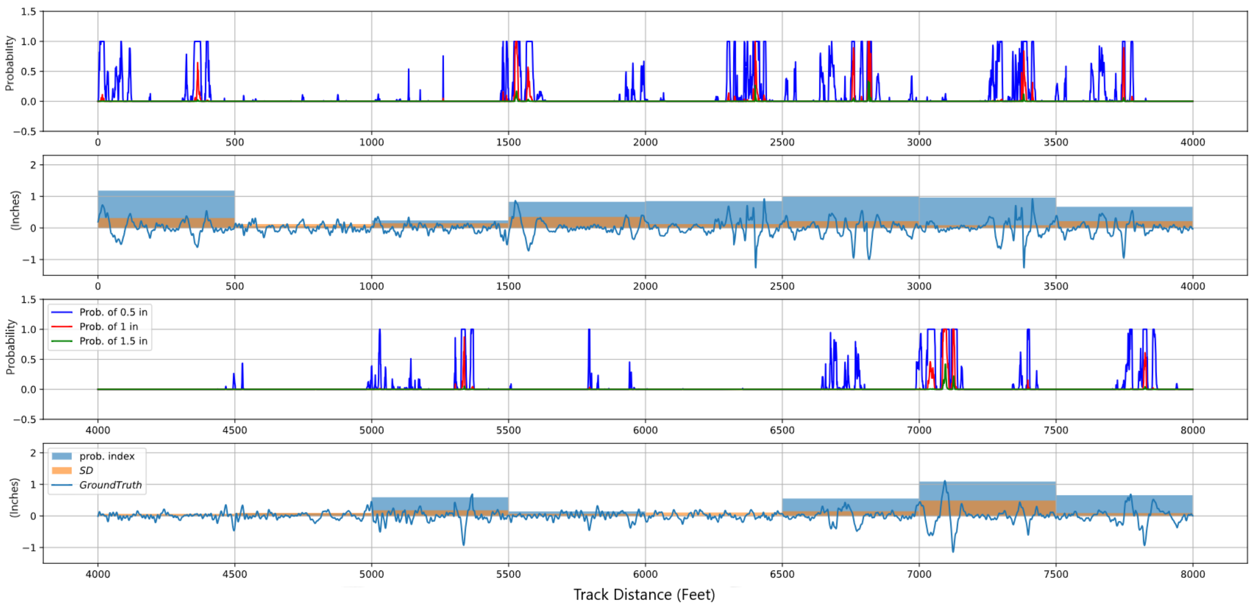
U.S. Department of Transportation

Federal Railroad Administration

Office of Research, Development and Technology
Washington, DC 20590

Multi-Dimensional Time-Based Track Safety and Quality Index in Support of Autonomous Track Geometry Inspection



Final Report | July 2026

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	**2. REPORT TYPE**	**3. DATES COVERED** (From - To)
01/01/2024	Technical Report	09/02/20-12/31/24

4. TITLE AND SUBTITLE	
Multi-Dimensional Time-Based Track Safety and Quality Index in Support of Autonomous Track Geometry Inspection	**5a. CONTRACT NUMBER** 693JJ619C000018
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER** FR19RPD31000000050
Joseph W. Palese, PhD, MBA, PE – ORCiD: 0000-0003-3946-3777 Allan M. Zarembski, PhD, PE, FASME – ORCiD: 0000-0002-4282-9330 Osman Mohamed- ORCiD: 009-0008-1533-6320	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
University of Delaware Newark, DE 19716	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
U.S. Department of Transportation Federal Railroad Administration 1200 New Jersey Ave. SE Washington, DC 20590	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)** DOT/FRA/ORD-26/05

12. DISTRIBUTION/AVAILABILITY STATEMENT
This document is available to the public through the FRA website.

13. SUPPLEMENTARY NOTES
COR: Hugh Thompson

14. ABSTRACT
In this research, the Federal Railroad Administration (FRA) sponsored a team from the University of Delaware to develop a multidimensional, time-based Track Safety and Quality Index (TSQI) for assessing and monitoring the condition of railway tracks. The index captures the temporal dependencies and uncertainties associated with track parameters by incorporating machine learning techniques and a multivariable normal distribution. The developed 3D track quality index (TQI) incorporates probabilities of exceeding predefined safety limits for the mid-chord offset of the track geometry data set by FRA. A comprehensive assessment of the track's condition is obtained by calculating the probability of exceedance for each measured parameter based on the multivariable normal distribution. The index provides an overall measure of the likelihood of the track quality exceeding specified thresholds for multiple parameters simultaneously. The 3D TQI is a valuable tool for assessing and monitoring the condition of railway tracks. It integrates machine learning techniques, probability analysis, and customization options to offer a comprehensive and customizable approach to track evaluation. The index enhances safety, optimizes maintenance practices, and contributes to the reliable operation of railway infrastructure.

15. SUBJECT TERMS
Autonomous track geometry, track quality index, ride quality, defect growth rate, profile, long short term memory, mixture density networks

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**		61	**19b. TELEPHONE NUMBER** (Include area code)

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std.

Contents

Illustrations

Tables

Executive Summary

Reliable transportation networks rely on effective maintenance activities. The increasing demand for railway transportation means infrastructure managers must develop maintenance plans and strategies to meet new requirements and ensure high service quality and capacity. Railways currently use analytic and autonomous track geometry assessments (e.g., autonomous inspection cars) to improve safety and maintenance decision-making. However, traditional techniques based on discrete threshold processing and track quality indices are limited and cannot adequately apply advanced analytics. Therefore, it is important to enhance these techniques to better leverage the vast amount of data collected by autonomous inspection cars to make more informed decisions regarding safety and maintenance.

In this research, the Federal Railroad Administration (FRA) sponsored a team from the University of Delaware to develop a multidimensional, time-based Track Safety and Quality Index (TSQI) for assessing and monitoring the condition of railway tracks. The index captures the temporal dependencies and uncertainties associated with track parameters by incorporating machine learning techniques and a multivariable normal distribution.

The inclusion of a normal distribution layer allows researchers to model uncertainties in track parameters, providing insights into the range of potential values and associated uncertainties. Separate models are trained for each track parameter channel, accommodating their distinct characteristics and variability and enabling a more accurate distribution prediction for each specific channel. This provides a comprehensive understanding of the track's condition.

Researchers can use multiple models and their corresponding output distributions to derive valuable insights into the variation and uncertainty associated with track parameters. These distributions are combined into a multivariable normal distribution, capturing the joint behavior and interdependencies of the track quality parameters. Statistical measures such as confidence intervals, percentiles, and probabilities can be derived from this distribution, allowing for a nuanced assessment of track quality.

The developed 3D track quality index (TQI) incorporates probabilities of exceeding predefined safety limits for the mid-chord offset of the track geometry data set by FRA. A comprehensive assessment of the track's condition is obtained by calculating the probability of exceedance for each measured parameter based on the multivariable normal distribution. The index provides an overall measure of the likelihood of the track quality exceeding specified thresholds for multiple parameters simultaneously.

The effectiveness of the 3D TQI is demonstrated through its application to eight segments of railway track. Visual representations, such as 3D plots, facilitate the identification of segments with varying levels of irregularity and aid in prioritizing maintenance efforts. The index offers a customizable approach by assigning weights to each parameter, allowing for the prioritization of specific parameters based on shareholders' needs.

The 3D TQI is a valuable tool for assessing and monitoring the condition of railway tracks. It integrates machine learning techniques, probability analysis, and customization options to offer a comprehensive and customizable approach to track evaluation. The index enhances safety, optimizes maintenance practices, and contributes to the reliable operation of railway infrastructure.

1. Introduction

In this research, the Federal Railroad Administration (FRA) sponsored a team from the University of Delaware to develop a 3D, time-based Track Quality Index (TQI) for assessing and monitoring the condition of railway tracks.

1.1 Background

Any reliable and resilient transportation network is characterized by effective, well-timed maintenance activities [1], [2], [3], [4]. Over the years, railroads have used analytic and autonomous track geometry assessments more frequently to improve safety-related plans and maintenance procedures. These improvements help railroads make better safety and maintenance decisions that benefit from the enormous volume of data collected by inspection cars.

Historically, the rail industry employed discrete threshold processing of measurements to identify track areas that either are or are close to being unsafe [5], [6], [7]. Moreover, classifying track segments can be based on their general maintenance condition using consecutive measurements. TQI is a metric or measure used to quantify the roughness (i.e., variability from norm) of railway track segments, typically using the standard deviation of values over a defined segment length. Although these techniques have been effective, they frequently cannot capture multifaceted components of track safety and quality and use these advanced analytics to their fullest extent [8].

Figure 1 and Figure 2 are derived from FRA safety database queries [9] from 2015 to 2023 and provide the counts/costs of accidents broken down by Human Factors, Track-related, Equipment-related, Signal-related, and Miscellaneous. The data indicates that out of 7,521 train accidents during this period, 1,545 were directly linked to track-related issues, accounting for almost 20 percent of all accidents that occur. These track-related incidents resulted in a substantial financial burden of about $315M, or almost 30 percent of total accident expenditures ($1.1B). It is critical to better understand the underlying causes, contributing variables, and potential solutions of track-related accidents by addressing and enhancing track safety, resulting in increased safety and economic efficiency in the rail transportation system.

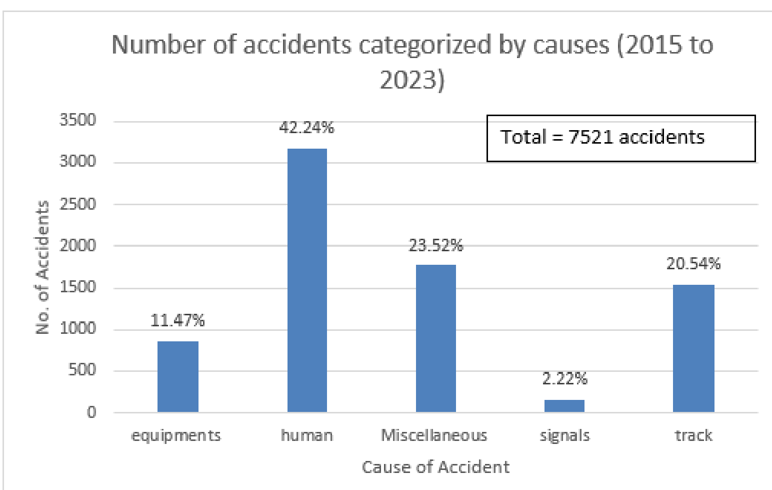


Figure 1. Number of accidents categorized by the cause, from 2015 to 2023 [9]

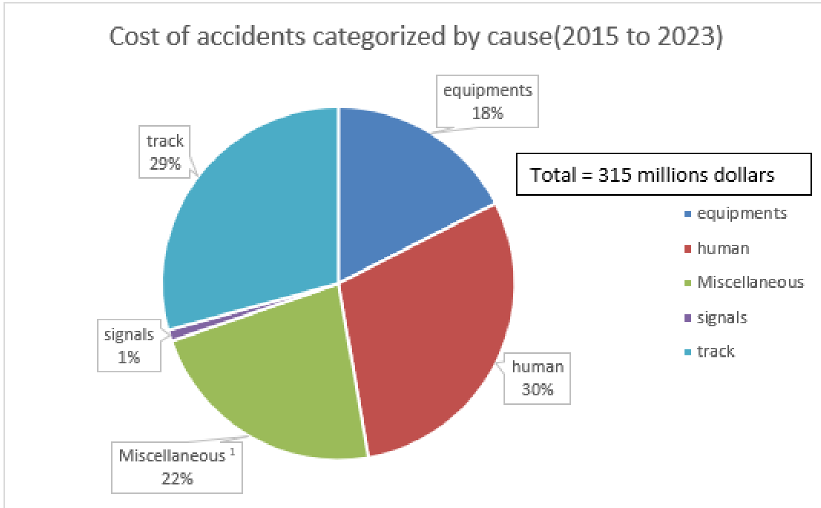


Figure 2. Cost of accidents categorized by cause, from 2015 to 2023[9]

1.2 Objectives

The objective of this research was to use frequently collected track geometry data, such as that collected by an autonomous track geometry car, and develop an algorithm that classifies the track condition in a 3D space, as well as a time based component indicative of degradation.

In this study, the research team created a 3D, time-based Track Safety and Quality Index (TSQI) using a defect risk model to identify areas of the track that are at high probability of a defect occurring based on derived defect growth characteristics. The new TSQI provides a more thorough and impartial assessment of track conditions and improves safety and maintenance decision-making by incorporating multiple dimensions and employing advanced analytic techniques.

This study is significant because it has the potential to completely alter how railroads evaluate the safety and maintenance of their tracks. To better understand how autonomous track geometry inspection data may be used to identify and reduce potential hazards, this study incorporates sophisticated analytics into the TSQI and defect risk model. The proposed methodology can allow railroads to optimize maintenance efforts and resources by concentrating on high-risk locations, lowering operational costs and boosting overall system reliability.

1.3 Overall Approach

The overall approach was to use track geometry car data collected monthly and apply Long-Short Term Memory (LSTM) network modeling as well as Mixture Density Network (MDN) modeling to forecast a probabilistic representation of actual track geometry in the future (based on a defined period). The team performed this modeling for each track geometry channel. The results were then used to develop a multivariate normal distribution, by which the probability and uncertainty of a defect or unsatisfactory condition could be determined. These results then were used to calculate a segment-specific index (the TSQI), based on the probability and uncertainty of a defect occurring in a defined period. This approach allows for predicting track condition, as well as defining current condition.

The TSQI was calculated using data from autonomous track geometry inspections, and defect risks were evaluated using mathematical models and algorithms. The proposed TSQI and defect risk model were compared with real data using the standard deviation as a default track quality index.

1.4 Scope

The research evaluated individual track geometry channels, including left and right profile, left and right alignment, and cross-level. These channels were chosen for their impact on ride quality and track surfacing. The same process can be used to incorporate additional channels that were not included in this research. Degradation rates were included by developing track geometry perturbation forecasting methods; thus, a time component is included.

The approach was developed and tested on several miles of high-speed track, however a full application to larger track segments with differing operations was not conducted.

The study was conducted in a specific geographic region; more research will be needed to determine whether the suggested approach can be used with various railway networks. Implementing the suggested technique may be hampered by issues with data accessibility, availability, and quality; researchers attempted to resolve these issues during the study.

1.5 Organization of the Report

This report is organized into several sections. Section 2 briefly describes traditional methods of track geometry data processing as background to the approaches developed in the research. Section 3 provides data analysis. Section 4 presents the models developed and the research results. Section 5 provides a summary of the research and thoughts on future work. Section 6 presents the research conclusions.

2. Traditional Track Geometry Data Processing

This section briefly describes traditional methods of track geometry data processing as background to the approaches developed in this research.

2.1 Threshold Processing/Exceptions (Safety/Maintenance)

Railroads typically evaluate directly measured channels of inspection data (e.g., gage, surface, alignment, cross-level, etc.) and first level derived channels of inspection data (e.g., twist, warp, maximum speed, etc.) against defined thresholds to evaluate the condition of the track. These thresholds are a function of the allowable operating speed of the track defined by FRA Track Classification. Maximum speed limits (i.e., larger deviations) are allowed at slower speeds, and railroads generally evaluate the data against two levels of thresholds: RED = FRA safety limit, and YELLOW = railroad maintenance limit, typically one FRA track class above the defined operating class [10]. This results in a list of locations (variable in length depending on the number of consecutive exceedances) that require either immediate attention (RED), or attention soon (YELLOW). Attention can be defined as maintenance (e.g., tamping) to bring the track into acceptable limits, or reduced operating speed such that the measurements are below an acceptable threshold limit. Note that some railways use more than two levels of thresholds to classify the condition of their track.

The goal of course is to have minimal or no RED exceptions. To achieve this, an inspection interval that flags a location before it grows into an exception should be defined and used. Research has shown that an increased inspection frequency has resulted in a decrease in track geometry safety exceptions and associated derailments. This is directly attributed to frequent inspection providing information to allow for intervention in the growth cycle of a defect, without fully knowing the defect growth parameters. In some instances, the growth rate of a particular measurement channel has been studied and has shown to take on both linear and nonlinear growth characteristics, depending on the specific segment. This requires consistent data at a high frequency that must be longitudinally aligned to obtain a representative rate of growth. Typically, analysis of the data has been limited to individual channels, or simple combinations of channels, without fully understanding the three-dimensional surface of the track.

2.2 TQI (Maintenance)

Many railroads use TQIs to classify the condition of track segments. TQIs are typically a statistical measure of the variation in a particular measurement channel (e.g., standard deviation, root mean square, etc.). Past research has been performed on developing various TQIs for individual measurement channels, and these values are generally representative of the roughness of the track [11]. The generated TQI values for individual channels, or multiple channels (typically combined based on some weighted average), can then be trended over time or million gross tons (MGT) to show the growth in the roughness/condition of the track. The time or MGT to maintenance can then be determined based on a TQI threshold and the observed rate of TQI degradation. Both linear and nonlinear rates have been documented.

Application of this technique has proved valuable and effective in the past but has several limitations, including lack of sufficient data, computing power, and appropriate analytic techniques to extract track degradation behavior from the data.

2.3 Limitations to Current Approach

While the two-level approach defined above has been effective, there are significant limitations in this approach that can be addressed using autonomous inspection data and emerging analytic techniques. Ultimately, the measurements acquired by the track geometry system are compared to thresholds that are defined to minimize the risk of derailment due to the dynamic interaction of the vehicle running over the track at a defined speed. There are many classes of vehicles with varying loads, car characteristics, and component conditions that result in varying dynamic responses to the track that can lead to a derailment. Thus, development of a single threshold value for all traffic types and speeds is difficult and often results in conservative standards.

Several significant limitations of the current approach are listed below.

- Safety Defects vs. Track Quality: Current FRA track geometry safety standards are focused on individual measurement channels. TQIs are currently used for long stretches of track to provide an indication of the ride quality or roughness of the track. Current TQIs tend to "wash-out" safety defects in favor of longer roughness conditions, which are maintenance and ride-quality related.
- Use of Independent Channels: Current FRA standards are for thresholds on individual channels. An entire class of derailments, specifically vehicle-track dynamics related derailments, can be attributed to multiple measured channels simultaneously being near a threshold without exceeding a particular threshold.
- Growth Rate Characteristics: Track settles differentially resulting in a top of rail geometry variation. The variable rate of settlement at different locations along the track (transversely and longitudinally) is related to the rate of track geometry variation. The rate of change of this variation with accumulated traffic is site specific. Traditional measurement intervals and analytic techniques make it difficult to define the rate of change in variation at specific locations. The current approach is to trap a specific location by lowering the safety threshold limit. This does not allow for differentiating between slow growth locations (e.g., maintaining too soon) or fast growth locations (e.g., allowing a safety defect to form).
- Interrelationship of Channels: While certain measurement channels can stand alone and provide insight to the track condition, there is a well-defined interrelationship between certain channels that must be considered. For example, alternating surface variation for the left and right rail will also appear as a cross-level variation.
- Dimensional Analysis: Use of individual channels does not provide a complete picture of the surface of the track, which is three dimensional and changes with time, representing a four dimensional "surface."
- Fixed Segmentation: TQI values are generally determined for a fixed track segment length. Variation can change significantly within a segment, resulting in a filtered representation of roughness. For example, a segment with constant moderate roughness and a segment that is 50 percent smooth and 50 percent very rough could result in the same TQI value. Another example would be a length of track with several defects surrounded by smooth track having the same TQI value as a segment with very moderate roughness.

3. Data and Exploratory Data Analysis

3.1 Data

The research team obtained the track geometry data used in this study from Amtrak. The data contained several critical measurements typically collected by track geometry inspection cars and was provided for the same 10 miles of track inspected monthly for two years (2018-2020). The data consisted of several independent and derived channels highlighted in Table 1 [12], [13].

Table 1. Data channels

	Variable	Description
1	Mile	Integer MP
2	Feet	Integer Ft
3	Track	Track Number
4	Line	Line Designation
5	Gage	TG Data
6	CrossLevel	TG Data
7	CrossLvlRate	TG Data
8	RightProfile	TG Data
9	RightProf62	TG Data
10	RightProf124	TG Data
11	LeftProfile	TG Data
12	LeftProf62	TG Data
13	LeftProf124	TG Data
14	RAlignment	TG Data
15	RAlign62	TG Data
16	RAlign124	TG Data
17	LAlignment	TG Data
18	LAlign62	TG Data
19	LAlign124	TG Data
20	Curvature	TG Data
21	CTS	Curve/Spiral Flag
22	Speed	Test Car Speed
23	ALD	Automated Location Detector
24	Class	FRA Track Class
25	Warp62	TG Data
26	LProfSC	TG Data
27	RProfSC	TG Data
28	LAlignSC	TG Data
29	RAlignSC	TG Data
30	SyncCnt	Synchronization Value
31	SyncFt	Synchronization Value

Moreover, the dataset consists of 24 files containing the track geometry data for each recorded monthly inspection. These files provide a wealth of information regarding the condition and characteristics of the track over time, since they are repeat measurements of the same location, and roughly synchronized in space.

Each file gives access to numerous channels (i.e., variables) of track geometry data. These channels record various characteristics of the track's state (e.g., gage, profile, alignment, cross-level, etc.) [14]. The profile data, for instance, includes information regarding the track's vertical deviation, whereas the alignment data concentrates on the track's horizontal alignment.

