1. Report No. RailTEAM UD-3	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Predicting Track Geometry Using Machine-Learning Methods		5. Report Date: October 6, 2023 6. Performing Organization Code:
7. Author(s) Mohammed Ahmed, Joseph W. Palese https://orcid.org/0000-0003-3946-3777, Allan Zarembski https://orcid.org/0000-0002-4282-9330		8. Performing Organization Report No. RailTEAM UD-3
9. Performing Organization Name and Address Department of Civil & Environmental Engineering University of Delaware 301 DuPont Hall Newark, DE 19716		10. Work Unit No. 11. Contract or Grant No. 69A3551747132
12. Sponsoring Agency Name and Address Office of Research, Development and Technology (RD&T) US Department of Transportation 1200 New Jersey Avenue, SE Washington, DC 20590		13. Type of Report and Period 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This study centers on predictive maintenance, an approach that foresees maintenance requirements based on anticipated defect occurrences. The research aims to create a model that accurately forecasts track geometry irregularities, empowering engineers to proactively address maintenance needs.

Traditionally, maintenance decisions relied on experience, manual inspections, and cyclical upkeep, leading to potential safety concerns and cost escalations. This research takes a novel approach by integrating mechanical and data-driven models, utilizing functional networks, Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks. RNNs capture sequences effectively, while LSTM networks excel in tracing long-term dependencies, making them apt for predicting track degradation patterns.

The study employs historical track geometry data collected over two years, undergoing Exploratory Data Analysis (EDA) to unveil insights and patterns. Data preprocessing ensues to ensure alignment and address missing values. The Track Quality Index (TQI) factors in past maintenance interventions, an essential step to enhance model accuracy. Subsequent steps involve generating and evaluating machine learning models using the processed data.

Incorporating functional networks and LSTM, the machine learning models adeptly forecast localized track irregularities, like profile values, by considering domain knowledge and data dynamics. The functional network model showcases superior predictive accuracy and interpretability, while the LSTM model excels in capturing underlying patterns for precise maintenance forecasts. This research thereby contributes to safer and more efficient railway operations through proactive maintenance planning.

17. Key Words	18. Distribution Statement
track geometry data, mechanical and data-driven models, utilizing functional networks, Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161. http://www.ntis.gov

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	75	

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

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

Final Report UD-3

PREDICTING TRACK GEOMETRY USING MACHINE-LEARNING METHODS

By

Mohammed Ahmed
Graduate Research Assistant
Department of Civil & Environmental Engineering
University of Delaware

Joseph W. Palese, PhD, MBA, PE
Research Assistant Professor
Department of Civil & Environmental Engineering
University of Delaware
palesezt@udel.edu

and

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

Date: October 6, 2023

Grant Number: 69A3551747132

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

ABSTRACT

The railway track stands as a pivotal component of the transportation system, ensuring the secure and smooth movement of people and goods. The Federal Railroad Administration reported a concerning trend between 2018 and 2021, with 37% of the 2097 train accidents stemming from track defects. To enhance safety and efficiency, vigilant monitoring and maintenance are imperative. This study centers on predictive maintenance, an approach that foresees maintenance requirements based on anticipated defect occurrences. The research aims to create a model that accurately forecasts track geometry irregularities, empowering engineers to proactively address maintenance needs.

Traditionally, maintenance decisions relied on experience, manual inspections, and cyclical upkeep, leading to potential safety concerns and cost escalations. This research takes a novel approach by integrating mechanical and data-driven models, utilizing functional networks, Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks. RNNs capture sequences effectively, while LSTM networks excel in tracing long-term dependencies, making them apt for predicting track degradation patterns.

The study employs historical track geometry data collected over two years, undergoing Exploratory Data Analysis (EDA) to unveil insights and patterns. Data preprocessing ensues to ensure alignment and address missing values. The Track Quality Index (TQI) factors in past maintenance interventions, an essential step to enhance model accuracy. Subsequent steps involve generating and evaluating machine learning models using the processed data.

Incorporating functional networks and LSTM, the machine learning models adeptly forecast localized track irregularities, like profile values, by considering domain knowledge and data dynamics. The functional network model showcases superior predictive accuracy and interpretability, while the LSTM model excels in capturing underlying patterns for precise maintenance forecasts. This research thereby contributes to safer and more efficient railway operations through proactive maintenance planning.

Chapter 1 INTRODUCTION

1.1 Importance of track geometry data

Railways play a crucial role in transportation, offering efficient and environmentally friendly means of moving people and freight across vast distances. They can transport millions of individuals and tons of cargo across countries and continents. In comparison to other modes of transportation, such as trucks, railways are more cost-effective.

