	Maximum	0.1074	0.0311	0.1390	
Average	0.0515	0.0081	0.0479	Average	0.0932	0.0211	0.1273	
Std Dev	0.0036	0.0028	0.0011	Std Dev	0.0069	0.0053	0.0060	

The magnitudes of the wear rates show that head wear significantly outpaces the gage wear in the high rail, and that the wear rates in the 7.25-degree curve are higher than those of the 5 degree curve. In particular, the low rail head wear rate is nearly three times higher in the sharper curve.

MDN Process

The mixture density network models were run for these two curves. The conditional density distributions were determined for the characteristics available for these segments (year installed, degree of curvature, super elevation, speed, etc.).

The resulting conditional distribution parameters (μ, b) are shown in Table 14, along with the least squares fit (Actual) of the wear data, corresponding to the average wear rate in the full body of the curve. The average rate from the ARIMA process is included for comparison purposes.

Table 14 MDN Conditional Density Parameters[15]

					Head Wear Rates (in/100MGT)						Gage Wear Rates (in/100MGT)				
Deg	SE (in)	Ann MGT	Rail		ARIMA	Actual	μ	b	C		ARIMA	Actual	μ	b	C
5	1.5	31.35	High		0.0515	0.0523	0.0511	0.0174	46.9%		0.0081	0.0096	0.0341	0.0058	99.3%
7.25	1.75	31.35	High		0.0932	0.1022	0.1013	0.0182	47.6%		0.0211	0.0502	0.0496	0.0060	45.2%
5	1.5	31.35	Low		0.0479	0.0476	0.0482	0.0119	52.6%						
7.25	1.75	31.35	Low		0.1273	0.1150	0.0766	0.0126	2.4%						

[15] C is the classification parameter and is discussed in the next section.

Table 14 shows that the ARIMA wear rates and least squares fit of the full body averages (Actual) are very similar, as expected. The minor differences are expected due to the difference in averaging techniques; ARIMA averages fit rates, Least squares averages wear measurements and fits a rate. In addition, the conditional probability density parameters can be used to show the bounded wear rates and how the Actual rates relate to the expected rates based on the input independent variables.

Classification

The classifications[16] (C) of the wear rates, based on Equation (5.20), are also shown in Table 14. While the majority of the calculated wear rates are near the predicted conditional rates (+/- 5%), the head wear on the low rail of the 7.25 degree curve shows C = 2.4%. this means that the segment is behaving very poorly as a result of the unexpected severe head wear rate. In addition, the gage wear rate on the high rail of the 5 degree curve shows C = 99.3%, indicating that the rail is behaving extremely well with a much lower gage wear rate than expected based on similar segments (Actual << Mu).

Thresholds will be discussed in the next major section, as extreme values may indicate excellent/poor performance, or may in fact be outliers to be analyzed further.

Forecast

While the ARIMA process is a step forward process, and by definition provides a forecast, the wear rates determined in the MDN approach can also be used to forecast maintenance/renewal. Recognizing that the head wear rates are much greater than the gage wear rates, the forecasts for head wear for both the low and high rails for both curves are shown in Table 15[17].

This table has two parts; the first is the estimated life of the rail in accumulated MGT, and the second is the years to replacement, based on the life in MGT, the accumulated MGT to date, and the last annual MGT on the rail. Equation (5.18) was used to develop forecasts for each determined wear rate, and the ARIMA forecasts are included, for a replacement limit of 11/16".

The forecasts are explained as follows:

- Actual: least squares rate based on averages of wear measurements in the full body of the curve
- ARIMA
    - First: first instance a wear measurement will exceed the wear limit
    - 50%: 50% of the full body will exceed the wear limit

[16] C = 50% indicates the calculated wear rate is the same as the conditional target wear rate. As C goes to zero the segment performs worse, and as C goes to 100, the segment performs better.

[17] Note that the model evaluates the head and gage wear independently, and the mode that controls replacement is determined. For purposes of discussion, the gage results are left off.

- 100%: 100% of the full body will exceed the wear limit
- MDN: using middle 50% probability
  - μ-Δμ: lower 50% probability band
  - μ: conditional probability
  - μ+Δμ: upper 50% probability band

Table 15 Comprehensive Forecast for Head Wear for Two Segments

						Rail Lives in MGT						
			Head			ARIMA			MDN			
Deg	Ann MGT	Rail	MGTo	Wo (in)	Actual	First	50%	100%	μ-Δμ	μ	μ+Δμ	
5	31.35	High	477	0.244	1325	1257	1347	1577	1612	1344	1179	
7.25	31.35	High	477	0.418	741	706	796	866	781	743	714	
5	31.35	Low	557	0.255	1466	1346	1416	1486	1640	1454	1322	
7.25	31.35	Low	416	0.598	494	434	484	534	548	533	521	
					Years to Replacement based on Last Annual MGT							
			Head			ARIMA			MDN			
Deg	Ann MGT	Rail	MGTo	Wo (in)	Actual	First	50%	100%	μ-Δμ	μ	μ+Δμ	
5	31.35	High	477	0.244	27.1	24.9	27.7	35.1	36.2	27.7	22.4	
7.25	31.35	High	477	0.418	8.4	7.3	10.2	12.4	9.7	8.5	7.5	
5	31.35	Low	557	0.255	29.0	25.2	27.4	29.6	34.5	28.6	24.4	
7.25	31.35	Low	416	0.598	2.5	0.6	2.2	3.8	4.2	3.7	3.3	

It can be seen from Table 15 that the projected lives and replacement dates for the Actual, ARIMA at 50% and conditional probability (μ) are relatively close to one another, however, the differences can amount to large savings. While the entire model results are presented, it is important to note that in some case, actual inspection data may not be available. Thus, the MDN results can be used to show a stochastic forecast of when the rail will be required for replacement. When data is available, the Actual and ARIMA forecasts can be utilized. The ARIMA forecast range allows for the determination of risk, by evaluating what percentage of the rail will exceed the limit at any given point in the future.

Summary

The model was applied on two specific rail segments for both the high and low rails. Since sufficient data was available to employ all processes, from data alignment to forecast, the entire model was utilized. However, if rail segments do not have sufficient inspection data to forecast using ARIMA, or even have no inspection data, the MDN process can be used to develop conditional probability densities. The probability density distribution can then be utilized to forecast rail wear and provide a stochastic range of dates for rail replacement.

The ability to quantify performance through the stochastic classification method allows for maintenance interventions, such as lubrication or rail grinding, to extend the life of the rail when it is in the early stages of rapid wear.
The results show consistency between methods as well as reasonable values, as will be discussed later in this chapter

System Analysis

While the previous section showed the analysis framework deployed in detail (on a micro level), railroads often need to make system wide, or route wide (macro level) decisions. This section utilizes a larger subset of data to provide comprehensive results for the line segment being analyzed. The primary results used in this analysis are from the MDN component of the model. Each rail segment (defined as a curve or tangent stretch of track, averaging 0.25 mile for curves and 0.5 mile for tangents) was analyzed to provide an average wear rate in the segment (tangents between curves and full bodies of curves). The MDN component was then applied to provide a conditional density for each segment. These values were then used to classify the segment and forecast annual rail requirements.

Classification

Utilizing the calculated wear rates and the calculated conditional densities for each rail segment, a classification parameter was calculated using Equation (5.19). The distribution of classification for each rail and wear location are shown in Figure 62.

Recall that $C > 50\%$ is well performing rail (lower wear rate), and $C < 50\%$ is poor performing rail (higher wear rate), with respect to the conditional (expected) wear rate. Rails at the extreme tails of the distribution are likely outliers. For example, the high population of 100% rails for the high rail gage wear represent calculated wear rate less than zero, which fall very far from the conditional wear rate. This calculated wear rate of <= 0 is often a function of the head wear outpacing the gage wear, since the location of measurement of the gage wear point changes, and may result in an appearance of growth, i.e., appearing as if metal has been added, resulting in a negative rate.