For some channels, different chord values are available (e.g., 31, 62, and 124 foot chords). These chord values represent the sampling intervals or the distance between data points along the track. By using multiple chord options, the analysis can evaluate track condition at different granularity, which is important when considering speed of operations.

By studying these variables across different chords and over time, patterns, trends, and potential issues that may affect the track's performance and safety can be identified.

3.2 Exploratory Data Analysis

Understanding the data provided and extracting its features is a core element of this research. Exploratory Data Analysis (EDA) is a common data analytics approach that provides not only a first assessment of the data and the identification of any potential correlations within the dataset, but also sufficient insight into the dataset and its underlying structure. The team applied various EDA techniques using different types of graphical methods to better understand the track geometry data provided, and its behavior and characteristics.

Multiple techniques were used to assess the data employed in this study. For instance, a line graph was used to determine locations with a high degradation rate or whether a location required immediate maintenance during the recorded duration.

Figure 3 presents an abstract regarding the behavior of a 62' chord-based profile for the left rail of the track, in the segment between 2000 and 5000 feet. Although the FRA limit for the profile is 1 inch in a Class 6 track, it is evident in this segment that two locations started to grow between February 2018 and March 2019 and crossed the FRA threshold, then a maintenance activity occurred. Moreover, the line graph when comparing segments 2000-3000 and segment 4000-5000 shows the first segment has a rough track with low ride quality, and the latter with minimal roughness with higher ride quality.

In addition, a box plot is a valuable tool for visualizing and summarizing the distribution of a dataset. Like the line graph, box plots can provide insights into the condition of the track. In the specific context of this research, box plots are used to analyze the changes in the distribution of specific parameters over time.

For example, in Figure 4, the box plots represent the changes in distribution from February 2019 to March 2019 for parameters such as cross-level rate, right profile 62', left profile 62', right alignment 62', and left alignment 62'. These box plots allow researchers to compare the distributions between the two time periods, and provide insights into the overall track condition.

These box plots show that, for instance, the right profile parameter exhibited a worsening trend from February to March. The box plot shows a shift in the distribution, indicating a degradation in the track's right profile. Furthermore, some data points in March exceeded the allowable limit.

These observations highlight the usefulness of box plots in identifying changes in the distribution of track parameters and assessing compliance with safety standards. They provide a concise data summary, allowing for quick identification of trends and potential issues.

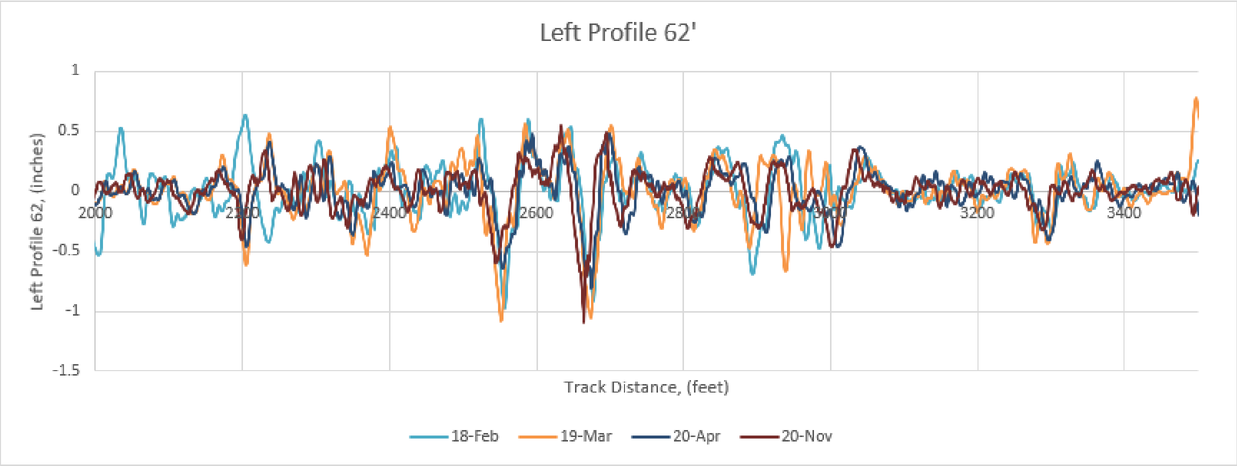
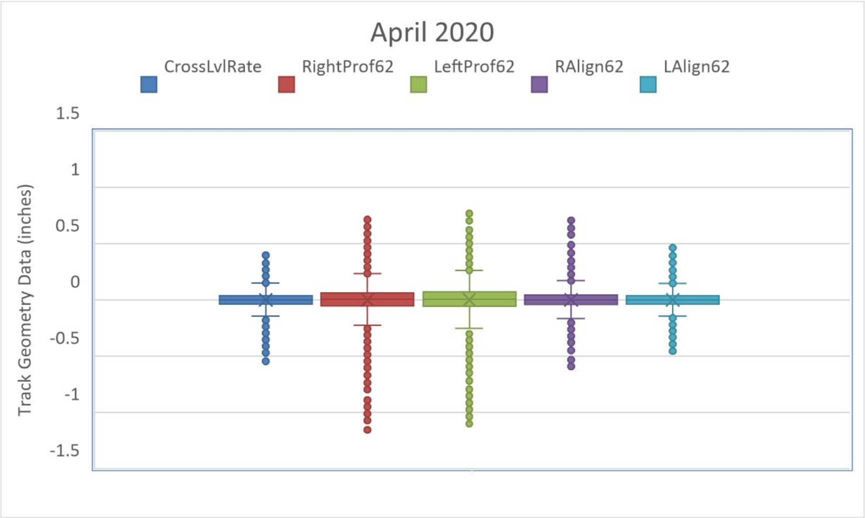
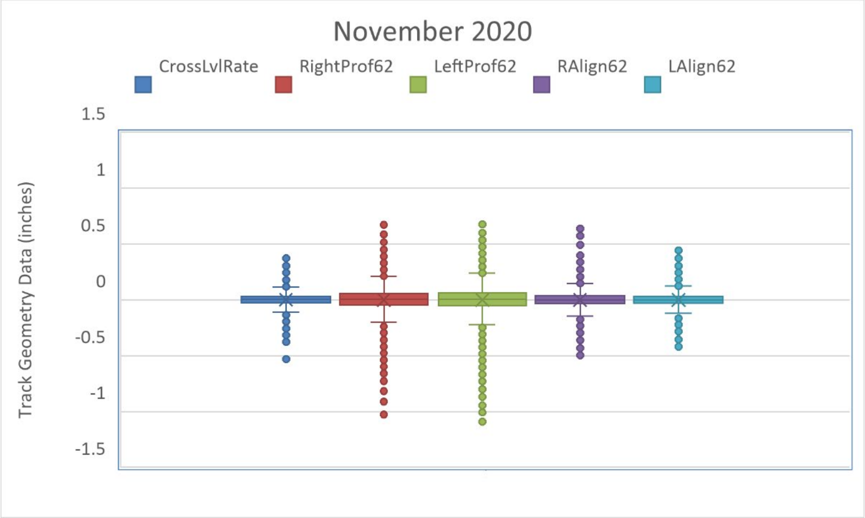


Figure 3. Line graph for the left profile 62 ft mid-chord



(a) Box plot for April 2020



(b) Box plot for November 2020

Figure 4. Box plots for April 2020 and November 2020

Incorporating box plots in the research visually presents the changes in distribution over time and emphasizes the significance of these changes in relation to track safety and maintenance. This supports the research objective of understanding the track's condition and identifying areas that require attention or improvement.

Finally, since the project's objective was to find a comprehensive 3D TQI, the first focus was on variables such as profile and alignment. Moreover, some of the preliminary analyses were conducted on the left and right profiles. The profile mean value for any segment should be equal to zero, as the data varies around a mean of zero. Thus, any distortion from this value can be easily captured.

A histogram is a graphical representation of the distribution of a dataset. It is a way to visualize the frequency or occurrence of different values or ranges of values in a dataset. As a starting point for exploratory analysis, the team created a histogram for the profile (left and right) and alignment (left and right) of the data for interrogation. Figure 5 captures the distributions of right profile (62' MCO) for the first 500 feet of the track and the first 12 months of data recorded. This figure clearly shows the change in shape of the distribution over time, where the wider the distribution, the rougher the track.

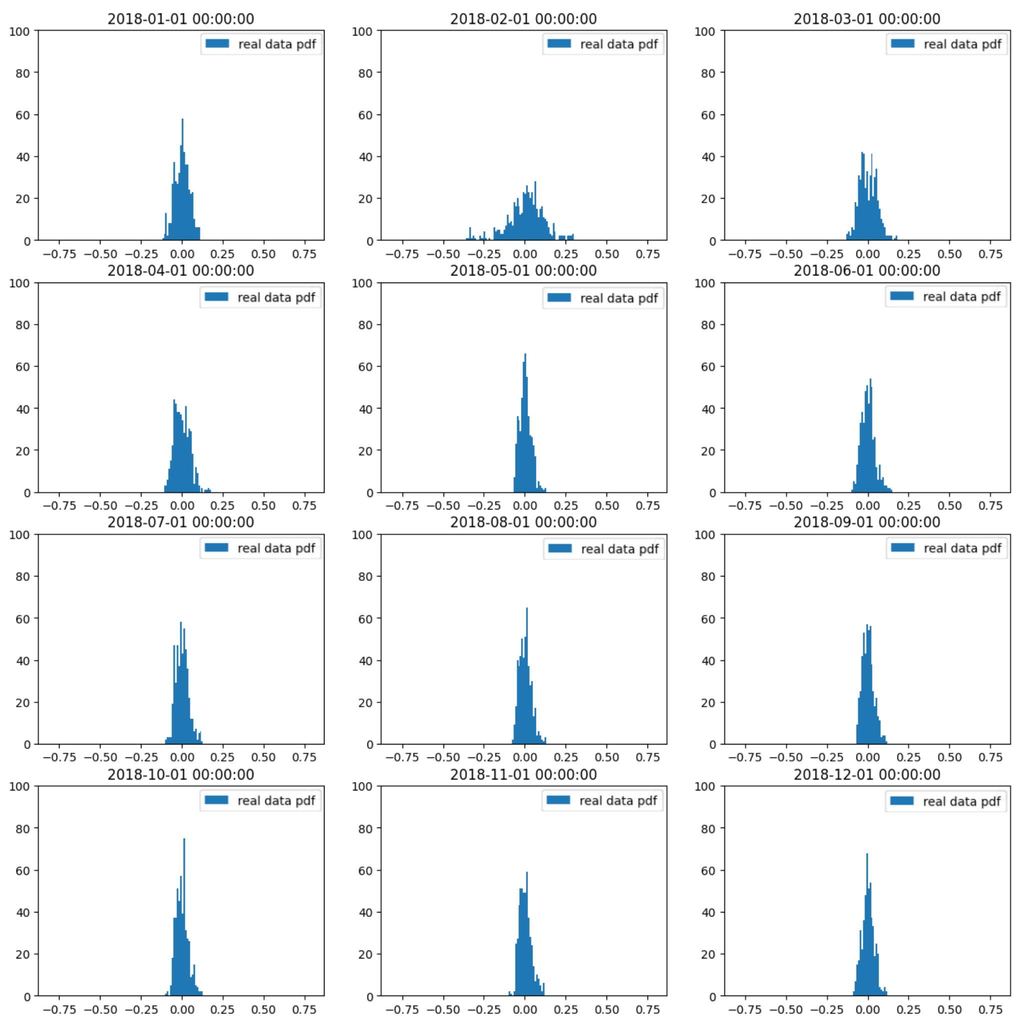


Figure 5. Right profile histogram for segment 0 to 500 feet over a 12-month duration

3.3 Best Fit Distribution

It is generally understood that track geometry data, such as profile and alignment, when plotted into a histogram, takes the shape of a Gaussian distribution. Therefore, the first step in the analysis is determining whether this assumption would hold for the data obtained or must be adjusted to different distributions. At first, the entire 10 mile track segment was split into equal segments of 1000 feet, but because the chosen segment length is quite long and might have different characteristics within the segment, a smaller segment was considered, using 500 feet for each subsegment. Four distributions were fitted into each subsegment over 2 miles for 24 months: Gaussian (i.e., normal) distribution, t-student distribution, exponential distribution, and Laplace distribution. Figure 6 and Figure 7 illustrate the shape of each distribution.

The Gaussian distribution is a continuous probability distribution used in statistics and probability theory. This type of distribution consists of a bell-shaped curve with a symmetric and unimodal shape. Two parameters define any Gaussian distribution – the mean and the standard deviation. The mean value is the center of the distribution (50 percent of the data lies to each side of the mean), while the standard deviation represents how spread the data is from the mean value. Generally, the Gaussian distribution is used to describe naturally occurring events, as this is how real-world variable events tend to be shown, and hereafter is referred to as the normal distribution.

There are several properties of the normal distribution, such as the 68, 95, and 99.7 rule. This rule states that approximately 68 percent of the data falls within one standard deviation, 95 percent within two standard deviations, and 99.7 percent within three standard deviations.

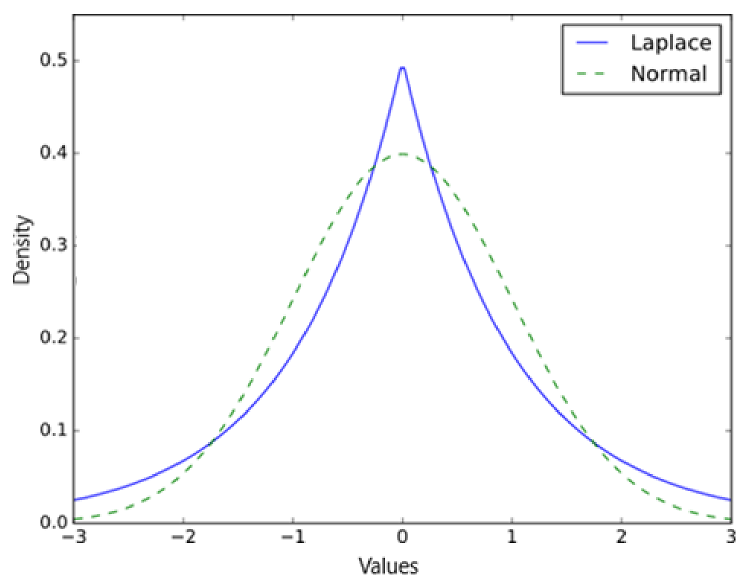


Figure 6. Laplace and Normal distribution [15]

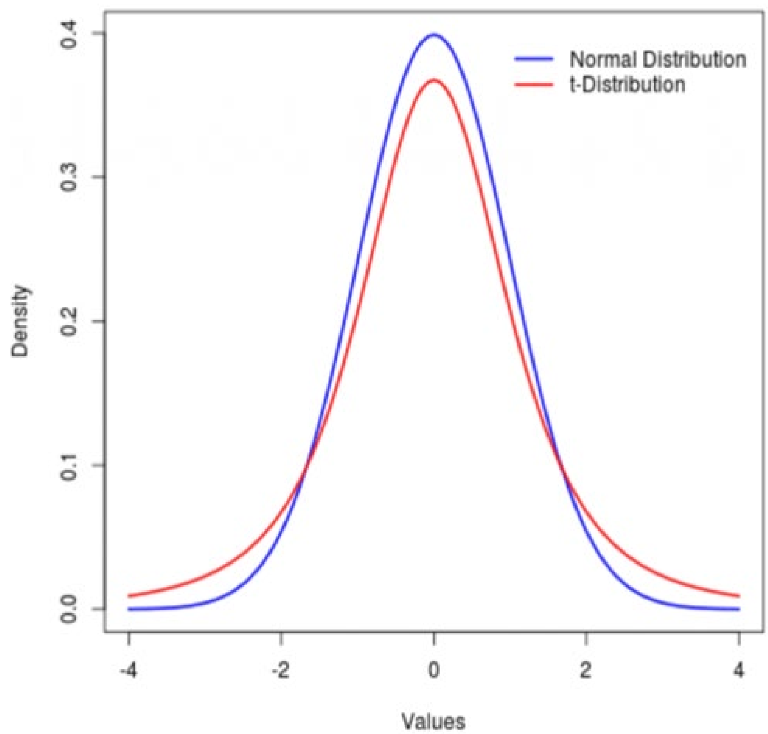


Figure 7. Normal and t-student distribution [16]

The t-student distribution is a probability distribution used when dealing with small sample sizes or when the population standard deviation is unknown. Its shape is close to normally distributed data, but it contains heavy tails which allow for more extreme values. Only one parameter defines this distribution, the degree of freedom (df). The df represents the sample size minus one. The more the df increases, the shape of the t-student distribution becomes closer to the normal distribution. The t-student distribution can be used in hypothesis testing, confidence interval estimation, and when working with small or less well-known datasets.

The exponential distribution is also a continuous probability distribution that describes the time between events occurring in a Poisson process, where events exist independently and at a constant average rate. It is frequently used to model the duration of specific systems or the time before an event occurs. The rate parameter (λ) represents the average number of events happening per unit of time, as it is the only parameter used for this type of distribution. Generally, it is illustrated by a rapid initial fall followed by a slight decrease in the exponential distribution's tail, which has a positively skewed shape. The exponential distribution usually is used in survival analysis, queuing theory, and reliability analysis.

Finally, the Laplace distribution, which is also a double-exponential distribution, is a continuous probability distribution. A heavier tail characterizes the shape of the Laplace distribution compared to the normal distribution. Two parameters define Laplace distribution: the location parameter and the scale parameter. Like the normal distribution, the location parameter determines where the center of the distribution is, while the scale parameter defines the spread or variability of the data. A sharp peak at the location parameter and exponential-like tails are notable features of the Laplace distribution. Those features make the distribution suitable for modeling data with outliers or heavy-tailed characteristics.

Figure 5 shows the need for a better representation for track segments other than the normal distribution. A preliminary analysis was conducted and algorithms were created to find the best-fit distribution for each segment out of the four mentioned distributions. The distribution with the least error represents the segment of the track.

A histogram was created for the segment given their geometry measurements for a certain channel (e.g., profile, alignment, cross-level, etc.). The midpoints of the histogram bins are calculated by taking the average between each bin edge and its adjacent edge according to:

$$x_i = \frac{x_{i-1} + x_{i+1}}{2} \quad for\ i = 1\ to\ n, n = number\ of\ measurements$$

This step ensures that the x-values align with the center of each bin. Furthermore, for each considered distribution, the probable density function (PDF) is evaluated at the midpoint "x" using the distribution parameters obtained from fitting the distribution to the data according to the functions presented in Table 2.

Table 2. PDF functions

Distribution Name	PDF Equation	Parameters
Gaussian/Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	x = measurement μ = mean σ = standard deviation
t-student	$f(x) = \frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2}$	x = measurement υ = number of degrees of freedom Γ = Gamma function
Exponential	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$	x = measurement λ = shape and scale parameter
Laplace	$f(x) = \frac{1}{2b} e^{-\left(-\frac{[x-\mu]}{b}\right)}$	x = measurement μ = median b = scale parameter

The parameters for each associated distribution presented in Table 2 are determined using a least squares fit to the actual histogram data. The error used to determine the best-fit distribution is the sum of squared errors (SSE) as follows:

$$SSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - f(x_i))^2$$

The PDF function with the lowest SSE will identify the distribution that best fits the actual measurement data of the segment. A comparison was made between the SSE value along different track segments. Table 3 provides the results of the SSE calculations for each 500 foot segment for each distribution. Figure 8 provides a plot of this data to easily visualize the results. It can be seen from Table 3 and Figure 8 that the 500 foot track segments can typically be classified as normal or Laplace.

Table 3. SSE values for each segment for each fit distribution

	norm	laplace	exponnorm	t
0-500	32.48	58.3	32.48	32.65
500-1000	16.78	40.88	14.56	16.94
1000-1500	9.32	8.11	7.38	6.55
1500-2000	48.03	84.77	47.28	48.03
2000-2500	5.17	0.92	5.17	2.2
2500-3000	2.23	1.67	2.23	1.5
3000-3500	8.35	5.07	7.04	5.99
3500-4000	2	0.84	2.01	0.97
4000-4500	4.19	8.22	4.19	3.4
4500-5000	7.75	18.46	7.76	7.13
5000-5500	33.55	46.3	33.56	33.54
5500-6000	11.8	20.14	11.81	11.23
6000-6500	3.12	7.34	3.12	2.94
6500-7000	7.56	6.77	7.56	6.76
7000-7500	12.95	10.18	11.44	7.83
7500-8000	14.41	7.05	10.42	6.45
8000-8500	20.69	2.39	20.7	0.74
8500-9000	11.7	3.44	11.71	1.12
9000-9500	14.33	20.25	14.34	12.53
9500-10000	4.95	4.83	4.95	4.83

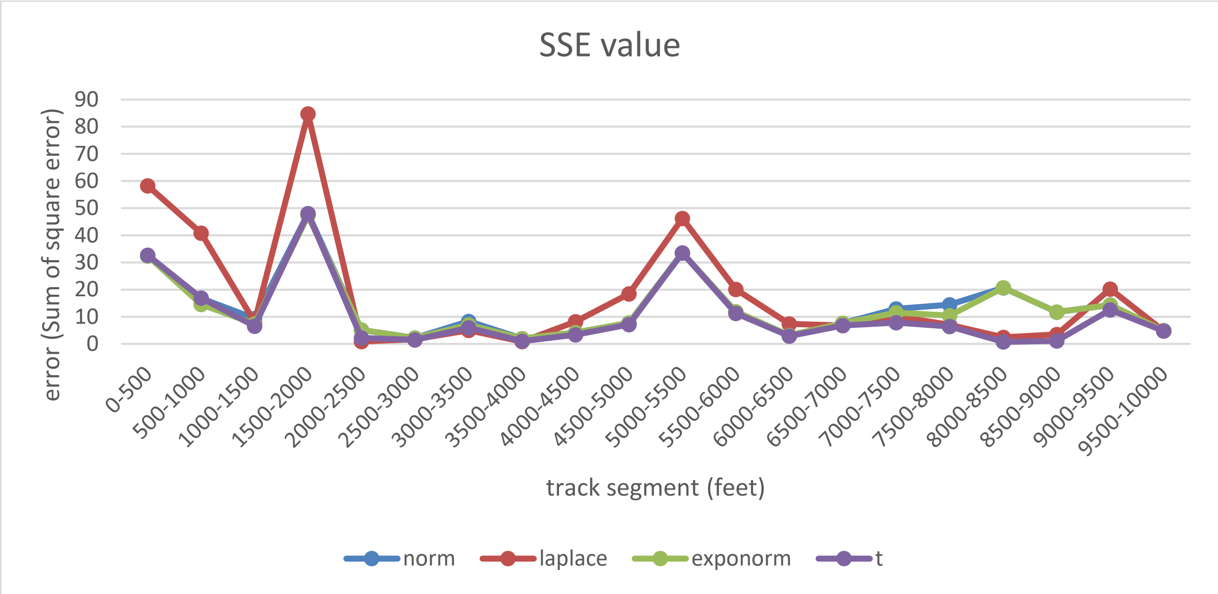


Figure 8. SSE value for each track segment for each distribution fit

A 3D graph was drawn to replicate the interchangeable characteristics of the track segment between the Laplace and the normal distribution. Figure 9 and Figure 10 represent the change in distribution type and values throughout the 24-month duration. Note that the increase and decrease in the values of the location and scale don't necessarily relate to certain distributions.