The significance of track geometry data becomes apparent when considering track maintenance. The primary objectives of track maintenance are to ensure the rail system's safety and reliability and guarantee the availability and suitability of railway services. Achieving these goals requires a comprehensive understanding of all aspects of the railway and its condition periodically. This is where track geometry data becomes essential, providing crucial insights into the track's condition and operational fitness.

Track geometry serves as an indicator of the degradation that occurs within track structures. By monitoring track geometry, railway companies can effectively assess the condition of the track and identify areas that require maintenance. Not all track sections deteriorate at the same rate, making it necessary to tailor maintenance efforts accordingly. Applying uniform maintenance measures solely based on the weakest part of the track can lead to unnecessary maintenance activities and increased costs. Furthermore, neglecting the heterogeneous nature of degradation can result in hazardous conditions that compromise train safety. Therefore, monitoring the entire track using track geometry data is crucial.

By utilizing track geometry data, railway operators can proactively identify areas of concern, prioritize maintenance efforts, and ensure the safe and reliable operation of the rail system. This data-driven approach enables informed decision-making, reduces maintenance costs, and minimizes the risk of accidents or disruptions. Moreover, track geometry data provides valuable insights into the overall health of the track, helping railway companies optimize their maintenance strategies and effectively allocate resources.

In summary, track geometry data plays a vital role in track maintenance within the railway industry. By utilizing this data, railway operators can monitor the condition of the track, identify areas requiring maintenance, and ensure the safety and reliability of their rail systems. The heterogeneous nature of degradation necessitates tailored maintenance approaches, which can be efficiently achieved by analyzing track geometry data. Ultimately, the effective utilization of track geometry data leads to enhanced operational efficiency, cost savings, and improved safety within the railway network.

1.2 Safety and Maintenance

The track is an essential and critical component of the railway system, playing a vital role in ensuring the safe and efficient transportation of people and goods. How- ever, according to data from the Federal Railroad Administration (FRA 2023), between 2018 and 2021, there were 2,097 train accidents (excluding those caused by human factors), with 781 of them attributed to track defects (37% of the total). These statistics highlight the significant challenge of track defects and emphasize the importance of adequate maintenance and continuous monitoring to enhance safety and prevent potential accidents.

Given the potential risks connected with even minor track problems, significant financial resources are committed to maintenance and safety procedures yearly. For example, the United Kingdom spent more than £3 billion in 2015 only on operating and maintaining the rail network, ignoring expenses for renewals or enhancements (Sheeran, et al. 2015). Surprisingly, this amount exceeds highway operating and maintenance expenditures,

demonstrating the importance and size of investment required for railway track repair. Similarly, track maintenance accounts for more than half of overall maintenance spending in the United States, highlighting the financial commitment required to ensure the safety and reliability of railway infrastructure (López-Pita et al. 2008).

In Europe, maintenance costs for high-speed lines vary depending on the type of trains operating on them. For high-speed lines dedicated exclusively to passenger trains, the maintenance cost per kilometer of track per year is estimated to be around $56,356 (López-Pita 2008). However, when high-speed lines serve passenger and freight trains, the maintenance costs can escalate to approximately $73,000 per kilometer of track per year (López-Pita et al. 2008). These figures highlight the substantial financial investments required to uphold high-speed rail networks' integrity and operational efficiency.

The considerable expenditures committed to track maintenance reflect the tracks' vital role in assuring railway systems' safety, dependability, and lifespan. If track problems are not addressed, they can lead to catastrophic accidents that jeopardize the lives of passengers, workers, and the general public. Railway authorities and operators hope to limit the hazards associated with track degradation and ensure the continuing and safe operation of the railway network by investing in maintenance and safety measures.

To efficiently manage rail maintenance, a comprehensive approach is required. This strategy includes routine inspections, on-time repairs, and constant monitoring of track conditions. Several approaches and technologies, such as track geometry measurement systems and advanced sensors, are used to assess the health and integrity of the track. Maintenance teams can use these technologies to discover problems, abnormalities, or potential areas of concern and take appropriate action to address them as soon as possible.

Moreover, preventive and predictive maintenance strategies are increasingly being adopted to optimize maintenance efforts and reduce costs. Preventive maintenance involves scheduled maintenance activities performed before track failures occur, aiming to proactively address potential issues and extend the lifespan of track components. On the other hand, predictive maintenance relies on data analysis and predictive models to forecast the occurrence of defects and plan maintenance activities accordingly. By leveraging historical data and advanced analytics techniques, predictive maintenance enables proactive decision-making, minimizing disruptions and optimizing maintenance schedules.