Figure 62 shows that the majority of high and low head wear rates are well behaved, falling between 40% and 60%, with a mean slightly on the well performing side. There are a number of segments at 100% associated with bad data showing negative wear rates, and there are several segments < 40% and > 60% that show unexpected performance.

The high rail gage wear rates are flatly distributed and shifted slightly to the well performing side, with a large population of outliers. The tangent rail head wear rates are uniformly distributed with a slight shift to the poor performing side. The further away from 50% a classification is, the more prone it is to require maintenance intervention. Thus, segments should be identified based on their classification and investigated.

Thresholds were arbitrarily selected as shown in Table 16, along with counts of segment in the respective bands. These thresholds will require verification with railway personnel to ascertain

validity. However, it is useful to see the number of segments that require additional attention and possible intervention.

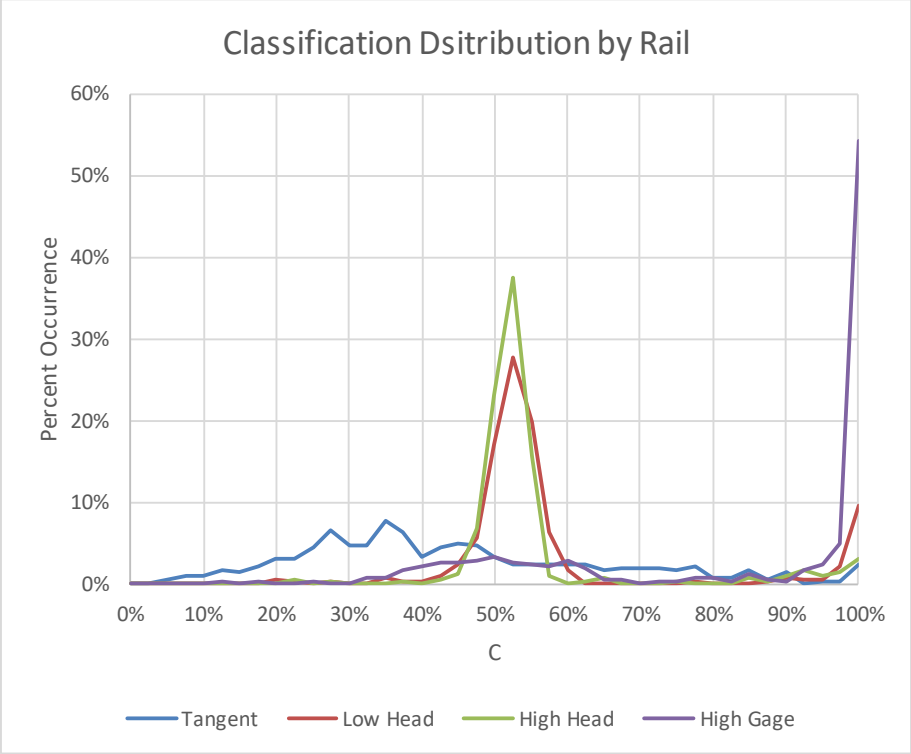


Figure 62 Distribution of wear rate classifications by rail and wear location.

Table 16 Distribution of Segments Based on Classification Thresholds and Occurrence

		Occurrence			
C	Description	Tangent	Low Head	High Head	High Gage
0 - 5	Outlier	2	0	0	2
5 - 20	Intervene	50	4	4	5
20 -40	Analyze	194	15	10	40
40 - 60	As Exepcted	128	498	516	134
60 - 80	Analyze	68	6	12	32
80 - 95	Intervene	22	10	24	24
95 - 100	Outlier	13	75	34	383

Forecast

Utilizing the calculated rate as well as the conditional density rates, several long range forecasts can be determined. The same process used in Section 0 is followed, and the results are summarized for all the segments. The stochastic forecast for tangent rail is shown in Figure 63.

The majority of the tangent rail still has more than 30 years of service life remaining with respect to wear. This is as expected since tangent rail is most often replaced due to fatigue. While the stochastic forecast using the expected rate compares well with the actual rate forecast, probability band forecast allows for understanding the possible fluctuation in rail replacement requirements over time.

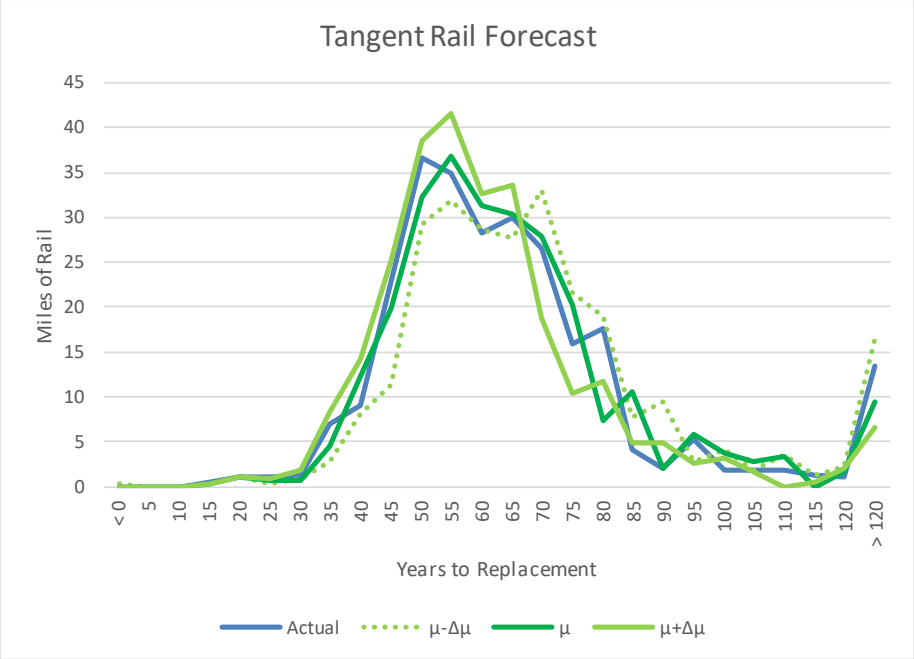


Figure 63 Forecast for tangent rail for entire route.

The low rail forecast is shown in Figure 64. This forecast shows that the forecast from the actual rate and expected rate are nearly identical, thus, the model can be used for a similar route to make predictions where no inspection data may be available. This forecast also shows a large amount of rail due within the next 10 years, and in fact, in year 10, while 24 miles is expected to be required, this could be between 21 and 29 miles of rail.

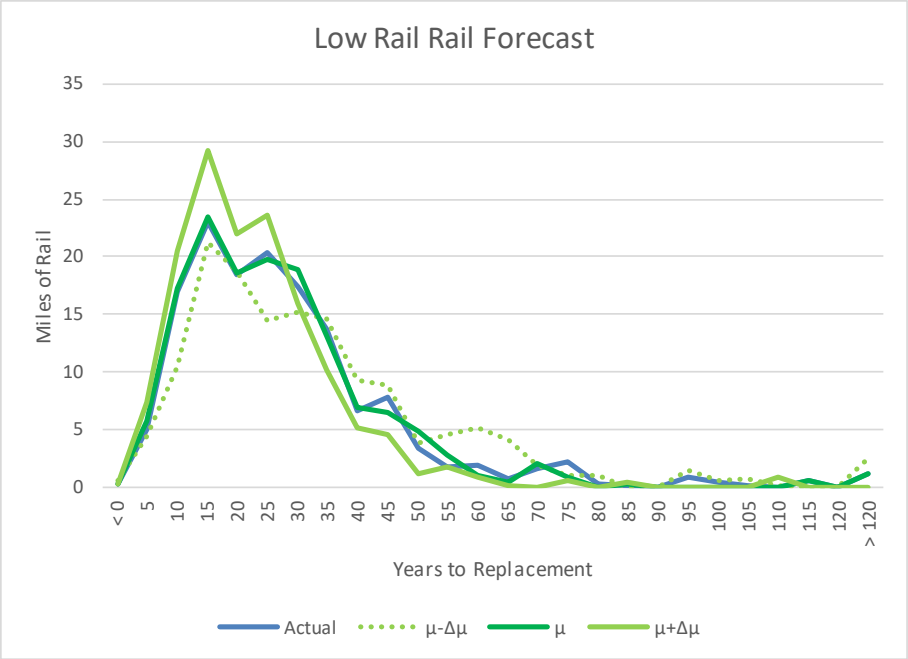


Figure 64 Forecast for low rail for entire route.