Decreasing the standard deviation value does not necessarily mean a change from normal to Laplace distribution, but appears to exhibit a more arbitrary behavior.

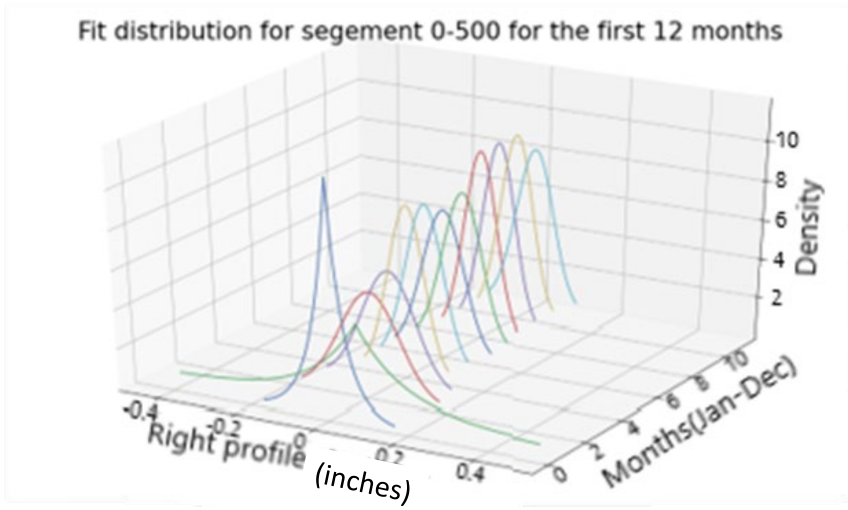


Figure 9. Fit distribution over time (1st to 12th month) for segment 0-500 feet

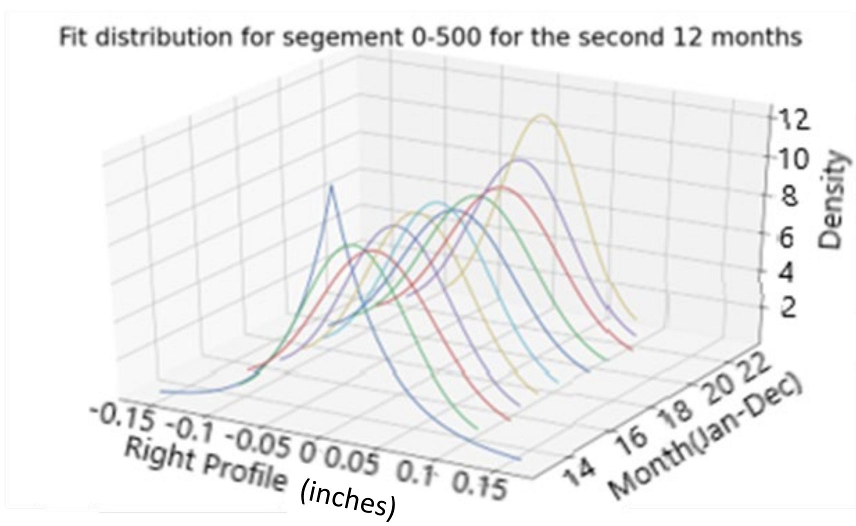


Figure 10. Fit distribution over time (13th to 24th month) for segment 0-500 feet

From the EDA analysis and the best-fit distribution, researchers found that a model was needed that would differ from the common practice (e.g., the assumption that the normal distribution is the only representation for track segments) and depict the interchangeable behavior of the captured track segments over time (i.e., a segment can exhibit various distribution behavior over its life). These results and findings are the core of the proposed models in the following section.

4. Models and Results

4.1 Deterministic vs. Probabilistic Deep Learning

Many machine learning (ML) tasks have been widely solved using deep learning. The popularity of deep learning fields such as reinforcement learning and natural language processing arose from its ability to learn complex, non-linear functions from large amounts of data. The historical trend of deep learning models is deterministic, meaning that predictions are the same if given the same inputs [17].

The deterministic deep learning models' goal is to learn a certain path between inputs and outputs without accounting for any form of uncertainty using Artificial Neural Networks (ANNs), computation models inspired by the function and structure of biological neural networks in the human brain. In ANNs, after the application of an activation function and the processing of weighted input sums, neurons convey the output to subsequent layers. Deterministic models learn to estimate complex functions and make precise predictions by training ANNs with large, labeled datasets and optimization methods like backpropagation (i.e., supervised learning) [18].

Probabilistic deep learning models (e.g., Generative Adversarial Networks (GANs), Bayesian modeling, uncertainty estimation, and reinforcement learning) are trained to improve probabilistic objective functions (e.g., negative loglikelihood of the data or an approximate posterior distribution). Such models highlight the uncertainty in their predictions.

Probabilistic deep learning uses techniques such as Bayesian Neural Networks (BNNs), variational inference, and Monte Carlo methods. BNNs are essentially Bayesian rules built into models to capture the distributions over network weights instead of fixed weights. This is where the concept of uncertainty is estimated, and then probabilistic predictions are made. Moreover, Monte Carlo methods, like Markov Chain Monte Carlo (MCMC), ease the process of sampling from complex probability distributions.

The superiority of probabilistic models is illustrated in their ability to create new outputs from training data and new image generators. They also are used for unsupervised learning, where learning the pattern or compact representation of the data without any leads is the end goal.

4.2 Deterministic Models

In degradation modeling, neural networks are presented as a powerful tool, especially in estimating track geometry degradation and assessing the railway's infrastructure condition. These models use the concept of neural networks to learn complex relationships and patterns in degradation data, improving the accuracy and reliability of predictions.

ANNs are computing systems consisting of interconnected processing units that dynamically respond to external inputs and operate like biological neural networks in the human brain. ANNs process information by passing it through input nodes, performing intermediate processing in hidden nodes, and generating the final output via output nodes. The strength and information flow of each connection is determined by the weights and connections between nodes.

In degradation modeling, the configuration of the neural network is customized to suit the specific requirements of any task. Generally, neural networks consist of input nodes, hidden nodes, and output modes. Input nodes receive the data, hidden nodes apply transformations and computations on the data, and the output nodes connect the processed information to the desired format by using the activation function.

Information is carried from one neuron to another through these connections with consideration for the weight of each connection, depending on their relative importance. The adjustment of weights occurs during the learning process by repeating steps called epochs. During the learning process, the weight of the model is updated based on prediction errors and the minimization of the difference between actual data and predictions.

Historical degradation data is used to train the network and create a degradation model. This information provides track geometry related input properties such as alignment, profile, cross-level rate, and other pertinent factors. The training data also includes the associated output values that represent the observed degradation or track condition.

The neural network learns the underlying patterns and connections between the input features and the related deterioration throughout the training phase. After being trained, the network can be used to forecast track geometry for new data or periods.

4.2.1 Data Preparation

Section 2 details how the team applied a distribution technique to 2 miles of track divided into 53 segments, each 200 feet in length. Data was available monthly for a two year period. Each of these segments was then analyzed to determine the location and scale parameters for a normal and Laplace distribution for each month of data collected.

The Laplace and normal distributions were used to describe the characteristics of each segment. The results of this analysis, including the location (i.e., the mean in case of normal distribution) and shape parameter (i.e., standard deviation in case of normal distribution) for each distribution, were calculated.

Table 4 and Table 5 present a subset of the calculated location and scale values obtained for the normal and Laplace distributions, respectively, showing results for 24 segments over 6 months for the left profile. These values were collected and organized in spreadsheets to facilitate further analysis. The data covers the first two miles of the track (ten miles overall) with a recorded duration of 24 months and includes 53 segments, each 200 feet long. This totals a dataset of 10,600 feet, which was used to simplify the calculations and ensure comprehensive track coverage.

Examining the location and scale parameters obtained from the fit distributions, valuable information can be gained regarding the characteristics of each segment, such as the central tendency and spread of the data, which will allow for a better understanding of the track's condition over time. The recorded location values represent each segment's mean or average profile measurements. This provides information about the typical or average behavior of the track within each segment. The scale values, represented by the shape parameter (standard deviation for the normal distribution), indicate the variability or dispersion of the data within each segment. A higher scale value suggests more significant fluctuations or irregularities in the track's geometry measurements. The cells in Table 4 and Table 5 are color coded to visually describe their magnitudes, with red showing larger positive and negative values, and green indicating proximity to zero.

Trends and patterns can be identified by analyzing the location and scale parameters across the entire dataset. Deviations from the norm or significant variations in these parameters may indicate areas of the track that require closer inspection or maintenance interventions. These findings contribute to a comprehensive understanding of the track's quality and performance characteristics.

Table 4. The recorded location and scale for the left profile (fitting normal distribution)

	Loc Jan	Scale Jan	Loc Feb	Scale Feb	Loc Mar	Scale Mar	Loc Apr	Scale Apr	Loc May	Scale May
Seg 1	0.00315	0.03485	0.00065	0.06376	0.00304	0.04358	0.0018	0.04034	0.00438	0.02562
Seg 2	0.00894	0.05349	0.01603	0.14498	0.00691	0.06307	0.00703	0.05871	0.01167	0.04201
Seg 3	-0.005	0.05875	-0.0098	0.08453	-0.0055	0.06041	-0.0054	0.04958	-0.0069	0.02955
Seg 4	0.00183	0.05404	0.01053	0.07018	0.00254	0.08217	0.00232	0.07777	-0.001	0.02982
Seg 5	-0.0013	0.07459	-0.0094	0.05821	-0.0036	0.10198	-0.0041	0.09789	-0.0015	0.03996
Seg 6	-0.0013	0.12464	0.00868	#######	0.00416	0.11583	0.00556	0.1054	0.00196	0.05191
Seg 7	0.00031	0.03676	-0.0015	0.06351	-0.0005	0.10077	-0.0003	0.08675	0.00275	0.05839
Seg 8	0.00495	0.05331	-0.0095	0.14904	0.00338	0.05785	0.00403	0.05432	0.00173	0.02873
Seg 9	0.00207	0.07616	0.01137	0.08799	0.00669	0.09306	0.00586	0.08409	-0.0033	0.02778
Seg 10	-0.0052	0.04227	0.00077	0.11118	-0.0137	0.08087	-0.0161	0.08508	0.0027	0.0301
Seg 11	0.01068	0.07627	-0.012	0.44329	0.01596	0.0714	0.02245	0.07263	0.0081	0.053
Seg 12	-0.0422	0.29764	-0.1215	0.44608	-0.0251	0.31727	-0.0219	0.30724	-0.0156	0.2539
Seg 13	0.02981	0.14173	0.12835	0.60696	-0.0017	0.18327	-0.0107	0.1913	-0.008	0.17316
Seg 14	0.00476	0.16063	0.00082	0.16722	0.01491	0.19349	0.01673	0.18257	0.01494	0.17105
Seg 15	0.01247	0.16087	0.01349	0.15104	0.01333	0.1856	0.01137	0.19223	0.01616	0.19241
Seg 16	-0.022	0.10286	-0.0201	0.0899	-0.0193	0.1004	-0.0193	0.10142	-0.0263	0.09681
Seg 17	0.01147	0.08135	0.00929	0.07423	0.00689	0.08391	0.01068	0.08911	0.01099	0.07582
Seg 18	-0.0053	0.08233	-0.0033	0.08496	-0.0028	0.09705	-0.0081	0.09014	-0.0074	0.06736
Seg 19	0.0091	0.09514	0.00754	0.08316	0.00606	0.11597	0.01348	0.09598	0.01053	0.08264
Seg 20	-0.0074	0.08088	-0.0055	0.06116	-0.0031	0.08187	-0.0084	0.07516	-0.003	0.06439
Seg 21	-0.0016	0.09188	-0.0021	0.08899	-0.003	0.10335	-0.0013	0.09919	-0.0045	0.08934
Seg 22	-0.0035	0.06806	-0.0036	0.05582	-0.0039	0.07491	-0.0045	0.06459	-0.0041	0.05554
Seg 23	0.01003	0.05815	0.01107	0.05516	0.01262	0.06585	0.01243	0.06144	0.01121	0.0545
Seg 24	-0.0036	0.05011	-0.0049	0.04349	-0.0057	0.05104	-0.0048	0.04731	-0.0037	0.04333

Table 5. The recorded location and scale for the right profile (fitting Laplace distribution)

	Loc Jan	Scale Jan	Loc Feb	Scale Feb	Loc Mar	Scale Mar	Loc Apr	Scale Apr	Loc May	Scale May
Seg 1	0.00382	0.02385	0.01831	0.06293	0.00351	0.0504	0.00519	0.04557	0.00031	0.04354
Seg 2	-0.006	0.06291	0.02914	0.13706	-0.0096	0.08078	-0.0041	0.07456	-0.0037	0.02886
Seg 3	-0.0015	0.0648	0.02075	0.13105	-0.024	0.06096	-0.0139	0.05138	0.01434	0.04803
Seg 4	0.00092	0.05465	-0.0206	0.16274	-0.0043	0.07183	0.0029	0.06883	-0.0002	0.03595
Seg 5	-0.0003	0.09249	0.01007	0.14595	0.00778	0.08879	0.01007	0.09566	-0.0067	0.04291
Seg 6	-0.0047	0.13511	-0.0166	0.09494	0.01129	0.14148	0.02426	0.14972	0.0206	0.10304
Seg 7	0.009	0.06321	0.01373	0.05591	0.04181	0.06709	0.04105	0.07067	0.0058	0.05864
Seg 8	0.0058	0.04393	0.00427	0.07043	-0.0087	0.0469	-0.0075	0.04248	-0.0058	0.04272
Seg 9	0.02869	0.07516	0.0061	0.04665	0.00107	0.05096	-0.0015	0.04512	-0.0031	0.0391
Seg 10	-0.0226	0.06006	-0.0212	0.07066	-0.0032	0.05314	0.00336	0.05493	-0.007	0.03948
Seg 11	0.01221	0.15118	-0.0298	0.27407	-0.0005	0.09128	-0.0195	0.10278	-0.0107	0.09586
Seg 12	0.01907	0.16942	0.03113	0.3231	-0.0122	0.13291	-0.0116	0.12367	0.01267	0.13898
Seg 13	0.04334	0.21118	-0.0475	0.26343	0.07889	0.23902	0.0621	0.22864	0.08957	0.24376
Seg 14	-0.0249	0.14489	-0.0171	0.13929	0.05341	0.18047	0.05096	0.16331	0.05814	0.18623
Seg 15	0.05631	0.17607	0.04608	0.16676	0.04868	0.2075	0.04349	0.22011	0.05203	0.22504
Seg 16	0.01144	0.07319	0.00412	0.06072	0.01511	0.08354	0.00122	0.06311	0.00107	0.07466
Seg 17	-0.0064	0.06131	0.00031	0.06657	0.00488	0.07907	-0.0105	0.07106	0.02884	0.08652
Seg 18	0.0029	0.14559	0.00351	0.19856	-0.0034	0.21959	0.0058	0.2426	0.0029	0.20012
Seg 19	0.0325	0.14172	-0.0053	0.19636	-0.0079	0.20967	0.03311	0.21957	-0.0005	0.21807
Seg 20	-0.0005	0.13038	-0.0084	0.12835	-0.0061	0.13205	-0.0064	0.13497	0.00382	0.13256
Seg 21	0.00534	0.08117	0.01511	0.07042	0.01434	0.09007	0.00778	0.07859	0.01343	0.06862
Seg 22	-0.0023	0.07337	-0.007	0.07491	-0.0035	0.07393	-0.0044	0.07551	-0.0053	0.0809
Seg 23	0.01541	0.07745	0.01312	0.07858	0.02792	0.08317	0.01267	0.08046	0	0.07685
Seg 24	0.00061	0.04118	0.00137	0.03936	-0.0005	0.04588	0.00275	0.04351	0.00671	0.04767

4.2.2 Neural Network Model Configuration

The neural network model in this study is a feed-forward neural network architecture designed to predict the distribution of the segment in the following month based on previous distribution characteristics and behavior (see Figure 11). There are two major layers in the model, the hidden layer and the output layer. The hidden layer construction is based on 100 neurons, each activated using the rectified linear unit (ReLU) activation function. The ReLU activation function allows for non-linearity and to capture complex patterns in the data.

The input layers of the neural network, constructed of two neurons, represent the scale and location parameter for each previous month. In a time-series form, the inputs are organized so that the change in distribution from month to month can be captured. Note that the distribution type is assumed to be the same for the segments and duration of inspection history. Also, note that the value of input layers enables the network to learn the relationship and pattern between the input values and the target outputs.

The output layer consists of two neurons as a representation of the desired output. A linear activation function is used to allow the model to produce continuous outputs. The team focused on minimizing the discrepancy between the model-predicted outputs and the actual values of the output during the training process.

Finally, to optimize the performance of the neural networks and improve their learning capability, an Adam optimizer is employed. The Adam optimizer is a stochastic gradient descent (SGD) version that modifies the learning rate according to the first and second moments of the gradient. The model converges faster and avoids potential problems like vanishing or expanding gradients thanks to its adjustable learning rate.

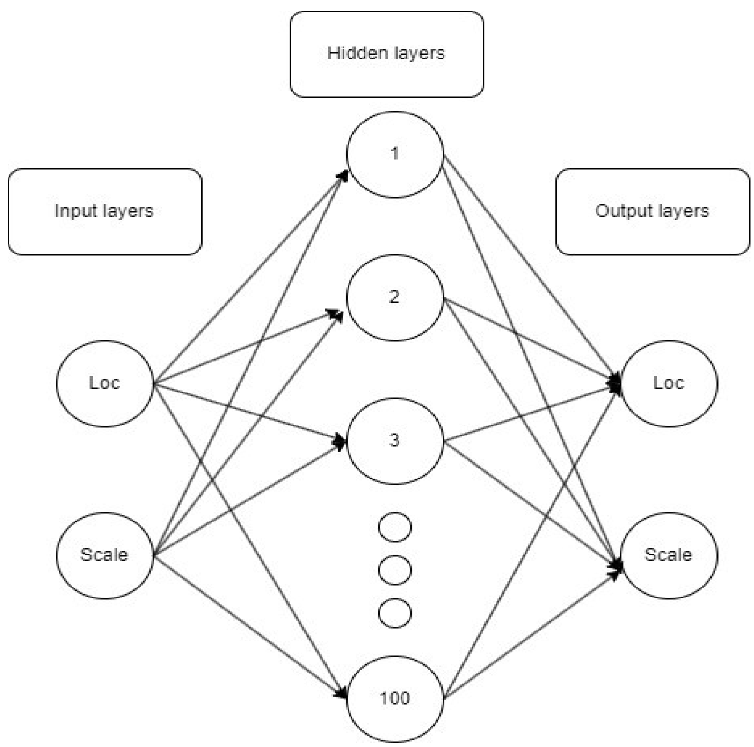


Figure 11. The neural network configuration

4.2.3 Model Training and Validation

This study used a feed-forward neural network architecture as a predictive model for forecasting the distribution parameters (e.g., location and scale) of track segments in the subsequent month. The data was split into training and testing datasets. The training dataset comprised 75 percent of the entire dataset, while the remaining 25 percent was allocated for testing purposes. This division allowed for evaluating the model's performance on unseen data.

During the training process, the primary objective of the neural network model is to minimize the mean squared error (MSE) loss function. The MSE loss function quantifies the average squared difference between the predicted outputs generated by the model and the true target outputs provided in the training dataset. By minimizing this loss function, the model improves the accuracy of its predictions. Based on the MSE value, the Adam optimizer adjusts the parameters and updates the model's weights and biases to improve the overall model performance.