To minimize costs and prevent accidents, efficient maintenance plans must be implemented. Various types of maintenance strategies exist, each with its own unique properties. These include corrective, preventive, condition-based, and predictive maintenance (Budai-Balke 2009, Jardine et al. 2006, Jezzini et al. 2013), as shown in Figure 1.1. Figure 1.1 shows the P-F curve for the track health over time to identify the interval between potential failure and functional failure.

- Corrective maintenance is performed when the track requires immediate repair due to the occurrence of a defect. This type of maintenance is costly, interrupts railway services, and necessitates alternative transport arrangements.
- Preventive maintenance involves scheduled maintenance before track failures occur. However, one drawback of this approach is the reduced lifespan of track components due to early replacement, resulting in additional costs.
- Condition-based maintenance utilizes advanced technologies for real-time monitoring of track conditions. Repairs or replacements are conducted only when conditions exceed predetermined thresholds.
- Predictive maintenance involves designing maintenance schedules by predicting and estimating when defects will likely occur. This approach allows for proactive scheduling based on predicted track conditions, saving time and minimizing the need for railway shutdowns. Predictive maintenance is highly desirable due to its convenience and efficiency. However, its effectiveness heavily relies on data collection, processing, and analysis.

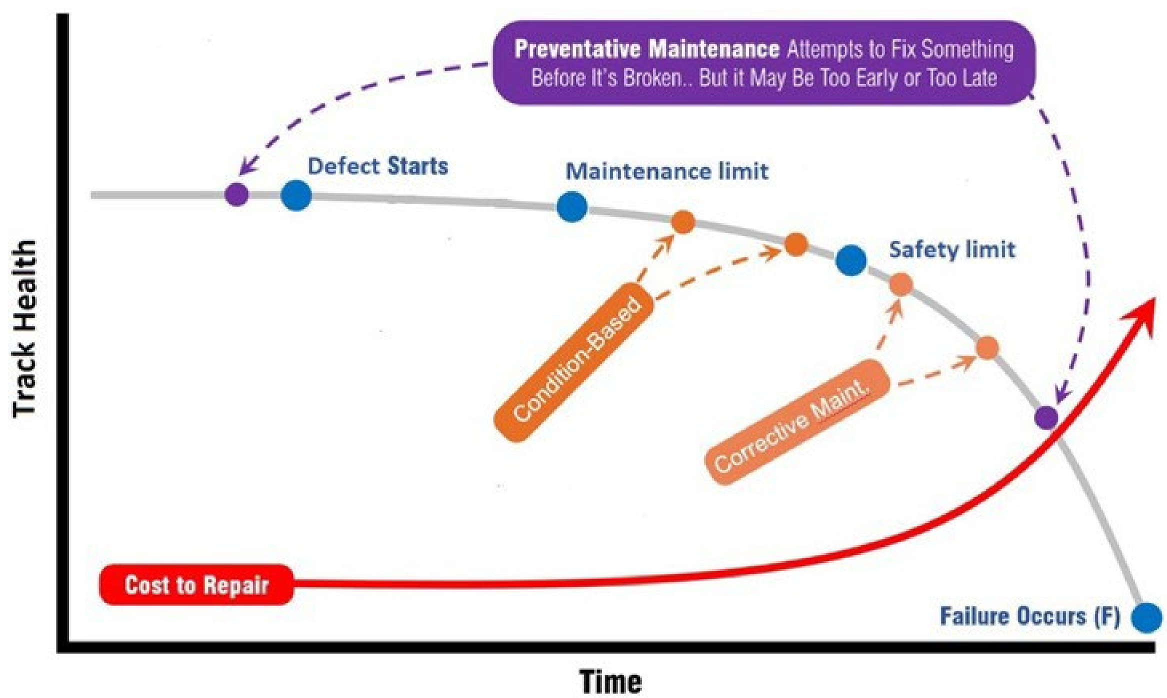


Figure 1.1: Different Types of Maintenance

1.3 Railway's Data

In general, Data on railways are gathered from various sources, including walking patrols (Marino et al. 2007), mechanized track patrols (Mohammadi et al. 2019), wayside detectors (Li and He 2015), and autonomous. These sources can detect defects, such as broken rails and track geometry issues, resulting in a wealth of data in the railway industry. This data exhibits unique characteristics, including large volume, multiple sources, high imbalance towards normal behavior, and high noise levels.