The forecast for the high rail is provided in Figure 65. Note that this forecast was developed considering both head wear rate and gage wear rate to determine the controlling wear rate. The results showed that 85% of the segments required replacement due to head wear. Rail requirements for the high rail are similar to those of the low rail.

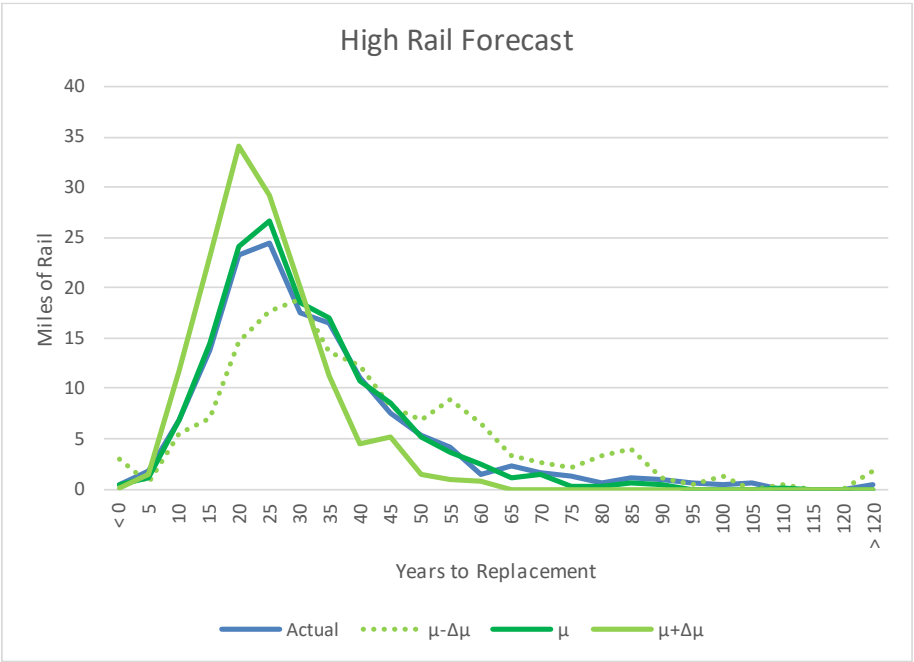


Figure 65 Forecast for high rail for entire route.

To better understand the overall requirements by rail, Figure 66 shows the forecast for each rail position (tangent, low and high) utilizing the conditional rate (μ). The forecast shows that the low rail is requiring replacement slightly sooner than the high rail, and that the tangent rails are pushed further out in time. The proximity of the high and low rail forecasts allows for determining if both rails in a curve should be replaced simultaneously, or if one of the rails should remain in track a few years longer to maximize asset life.

Finally, the forecast for the tangent, low and high rails are summed to give an overall rail requirement by year (See Figure 67). This figure clearly shows the variation in rail requirements over time, as well as the stochastic nature of the forecast.

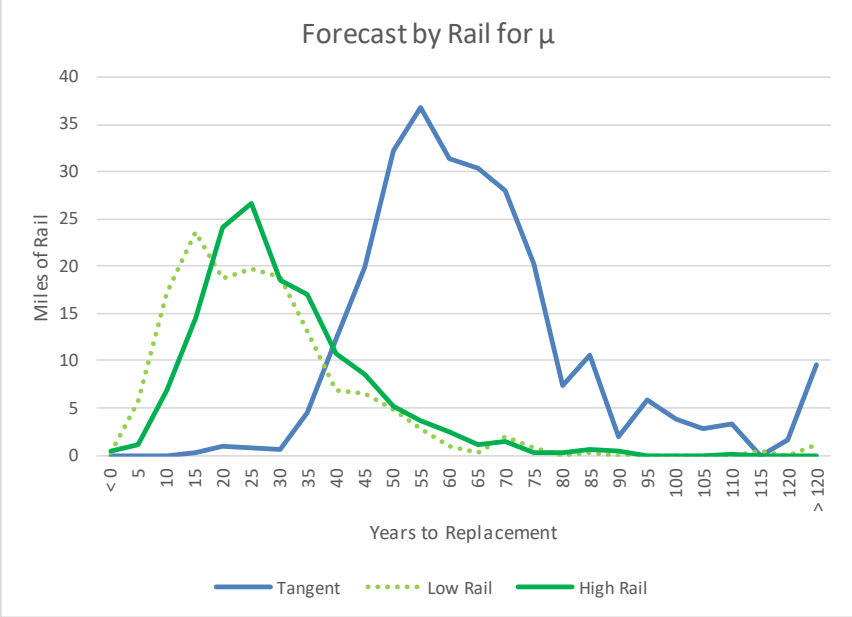


Figure 66 Forecast by rail location for the conditional wear rate, μ.

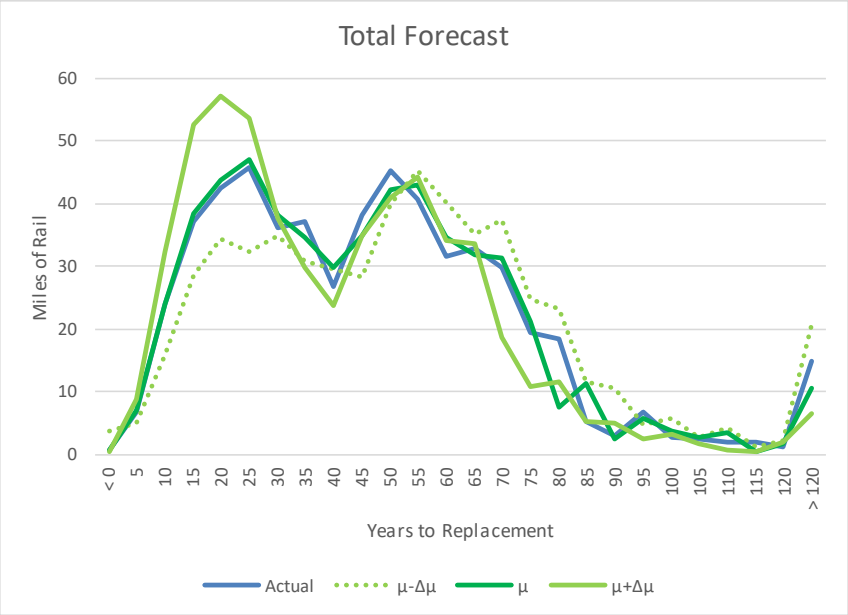


Figure 67 Total forecast for entire route.

Discussion

Utilizing the stochastic processes to determine rail requirements for an entire route (or railroad) as well as classify the rail segments as to their wear performance allows for better decision making with regard to rail maintenance and renewal. The similarities in forecast with the actual wear rate and the conditional wear rate lend validation to using the MDN model as a first step in forecasting when inspection data is limited.