As the model progresses through each iteration, it calculates the prediction error, which is the discrepancy between the predicted and true target outputs. This error is then backpropagated through the network, allowing the model to update its parameters and refine its predictions. The model gradually improves its predictive capabilities by iteratively minimizing the MSE loss function and updating its internal weights and biases. This process is commonly known as training or learning. The goal is to achieve a state where the model can accurately predict the distribution of the track segment for the following month.

It is important to note that the training process is typically performed on the training dataset, while the testing dataset remains unseen until the model is fully trained. The testing dataset then evaluates the model's performance on new, independent samples. This evaluation helps assess the model's generalization ability and capability to make predictions on unseen data. The training was done using Python, with 2000 epochs.

4.2.4 Results and Discussion

Each track segment's scale and location values were recorded after applying the Laplace and normal distributions independently within the model, regardless of which distribution was more suitable. This approach allowed the team to make a comprehensive analysis of the segment's characteristics under both distribution assumptions and gain insights into the range and central tendency of the data under different distributional assumptions.

This information can be valuable in understanding the behavior of the track segments and comparing their characteristics across different distribution types, provides a more comprehensive view of the data, and allows for a more thorough analysis of the degradation patterns. By considering both distributions, the fit for each distribution can be assessed to determine which better represents the data, and researchers can choose the most appropriate distribution for accurately modeling the degradation process. By recording the scale and location values for each track segment under both the Laplace and normal distributions, no potential information is overlooked, and a more robust understanding of the data's characteristics can be realized. This information can be further used in developing degradation models, assessing track quality, and making informed decisions regarding maintenance and repair strategies.

Two deterministic models were used to predict the segment distribution parameters: one assuming all segments follow a Laplace distribution shape and the other assuming a normal distribution shape. These models estimate the distribution parameters based on the available data and compare the predicted values with the actual PDF of the segments.

Figure 12 and Figure 13 show the comparison between the actual histogram and the predicted normal and Laplace distributions for two example segments. These figures provide visual representations of how well the models capture the distribution characteristics of the segments. By examining the figures, it is possible to assess the performance of each model in capturing the shape and characteristics of the segment distributions. A close alignment between the predicted values and the actual PDF indicates a good fit of the model to the data. Significant deviations or discrepancies suggest areas where the models may require further refinement or adjustments.

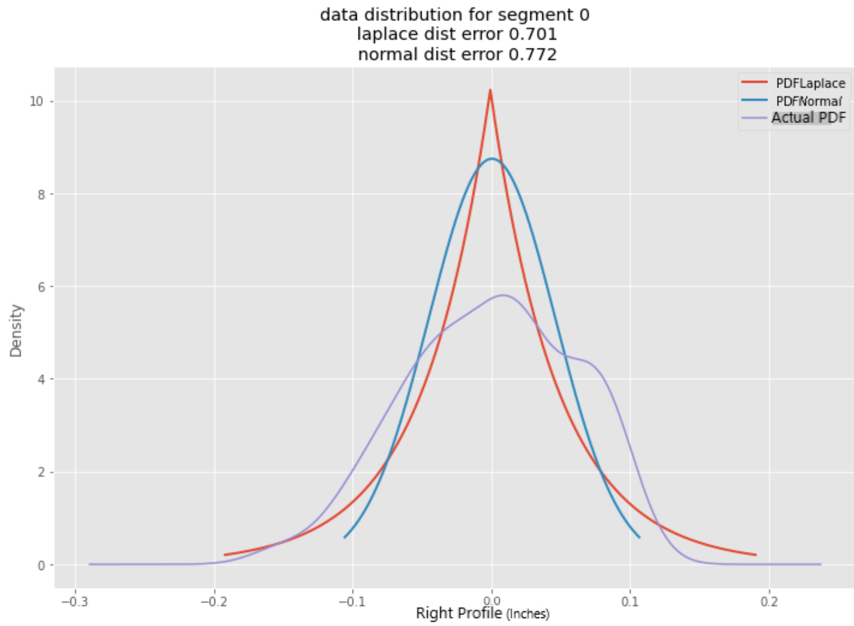


Figure 12. Data distribution for Segment 0

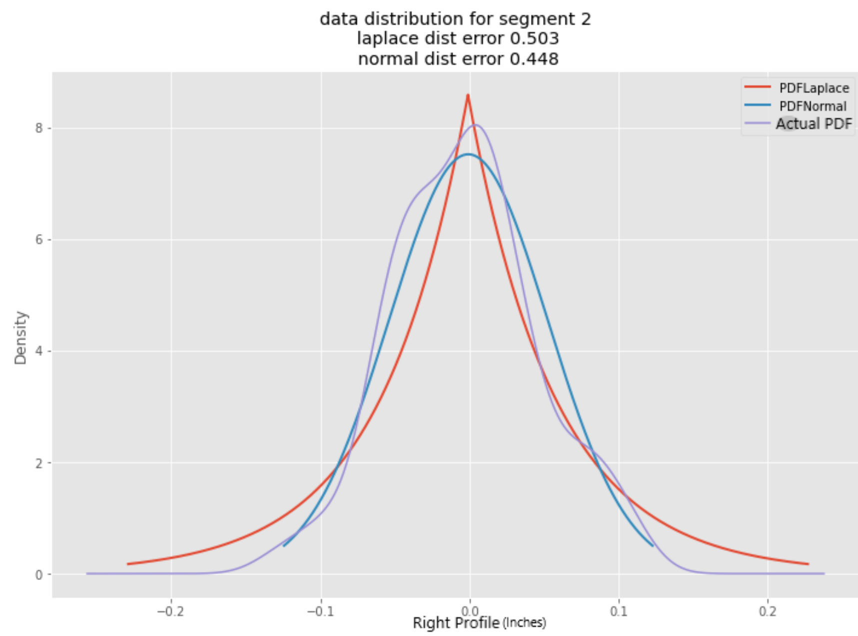


Figure 13. Data distribution for Segment 2

Figure 12 indicates that a Laplace distribution (MSE = 0.701) is a better fit than a normal distribution (MSE = 0.772), although both have a relatively large error. In the actual histogram, the distribution appears to be mixed (which will be addressed in later sections). Figure 13 indicates that a normal distribution (MSE = 0.448) is a better fit than a Laplace distribution (MSE = 0.503), and the actual data appears to fit a normal distribution.

The comparison between the predicted values and the actual PDFs serves as a vital validation step in evaluating the accuracy and reliability of the deterministic models. It allows researchers and practitioners to assess the models' effectiveness in representing the underlying distribution patterns of the segments and gain insights into the potential strengths and limitations of each modeling approach. These results provide valuable information for decision-making processes related to track maintenance and management. Understanding the distribution characteristics of the segments and accurately predicting their parameters makes it possible to assess the likelihood of certain degradation patterns and plan appropriate maintenance interventions.

Considering the varying distribution results, the accuracy of the proposed neural network, and several other pertinent factors, it can be inferred that the deterministic model approach has limitations in effectively handling the data. Some of the limitations of the deterministic models proposed are:

- Inability to capture the change in distribution over time: The data analysis has shown that the distribution that best fits a segment can change over time. However, the deterministic models assume a fixed distribution for each segment throughout 24 months. This approach fails to capture the interchangeability of distributions for segments. As a result, the deterministic models may not accurately represent the evolving nature of track degradation.
- Limited ability to detect past track segment behavior: Historical data provides valuable insights into the behavior of track segments over time, including maintenance cycles and degradation patterns. However, the deterministic models only consider one step at a time when predicting new distribution values. This limitation restricts their ability to leverage the full historical context and may result in less accurate predictions.
- Lack of consideration for uncertainty in predictions: When dealing with distribution and probabilities, it is important to account for their uncertainty. Deterministic models provide more certain value predictions, disregarding the inherent uncertainty associated with real-world scenarios. Uncertainty can arise due to various factors, such as measurement errors, variability in track conditions, and other external factors. Neglecting uncertainty can lead to overconfident predictions and unrealistic expectations.

In conclusion, as illustrated in Figure 14, and from the analysis of track segment behavior, the team found that track segments do not adhere to a single distribution family, instead exhibiting interchangeability between Laplace and normal distributions over different time periods. In the deterministic models, the assumption is that all segments follow the same distribution family and neglect this fluctuation. These results address the need for a new model to address these limitations and focus on the probabilistic aspect of the predicted track geometry condition.

A proposed new model that would overcome the deterministic limitations and adopt a probabilistic approach in assessing track geometry conditions is presented in the next subsection. These models' goal is to capture the fluctuating nature of track segments' distribution patterns over time, enabling a more accurate portrayal of the track's actual behavior.

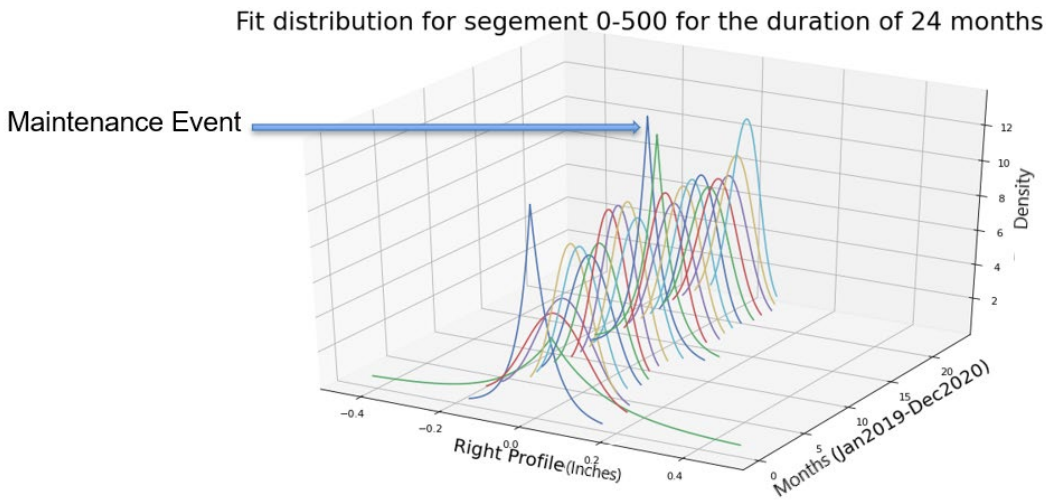


Figure 14. Best-fit distribution of 0-500 feet segment during a 24 month period

The suggested models consider the inherent uncertainty and unpredictability associated with track degradation by shifting the emphasis toward probabilistic modeling. By considering the probabilistic character of the distribution, they seek to provide a more thorough picture of the track state, enabling more accurate projections and well-informed track management and maintenance decisions.

4.3 Probabilistic Deep Learning Model

Given the limitations of deterministic models in that they cannot capture the uncertainty in predictions, more sophisticated models are necessary. In this research, two probabilistic models were considered: Long Short-Term Memory (LSTM) networks and Mixture Density Networks (MDN).

4.3.1 Long Short-Term Memory Networks and Mixture Density Networks

The limitations of the Deterministic Neural Network (DNN) include: the reliance on fitted distributions, which may not perfectly capture the true distribution of segments; the inability to handle multimodal distributions, where a segment can exhibit multiple modes; some segments may exhibit multimodal behavior, meaning they can have multiple inherent distributions, instead of a single distribution describing the data; and the DNN model only predicts the distribution for the following month, providing a short-term outlook which restricts its ability to capture longer-term trends or changes in the distribution of track segments over time.

Researchers developed a combined neural network model to address these limitations, incorporating both Long Short-Term Memory (LSTM) networks and Mixture Density Networks (MDNs). The LSTM network is a specialized type of recurrent neural network (RNN) that is particularly effective in capturing and modeling temporal dependencies in sequential data. Unlike the DNN, which only considers the current input and predicts a single outcome, the LSTM network considers historical information by maintaining an internal memory state. This memory state allows the LSTM network to retain and use information from previous time steps, enabling it to capture long-term dependencies and patterns in the data.

By leveraging the sequential nature of the track geometry information, the LSTM network can effectively learn the underlying dynamics and trends in the track segments' distributions. It considers historical data, such as past distributions and their evolution over time, to predict the future. This ability to incorporate temporal dependencies overcomes the short-term limitation of the DNN, which only looks at the immediate input without considering the context or history.

MDNs were introduced in 1994 as an alternative to DNNs. MDNs output mixture probability distributions where each component of the mixture represents a different possible outcome or mode. MDNs represent the output distribution using mixture models, particularly Gaussian mixture models (GMMs). Within an MDN, the neural network component uses the input data to generate the parameters for the mixture model, including the mean, variance, and mixing coefficients. These parameters define the characteristics and properties of the distributions within the mixture. During the training process, MDNs optimize the parameters to maximize the likelihood of the observed data. This involves adjusting the neural network weights and biases to improve the fit between the predicted mixture distributions and the actual data. By optimizing the likelihood, MDNs learn the underlying patterns and relationships within the data and produce more accurate and probabilistic predictions.

By using MDNs, the model cannot provide a single point estimate but can capture the uncertainty and variability in the data through mixture distributions. This allows MDNs to capture complex and multimodal distributions, making them particularly suitable for problems with multiple possible outcomes or modes.

By combining the strengths of LSTM and MDN, the LSTM-MDN model offers more accurate and probabilistic forecasting of track segment distributions. The LSTM component captures the temporal dependencies and patterns in the track data, allowing for trajectory forecasting over multiple time steps, and the MDN component enables the representation of multiple possible outcomes or modes within the predicted distributions.

The LSTM-MDN model provides a more comprehensive and flexible approach to track geometry prediction. It overcomes the limitations of the DNN by incorporating the probabilistic nature of MDNs and the ability to capture multimodal distributions. This enables the model to handle scenarios where a track segment may exhibit different distribution modes simultaneously.

Overall, the combination of LSTM and MDN in the LSTM-MDN model provides a powerful solution to address the limitations of deterministic approaches, enabling accurate and probabilistic multimodal forecasting for track geometry prediction and track segment classification.

4.3.1.1 Data Preparation

The team processed the raw data for input into the LSTM-MDN model. The data consisted of two contiguous miles of track inspected each month for two years. The data was segmented into 200 foot lengths such that each segment could be compared month to month. This resulted in 53 individual track segments, each with 24 inspections over time. This segmentation allowed more localized predictions and analysis.

Next, the data was split into training and testing batches. The training data comprised 75 percent of the entire dataset, while the remaining 25 percent was used for validation and testing purposes. This split was performed randomly so the model would be trained on sufficient data while providing a separate dataset for evaluating its performance.

The input to the LSTM-MDN model was structured as a 3D array to represent the temporal aspect of the data. Each input consists of 200 foot segments of the track recorded in a sequence of 12 months (time steps), for a total of 53 segments in the spatial domain, with 24 corresponding repetitions for that segment in the time domain. This representation captures the historical information needed to predict the future distribution of the segment.

The model's output is the predicted distribution for the segment for the next six months. It is represented as a mixture of normal distributions to model multiple possible outcomes. The parameters of the mixture model, including the means, variances, and mixing coefficients, are generated by the neural network component of the model.

The LSTM-MDN model was trained on the prepared data and learned to capture the patterns and dependencies in the track geometry channels. It then used this learned knowledge to make probabilistic predictions of the segment's distribution for six months.

In summary, the LSTM-MDN model uses a 3D input structure and a mixture of normal distributions as output to predict the distribution of track geometry channels for the next six months. This approach leverages historical data and the power of neural networks to capture complex patterns and uncertainties, ultimately enhancing the accuracy and reliability of track condition predictions.

4.3.1.2 Model Architecture

The model's architecture consists of two main parts: the LSTM Encoder-Decoder network and the MDN layer (see Figure 15). This architecture is designed to combine the deterministic nature of the LSTM network with the probabilistic capabilities of the MDN.

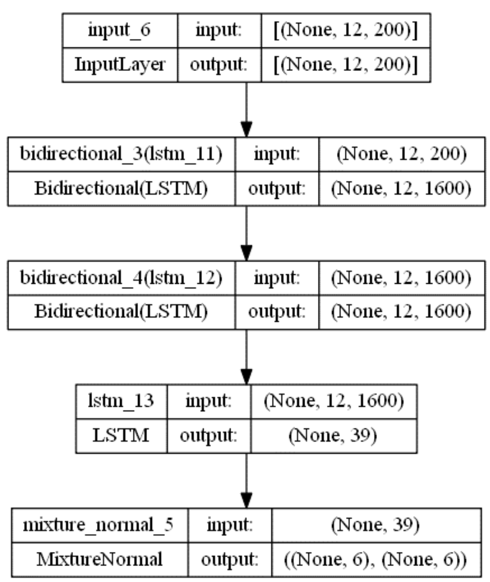


Figure 15. LSTM and MDN configuration

The LSTM Encoder-Decoder network is responsible for encoding the input data and decoding it to the desired dimension for the MDN layer. The encoder consists of two bidirectional LSTM layers, which help capture temporal dependencies in both forward and backward directions. The encoder processes the input data and generates hidden states. The decoder portion contains an LSTM layer that down samples the hidden states from the encoder into the desired dimension for the MDN layer. The LSTM decoder further captures the temporal information and prepares it for probabilistic modeling.

The MDN layer inputs the hidden states from the LSTM decoder. It generates the parameters for the mixture model, including the means, variances, and mixing coefficients for each time step in the target sequences. The output of the MDN layer is a set of components for the GMM for each time step in the target sequences. These components represent different modes or possible outcomes of the predicted distribution.

By combining the LSTM Encoder-Decoder network and the MDN layer, the model is able to capture temporal dependencies in the input data while providing a probabilistic representation of the predicted distribution.

4.3.1.3 Loss Function

A custom loss function known as the negative log-likelihood (NLL) is employed to train the model. This loss function is well-suited for dealing with distribution predictions because it measures the dissimilarity between the predicted distribution and the actual observed values. By optimizing the NLL, the discrepancy between the predicted distribution and the ground truth distribution is effectively minimized.

The NLL loss considers the normal distribution's mean and standard deviation predictions. It penalizes the model more heavily for predictions that deviate significantly from the observed values, leading to more accurate and calibrated distribution forecasts. Using this loss function helps capture the uncertainty inherent in the time series data and provides a more realistic representation of the predictive uncertainty. This is derived below.

$$NLL = -\log\big(P(x|\mu,\sigma)\big)$$

Where:

NLL is the negative log-likelihood loss.

$P(x|\mu, \sigma)$ is the PDF of the normal distribution.

x is the observed value or target variable.

μ is the mean of the normal distribution.

σ is the standard deviation of the normal distribution.

The PDF of the normal distribution is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

To compute the NLL loss, we take the logarithm of the PDF and multiply it by -1, which yields:

$$NLL = -\log\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\right)$$

Simplifying further:

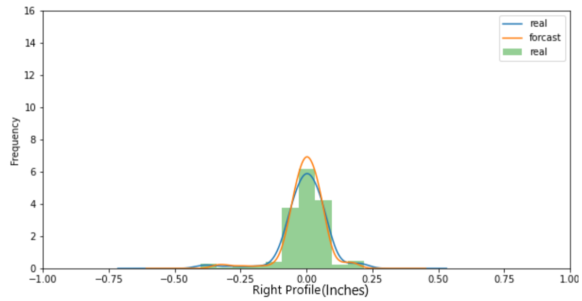
$$NLL = \log\big(\sigma\sqrt{2\pi}\big) + \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$$

In practice, the NLL loss is calculated for each predicted value and then averaged over the entire dataset during training. Minimizing this loss helps the model learn to generate more accurate mean and standard deviation predictions that better fit the observed data.

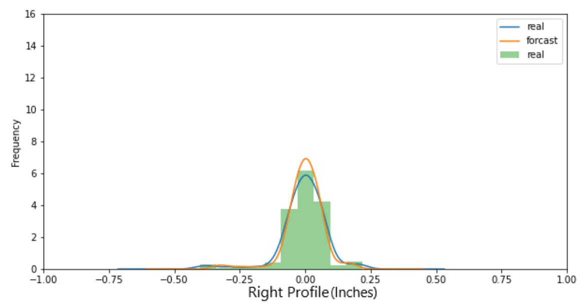
4.3.1.4 Model Training and Validation

During the model-building phase, the loss function, optimizer, and evaluation metrics are specified to train the model. The chosen loss function is the NLL. The model is trained using the fit function with input and output sequences, and various callbacks (e.g., reducing the learning rate on the plateau, early halting, and model checkpointing) are employed to track results and prevent overfitting.

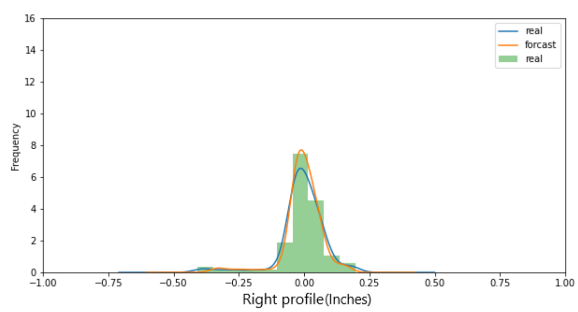
Once the model is trained, it is evaluated using a test sample. The best-performing model is loaded from a saved checkpoint file based on validation loss and predictions are generated using this model. To assess the precision and uncertainty of the predictions, the predicted distributions were compared with the actual data, showing their contour and spread for the first month of segment 3200 – 3400 feet where the actual histogram, actual distribution, and forecast distribution are depicted, as shown Figure 16 (a).