The data collected for railway maintenance exhibits the following characteristics:

1. Large volume: The data collected for railway maintenance involves massive amounts of information. This includes data from both the time domain, such as real- time or near real-time updates, and the space domain, which covers thousands of miles of railway tracks.
2. Multi-source: The data is generated from various measurement methods and sources. These sources may utilize different techniques and technologies to gather data, resulting in multiple types of data being collected for analysis.
3. Highly imbalanced: The distribution of rail defects within the collected data sets is highly skewed. Most observations correspond to normal states, while only a small portion of the data relates to actual defects or anomalies in the track condition.
4. High noise: Noise in the data arises from two primary sources during the data collection. Firstly, it stems from the inherent environmental uncertainties along the railway tracks, such as variations in soil type, climate conditions, track profiles, and materials used. Secondly, the precision and accuracy of the sensors employed to collect the data contribute to the noise levels present in the dataset.

These characteristics pose challenges in analyzing and processing railway maintenance data. Approaches and techniques need to be employed to handle the large volume of data from multiple sources, address the imbalanced nature of the data, and mitigate the impact of noise for accurate analysis and effective decision-making in maintenance operations.

Track condition analysis and evaluation are crucial aspects of railway maintenance. Two major methods employed for this purpose are mechanical models and data-driven models. Mechanical models utilize simplified approaches based on the understanding of tracking behavior and mechanistic knowledge. These models rely on theoretical principles and equations to assess the condition of the track. These models are often based on well-established engineering principles and offer interpretability in understanding the underlying mechanisms affecting track conditions. On the other hand, data-driven models aim to detect patterns and trends by analyzing observed data. These models leverage the vast amount of data collected from various sources, such as sensors, inspection records, and maintenance logs. Data-driven models can be categorized into statistical models and machine learning models (Xie et al. 2020).

Statistical models focus on establishing relationships between variables through statistical analysis. These models identify the data's correlations, dependencies, and trends to infer the track condition. By examining historical data, statistical models can provide insights into the probability of track defects and potential maintenance needs. These models are useful for understanding the statistical significance of different factors affecting track conditions. Machine learning models, on the other hand, aim to achieve the most accurate predictions by learning from the data. These models employ advanced algorithms to automatically extract hidden features and patterns from railway data's large volume and complexity. Machine learning models can make predictions and highly accurately classify track conditions by training on labeled data. These models are particularly effective in handling non-linear relationships and complex datasets.

Both data-driven models and mechanical models have their strengths and limitations. Mechanical models offer a fundamental understanding of track behavior and are well-suited for interpreting the physical processes involved. However, they may rely on simplifications that might not capture all real-world complexities. On the other hand, data-driven models can leverage the richness of data and handle large-scale datasets. They can uncover intricate relationships and provide accurate predictions.

However, the performance of data-driven models heavily relies on appropriate data pre-processing techniques and the selection of suitable analysis models.

In practice, a combination of both mechanical models and data-driven models can be beneficial. By integrating the strengths of each approach, a comprehensive understanding of track conditions can be obtained. This can lead to effective maintenance planning, timely defect detection, and improved safety and reliability of the railway system.

1.4 Models' Background

There are two methods used for scientific discovery nowadays: Theory-based Models (Physics-based Models) and Data Science Models (Machine Learning or ML), and each model has its strengths and limitations. It highlights that these models have distinct strengths and limitations. While data science models have been successful in commercial fields, they face challenges in solving scientific problems characterized by complexity, insufficient data, high-dimensional parameter spaces, and model uncertainty (Karpatne et al.2017).

One of the key issues is the complex, nonlinear nature of physical processes that may be challenging to deduce solely from data. Furthermore, ML models may not fully capture the complexity of scientific hypotheses due to gaps in knowledge, potentially leading to suboptimal performance.

The data science models fail to improve scientific discovery, although they have huge potential and have witnessed success in non-scientific areas such as advertising. Karpatne suggests that ML modes fail in scientific problems due for several reasons. One solution to these problems by combining Physics-based and ML models (Karpatne et al.2017, Wadud et al. 2015). The new approach is called Theory-Guided Data science models (TGDS). One of these models' advantages is using scientific knowledge to solve the shortcomings of ML. Also, by using TGDS models, we can ensure better generalizability by leveraging these models to provide a way to integrate physics-based modeling with Machine Learning and improve the effectiveness of data science models.

1.4.1 Data-Driven Models in Railway Track

Railway track monitoring involves collecting diverse types of data to evaluate the track's condition and its components. These data vary in nature and are acquired through different measurement methods. Some components necessitate physical inspections, like walking patrols to assess ballast sections, ties, and fasteners. In contrast, mechanized track patrols, typically conducted by inspection cars, are used for examining track geometry data and rail heads.