It is important to understand the life of the rail. The conditional model allows for the comparison of rail wear lives. While the many input factors have a direct effect on the wear rate and its resulting life, a simple calculation of the median rail wear life is useful, and is shown in Table 17. This table shows median rail wear lives in accumulated MGT. Of course, evaluating the effects of the causal factors, and defining maintenance interventions to prolong rail life is of paramount importance to the railways.

Table 17 Median Rail Wear Lives by Rail

	Median Rail Wear Lives (MGT)			
	Actual	μ-Δμ	μ	μ+Δμ
Tangent Rail	2974	3158	3000	2888
Low Rail	1312	1444	1290	1193
High Rail	1395	1628	1368	1210

Taking advantage of the stochastic nature of the model allows for changing the central probability band and understanding the risks associated with the unknown latent variables, and their effects on rail requirements.

Utilizing the forecast lives, a comprehensive maintenance and renewal plan can be developed. This can often be challenging, in that practical and logistical concerns must be accounted for.

CONCLUDING REMARKS

The focus of this research was developing a new approach to rail wear modeling utilizing available data and emerging data science analysis techniques. Wear, including wear of rail steels, has been the topic of extensive research for well over 100 years. Rail steel is expensive and is a vital asset to the railway in order to maintain safe and reliable operations.

Much research has been performed in the laboratory to understand the relationships of load, contact geometry, metallurgy and lubrication to understand the effects of these components on wear rate, predominantly in the form of wear coefficients. Several in depth field studies have been conducted to monitor wear and relate the rate of wear to everyday railway conditions and operations. However, the many factors that affect rail wear are not, or cannot, be measured in practice. Thus, all things being equal, a rail may wear at different rates even at the same location.

Inspection data is readily available to railways in the form of rail profiles and corresponding lineal wear measurements at specific locations across the railhead. These inspections can be quite frequent (weekly) and result in a tremendous amount of wear data. Significant factors that affect rail wear are also known or measured during inspections. This data has been used in the past to develop rail wear models, as well as linearly forecast rail wear with some success. However, shortcomings due to unknowns made the forecasts less reliable.

The focus of this research was to address some of these shortcomings using stochastic processes to determine ranges of more reliable forecasts, thus taking advantage of the variation in the data, which was directly related to the unknowns. With the continued development of low cost computing power, and emergence of advance data science techniques, the ability to better understand the variance in measured data was made possible.

Stochastic techniques were adapted, incorporating Laplace distribution parameters, formally unknown to the industry, and were made immediately apparent in the Exploratory Data analysis. ARIMA processing of segments for forecasting, and MDNs for developing conditional density functions for each rail segment resulted in stochastic classification and forecasts that can be used to make better maintenance and renewal decisions.

Impact of the Research

The results of the research have direct impact on the railway industry for optimizing rail maintenance and renewal. In particular, the ability to classify a rail segment's wear performance allows the railway engineer to make more informed decision regarding maintenance. Some examples include:

- Determine why a rail segment is performing better than expected:
    - Better rail emetallurgy?
    - Better lubrication practices?
    - Optimal grinding?
    - Better vertical support?
    - Proper curve geometry; elevation, length of spirals?

- Determine why a rail segment is performing worse than expected:
    - Poor track geometry, rough track?
    - Improperly elevated curves for speed of operations?
    - Lubrication ineffective?
    - Improper grinding, or transverse rail profile?
- Monitor the as expected wear rates:
    - Compare routes
    - Compare over time to see if improving in general

In addition, the approaches developed can be applied to other readily available inspection data for various railway assets.

Challenges

The data that was used for this research was in poor condition. Some of the inspection data was obviously out of calibration. While algorithms were put into the code to clean up the data, they were not 100% effective.

Only 277 miles of data were used for this research, from a railway that has 20,000+ miles of track. More data (and cleaner data) would help to improve the models. In addition, access to the data was difficult, and code scripts to import and manage the data would accelerate the research.

The data is quite large. Analyzing an entire railway may be beyond the capabilities of personal computer, thus, access to higher power computing resources is required.

Conclusions

The results of the stochastic modeling proved to be useful and reliable when validating the techniques developed and adapted herein.

The data showed explicitly that rail wear rates do in fact vary significantly based on many factors, and that for like segments (all known factors being equal), there is still variation with the measured wear data (and corresponding wear rates). This is due to the unknown factors manifesting themselves into the inspection data. The wear rate data was observed to follow a Laplace distribution, with defined mean and variance.

Traditional modeling techniques have focused on predicting target values based on independent variables, calibrated to laboratory and field experience. The stochastic approach developed herein (Rail-WASP) derives suitable ranges of results that are a direct function of the knowns and unknowns, in particular, the conditional mean and conditional variance. While traditional methods allow for developing expected values, the stochastic methodology allows for understanding a range of expected values, thus allowing practitioners to better understand time to failure, and associated risk.

As a comparison, Rail-WASP was evaluated against the Reiner-Staplin model (Reiner & Staplin, 1984), as this model provided related results for primary inputs used in Rail-WASP. Specifically, the Reiner-Staplin model was evaluated for the following parameters:

- Axle Load: P = 32.9 kips
- Annual MGT: T = 30 MGT/year
- Wear Limit: $W_t = W_c$ = 5/8 inch
- Grade: G = 0% - 1%
- Curvature: $D = 0^{\circ} - 14^{\circ}$
- Constants: $K_t = K_c = 0.3$
- Note that Reiner-Staplin used a value of 1.0 for standard carbon rail with no lubrication for rail rolled in the 1960's and 1970's, while Rail-Wasp was evaluated for rail rolled in 2000. Rail hardness has increased significantly over that time period. The FAST report (Steele, 1982) concluded that the effect of hardness (H) for non-lubricated rail on wear rate (W) is of the form: $\log(W) = -0.014-0.0066H$. Per the FAST report, assuming standard carbon rail to have a hardness $H = 260$[18], and 2000 rail to have a hardness, $H = 340$[19], results in a reduction of the coefficient from 1.0 to 0.3.
- The corrosion component of the Reiner-Staplin equation was ignored (set to zero) as the wear rates developed in Rail-Wasp were over a shorter timeframe (6 years), whereby significant corrosion would not contribute to the overall wear.

The results of this comparison are presented for 136RE rail at 20 mph operating speed, no lubrication and 30 MGT/year. The Reiner-Staplin model results in a rail life in accumulated MGT and was transformed to a wear rate (inches/MGT) by dividing the life in MGT into the wear limit in inches, to be consistent with Rail-WASP. **Figure 68** and Figure 69 show the comparison for of Rail-WASP and the Reiner-Staplin model for tangent rail and the low rail in a curve respectively.

Figure 68 shows that Rail-WASP has slightly more sensitivity to grade and predicts better wear performance of the tangent rail than the Reiner-Staplin model. In particular, it is important to note the conditional variance of the predicted wear rate.

[18] Representative of rail hardness for standard rail during the time period of the Reiner-Staplin model.

[19] Representative of modern rail steels.