(a) Segment (3200 feet - 3400 feet) month 1



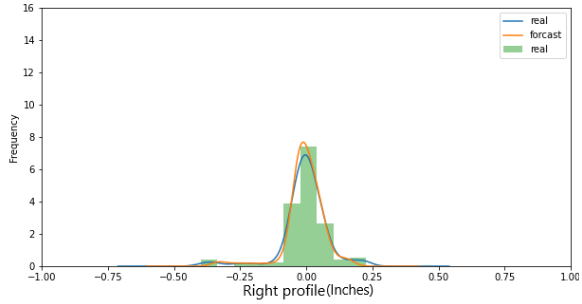
(b) Segment (3200 feet - 3400 feet) month 2



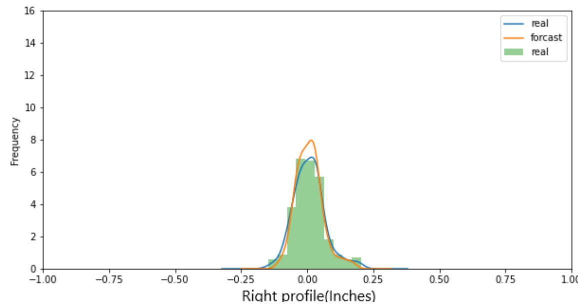
(c) Segment (3200 feet - 3400 feet) month 3

Figure 16. Segment (3200 feet - 3400 feet) months[1, 2 , 3]

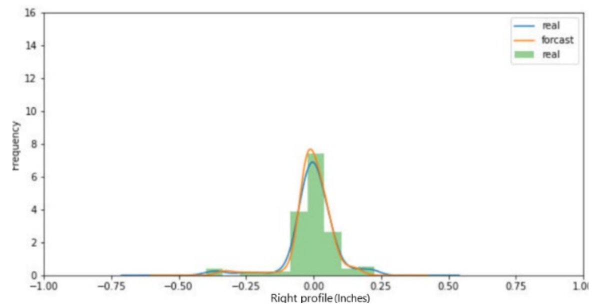
The results explore the predicted distributions in more detail, visualizing features like means and standard deviations to gain better insights into the predictions. These results are shown in the remaining plots of Figure 16 and Figure 17 for the same segment for six months of prediction.



(a) Segment (3200 feet - 3400 feet) month 4



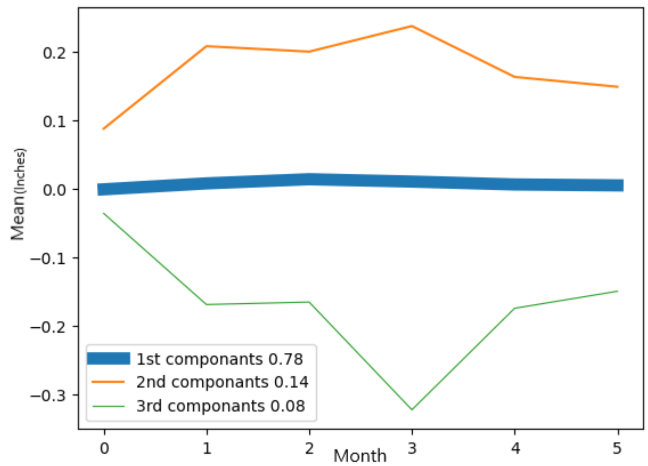
(b) Segment (3200 feet - 3400 feet) month 5



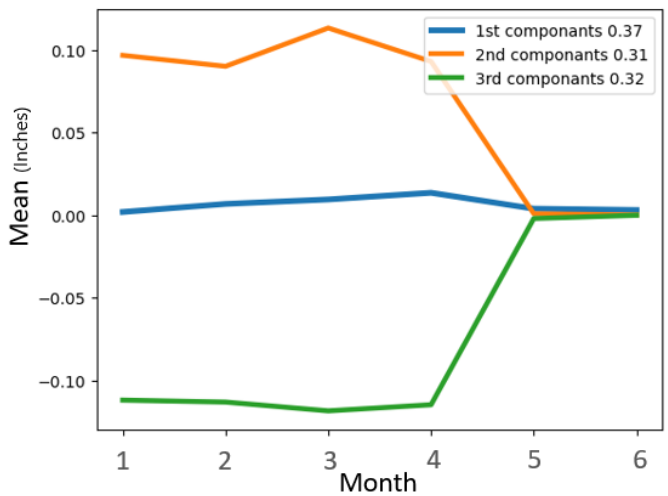
(c) Segment (3200 feet - 3400 feet) month 6

Figure 17. Segment (3200 feet - 3400 feet) months [4, 5 , 6]

Additionally, the results show how the projected weights, means, and standard deviations change with time (as shown in Figure 18), enabling a better understanding of the individual components of the mixture distribution. The individual line weights represent the magnitude of the mixing coefficient which are provided numerically in the legend.



(a) Forecast means and weights for LSTM-MDN in Segment 40



(b) Forecast means and weights for LSTM-MDN in Segment 60

Figure 18. Forecast means and weights for LSTM-MDN in Segments 40 & 60

4.3.1.5 Results and Discussion

Figure 16 and Figure 17 highlight the outcome of the LSTM-MDN model and allow for a comparison between the model's predictions and the PDF of each segment. This visual comparison enables a qualitative assessment of how well the model captures the distributional characteristics of the track segment data.

Figure 18 provide specific examples of how the projected weights and means change over time. Figure 18 (a) shows that there is a dominant normal distribution within the mixture of normal distributions by evidence of the 078 mixing coefficient of the first component. This suggests that for that segment, the LSTM-MDN model predicts a single primary distribution as the most likely outcome.

Conversely, Figure 18 (b) depicts a segment that exhibits approximately three distinct distributions. The mixing coefficients for the three components are all approximately one third. This indicates that the LSTM-MDN model has the capability to adapt to different numbers of distributions and can capture the multimodal nature of the data. It shows that the model can predict segments that may have multiple potential outcomes, each with its own mean and standard deviation.

The ability of the LSTM-MDN model to handle different numbers of distributions is a valuable feature, as it allows for more accurate and flexible predictions. It acknowledges the inherent variability and complexity of track geometry data, which can exhibit diverse patterns and behaviors.

Figure 19 presents an example of segment distributions, showcasing the presence of a mixture distribution. This highlights the importance of incorporating a mixture of distributions to capture the multimodal nature observed in this segment accurately. It becomes evident that a single distribution is insufficient to represent the complexity of the data, calling for a more comprehensive approach.

Several models with different numbers of components ranging from 1 to 5 were tested to determine the optimal number of components in the mixture density, which is crucial in achieving a balance between model complexity and accuracy. Too few components may lead to an oversimplified representation, while too many components can result in overfitting or unnecessary computational overhead.

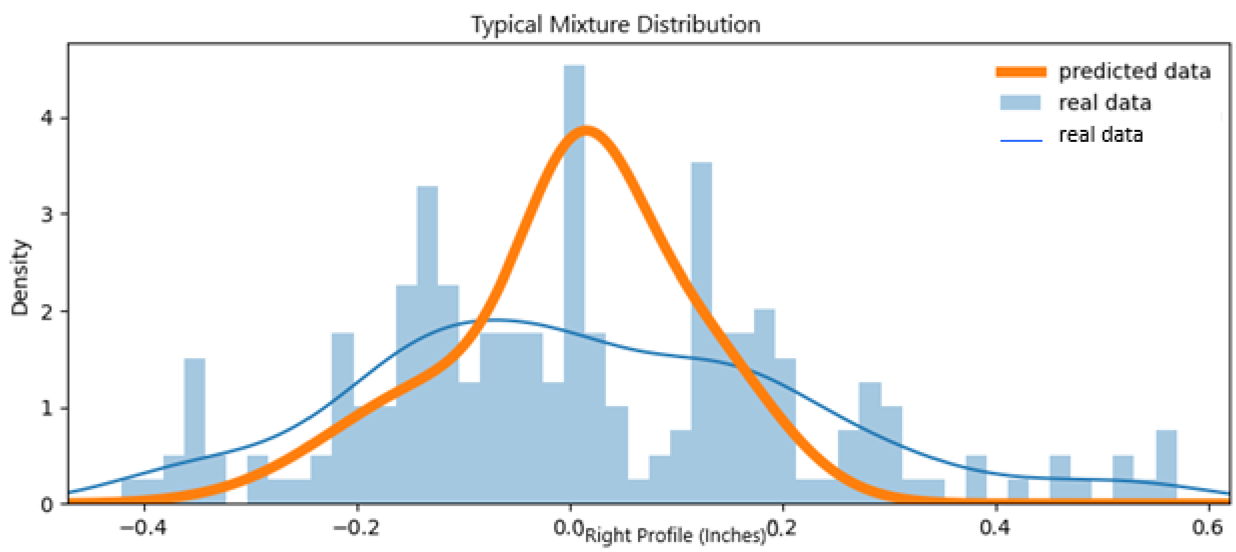


Figure 19. Typical mixture distribution

By selecting three components in the mixture density approach, the model strikes a balance between capturing the multimodal nature of the data and maintaining a manageable level of complexity. Overall, the examination of different numbers of components in the mixture density highlights the importance of finding the optimal balance between complexity and accuracy in modeling track segment distributions. In this case, it was determined that three components were sufficient to capture the observed multimodality, effectively representing the various modes in the data without excessive complexity.

As shown in Figure 18, the value and the weight of each of the three distribution components is represented and recorded. For instance, for each prediction of any segment, there are three distributions from any predicted distribution; each of these distributions has it is own weight, mean, and standard deviation, where the summation of the three weights is equal to 100 percent. Every segment does not necessarily have to be composed of three distributions. When a segment consists of one uniform distribution, the weight of the other two is equal to zero. This is helpful, especially for segments where the distribution is skewed to one side.

Furthermore, given the results of the six month predictions (shown in Figure 16 and Figure 17), the orange line represents the predicted values, the thin blue line presents the actual PDF of the real data, and the light blue blocks represent the histogram for any segment. Some of the graphs captured the existence of multiple distributions within segments, depict the behavior of some of the segments' probability density function, and capture the skewness of the distribution.

Considering these results, the LSTM-MDN model showed more refined and better outcomes than those of the deterministic model. However, the model has some limitations and still didn't provide the desired results. Although the model was able to depict some of the probability density functions of the actual data, some features are still not realized, as discussed below.

The first is the error of distribution build-up due to the inaccuracy of the probability density function of the actual data. The accuracy of the mixture distribution depends on increasing the likelihood of the prediction of the distribution of the segments. As seen in the graphs, the probability density function of each segment doesn't depict all the data histogram visually, since some of the points are disconnected. Moreover, whenever there is a sudden change in the track behavior (e.g., a maintenance event) the model will not be able to capture it accurately.

Secondly, given that there is a mix of three distributions to represent each segment, most of the segments' middle distribution mean is equal to zero because most of the points in each segment of the profile data are centered around the zero point. This affects the training progress, and the model starts to learn to fit the middle distribution mean to equal zero. However, this will wash out the segments with skewed distributions.

Overall, the LSTM-MDN model showed more promising results than the deterministic model. But when trying to predict the track segments and the uncertainty accompanying every segment, the error of predicting over each month and the existence of sudden changes in the track over time have an effect. That is why in the next proposed model, the focus is on predicting each point in the track foot by foot rather than looking at each segment.

Overall, the results from the MDN model demonstrate its effectiveness in capturing the behavior of some track segment probability density functions. By incorporating multiple components and accurately representing the multimodal nature of the data, the model provides a more comprehensive and accurate depiction of the track segments' degradation patterns.

4.3.2 Long Short-Term Memory Networks and Normal Distribution (point wise)

4.3.2.1 Model Architecture

Time series plays an important role in these models, as shown in the last two modeling approaches. Since the goal is for a model that captures the behavior of track segments over time, the LSTM is mainly incorporated in the two final models. However, in this section, rather than predicting segments of the track, the focus is on predicting the distribution of output as in individual points by combining the LSTM model and normal distribution as the last layer. The goal is to generate the output distribution for each individual point within any segment and see the uncertainty of predicting the mean and standard deviation of each point.

A LSTM normal distribution (LSTM-ND) based model was created to incorporate the historical track geometry data (i.e., profile or alignment) for the previous 12 months. The input data provides the necessary contextual information to seize the temporal dependencies and pattern within the data. The LSTM model is well suited for this type of prediction since it allows researchers to model and learn from sequential data in an efficient manner.

The final layer of the LSTM model is a normal distribution layer with multiple layers within, some of which are LSTM. The LSTM network can anticipate the characteristics of a normal distribution for each point, which results in an output with a probability distribution. The model can capture the central tendency and the spread of the expected values for any point along the track and in time, by estimating the mean and variance of the distribution.

The prediction of the next six months gives a thorough picture of the uncertainty accompanied by each predicted point. This probabilistic approach allows the capture of the inherent variability and potential range of values for the predicted output rather than relying on a single-point estimate.

The LSTM model is trained using suitable loss functions that consider the characteristics of the output's normal distribution. The model's parameters are improved to reduce disparities between the anticipated and actual distributions and increase the likelihood of the observed data. The LSTM network learns to produce precise and representative distributions for the specified prediction horizon because of this training process.

Once the model training is complete, it is applied to the test dataset to obtain new predicted points.

4.3.2.2 Data Preparation

In this model, the goal is to predict uncertainties point by point. In other words, when predicting the next six months in time using LSTM, the last layer of the model (i.e., normal distribution) can produce the mean and standard deviation of the predicted points.

The mean provides an indication of the average value for each data point, while the standard deviation reflects the uncertainty associated with predicting each point. It is important to note that each data point has its own unique meaning and standard deviation, capturing the individual characteristics and variations in the dataset.

The standard deviation values are presented in Figure 20, allowing for the identification of the confidence level in the predictions. By considering the standard deviation, confidence in the predictions can be assessed regarding the value of each point. For example, when accounting for one standard deviation (plus and minus the mean), there is 68.3 percent confidence in the prediction. Similarly, two standard deviations correspond to a confidence level of 95.5 percent, and three standard deviations represent a confidence level of 99.7 percent.

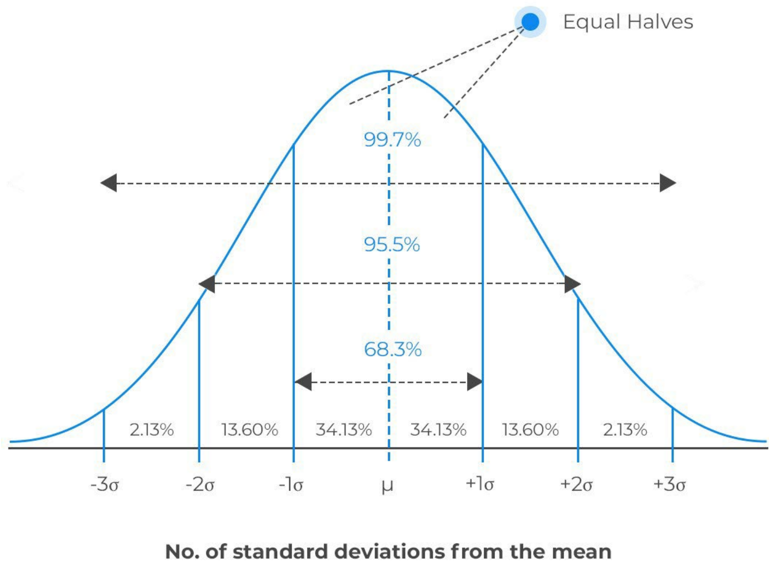


Figure 20. Illustration of the normal distribution and confidence percentage [19]

4.3.2.3 Prediction and Distribution Results

The team analyzed the results obtained from the LSTM- ND model using a normal distribution to predict points and the associated uncertainty using ±2 standard deviations.

Figure 21 depicts the prediction of profile values for a track segment spanning 0 to 2000 feet. This figure provides valuable insights into the performance and reliability of the LSTM-ND model. The following is a breakdown of the elements depicted in the figure:

Actual Left Profile Value: The straight red line in the figure represents the actual left profile value. This serves as a reference point to assess the accuracy of the model's predictions.

Anticipated Mean: The blue dotted lines illustrate the anticipated mean for the distribution of predicted values. The mean represents the central tendency of the predicted data points and provides a measure of the model's overall performance.

Variation from the Mean: The green shaded area captures the variation of the data points from the mean, considering two standard deviations. In other words, the blue shaded area represents the range of 95 percent confidence that the predicted geometry surface values will fall. It indicates that approximately 95 percent of the predicted values will be within this range, above or below the mean.

Using ±2 standard deviations offers valuable insights into the uncertainty associated with the model's predictions. By considering the range that encompasses most of the predicted values (95 percent confidence interval), a deeper understanding of the potential outcomes and the level of certainty can be attributed to the model's predictions.

This estimation of uncertainty is crucial in practical applications as it allows decision-makers to gauge the reliability of the predictions and make informed judgments. The ±2 standard deviations provide a reasonable margin for error, indicating the range within which the actual values are likely to fall.

Looking at the blue dotted points, the model-predicted mean values mostly capture the actual values of the profile with low variance, especially where there is a small variation in the profile values (Segment 500 to 1500 is an example). However, near 400, 1520, and 1540 feet, the standard deviation of these points is relatively larger, and the mean slightly misses capturing the peak in these locations. At the same time, the actual measurement falls within the distribution range.

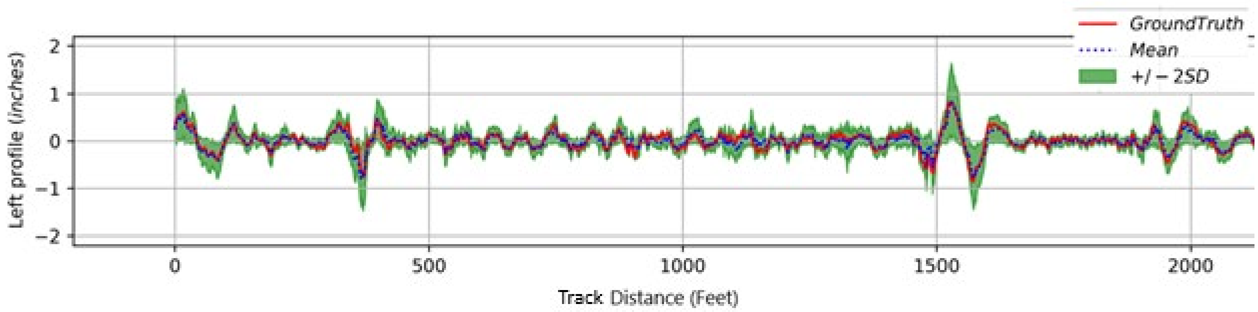


Figure 21. Segment from 0 to 2000 feet

Figure 22 through Figure 27 highlight an extended version of the results from 0 to 2000 feet. By visualizing the predicted values and the associated uncertainty, the figures offer a comprehensive view of the LSTM-ND model's performance, allowing the model to capture the underlying patterns in the data and make accurate predictions.

Using a normal distribution in the LSTM-ND model allows for a probabilistic approach to prediction. Instead of providing a single-point estimate, the model generates a distribution of values, with the means representing the most likely outcome. The probabilistic nature of the predictions aligns well with real-world scenarios where outcomes are often subject to various sources of uncertainty.

The analysis of the predicted points and the associated uncertainty provide valuable insights into the strengths and limitations of the LSTM-ND model. Researchers can assess the model's overall accuracy, understand the range of potential outcomes, and make informed decisions based on the level of certainty associated with the predictions.

Including uncertainty in the form of standard deviations aligns well with real-world scenarios, where uncertainties and variations are inherent in track conditions. This probabilistic approach allows for a more nuanced interpretation of the predictions and aids in making informed decisions regarding infrastructure management and maintenance activities.

Accurate prediction of the track condition is important for effective infrastructure management. The insights provided by the LSTM-ND model can contribute to better decision-making, enabling proactive maintenance and targeted interventions in areas that require attention. By understanding the potential outcomes and the associated uncertainty, stakeholders can allocate resources effectively, prioritize maintenance efforts, and optimize track safety.