The choice of algorithms for data analysis is profoundly influenced by the type of input data and the measurement techniques employed. Time series data, which is frequently gathered using track geometry recording cars, ultrasonic inspections, in-service vehicles, or fiber optic sensors, demands algorithms capable of capturing temporal dependencies and trends. Long Short-Term Memory (LSTM) and Autoregressive Integrated Moving Average (ARIMA) are commonly applied in this context. On the other hand, image and video data collected through cameras or ground-penetrating radar provides vital visual insights into the track's condition. Convolutional Neural Networks (CNNs) and object detection algorithms are extensively utilized in railway track inspection and monitoring to analyze images and videos effectively. For instance, research by Zhang has shown the effectiveness of CNNs, especially those with pre-convolution and residual structures, in identifying rail damage by analyzing rail vibration signals. Similarly, Zauner utilized a fully convolutional network to automatically detect critical rail areas by performing semantic segmentation on 3D scanner data (Xie et al. 2020, Zhang et al. 2022, Zauner et al. 2019).

Additionally, certain data types, such as temperature, weather conditions, rail conditions, and tonnage, fall into the category of discrete value data. To extract meaningful insights from these data types, machine learning (ML) algorithms come into play. Decision trees, random forests, and support vector machines (SVMs) are commonly employed to address classification or regression problems related to discrete value data. Moreover, ML techniques like Natural Language Processing (NLP) can effectively process textual data, encompassing records of accidents and maintenance. Text classification algorithms and sentiment analysis models can extract valuable information from these textual sources.

In summary, the specific data types collected in railway track monitoring significantly influence the selection of ML algorithms. It is crucial to consider the characteristics of the data and choose the most suitable algorithms based on the desired results and the nature of the data.

1.4.2 Time Series Type Models

Time series data essentially consists of sequential observations collected at regular intervals. In railway track monitoring, this data is instrumental because it helps in understanding how various track parameters evolve over time. The statistical techniques play a pivotal role in analyzing time series data, both in the time and frequency domains.

In the time domain, statistical measures such as the mean, standard deviation, skewness, and kurtosis are discussed. These measures provide insights into the central tendency and variability of the data. They are valuable for identifying patterns, trends, and anomalies in track geometry measurements. For instance, changes in the mean or standard deviation over time might indicate shifts or variations in the track's condition. Moreover, the autocorrelation function is mentioned as a tool for identifying dependencies between past and future values in time series data (Li et al. 2013, Tsui et al. 2015, Lederman et al. 2017, Jiang et al. 2019).

Moving beyond the time domain, it is worth mentioning the importance of frequency domain analysis in understanding the periodicity and frequency characteristics of time series data. Fourier transform is introduced as a key method for decomposing time series data into its constituent frequencies. This enables the detection of periodic patterns, trends, or recurring events in the data. The section mentions that Fourier transform has been used to detect unusual frequency components in the data, which can be indicative of rapid changes or anomalies in the track's underlying mechanism (El-Sibaie and Zhang 2004, Sadeghi and Askarinejad 2010).

However, it is noted that processing non-stationary time series data using Fourier transform has limitations. To address this, the Wavelet transform is introduced as a powerful alternative. Wavelet transform is capable of handling non-stationary time series data by breaking it down into different frequency bands with varying scales. This approach enables the localization of changes or occurrences in the data more effectively. In summary, the section underscores that signal processing techniques, encompassing both time and frequency domain analysis, are essential for extracting meaningful information from time series data in railway track engineering (Nadarajah et al. 2018, Zhang 2003, Heidarysafa et al. 2018).

Additionally, upon the application of machine learning models in analyzing time series data. It mentions that machine learning models leverage the statistical and frequency domain attributes of time series data as inputs for various tasks, such as classification, regression, anomaly detection, and forecasting. For instance, recurrent neural networks (RNN) and Long Short-Term Memory (LSTM) networks are recommended for multi-step time series forecasting due to their robustness and capacity to handle missing data.