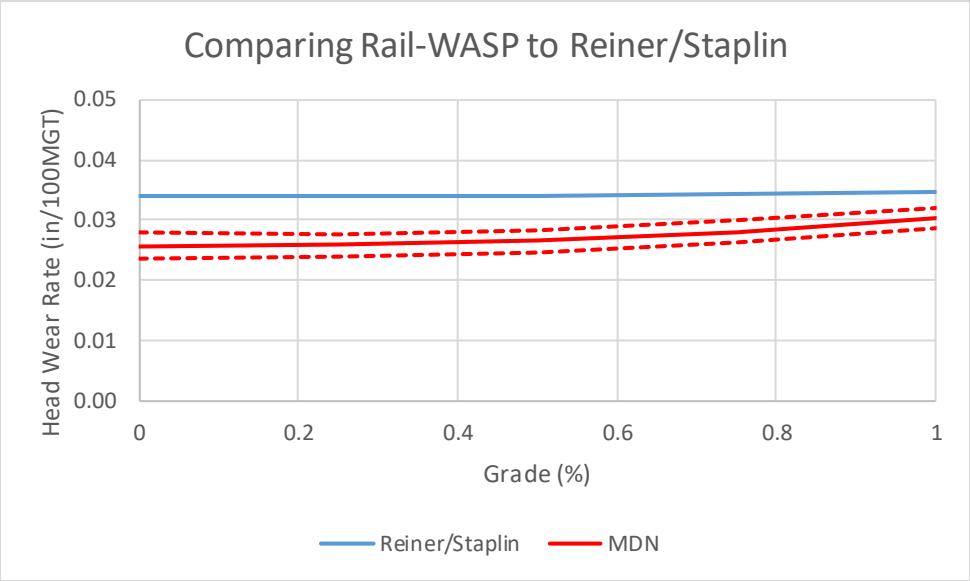


Figure 68 Comparison of Rail-WASP and Reiner-Staplin for tangent rail as a function of grade.

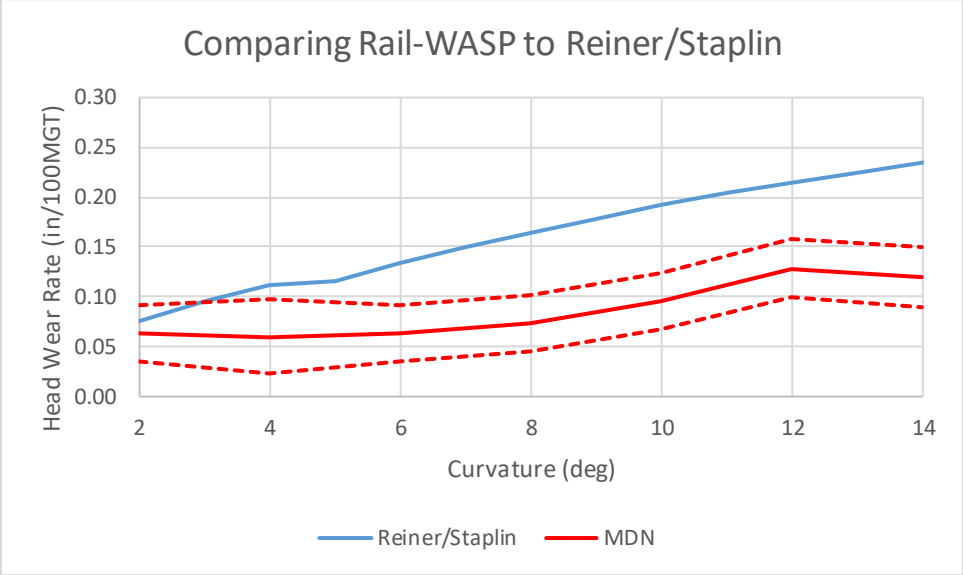


Figure 69 Comparison of Rail-WASP and Reiner-Staplin for the low rail as a function of curvature.

Figure 69 shows that the data used to develop Rail-WASP does not show a significant effect to curvature (as does the Reiner-Staplin model[20]), though the trend is clearly the same. Again, the conditional variance provides a range of expected wear rates.

Having an acceptable range of expected wear rates as output allows the railway engineer to develop of range of expected renewal dates, with corresponding probability (classification) using the Laplace distribution.

In addition, a measured wear rate can be compared to the expected conditional distribution to classify the performance of a particular rail segment. This can be extremely valuable for identifying rail segments that are candidates for maintenance intervention (lubrication, grinding) to extend the life of the rail, which has significant economic returns.

Finally, the approach developed herein allows the railway engineer to make more informed decisions as to rail maintenance and renewal, at both the system and subsystem level. This allows for better control of risk associated with the structural integrity of the rail, and is particularly true when budget constraints are imposed.

[20] Note that the Reiner-Staplin model provides a rail life value for curved rail and does not identify rail location or rail mode. For this comparison the low rail head wear model from Rail-WASP was utilized.

.

ACKNOWLEDGEMENTS

The authors wish to thank and acknowledge the US Department of Transportation, University Transportation Center Program (RailTEAM UTC) for funding support for this research. Additionally, the authors wish to acknowledge CSX Transportation for providing data for this research.

REFERENCES

1. Abbasnejad, S., & Mirabadi, A. (2017). Predicting the failure of railway point machines by using Autoregressive Integrated Moving Average and Autoregressive-Kalman methods, *0*(0), 1–10. https://doi.org/10.1177/0954409717748790
2. Aceituno, J. F., Wang, P., Wang, L., & Shabana, A. A. (2017). Influence of rail flexibility in a wheel/rail wear prediction model. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *231*(1), 57–74. https://doi.org/10.1177/0954409715618426
3. Adams, W. B. (1865). On the mechanical conditions of railway working to prevent destructive wear and risk. *Journal of the Franklin Institute*, *79*(4), 217–226. https://doi.org/http://dx.doi.org/10.1016/0016-0032(65)90094-3
4. Alarcon, G. I., Burgelman, N., Meza, J. M., Toro, A., & Li, Z. (2016). Power dissipation modeling in wheel/rail contact: Effect of friction coefficient and profile quality. *Wear*, *366–367*, 217–224. https://doi.org/10.1016/j.wear.2016.04.026
5. Archard, J. F. (1953). Contact and rubbing of flate surfaces. *Journal of Applied Physics*, *24*(8), 981–988. https://doi.org/10.1063/1.1721448
6. Association of American Railroads. (2016). *Total Annual Spending, 2015 Data*. Retrieved from www.aar.org
7. Association of American Railroads. (2017). Class I Railroad Statistics. *Association of American Railroads*, 1–3. Retrieved from www.aar.org
8. Attoh-Okine, N. (1994). Time series analysis for ground penetrating radar (GPR) asphalt thickness profile. *Applied Stochastic Models and Data Analysis*, *10*(3), 153–167. https://doi.org/10.1002/asm.3150100303
9. Babcock, M. W., Lu, X., & Norton, J. (1999). Time series forecasting of quarterly railroad grain carloadings, *35*, 43–57.
10. Bărbînţă, C. I., & Creţu, S. (2018). Influence of roughness and wear on pressure distribution and stress state at wheel - Rail contact. *IOP Conference Series: Materials Science and Engineering*, *444*(2). https://doi.org/10.1088/1757-899X/444/2/022003
11. Bishop, C. M. (1994). Mixture Density Networks.
12. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.
13. Box, G. E. ., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
14. Braghin, F., Lewis, R., Dwyer-Joyce, R. S., & Bruni, S. (2006). A mathematical model to predict railway wheel profile evolution due to wear. *Wear*, *261*(11–12), 1253–1264. https://doi.org/10.1016/j.wear.2006.03.025
15. Brouwer, R. K., & Pedrycz, W. (2003). Training a feed-forward network with incomplete data due to missing input variables. *Applied Soft Computing*, *3*, 23–36. https://doi.org/10.1016/S1568-4946(03)00003-6
16. Brouzoulis, J., Torstensson, P. T., Stock, R., & Ekh, M. (2011). Prediction of wear and plastic flow in rails-Test rig results, model calibration and numerical prediction. *Wear*, *271*(1–2), 92–99. https://doi.org/10.1016/j.wear.2010.10.021
17. Brownlee, J. (2017). A Gentle Introduction to the Box-Jenkins Method for Time Series Forecasting. Retrieved from https://machinelearningmastery.com/gentle-introduction-box-jenkins-method-time-series-forecasting/