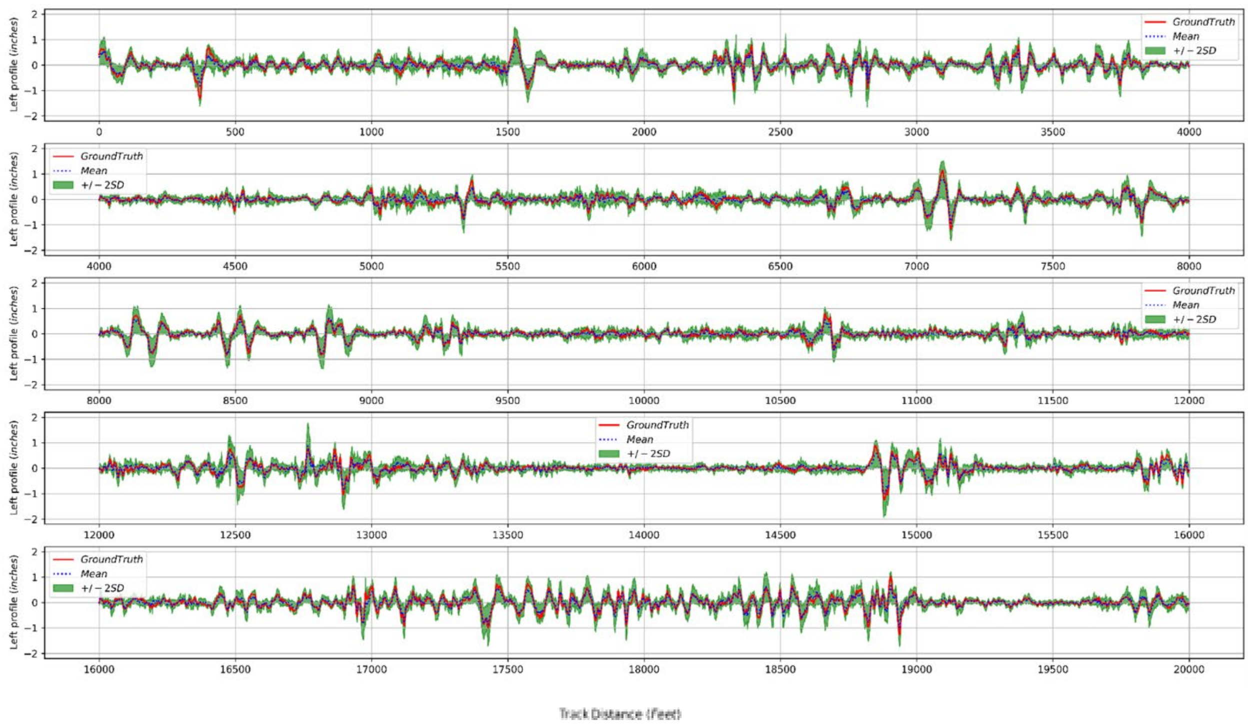


Figure 22. The results of the model including the uncertainty of predicted points for the first month

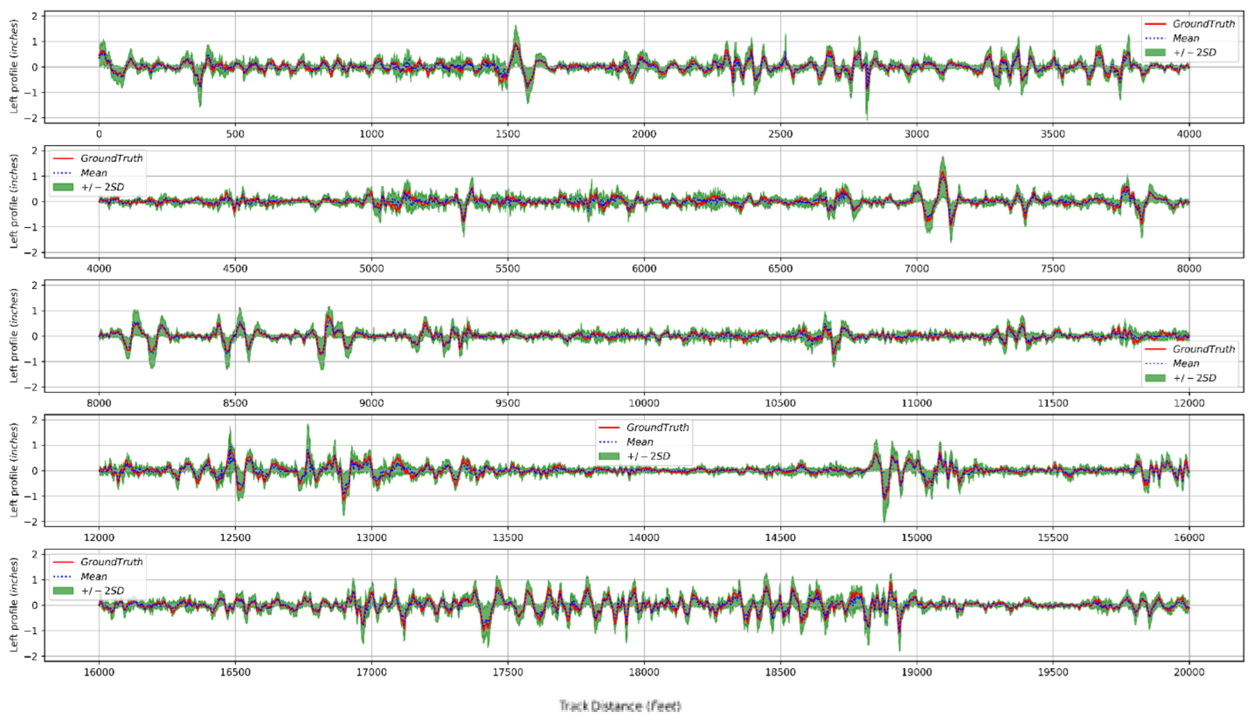


Figure 23. The results of the model including the uncertainty of predicted points for the second month

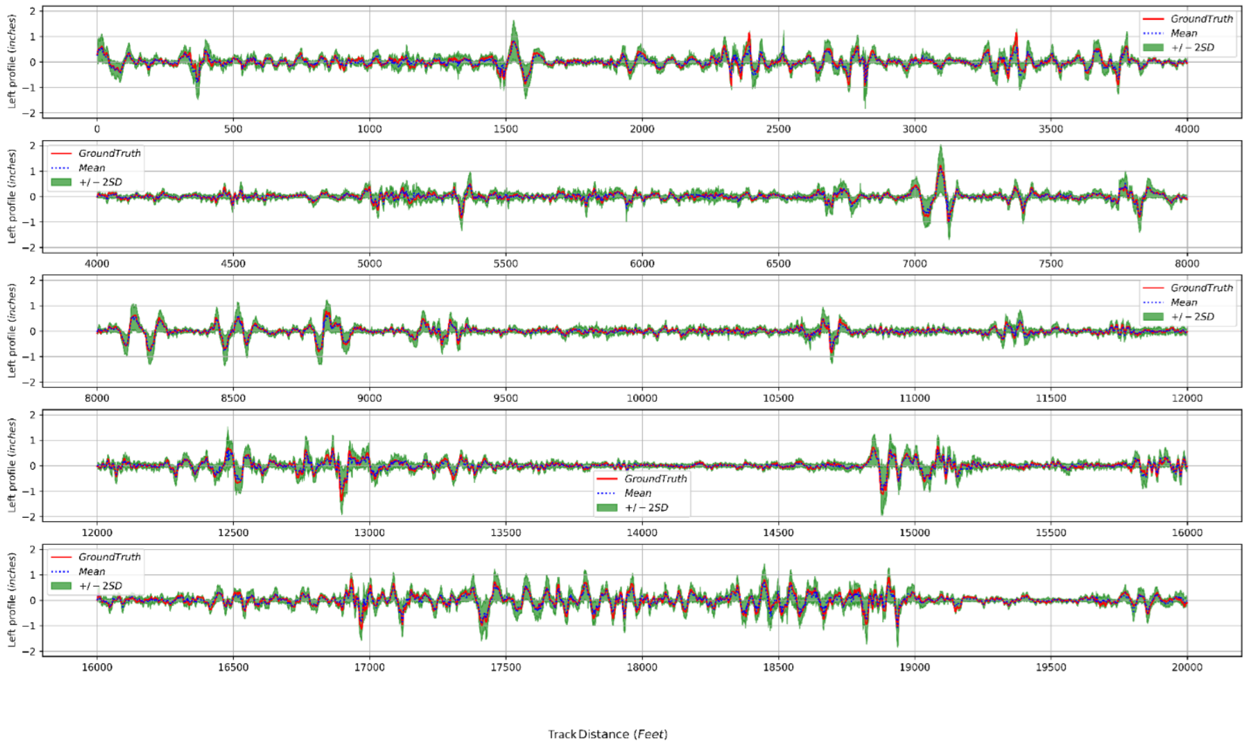


Figure 24. The results of the model including the uncertainty of predicted points for the third month

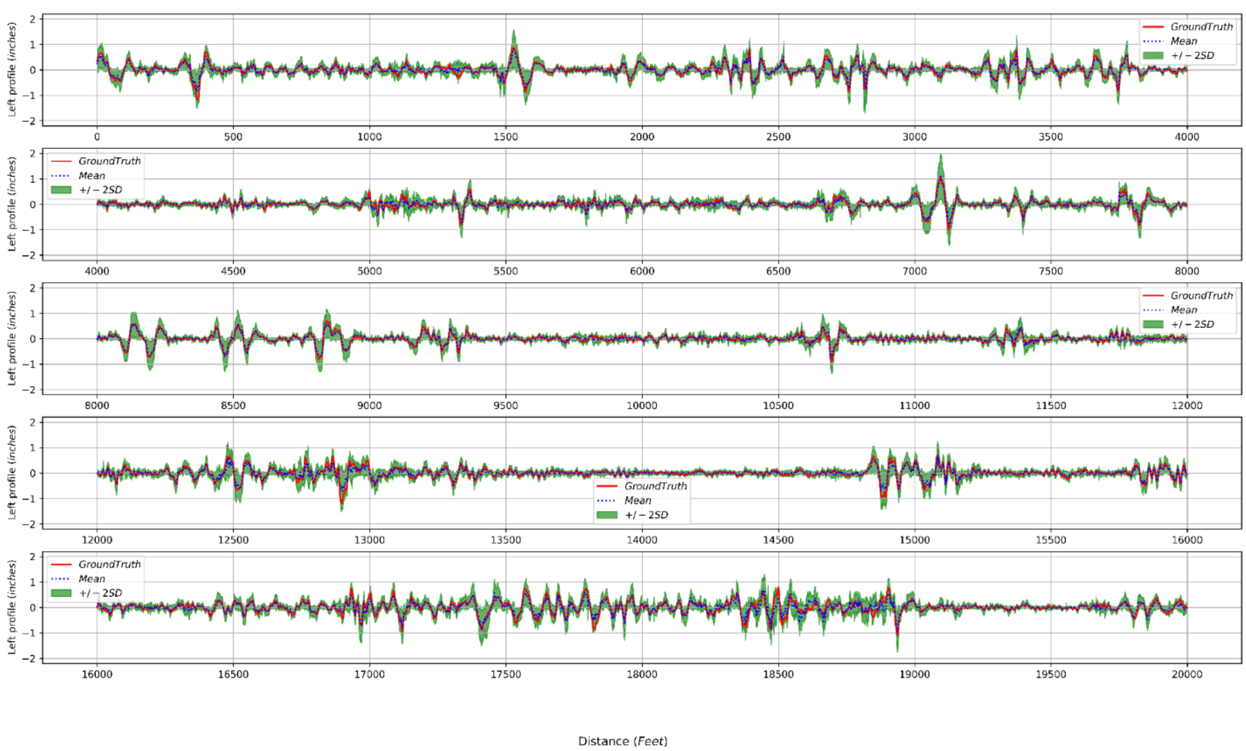


Figure 25. The results of the model including the uncertainty of predicted points for the fourth month

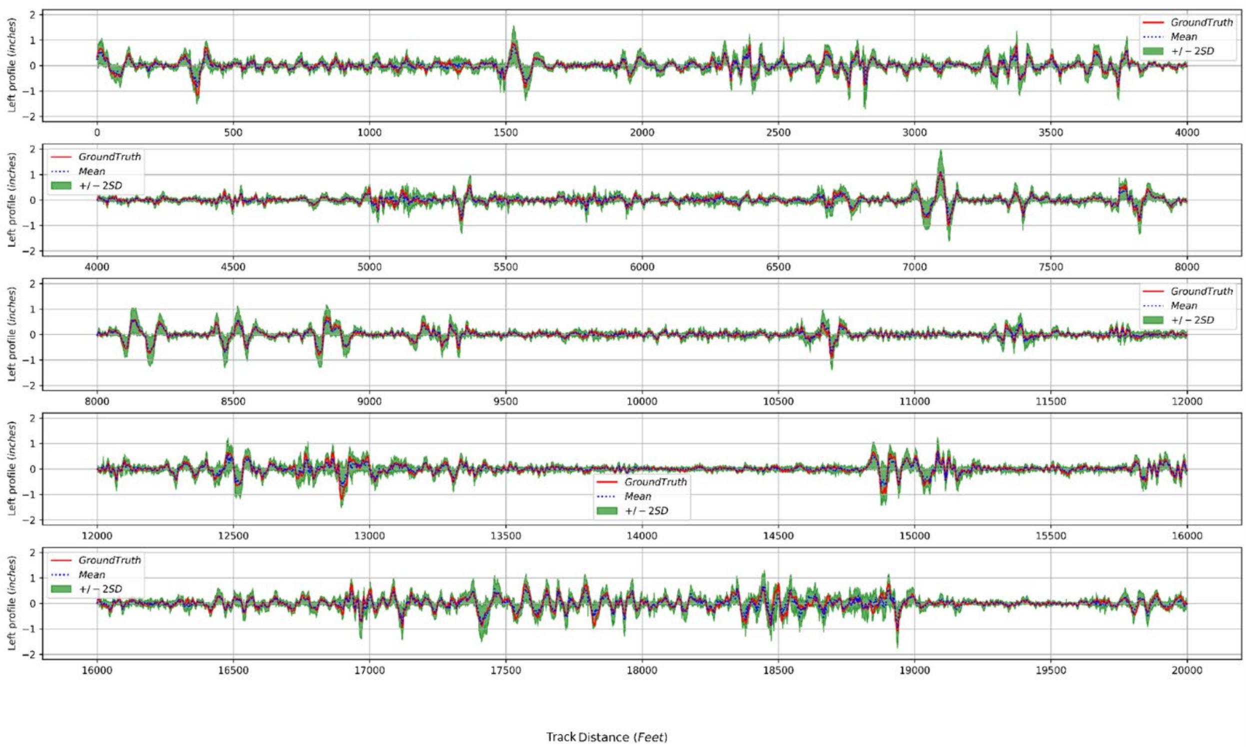


Figure 26. The results of the model including the uncertainty of predicted points for the fifth month

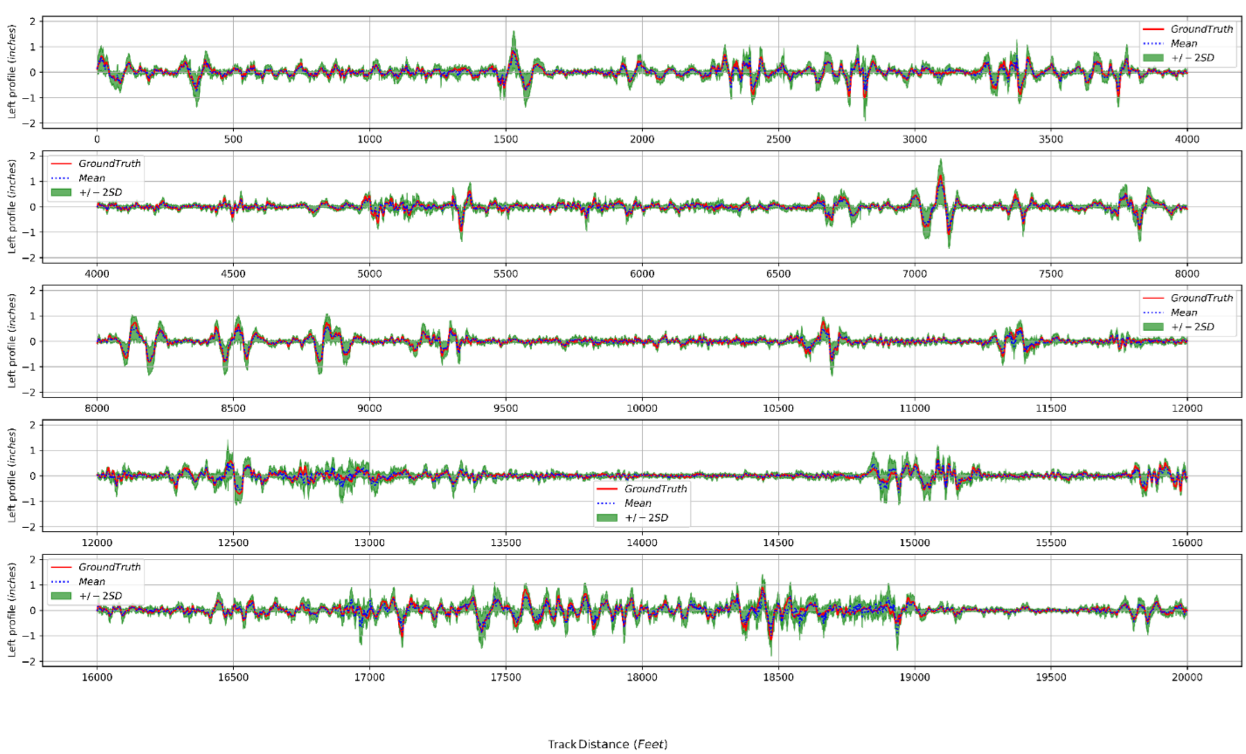


Figure 27. The results of the model including the uncertainty of predicted points for the sixth month

4.3.3 Probability of Exceedance

The team used the distribution of each point to get a sense of the percentage of points exceeding certain thresholds. Figure 28 demonstrates the probability of exceeding certain thresholds along the 1500 to 4000 feet segment: 0.5 inch (half the FRA limit for profile), 1.0 inch (the FRA limit for the profile), and 1.5 inch (one and half the FRA limit of the profile). Each predicted point is associated with a normal distribution that represents the uncertainty surrounding the mean value. The probability of exceedance refers to the percentage of predicted data expected to surpass these limits for each point.

In Figure 28 (left profile segment), the blue line represents the probability of exceeding 0.5 inch, the red line the probability of exceeding 1 inch, and the green line the probability of exceeding 1.5 inch, with a maximum probability of 100 percent. Moreover, there is a dominant presence of points exceeding 0.5 inch compared to 1 and 1.5 inch due to the minimal presence of points exceeding the latter limits, which are at or exceed the defined FRA safety limit.

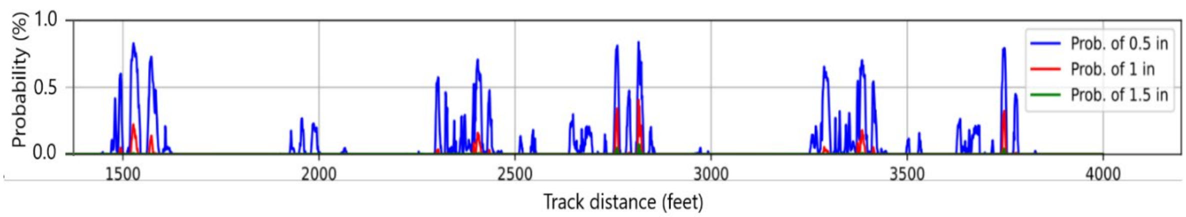


Figure 28. Segment from 1500 to 4000 feet

Figure 29 through Figure 34 provide the probability of exceeding 0.5, 1, and 1.5 inch of left profile for the entire two miles of track. Each figure represents the respective probabilities of exceedance for each of the six months of prediction. From these figures, locations can be identified with higher probability of exceedance that are stable (i.e., do not change much month to month), locations that are increasing in probability, locations that appear with small probability (i.e., having zero in previous months), and locations that are very stable for all six months (i.e., zero probability of exceedance).