1.4.3 Discrete Value-Based Models

In railway measurements, discrete value data are often not gathered using fixed frequency sensors. Instead, these data are usually collected after accidents or catastrophic events. They offer detailed information about the railway infrastructure, such as the age of the track or its type of break. These discrete values serve as vital data points when investigating the causes and consequences of accidents in the railway system. When selecting a model for analyzing discrete value data, the choice often depends on the size of the dataset. The dataset size is a common deciding factor when choosing a model to analyze discrete value data. Bayesian approaches are recommended because of their capacity to handle vast amounts of data (Andrade and Teixeira 2012). However, when dealing with continuous features and numerous dimensions, Support Vector Machines (SVMs) and neural networks perform effectively (Hu and Liu 2016). On the other hand, the K-nearest neighbors (KNN) method should be avoided when the dataset contains noisy features. In contrast, if interpretability is essential, tree-based and Bayesian methods are better options than neural networks or SVMs (Andrade and Teixeira 2012,. Hu and Liu 2016).

1.5 Objective

The scope of this study encompasses the application of machine learning methods to predict track geometric data and assist in planning maintenance activities for railway systems. The findings and methodologies developed in this research can be applied to various types of railway networks, including urban, suburban, and long- distance rail systems. In terms of track geometric data, the study aims to address a wide range of geometric characteristics that are crucial for maintenance planning.

This may include but is not limited to parameters such as Profile, track alignment, cross-level, gauge, twist, curvature, and other relevant geometric features. The study aims to explore the prediction of these geometric parameters using machine learning techniques.

The ability to predict the value of the profile or any value of the track geometry and how the track will behave in the next inspection is important for designing an efficient maintenance schedule. In this research, we aimed to integrate mechanical models and data-driven models by utilizing a functional network and long short-term memory networks, as opposed to the conventional neural network, to predict track degradation rates based on geometry data. The primary objective was to derive a precise equation that can effectively forecast future profile values using geometry data. This predictive capability will enable us to determine when the profile value is expected to exceed the threshold, indicating the necessity for maintenance.

Chapter 2 EXPLORATORY DATA ANALYSIS

2.1 Data

2.1.1 Data Source

The track geometry data used in this research was collected from 24 inspection runs conducted on a 10-mile track section from Ridley Park, PA to Wilmington, DE. The data were obtained between milepost 11 to milepost 20, with measurements taken at every foot, resulting in over 52,000 data points for each inspection. Figure 2.1 illustrates the location of the track section.

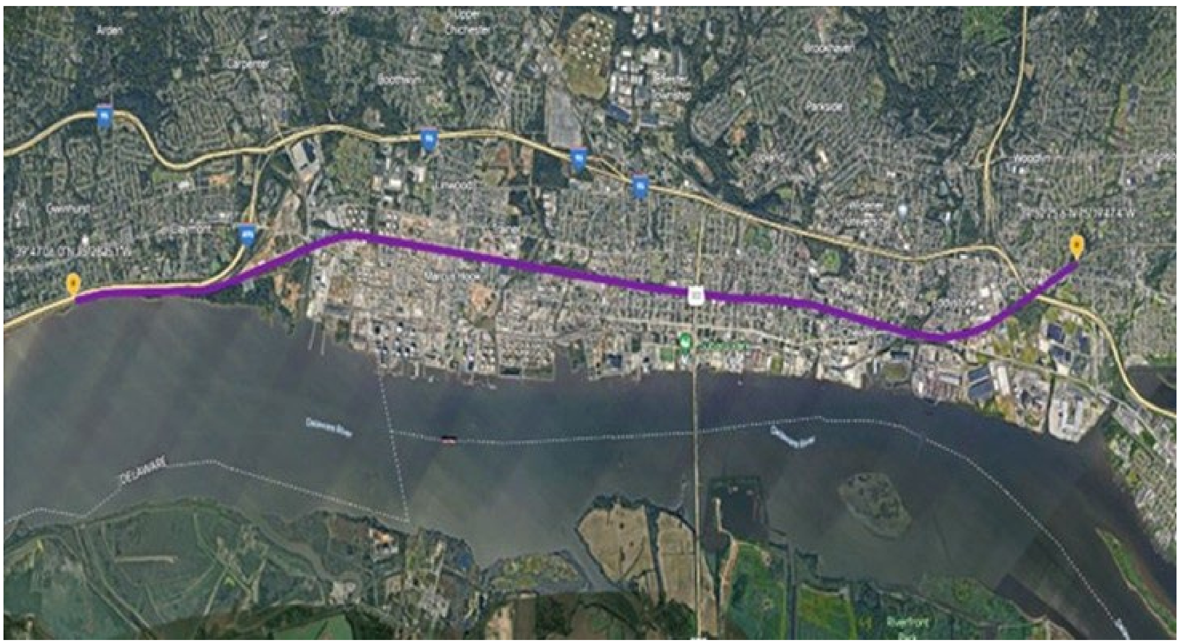


Figure 2.1: The location of the Track

2.1.2 Track Geometry Data

The collected track geometry data included both conventional and processed channels, encompassing 31 different geometry types. Conventional channels typically involve direct measurements or sensors placed on the track structure, such as vertical and horizontal alignment. Processed channels, on the other hand, entail applying specialized algorithms or filters to raw data collected, such as twist, and warp.