18. Bucher, F., Dmitriev, A. I., Ertz, M., Knothe, K., Popov, V. L., Psakhie, S. G., & Shilko, E. V. (2006). Multiscale simulation of dry friction in wheel/rail contact. *Wear*, *261*(7–8), 874–884. https://doi.org/10.1016/j.wear.2006.01.046
19. Busquet, M., Baillet, L., Bordreuil, C., & Berthier, Y. (2005). 3D finite element investigation on the plastic flows of rolling contacts - Correlation with railhead microstructural observations. *Wear*, *258*(7–8), 1071–1080. https://doi.org/10.1016/j.wear.2004.03.069
20. Cannon, A. J. (2012). Computers & Geosciences Neural networks for probabilistic environmental prediction : C ondition a l D ensity E stimation N etwork C reation and E valuation (CaDENCE) in R. *Computers and Geosciences*, *41*, 126–135. https://doi.org/10.1016/j.cageo.2011.08.023
21. Carney, M., Dowling, J., & Lee, C. (2005). Predicting Probability Distributions for Surf Height Using an Ensemble of Mixture Density Networks. In *Proceedings of the 22nd International Conference on Machine Learning*. Bonn, Germany.
22. Chongyi, C., Chengguo, W., & Ying, J. (2010). Study on numerical method to predict wheel/rail profile evolution due to wear. *Wear*, *269*(3–4), 167–173. https://doi.org/10.1016/j.wear.2009.12.031
23. Clayton, P. (1995). Predicting the wear of rails on curves from laboratory data. *Wear*, *181–183*(PART 1), 11–19. https://doi.org/10.1016/0043-1648(95)90003-9
24. Costello, S. B., & Premathilaka, A. S. (2011). Stochastic Rail Wear Model for Railroad Tracks, *620*(2289), 1–16. https://doi.org/10.3141/2289-14
25. Coxe, W. E. C. (1879). Note on the wear of an iron rail. *Journal of the Franklin Institute*, *108*(2).
26. Croarkin, C., & Tobias, P. (Eds.). (2013). *NIST/SEMATECH e-Handbook of Statistical Methods*. US Department of Commerce.
27. Daves, W., & Fischer, F. D. (2002). Modelling of the plastification near the rough surface of a rail by the wheel-rail contact. *Wear*, *253*(1–2), 241–246. https://doi.org/10.1016/S0043-1648(02)00107-2
28. Dirks, B., & Enblom, R. (2011). Prediction model for wheel profile wear and rolling contact fatigue. *Wear*, *271*(1–2), 210–217. https://doi.org/10.1016/j.wear.2010.10.028
29. Dirks, B., Enblom, R., & Berg, M. (2016). Prediction of wheel profile wear and crack growth – comparisons with measurements. *Wear*, *366–367*, 84–94. https://doi.org/10.1016/j.wear.2016.06.026
30. Enblom, R. (2009). Deterioration mechanisms in the wheel-rail interface with focus on wear prediction: A literature review. *Vehicle System Dynamics*, *47*(6), 661–700. https://doi.org/10.1080/00423110802331559
31. Enblom, R., & Berg, M. (2005). Simulation of railway wheel profile development due to wear influence of disc braking and contact environment. *Wear*, *258*(7–8), 1055–1063. https://doi.org/10.1016/j.wear.2004.03.055
32. Enblom, R., & Berg, M. (2008). Proposed procedure and trial simulation of rail profile evolution due to uniform wear. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *222*(1), 15–25. https://doi.org/10.1243/09544097JRRT173
33. Franklin, F. J., Widiyarta, I., & Kapoor, A. (2001). Computer simulation of wear and rolling contact fatigue. *Wear*, *250*(251), 949–955. https://doi.org/10.1016/S0043-1648(01)00732-3

34. Fries, R. H., & Dávila, C. G. (1987). Wheel Wear Predictions for Tangent Track Running. *Journal of Dynamic Systems, Measurement, and Control*, *109*(4), 397. https://doi.org/10.1115/1.3143873
35. Ghonem, H., & Kalousek, J. (1984). A Quantitative Model to Estimate Rail Surface Fatigue. *Wear*, *97*, 65–81.
36. Ghonem, H., Kalousek, J., Stone, D. H., & Dibble, D. W. (1982). Observations of Rail Wear on Heavy Haul Railway Lines. In J. Kalousek, R. V. Dukkipati, & G. M. L. Gladwell (Eds.), *Contact Mechanics and Wear of Rail/Wheel Systems* (pp. 249–270). Vancouver, BC: University of Waterloo Press.
37. Hay, W. W. (1982). *Railroad Engineering* (Second Edi). New York: John Wiley & Sons.
38. Herzallah, R., & Lowe, D. (2004). A mixture density network approach to modelling and exploiting uncertainty in nonlinear control problems. *Engineering Applications of Artificial Intelligence*, *17*, 145–158. https://doi.org/10.1016/j.engappai.2004.02.001
39. Husmeier, D., & Taylor, J. G. (1997). Predicting Conditional Probability Densities of Stationary. *Neural Networks*, *10*(3), 479–497.
40. Husmeier, D., & Taylor, J. G. (1998). Neural Networks for Predicting Conditional Probability Densities : Improved Training Scheme Combining EM and RVFL. *Neural Networks*, *11*(1), 89–116.
41. Ignesti, M., Malvezzi, M., Marini, L., Meli, E., & Rindi, A. (2012). Development of a wear model for the prediction of wheel and rail profile evolution in railway systems. *Wear*, *284–285*, 1–17. https://doi.org/10.1016/j.wear.2012.01.020
42. Innocenti, A., Marini, L., Meli, E., Pallini, G., & Rindi, A. (2014). Development of a wear model for the analysis of complex railway networks. *Wear*, *309*(1–2), 174–191. https://doi.org/10.1016/j.wear.2013.11.010
43. Innotrack. (2008). *D1 . 3 . 2 The state of the art of the simulation of vehicle track interaction as a method for determining track degradation rates Part One – Strategic Models*.
44. Jendel, T. (2002). Prediction of Wheel Profile Wear - Comparisons with Field Measurements. *Wear*, *253*, 89--99. https://doi.org/10.1016/S0043-1648(02)00087-X
45. Jeong, M.C., Kim, J. H., Lee, J. H., Kang, Y. ., & Kong, J. S. (2010). Track Life Evaluation through Regression Analysis of Subway Wear Data. *Journal of KSMI*, *4*(4), 86–93.
46. Jeong, Min Chul, Lee, S. J., Cha, K., Zi, G., & Kong, J. S. (2019). Probabilistic model forecasting for rail wear in seoul metro based on bayesian theory. *Engineering Failure Analysis*, *96*(December 2017), 202–210. https://doi.org/10.1016/j.engfailanal.2018.10.001
47. Jin, X., Wen, Z., Xiao, X., & Zhou, Z. (2007). A numerical method for prediction of curved rail wear. *Multibody System Dynamics*, *18*(4), 531–557. https://doi.org/10.1007/s11044-007-9073-3
48. Jin, Y., Ishida, M., & Namura, A. (2011). Experimental simulation and prediction of wear of wheel flange and rail gauge corner. *Wear*, *271*(1–2), 259–267. https://doi.org/10.1016/j.wear.2010.10.032
49. Kalker, J. J. (1990). *Three-Dimensional Elastic Bodies in Rolling Contact*. Dordrecht, Netherlands: Kluwer Academic Publishers.
50. Kalousek, J., & Bethune, A. E. (1978). Rail Wear Under Heavy Traffic Conditions. In D. H. Stone & G. G. Knupp (Eds.), *Rail Steels - Developments, Processing, and Use* (pp.