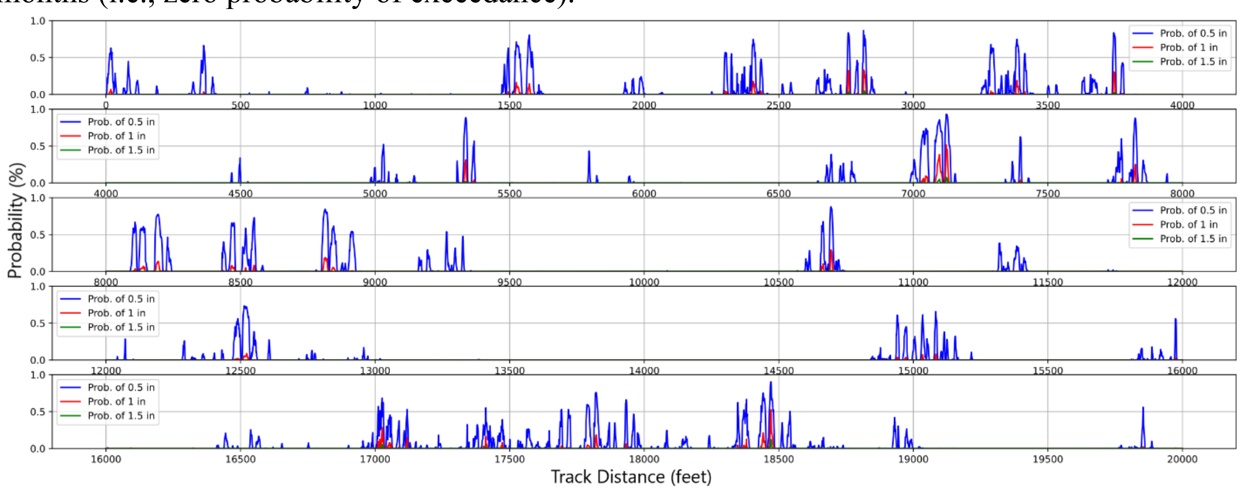


Figure 29. The results of the model including the uncertainty of predicted points for the first month

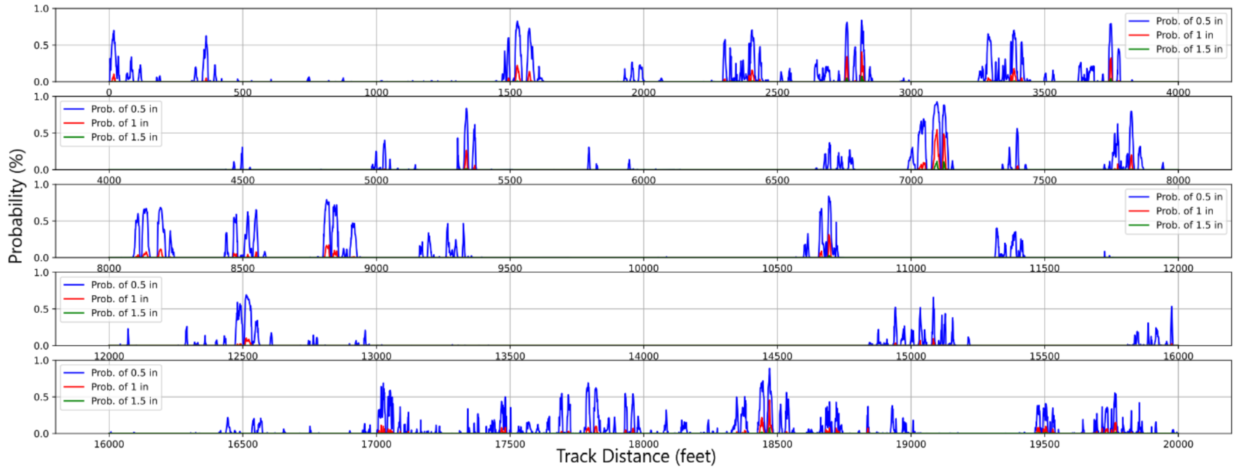


Figure 30. The results of the model including the uncertainty of predicted points for the second month

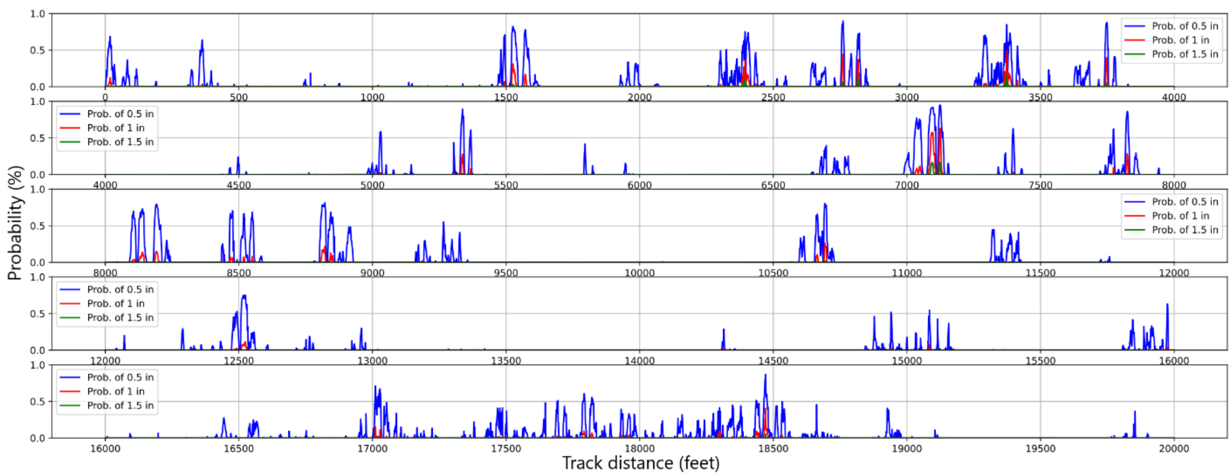


Figure 31. The results of the model including the uncertainty of predicted points for the third month

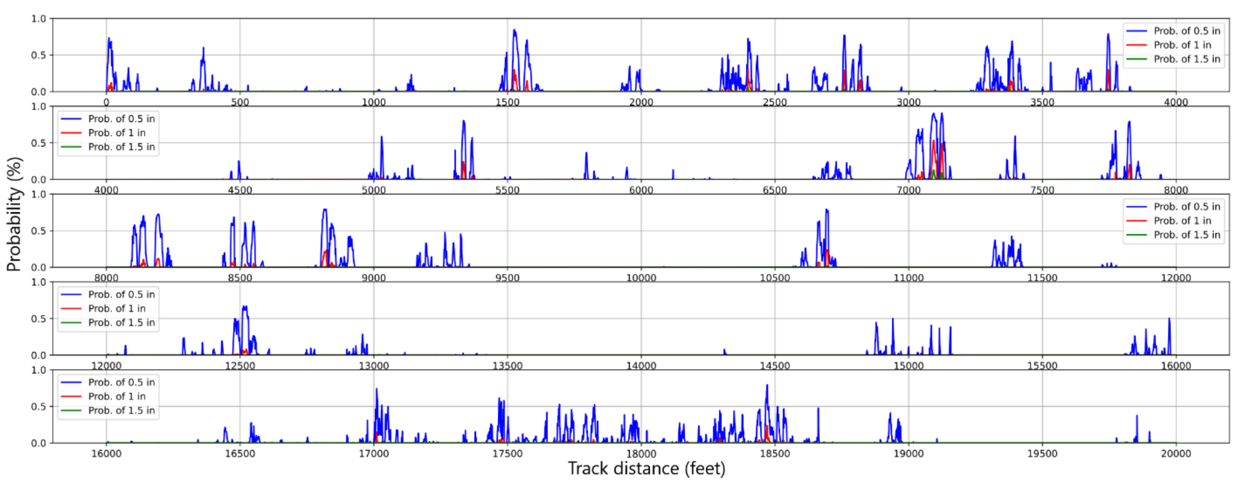


Figure 32. The results of the model including the uncertainty of predicted points for the fourth month

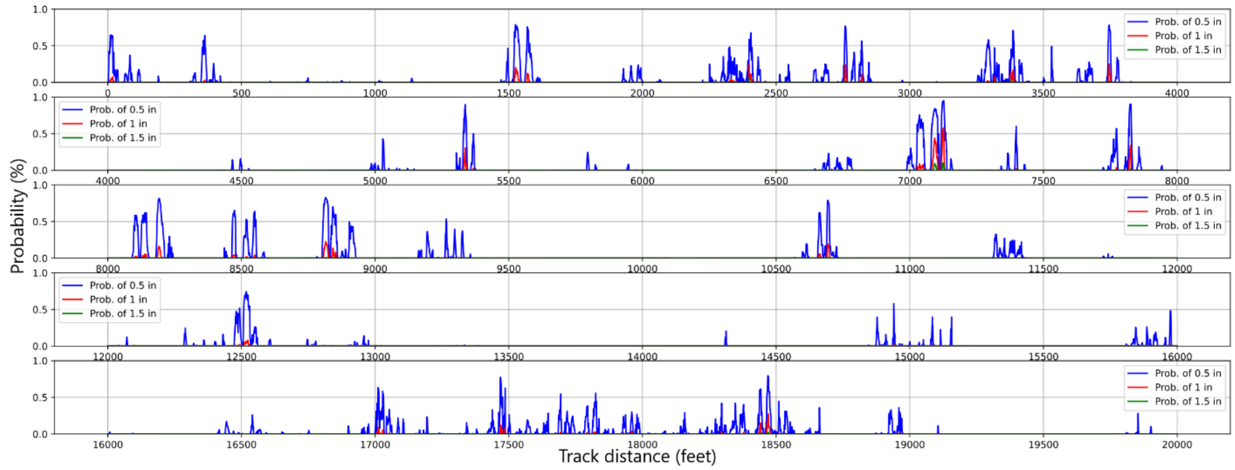


Figure 33. The results of the model including the uncertainty of predicted points for the fifth month

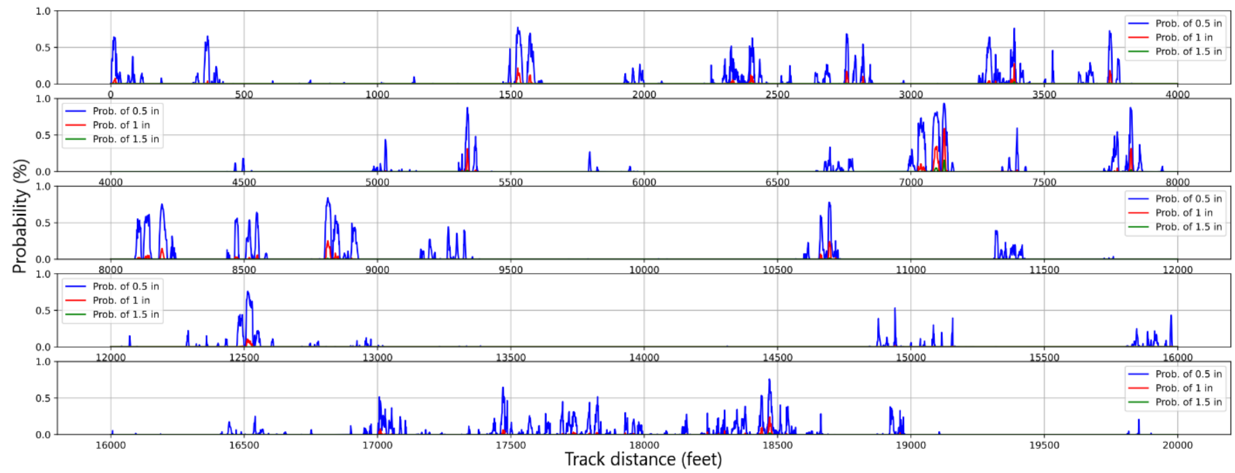


Figure 34. The results of the model including the uncertainty of predicted points for the sixth month

The previous figures show the locations of high risk for a geometry perturbation occurring. Figure 35 and Figure 36 show the spread of the uncertainty of the predicted points versus the actual reading for the first and sixth month, respectively.

The primary objective of this research was to develop a novel TSQ index that considers a time component, and takes advantage of the previously discussed ability to predict track geometry deviation with its associated uncertainty. A new probability-based index was proposed, which uses the probability of exceeding half the FRA track class limit for a defined channel. The equation below represents the defined approach.

$$NewTQI = \frac{\sum_{i=1}^{n} P(0.5L)_i \log(P(0.5L)_i)}{1.5}.$$

Where

n = number of points in a segment

L = FRA safety limit (define by track class) for the channel being analyzed

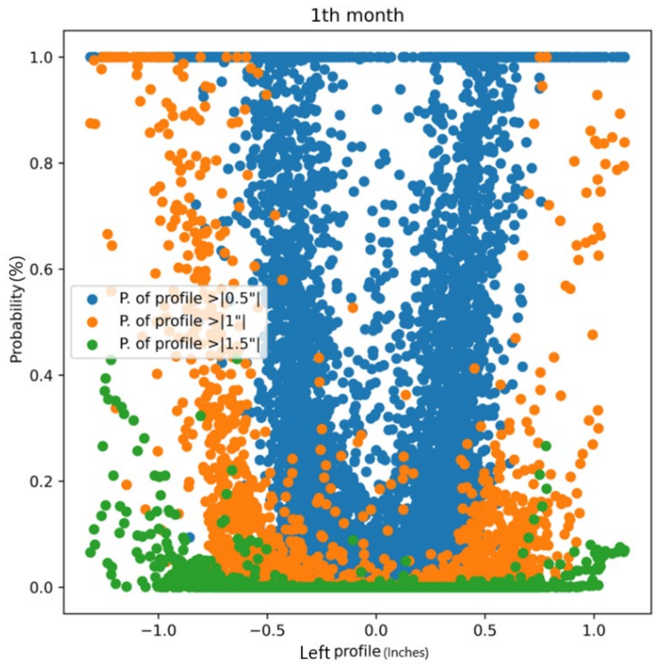


Figure 35. Spread of the uncertainty of predicted points for the first month

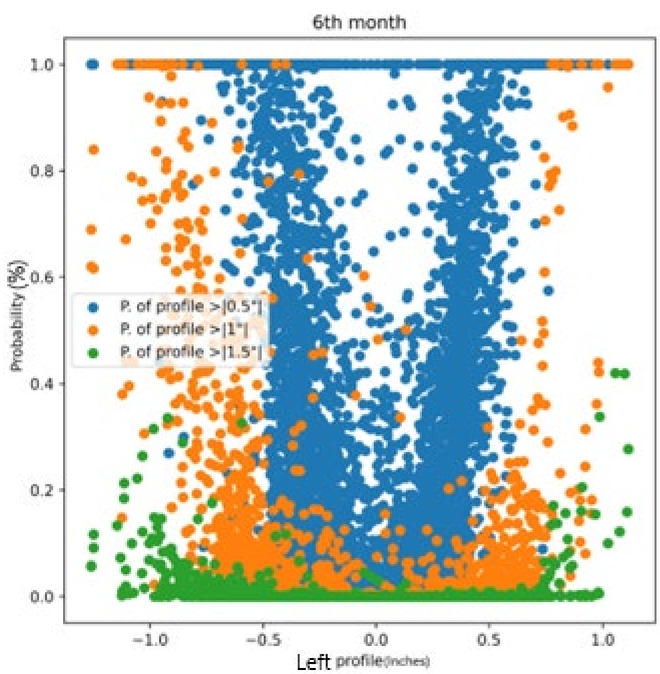


Figure 36. Spread of the uncertainty of predicted points for the sixth month

Using the equation above, the methodology is employed as follows: for any 500-foot segment[1], the probability of surpassing half the limit is multiplied by the logarithm of the same probability value, divided by a factor, and then combined to represent the probability for the 500 points in

[1] Note, the methodology allows for any segment length to be defined.

the segment. Application of this equation for left profile is illustrated in Figure 37 and Figure 38 for the first 4000 feet of track for the first and sixth months of prediction, respectively. The top panel of the plot includes the point-by-point probability of exceedance. In the center panel, the combined probability for the segment (NewTQI) is visually depicted as the shaded blue area. Additionally, this value is compared with the standard deviation of points within each segment (gold shaded area), which serves as a commonly used quality index. The measured left profile value for that month also is shown. Finally, the bottom panel shows the next 4000 feet of probability prediction.

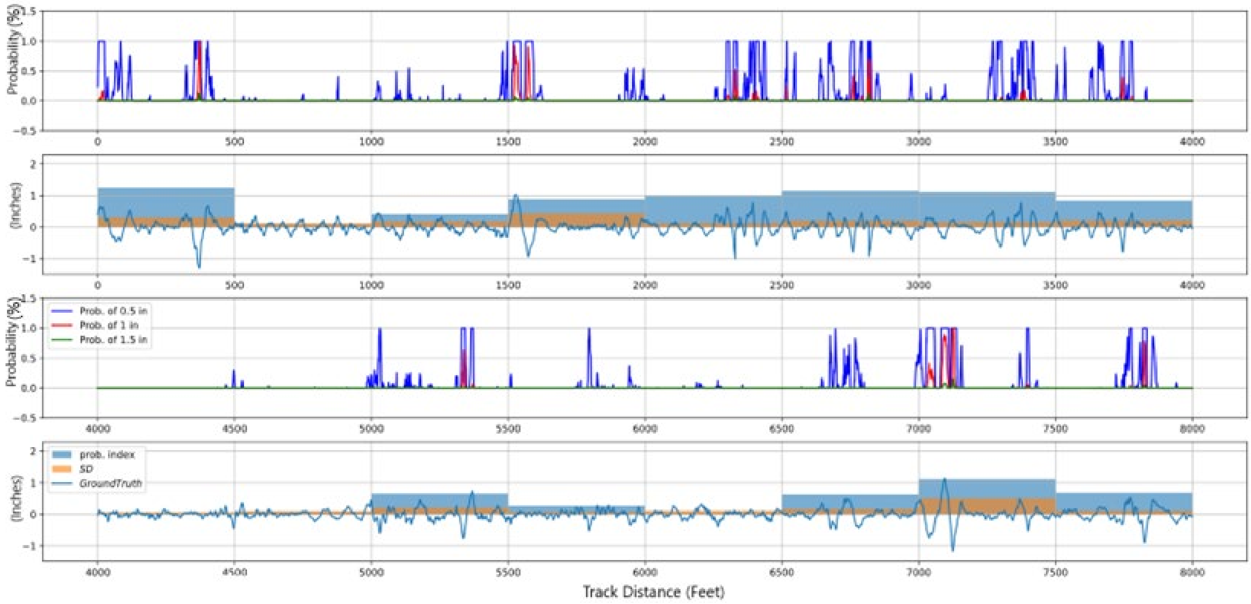


Figure 37. The results of the model including the uncertainty of predicted points for the first month

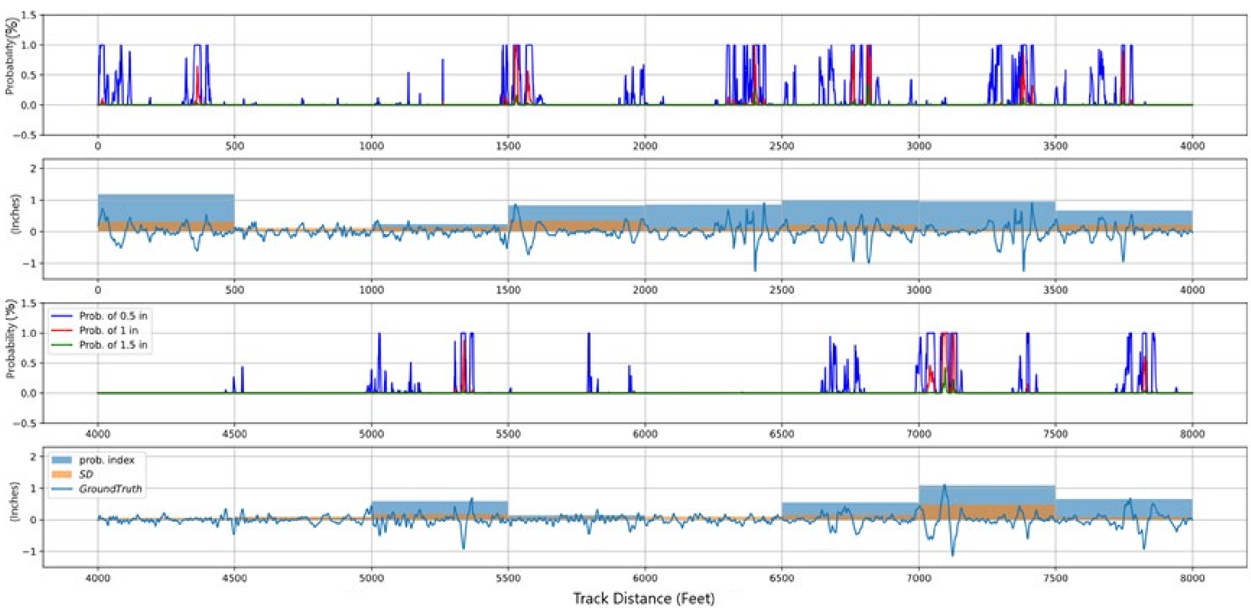


Figure 38. The results of the model including the uncertainty of predicted points for the sixth month

From the figures above, it is clear that the NewTQI exhibits greater sensitivity to various forms of roughness or degradation compared to conventional techniques, such as relying solely on standard deviation. By introducing the concept of probability of exceedance and incorporating it into the calculation of NewTQI, this research addresses the limitations of traditional quality indices, mainly by introducing a time component. The probability index considers both the uncertainty associated with each predicted point and the likelihood of surpassing predefined thresholds. This approach provides a more comprehensive evaluation of track quality, enabling a more nuanced understanding of potential issues.

It should be noted that the application of the methodology thus far has been constrained to a six month time horizon. Extension to further time horizons to allow for longer term planning can be achieved by extending the time horizon of prediction, which will still encompass uncertainty.

This approach can capture and quantify the likelihood of exceeding critical limits for track roughness or degradation. By leveraging probability distributions and combining them within segments, NewTQI offers a more reliable and accurate track quality representation than traditional methods. The inclusion of the logarithm in the calculation further emphasizes the importance of extreme values, allowing for a more balanced assessment of the track condition.

Moreover, using the standard deviation as a quality index serves as a benchmark for comparison. While standard deviation provides valuable insights into the variability of track data, it may not fully capture the critical aspects related to exceedance probabilities. The NewTQI accounts for the distribution of values and the likelihood of exceeding specific limits, resulting in a more comprehensive assessment of track quality.

Table 6 illustrates the traditional TQI (standard deviation) and the proposed TQI (NewTQI) for multiple 500 foot segments of left profile. The table provides a side-by-side view of TQI values, allowing for a comparison of the proposed TQI performance across different track sections.

Table 6. Traditional TQI vs Proposed TQI

Segment	Traditional TQI (standard deviation)	Proposed TQI (NewTQI)
0-500	0.3	1.18
500-1000	0.05	0
1000-1500	0.1	0.15
1500-2000	0.3	0.9
2000-2500	0.07	1
2500-3000	0.1	1.05
3000-3500	0.12	1.03
3500-4000	0.04	0.88
4000-4500	0.04	0
4500-5000	0.04	0
5000-5500	0.15	0.6
5500-6000	0.04	0.2
6000-6500	0.05	0
6500-7000	0.16	0.6
7000-7500	0.5	1.07
7500-8000	0.09	0.7

A notable pattern emerges in the table, highlighting consistent improvements in TQI values with the proposed methodology across several segments. For instance, using Figure 38 and comparing between segment 0-500 and 1500-2000, segment 0-500 generally has a rough track over the entire segment. On the other hand, segment 1500 to 2000 generally has a smooth track with two points with peak values. The traditional TQI treated both segments as equal. However, the proposed TQI showed a more accurate representation of the two segments with the first 500 feet having higher values than the latter segment with two peak values.

In conclusion, the consistent improvements observed across segments underscore the value of adopting more advanced techniques. Further research and field testing are recommended to validate the proposed TQI's real-world applicability and address any outliers or anomalies identified.

4.4 Three-dimensional Track Quality Index

To develop a 3D TQI, the team used the LSTM-ND model for each track geometry channel, including the left profile, right profile, left alignment, right alignment, and cross-level. Separate models were trained for each channel using the LSTM-ND architecture, allowing for channel-specific predictions.