These types, outlined in Table 2.1, provided specific measurements related to the track's geometric characteristics, such as profile, alignment, gage, twist, and warp. Additionally, the dataset included left and right profile space curves and left and right alignment space curves.

Table 2.1: Channels for 31 different parameters in the data

1	Mile	11	LeftProfile	21	CTS
2	Feet	12	LeftProf62	22	Speed

3	Track	13	LeftProf124	23	ALD
4	Line	14	RAlignment	24	Class
5	Gage	15	RAlign62	25	Warp62
6	CrossLevel	16	RAlign124	26	LProfSC
7	CrossLvlRate	17	LAlignment	27	RProfSC
8	RightProfile	18	LAlign62	28	LAlignSC
9	RightProf62	19	LAlign124	29	RAlignSC
10	RightProf124	20	Curvature	30	SyncCnt
				31	SyncFt

2.1.3 Consolidation of Data into the Common Database

The main objective of this study is to predict profile data in the time domain, specifically focusing on accurately forecasting how profile values will change over time. By using time series analysis, the study aims to understand and predict the temporal evolution of the profile data, allowing for anticipation of future trends and changes in the variable of interest. To achieve this, a time series dataset was created by consolidating left profile data from 24 geometry files into a single dataset. This consolidation facilitates a comprehensive analysis of the profile data and enables the application of time series techniques for analysis and prediction.

The main objective of this study is to predict profile data in the time domain, specifically focusing on accurately forecasting how profile values will change over time. By using time series analysis, the study aims to understand and predict the temporal evolution of the profile data, allowing for anticipation of future trends and changes in the variable of interest. To achieve this, a time series dataset was created by consolidating left profile data from 24 geometry files into a single dataset. This consolidation facilitates a comprehensive analysis of the profile data and enables the application of time series techniques for analysis and prediction.

2.1.4 Descriptive Statistics

A database for the 24-month profile must be built for practical exploratory data analysis (EDA) and subsequent statistical studies. The database offers a systematic framework for managing and organizing the data, making it easier to retrieve and analyze it quickly. Once the database is created, several statistical techniques can investigate and quantify different data features, such as mean, median, standard deviation quantiles, and interquartile ranges. These features can provide an understanding of the spread or dispersion of the data. Quantiles divide the data into equal parts, allowing us to identify values that represent specific proportions of the dataset. The interquartile range represents the range between the first and third quartiles, indicating the spread of the middle 50% of the data.

Table 2.2: The descriptive statistics for the 24 months

	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	Month 12
count	52185	52185	52185	52185	52185	52185	52185	52185	52185	52185	52185	52185
mean	0.000141	0.000156	0.000143	0.000137	0.00015	0.000125	0.000114	9.55E-05	0.000112	9.53E-05	0.00011	8.43E-05
std	0.169919	0.156947	0.167615	0.171581	0.16552	0.16875	0.173168	0.170901	0.175528	0.174462	0.176287	0.194886
min	-1.26892	-1.09222	-1.34064	-1.55457	-1.25885	-1.19568	-1.2854	-1.23657	-1.3739	-1.40259	-1.45294	-1.44897
25%	-0.06287	-0.05768	-0.06012	-0.06134	-0.06104	-0.05798	-0.05798	-0.05798	-0.05798	-0.05829	-0.05768	-0.07721
50%	0.005493	0.003662	0.004272	0.004272	0.004578	0.004272	0.005188	0.004578	0.004578	0.004883	0.004883	0.005188
75%	0.072632	0.065308	0.06958	0.070801	0.069275	0.068665	0.068054	0.068665	0.06897	0.06897	0.06897	0.085144
max	0.985107	0.979919	0.952454	0.947571	1.027832	1.096802	1.111145	1.069336	1.121826	1.083374	1.107483	1.489868
count	52185	52185	52185	52185	52185	52185	52185	52185	52185	52185	52185	52185
mean	0.000119	6.3E-05	6.48E-05	7.87E-05	7.04E-05	9.25E-05	5.59E-05	9.31E-05	4.91E-05	8.37E-05	-2.7E-05	1.82E-05
std	0.188657	0.183645	0.179654	0.17205	0.168077	0.167975	0.166854	0.15822	0.157204	0.154293	0.140937	0.136616
min	-1.36932	-1.33362	-1.43433	-1.49689	-1.14838	-1.25916	-1.20728	-1.23688	-1.17615	-1.24664	-1.1261	-1.09131
25%	-0.07202	-0.07233	-0.06561	-0.06195	-0.06042	-0.06134	-0.06256	-0.05493	-0.05402	-0.05493	-0.047	-0.04578
50%	0.003662	0.005188	0.004883	0.004272	0.004578	0.004883	0.005188	0.002747	0.002441	0.003052	0.002136	0.002136
75%	0.081177	0.079346	0.074463	0.071411	0.071411	0.071106	0.071716	0.061951	0.06134	0.061951	0.052185	0.052185
max	1.140137	1.172485	1.221619	1.103516	1.228943	1.113281	1.250916	1.148071	1.610107	1.212769	1.073608	1.03363