63–79). Denver, CO: American Society for Testing and Materials.
51. Kingma, D. P., & Ba, J. L. (2015). ADAM: A Method for Stochastic Optimization. *ICLR*, 1–15.
52. Kongcharoen, C., & Kruangpradit, T. (2013). Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) Model for Thailand Export, (June), 1–8.
53. Li, X., Yang, T., Zhang, J., Cao, Y., Wen, Z., & Jin, X. (2016). Rail wear on the curve of a heavy haul line—Numerical simulations and comparison with field measurements. *Wear*, *366–367*, 131–138. https://doi.org/10.1016/j.wear.2016.06.024
54. Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303. https://doi.org/10.1093/biomet/65.2.297
55. Magel, E., & Kalousek, J. (2002). The application of contact mechanics to rail profile design and rail grinding. *Wear*, *253*(1–2), 308–316. https://doi.org/10.1016/S0043-1648(02)00123-0
56. Magel, E., Kalousek, J., & Caldwell, R. (2005). A numerical simulation of wheel wear. *Wear*, *258*(7–8), 1245–1254. https://doi.org/10.1016/j.wear.2004.03.033
57. Magel, E., Mutton, P., Ekberg, A., & Kapoor, A. (2016). Rolling contact fatigue, wear and broken rail derailments. *Wear*, *366–367*, 249–257. https://doi.org/10.1016/j.wear.2016.06.009
58. Magnus, D. L. (2014). The Digital Railway, From Kilobytes to Terabytes, a Rail Measurement Perspective. In *Track Maintenance Planning in the Era of Big Data*. Newark: university of Delaware.
59. Mart´ın Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow,Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser,Manjunath Kudlur, Josh Levenberg, Dan Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, Google Research. (2015). {TensorFlow}: Large-Scale Machine Learning on Heterogeneous Systems. Retrieved from https://www.tensorflow.org/
60. Men, Z., Yee, E., Lien, F., Wen, D., & Chen, Y. (2016). Short-term wind speed and power forecasting using an ensemble of mixture density neural networks. *Renewable Energy*, *87*, 203–211. https://doi.org/10.1016/j.renene.2015.10.014
61. Miazhynskaia, T., Dorffner, G., & Dockner, E. J. (2003). Risk Management Application of the Recurrent Mixture Density Network Models. In Kaynak (Ed.), *ICANN/iCONIP* (pp. 589–596). Heidelberg, Germany: Springer-Verlag.
62. Milenkovic, M., Svadlenka, L., Melichar, V., & Bojovic, N. (2016). SARIMA modelling approach for railway passenger flow forecasting, The journal TRANSPORT (January). https://doi.org/10.3846/16484142.2016.1139623
63. Morton, K., Cannon, D. ., Clayton, P., & Jones, E. G. (1978). The Assessment of Rail Steels. In D. H. Stone & G. G. Knupp (Eds.), *Rail Steels - Developments, Processing, and Use* (pp. 80–98). Denver, CO: American Society for Testing and Materials.
64. Nikolaev, N., Tino, P., & Smirnov, E. (2013). Neurocomputing Time-dependent series variance learning with recurrent mixture density networks. *Neurocomputing*, *122*, 501–512. https://doi.org/10.1016/j.neucom.2013.05.014

65. Olofsson, U., & Telliskivi, T. (2003). Wear, plastic deformation and friction of two rail steels - A full-scale test and a laboratory study. *Wear*, *254*(1–2), 80–93. https://doi.org/10.1016/S0043-1648(02)00291-0
66. Pacella, M., & Anglani, A. (2008). A Real-Time Condition Monitoring System by using Seasonal ARIMA Model and Control Charting. In *Methods, Models and Information Technologies for Decision Support Systems*. Lecce: Università del Salento.
67. Palese, J.W., Zarembski, A. M., Hartsough, C. M., & Ozturk, S. (2015). Use of Switch Profile Data for Enhanced Analysis of Wheel Rail Behavior at the Switch Point. In *Proceedings of the 2015 IHHA Conference*. Perth.
68. Palese, Joseph W, Zarembski, A. M., & Attoh-Okine, N. O. (2019). *Railteam UTC Report - The Laplace Distribution in Railway Track Degradation – A Case Study for Rail Wear Rev 2*. Newark, DE.
69. Pearce, T. G., & Sherratt, N. D. (1991). Prediction of wheel profile wear. *Wear*, *144*(1–2), 343–351. https://doi.org/10.1016/0043-1648(91)90025-P
70. Pombo, J., Ambrósio, J., Pereira, M., Lewis, R., Dwyer-Joyce, R., Ariaudo, C., & Kuka, N. (2011). Development of a wear prediction tool for steel railway wheels using three alternative wear functions. *Wear*, *271*(1–2), 238–245. https://doi.org/10.1016/j.wear.2010.10.072
71. Prematilaka, A., & Costello, S. B. (2010). *The Development of a Rail Wear Prediction Model – Lessons, the Challenges, and the Way Ahead*.
72. R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria.: R Foundation for Statistical Computing.
73. Ramalho, A. (2010). A reliability model for friction and wear experimental data. *Wear*, *269*(3–4), 213–223. https://doi.org/10.1016/j.wear.2010.03.023
74. Ramalho, A. (2015). Wear modelling in rail-wheel contact. *Wear*, *330–331*, 524–532. https://doi.org/10.1016/j.wear.2015.01.067
75. Reiner, I. A., & Staplin, D. E. (1984). Empirical Rail Wear Model: AAR Report No. WP-104. *American Railway Engineering Association*, *85*(697), 237–267.
76. Roney, M. D. (1982). Economic Aspects of the Wear of Rail on Canadian Railways. In J. Kalousek, R. V. Dukkipati, & G. M. L. Gladwell (Eds.), *Contact Mechanics and Wear of Rail/Wheel Systems* (pp. 271–292). Vancouver, BC: University of Waterloo Press.
77. Rovira, A., Roda, A., Marshall, M. B., Brunskill, H., & Lewis, R. (2011). Experimental and numerical modelling of wheel-rail contact and wear. *Wear*, *271*(5–6), 911–924. https://doi.org/10.1016/j.wear.2011.03.024
78. Sandberg, C. P. (1869). The manufacture and wear of rails. *Journal of the Franklin Institute*, *87*(1), 17–27. https://doi.org/10.1016/0016-0032(69)90331-7
79. Schittenkopf, C., Dorffner, G., & Dockner, E. J. (1999). Forecasting Time-dependent Conditional Densities : A Semi-non- parametric Neural Network Approach. *Journal of Forecasting*, (February), 244–263.
80. Shebani, A., & Iwnicki, S. (2018). Prediction of wheel and rail wear under different contact conditions using artificial neural networks. *Wear*, *406–407*(March 2017), 173–184. https://doi.org/10.1016/j.wear.2018.01.007
81. Siri-Team. (2017). Deep Learning for Siri's Voice: On-device Deep Mixture Density Networks for Hybrid Unit Selection Synthesis. *Apple Machine Learning Journal*, *1*(4). Retrieved from https://machinelearning.apple.com/2017/08/06/siri-voices.html
82. Steele, R. K., & Devine, T. J. (1982). Wear of Rail/Wheel Systems. In J. Kalousek, R. V.