The output distribution of each channel at every point along the track was combined using a multivariable normal distribution. This means that at each point, there is a distribution that represents the uncertainty and variability in the track quality across the different channels. By incorporating multiple channels into the distribution, the model captures the interdependencies and correlations between different aspects of the track geometry.

Using this combined distribution, the probability of exceeding the thresholds set by FRA was calculated for each channel. The FRA thresholds define the acceptable limits for various track geometry parameters, considering that each of the mentioned channels has its own FRA limit. For a Class 6 track, the alignment safety limit is 0.75 inch, the profile limit is 1.0 inch, and the cross-level rate is 0.78 inch. The overall probability of exceeding the FRA thresholds is derived by assessing the probabilities for each channel.

The 3D TQI integrates these probabilities from multiple channels and provides a comprehensive measure of the track's quality. It offers insights into the likelihood of the track geometry parameters exceeding the regulatory thresholds, indicating potential areas of concern or the need for maintenance actions.

The PDF of the multivariate normal distribution is given by the equation:

$$f(x) = \frac{1}{\sqrt{(2\pi)^k \det(\Sigma)}} exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

where:

- $\boldsymbol{x}$ represents a vector of length k (the number of channels)
- μ is the mean vector, which also has a length of k
- $\boldsymbol{\Sigma}$ is the covariance matrix, a square matrix of size $k \times k$
- $det(\boldsymbol{\Sigma})$ denotes the determinant of the covariance matrix
- $\boldsymbol{\Sigma}^{-1}$ represents the inverse of the covariance matrix
- *exp()* denotes the exponential function
- $\boldsymbol{x}^T$ represents the transpose of the vector $\boldsymbol{x}$

This equation describes the likelihood of observing a specific combination of values for the track quality parameters, given their joint distribution. Mathematically, the multivariable normal distribution is defined by its mean vector and covariance matrix. The mean vector, denoted by μ, represents the average values of the variables, and the covariance matrix, denoted by Σ, captures the relationships and variances among the variables.

The mean vector μ is a fundamental component of the multivariable normal distribution used to model the track quality at each point. It represents the expected values for each track geometry parameter, indicating the average behavior or typical values of the track quality at that specific point.

Examining the mean vector μ provides insights into the central tendencies of the track geometry parameters. It explains the average track profile, alignment, cross-level, and other parameters at a given point. This information is valuable for establishing baseline expectations and identifying deviations from the norm.

The covariance matrix Σ plays a crucial role in capturing the relationships between the track geometry parameters. It provides a comprehensive description of the variances and covariances among the parameters, revealing how changes in one parameter may be correlated or independent of changes in other parameters. The covariance matrix offers insights into the interdependencies and interactions between track geometry channels. For example, it can reveal whether an increase in cross-level is associated with a specific profile deviation or alignment change.

Understanding these relationships provides a more nuanced understanding of the track's overall behavior and helps identify potential patterns or anomalies. Analyzing the covariance matrix allows consideration of the joint behavior of the track geometry parameters rather than examining them in isolation. It helps to understand how changes in one parameter may impact or relate to changes in other parameters, providing a more comprehensive view of the track quality.

For example, consider a Class 6 track with the following safety limits: alignment limit of 0.75 inch, profile limit of 1.0 inch, and cross-level rate limit of 0.78 inch. To calculate the probability of exceedance for each channel, the threshold is set to half of these limits, which are 0.375 inch for alignment, 0.5 inch for profile, and 0.39 inch for cross-level rate. At each point, the track geometry parameters (e.g., alignment, profile, and cross-level rate) have their own probability distributions. These distributions capture the uncertainty and variability associated with each parameter at that specific point. Combining these individual distributions into a multivariate normal distribution allows analysis of the joint behavior of the parameters.

To calculate the joint probability of exceedance, compare the combined multivariate normal distribution with the safety limits for alignment (0.75 inch), profile (1.0 inch), and cross-level rate (0.78 inch). The joint probability of exceedance represents the likelihood that all the track geometry parameters simultaneously exceed half their respective safety limits at a given point. This value is used as an input in the multivariate normal distribution equation for each point, and then combined for each segment to get the new TQI.

It is worth noting that the specific equation and methodology for combining probabilities may vary depending on the context and requirements of the track quality assessment. The equation provided here serves as a general example, and further customization or modifications may be necessary based on specific project requirements and objectives.

4.4.1 Results

Figure 39 illustrates the 3D TQI for eight segments, each spanning 500 feet, during the sixth month. The index includes the left and right profile, left and right alignment, as well as the cross-level channels (measured data shown below the calculate TQI bars). The results demonstrate the effectiveness of the index in capturing irregularities across these five channels.

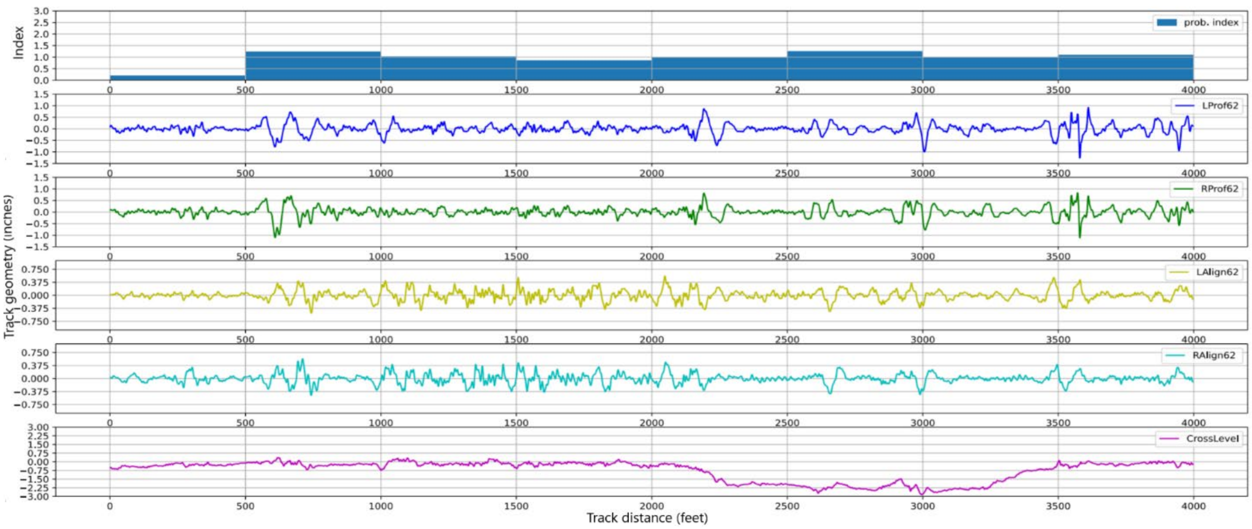


Figure 39. The results of the model including the uncertainty of predicted points for the sixth month

In the first segment, all channels exhibit a relatively smooth surface with minimal irregularities. This is reflected in the TQI, which indicates a generally favorable condition for the segment.

In contrast, the second segment displays significant irregularities, particularly in the profile channel. These irregularities have a pronounced impact on the overall TQI, reflecting the deteriorated condition of the segment. The index effectively captures the high level of irregularity in the profile channel, providing valuable insights into the segment's performance.

The flexibility of the 3D TQI allows for customization to meet the specific needs of a railroad. One way to customize the index is by assigning weights to each channel, thereby placing greater importance on certain parameters. This can include all channels, effectively zeroing out channels that are not of importance by placing a weight of zero on them. By assigning higher weights to channels that are deemed more critical or influential in assessing track quality, the customized index can prioritize those aspects that have a significant impact on overall track performance.

Note that the process of assigning weights should be based on careful consideration of operational requirements, safety standards, and maintenance priorities. It may involve consultation with experts, data analysis, and knowledge of the specific challenges and objectives of the railroad.

4.4.2 Conclusion

To develop ML models for predicting track geometry growth and uncertainty, the team trained five separate models, each dedicated to one specific track parameter channel. These models were designed to predict the distribution of each point along the track, resulting in five output distributions (one for each channel).

The LSTM model played a pivotal role in capturing temporal patterns and dependencies in the data. Due to its recurrent neural network architecture, it retained information from previous data points, enabling a more accurate prediction of future track conditions. By leveraging historical data, the LSTM model can effectively capture the dynamic behavior of track parameters over time.

In addition to capturing temporal dependencies, the model also incorporates a normal distribution layer for modeling the uncertainty associated with track parameters. This layer facilitated the estimation of the probability distribution for each predicted point, providing insights into the range of potential values and associated uncertainties.

The ML approach considers the distinct characteristics and variability of different track parameters by training separate models for each channel. This allowed for a more focused and accurate distribution prediction for each specific channel, leading to a more comprehensive understanding of the track's condition.

The use of multiple models for each channel, with their corresponding output distributions, offered valuable insights into the variation and uncertainty associated with the track parameters. This comprehensive information facilitated the development of the 3D TQI, which incorporated probabilities of exceedance based on predefined safety limits set by FRA.

After training and validating each individual model for the five channels (i.e., left profile, right profile, left alignment, right alignment, and cross-level rate), the team combined the output distributions into a single multivariable normal distribution for each point along the track. Combining the five distributions allows for a holistic representation of the track's condition at each foot. This multivariable normal distribution includes the joint probability distribution of the track quality parameters, considering their interdependencies.

Combination of the distributions was achieved by leveraging techniques such as multivariate statistics and probability theory. The mean and covariance matrices of the individual distributions were used to calculate the mean vector and covariance matrix of the multivariable normal distribution. The resulting multivariable normal distribution for each foot provides valuable information about the expected values and uncertainties associated with the track.

5. Conclusion

In this research, FRA sponsored a team from the University of Delaware to develop a 3D, time-based TQI for assessing and monitoring the condition of railway tracks. The research team evaluated TQIs by considering several characteristics categorized into three main groups: standard deviation-based, threshold-based, and miscellaneous indices.

Standard deviation-based indices quantify the assessment of track quality by measuring the dispersion of track geometry parameters. However, though this might provide a comprehensive assessment of roughness, it could lead to some disproportionately high values for certain parameters in sections with curves, which may not necessarily indicate track defects.

Threshold-based indices convert the number of threshold exceedances into a TQI. However, these indices do not consider the severity of defects and may have limited informative value.

Finally, other miscellaneous indices employ different concepts or unique approaches to assess track quality, including the 3D TQI that incorporates threshold exceedances and standard deviations.

Further investigation is required to ensure a full assessment of track quality. This involves creating indices that consider the severity and frequency of problems, providing a more comprehensive depiction of the track state. Improvements in evaluating raw data and standard deviations can also provide more significant insights, allowing for a more comprehensive understanding of track conditions and more targeted and effective maintenance programs.

The primary objectives of the EDA phase were to understand the data distribution, identify patterns or trends, detect outliers, and explore relationships between variables. For this analysis, the team used data from a subsegment of track owned by Amtrak. The data is collected monthly by track inspection cars and is representative of autonomous track inspection. It contains 31 parameters, (e.g., vertical alignment, horizontal alignment, cross-level, etc.). The team used histograms, line graphs, and other techniques to examine profile and alignment, which play an important role in creating the 3D TQI for the track running surface.

The team considered t-student distribution, normal distribution, Laplace distribution, and exponential distribution to find the best-fit distribution for the profile and alignment. Overall, the normal distribution and Laplace distribution were found to be the two best-fit distributions; for any segment, the distribution of the data fluctuated between these two over time. These results provided the basis for the proposed model development.

Using the results of the EDA, the team designed several models to estimate future track condition, leveraging ML techniques to capture the complex relationships between track quality parameters and accurately predict track condition.

The team conducted an in-depth evaluation of the difference between deterministic and probabilistic models. The deterministic model is designed to predict the profile values one month in the future, considering the value of the profile one month in the past. The model used the values of scale and location for each month as input. Two models were designed, each assuming that the track segments take Laplace distribution for one model and Normal distribution for the other. The model results were presented and matched with the actual data PDF. The model predicted the location and scale for the distribution of the segments with some degree of accuracy, but it lacked the ability to capture the fluctuation in distribution. Moreover, the model lacked consideration for the uncertainty in prediction.

The team used probabilistic models to address these shortcomings. Both probabilistic models relied on LSTM methods to capture the behavior of the past data and predict multiple steps in time. The LSTM-MDN and LSTM-ND models used input data from 12 months in the past to output data 6 months in the future, and a custom loss function was employed (NLL).

The LSTM-MDN employed MDN as the last layer in the model to consider uncertainty in predictions. Visual comparisons between the model's predictions and the actual data highlighted the model's ability to capture distributional characteristics. The predicted distributions exhibited single or multiple modes (i.e., number of distributions), depending on the segment's complexity. The optimal number of components in the mixture density was determined to strike a balance between model complexity and accuracy.

Overall, the LSTM-MDN model offered a powerful solution for track geometry prediction by addressing the limitations of deterministic approaches. It leveraged historical data, captured uncertainties, and provided probabilistic multimodal forecasting, enhancing the accuracy and reliability of track condition predictions.

However, to improve the accuracy of the LSTM-MDN model, the team proposed the LSTM-ND model. The goal of the model was to predict the distribution of output as individual points in a time series.

The model's predictions were accompanied by confidence levels based on standard deviation values. For instance, one standard deviation corresponds to approximately 68.3 percent confidence, while two standard deviations represent around 95.5 percent confidence. The results were visualized by comparing predicted values with the actual track segment data, where the mean indicates the central tendency and the range within two standard deviations includes approximately 95 percent of the predicted values.

This visualization provided valuable insights into the model's performance, accuracy, and reliability. Decision-makers can evaluate data patterns, assess the reliability of predictions by considering uncertainty, and make well-informed judgments. By leveraging this probabilistic approach, the LSTM-ND model can enhance track condition prediction, aligning with real-world scenarios that involve inherent uncertainties and variations. It can empower stakeholders to allocate resources effectively, prioritize maintenance efforts, and optimize track safety based on a comprehensive understanding of potential outcomes and associated uncertainty.

Finally, the development of a 3D TQI involved training separate LSTM-ND models for each track geometry channel and combining their output distributions using a multivariate normal distribution. The index integrated probabilities of exceeding the value of the safety thresholds set by FRA, providing a comprehensive measure of track quality. The 3D TQI incorporated the mean vector and covariance matrix of the multivariable normal distribution, allowing insights into the average values and interdependencies between track geometry parameters.

The 3D TQI incorporates probabilities of exceeding predefined values of the safety limits set by FRA. The use of multiple models for each channel, with their corresponding output distributions, offers valuable insights into the variation and uncertainty associated with track parameters. The findings for eight segments are shown in this report, indicating how the proposed approach can detect anomalies in various channels. The index provides a thorough assessment of the track state, enabling informed decision-making in asset management and maintenance tasks.

The 3D TQI is a valuable tool for assessing and monitoring the condition of railway tracks. It combines ML techniques, probability analysis, and customization options to offer a comprehensive and customizable approach to track evaluation. The index enhances safety, optimizes maintenance practices, and contributes to the reliable operation of railway infrastructure.

The team identified a shortcoming in the approach used to calculate the probability of exceeding a threshold. Currently, the probability of exceedance is based on the probability of the predicted outcome exceeding a certain threshold, rather than the probability of each individual data point exceeding the threshold. This approach may not accurately capture the actual risk associated with specific data points exceeding the threshold. By considering the probability of the predicted outcome exceeding the threshold, the analysis provides an estimate of the overall likelihood of the track condition exceeding the specified threshold. However, it may overlook the variation and uncertainty associated with individual data points within the predicted outcome.

This research illustrated how the proposed 3D TQI can better assess track geometry. However, because the data used in this research belong to a single Amtrak track, the team recommends extending the application of the proposed index to different railway systems with various characteristics, from high-speed rail to freight railways, to assess how the approach is generalized and adaptable across diverse systems.

In addition, the effectiveness and long-term viability of track maintenance and management techniques performed in accordance with the recommendations of the index can be better understood by conducting long-term performance evaluations of the index. By monitoring the track condition over a prolonged period and assessing the effects of maintenance operations, the index can be improved, and track maintenance procedures can be made more effective.

Future research should include the development of approaches and algorithms to maximize the customization of the index based on operational requirements and objectives. This includes choosing the proper weightings for various track factors, looking into alternative techniques for combining probability, and evaluating how customization affects the index's performance.

6. References

[1] Fourie, C. J. & Zhuwaki, N. T. (2017). A Modelling Framework For Railway Infrastructure Reliability Analysis. *The South African Journal of Industrial Engineering, 28*(4).

[2] Yuan, F. Q. (2016). Overall reliability index development for railway infrastructure and rolling stock with case study. *2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Bali, Indonesia: pp. 1264-1268.

[3] Hidirov S. & Guler, H. (2019). Reliability, availability, and maintainability analyses for railway infrastructure management. *Structural and Infrastructure Engineering, 15*(9), pp. 1221-1233.

[4] Wang, X., Yang, J., Wang, J., Wang, Y., & Xu, G. (2020). An Incentive Factor-Based Dynamic Comprehensive Evaluation on a High-Speed Railway Track. *Applied Sciences, 10*(16), p. 5546.

[5] Soleimanmeigouni, I., Ahmadi, A., & Kumar, U. (2018). Track geometry degradation and maintenance modelling: A review. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 232*(1), pp. 73-102.

[6] Landgraf M. & Hansmann, F. (2019). Fractal analysis as an innovative approach for evaluating the condition of railway tracks. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 233*(6), pp. 596-605.

[7] Fellinger, M., Neuhold, J., & Marschnig, S. (2020). Ballast Condition Monitoring for Turnouts Using Power Spectral Density. *Journal of Transportation Engineering, Part A: Systems, 146*(9), p. 04020099.

[8] Satiennam, T. (2018). Track-quality index and degradation of railway track structure: The construction track doubling project of northeast line from thanon chira junction to khon kaen station, Thailand. *MATEC Web of Conferences, 192*, p. 02022.

[9] Accident/Incident Dashboards & Data Downloads | FRA. Accessed: Jun. 26, 2023.

[10] Ekberg, A., Kabo, E., & Lundén, R. (2023). Rail and wheel health management. *Wear, 526–527*, p. 204891.

[11] Offenbacher, S., Neuhold, J., Veit, P., & Landgraf, M. (2020). Analyzing Major Track Quality Indices and Introducing a Universally Applicable TQI. *Applied Sciences, 10*(23), p. 8490.

[12] Neuhold, J., Vidovic, I., & Marschnig, S. (2020). Preparing Track Geometry Data for Automated Maintenance Planning. *Journal of Transportation Engineering, Part A: Systems, 146*(5), p. 04020032.

[13] Soleimanmeigouni, I., Ahmadi, A., Khajehei, H., & Nissen, A. (2020). Investigation of the effect of the inspection intervals on the track geometry condition. *Structural and Infrastructure Engineering, 16*(8), pp. 1138-1146.

[14] Austrian Standards Institute (2014). Railway Applications - Track - Track Geometry Quality - Part 6: Characterization of Track Geometry Quality (Report No. EN 13848-6). *Austrian Standards Institute*: Vienna, Austria.

[15] Cook, J. (2023). Normal and Laplace distributions in differential privacy. Accessed: Jul. 14, 2023.

[16] Hayes, A. (2023). What Is T-Distribution in Probability? How Do You Use It? *Investopedia*. Accessed: Jul. 14, 2023.

[17] Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. (2017). Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *arXiv:1703.09039*. Accessed: Jun. 26, 2023.

[18] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature, 521*, pp. 436–44.

[19] Key Properties of the Normal distribution | CFA Level 1 - AnalystPrep. Accessed: Jun. 22, 2023.

7. Abbreviations and Acronyms

ACRONYM	DEFINITION
ALD	Automated Location Detector
ANN	Artificial Neural Network
BNN	Bayesian Neural Network
CrosslvlRate	Cross Level Rate
CTS	Curve/Spiral Flag
df	Degree of Freedom
DNN	Deterministic Neural Network
EDA	Exploratory Data Analysis
FRA	Federal Railroad Administration
GAN	Generative Adversarial Network
GMM	Gaussian Mixture Model
LAlign	Left Alignment for 31 chord length
LAlign62	Left Alignment for 62 chord length
LAlign124	Left Alignment for 124 chord length
LAlignSC	Left Alignment for the space curve data
LProf	Left profile for 31 chord length
LProf62	Left profile for 62 chord length
LProf124	Left profile for 124 chord length
LProfSC	Left profile for the space curve data
LSTM	Long Short Term Memory
LSTM-ND	Long Short Team Memory – Normal Distribution model
MCMC	Markov Cain Monte Carlo
MCO	Mid Chord Offset
MDN	Mixture Density Network
MGT	Million Gross Tons
MSE	Mean Squared Error
NLL	Negative Log-Likelihood Loss

ACRONYM	DEFINITION
PDF	Probability Density Function
RAlign	Right Alignment for 31 chord length
RAlign62	Right Alignment for 62 chord length
RAlign124	Right Alignment for 124 chord length
RAlignSC	Right Alignment for the space curve data
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
RProf	Right profile for 31 chord length
RProf62	Right profile for 62 chord length
RProf124	Right profile for 124 chord length
RProfSC	Right profile for the space curve data
SGD	Stochastic Gradient Descent
SSE	Sum of Squared Errors
SyncCnt	Synchronization Value
SyncFt	Synchronization Value
TQI	Track Quality Index
TSQI	Track Safety and Quality Index