Dukkipati, & G. M. L. Gladwell (Eds.), *Contact Mechanics and Wear of Rail/Wheel Systems* (pp. 293–316). Vancouver, BC: University of Waterloo Press.
83. Steele, R. K., & Reiff, R. P. (1982). Rail: Its Behavior and Relationship to Total System Wear. In *Second International Heavy Haul Railway Conference* (pp. 227–276). Colorado Springs, CO: International Heavy Haul Committiee.
84. Stone, D. H., & Steele, R. K. (1978). The Effect of Mechanical Properties Upon the Performance of Railroad Rails. In D. H. Stone & G. G. Knupp (Eds.), *Rail Steels - Developments, Processing, and Use* (pp. 21–62). Denver, CO: American Society for Testing and Materials.
85. Sun, Y., Guo, Y., & Zhai, W. (2019). Prediction of rail non-uniform wear – Influence of track random irregularity. *Wear*, *420–421*(October 2018), 235–244. https://doi.org/10.1016/j.wear.2018.10.020
86. Szabo, A., & Zobory, I. (1998). Deterministic and Stochastic Simulation of Wheel and Rail Profile\Wear Processing. *Periodica Polytechnica Se. Transp. Eng.*, *28*(1–2), 3–17.
87. Tao, G., Wen, Z., Zhao, X., & Jin, X. (2016). Effects of wheel???rail contact modelling on wheel wear simulation. *Wear*, *366–367*, 146–156. https://doi.org/10.1016/j.wear.2016.05.010
88. Telliskivi, T. (2004). Simulation of wear in a rolling-sliding contact by asemi-Winkler model and the Archard's wear law. *Wear*, *256*(7–8), 817–831. https://doi.org/10.1016/S0043-1648(03)00524-6
89. Telliskivi, T., & Olofsson, U. (2004). Wheel-rail wear simulation. *Wear*, *257*(11), 1145–1153. https://doi.org/10.1016/j.wear.2004.07.017
90. Tournay, H. M. (2008). A future challenge to wheel/rail interaction analysis and design: Predicting worn shapes and resulting damage modes. *Wear*, *265*(9–10), 1259–1265. https://doi.org/10.1016/j.wear.2008.03.029
91. Tournay, H. M., & Mulder, J. M. (1996). The transition from the wear to the stress regime. *Wear*, *191*(1–2), 107–112. https://doi.org/10.1016/0043-1648(95)06693-4
92. Urban, C., Barbosa, R. S., Santos, G. F. M. dos, Joy, R., & Shu, X. (2015). Characterization and Modelling of a New Heavy Axle Load Freight Wagon for Wheel Rail Wear Prediction. *Journal of Mechanical Engineering and Technology*, *2*, 17–28. https://doi.org/10.18005/jmet0203001
93. Van, K. D., & Maitournam, M. H. (2002). On some recent trends in modelling of contact fatigue and wear in rail. *Wear*, *253*(1–2), 219–227. https://doi.org/10.1016/S0043-1648(02)00104-7
94. Villani, M., Kohn, R., & Giordani, P. (2009). Regression density estimation using smooth adaptive Gaussian mixtures. *Journal of Econometrics*, *153*(2), 155–173. https://doi.org/10.1016/j.jeconom.2009.05.004
95. Vuong, T. T., & Meehan, P. A. (2009). Wear transitions in a wear coefficient model. *Wear*, *266*(9–10), 898–906. https://doi.org/10.1016/j.wear.2008.12.006
96. Wang, J., Chen, X., Li, X., & Wu, Y. (2015). Influence of heavy haul railway curve parameters on rail wear. *Engineering Failure Analysis*, *57*, 511–520. https://doi.org/10.1016/j.engfailanal.2015.08.021
97. Wang, L., Zhao, W., Xu, H., & Na, W. (2015). Wear prediction of metro wheels based on the ARIMA model. In *2015 27th Chinese Control and Decision Conference (CCDC)*. Qingdao, China: IEEE.
98. Wang, P., Xu, J., Xie, K., & Chen, R. (2016). Numerical simulation of rail profiles

evolution in the switch panel of a railway turnout. *Wear*, *366–367*, 105–115. https://doi.org/10.1016/j.wear.2016.04.014
99. Western Safety Products. (2018). Rail Wear Gauge. Retrieved from http://www.westernsafety.com/products/aldon2012/aldon2012pg9.html
100. Zakeri, J. A., & Shahriari, S. (2012). Developing A Deterioration Probabilistic Model for Rail Wear. *International Journal of Traffic And Transportation Engineering*, *1*(2), 13–18. https://doi.org/10.5923/j.ijtte.20120102.02
101. Zakharov, S., & Zharov, I. (2002). Simulation of mutual wheel/rail wear. *Wear*, *253*(1–2), 100–106. https://doi.org/10.1016/S0043-1648(02)00088-1
102. Zarembski, A. M. (1992). *Tracking R&D*. Omaha, NE: Simmons-Boardman Books, Inc.
103. Zarembski, A. M. (2005). *The Art and Science if Rail Grinding*. Omaha, NE: Simmons-Boardman Books, Inc.
104. Zarembski, A. M. (2010). Autmoated Planning of Wear Based Rail Replacement. *Railway Track & Structures*.
105. Zarembski, A. M., & Bohara, A. P. (1993). Controlling Rail and Wheel Wear on Commuter Operations. *American Railway Engineering Association*, *94*(742).
106. Zarembski, A. M., Thornton, D., Palese, J. W., & Forte, N. (1998). Development and Implementation of RailGraph: A Field Deployable Maintenance Management Tool. In *Proceedings of the nnual AREMA Conference*. Chicago.
107. Zen, H., & Senior, A. (2015). *Deep Mixture Density Networks for Acoustic Modeling In Statistical Parametric Speech Synthesis*. Retrieved from google.com
108. Zobory, I. (1997). Prediction of Wheel/Rail Profile Wear. *Vehicle System Dynamics*, *28*(2–3), 221–259. https://doi.org/10.1080/00423119708969355

ABOUT THE AUTHORS

Joseph W. Palese, MCE, PE

Mr. Palese is a Senior Scientist and Program Manager: Railroad Engineering and Safety Program at the University of Delaware. He has over 29 years of experience in track component design and analysis, failure analysis and component life forecasting algorithm specifications, and development of inspection systems. Throughout his career, Mr. Palese has focused on acquiring and utilizing large amounts of track component condition data for planning railway maintenance activities.

Mr. Palese has a Bachelor's Degree of Civil Engineering, and a Master's Degree of Civil Engineering, both from the University of Delaware, along with a MBA from Rowan University. He received his PhD in Civil Engineering at the University of Delaware in 2019. He is a registered Professional Engineer in the state of New Jersey.

Dr. Allan M. Zarembski, P.E., Hon. Mbr. AREMA, FASME

Dr. Zarembski is an internationally recognized authority in the fields of track and vehicle/track system analysis, railway component failure analysis, track strength, and maintenance planning. Dr. Zarembski is currently Professor of Practice and Director of the Railroad Engineering and Safety Program at the University of Delaware's Department of Civil and Environmental Engineering, where he has been since 2012. Prior to that he was President of ZETA-TECH, Associates, Inc. a railway technical consulting and applied technology company, he established in 1984. He also served as Director of R&D for Pandrol Inc., Director of R&D for Speno Rail Services Co. and Manager, Track Research for the Association of American Railroads. He has been active in the railroad industry for over 40 years.

Dr. Zarembski has PhD (1975) and M.A (1974) in Civil Engineering from Princeton University, an M.S. in Engineering Mechanics (1973) and a B.S. in Aeronautics and Astronautics from New York University (1971). He is a registered Professional Engineer in five states. Dr. Zarembski is an Honorary Member of American Railway Engineering and Maintenance of way Association (AREMA), a Fellow of American Society of Mechanical Engineers (ASME) , and a Life Member of American Society of Civil Engineers (ASCE). He served as Deputy Director of the Track Train Dynamics Program and was the recipient of the American Society of Mechanical Engineer's Rail Transportation Award in 1992 and the US Federal Railroad Administration's Special Act Award in 2001. He was awarded The Fumio Tatsuoka Best Paper Award in 2017 by the Journal of Transportation Infrastructure Geotechnology

He is the organizer and initiator of the **Big Data in Railroad Maintenance Planning Conference** held annually at the University of Delaware. He has authored or co-authored over 200 technical papers, over 120 technical articles, two book chapters and two books.