On the other hand, Measures such as the mean and median are commonly used to analyze the central trends within the data. The mean represents the average value of the data, while the median represents the middle value when the data is sorted in ascending or descending order. These measurements shed light on the data distribution's geographic center and aid in comprehending the overall pattern. By leveraging these basic statistical measurements, we can deeply understand the data's distribution and characteristics, which will help find trends, identify outliers, and make informed decisions during the data analysis.

Table 2.2. show the descriptive statistics for the complete dataset, such as count, mean, standard deviation, quartiles, and max, encompassing all 24 inspections conducted in this analysis.

Overall, the descriptive statistical analysis indicates that, over the duration of 24 months, the mean value demonstrates that the track values averagely evolve around zero. However, the existence of minimum and maximum values exceeding the FRA threshold for our track (± 1) means that at least one of the points exceeded the safety limits. Finally, the standard deviation illustrates the condition of the track from month to month. For instance, from the first to the second month, the condition of the whole track got relatively better.

2.2 Exploratory Data Analysis

The goal of exploratory data analysis (EDA) is to gather insights and understanding from raw data prior to formal statistical modeling or hypothesis testing. EDA assists in exploring the structure, patterns, and relationships within data, identifying probable outliers or anomalies, and formulating hypotheses for further examination. EDA attempts to reveal hidden patterns, trends, and underlying distributions within data using data visualization, summary statistics, and data exploration approaches, giving a solid platform for subsequent analysis and decision-making.

2.2.1 Bivariate Visualization

Bivariate visualization is a fundamental technique for exploring the relationship between two variables, providing insights into their interactions and potential impacts, and giving a better understanding of the patterns and relationships between them. One of these techniques is scatter plots, which display data points. Each of these points represents the combination of values from two variables. In Figures [2.2-

2.5] the scatter plot shows the correlation between the profile variable and the preceding or subsequent months. The figure suggests that there are predominantly positive linear relationships, indicating a tendency for both variables to increase together. As the profile increases, there is a tendency for the previous or following months' values also to increase. The tightness and direction of the points can be used to determine how strong the linear relationship is. The variables are positively correlated when the points form a tight circle around a straight line that slopes upward from left to right.

2.2.2 Histograms Combination of a Histogram and Kernel

Histograms are a powerful graphical tool for finding information on the data's distribution and analyzing and visualizing the distribution of various variables related to railway profiles and track geometry. It divides the data into bins or intervals along the x-axis, and the height of the corresponding bar on the y-axis shows the frequency or proportion of data points falling into each bin, displaying the frequency or probability distribution. The shape of the histograms can identify patterns, central tendencies, and the spread of the data.

Histograms for the left profile in the railway dataset are shown in Figures [2.2- 2.5]. The focus on the profile is due to the fact that the profile values need to be and evolve around zero, in this case, when predicting the profile values. These histograms show that the profile has a roughly Gaussian or normal distribution or a roughly even distribution. This normality assumption is important for further investigation, particularly when using machine learning techniques. However, confirming this supposition with further analysis, such as QQ plots, is crucial. Moreover, A histogram and Kernel Density Estimation (KDE) can be combined to produce a smoothed distribution representation. KDE is a well-known technique for calculating a random variable's probability density function, and this allows for refined visualization of the distribution. by overlaying the density curve calculated by KDE onto the histogram. In railways, these techniques can be employed to examine the conditions of the track by observing the spread or variability of distributions; a wider spread or more variability may indicate a higher likelihood of encountering defects or irregularities.

Histogram and Kernel Density Estimation (KDE) are not limited to profile or even geometry data, and they can be employed to analyze various variables in the railway domain, such as the ballast fouling index or other measurements related to railway infrastructure and performance.

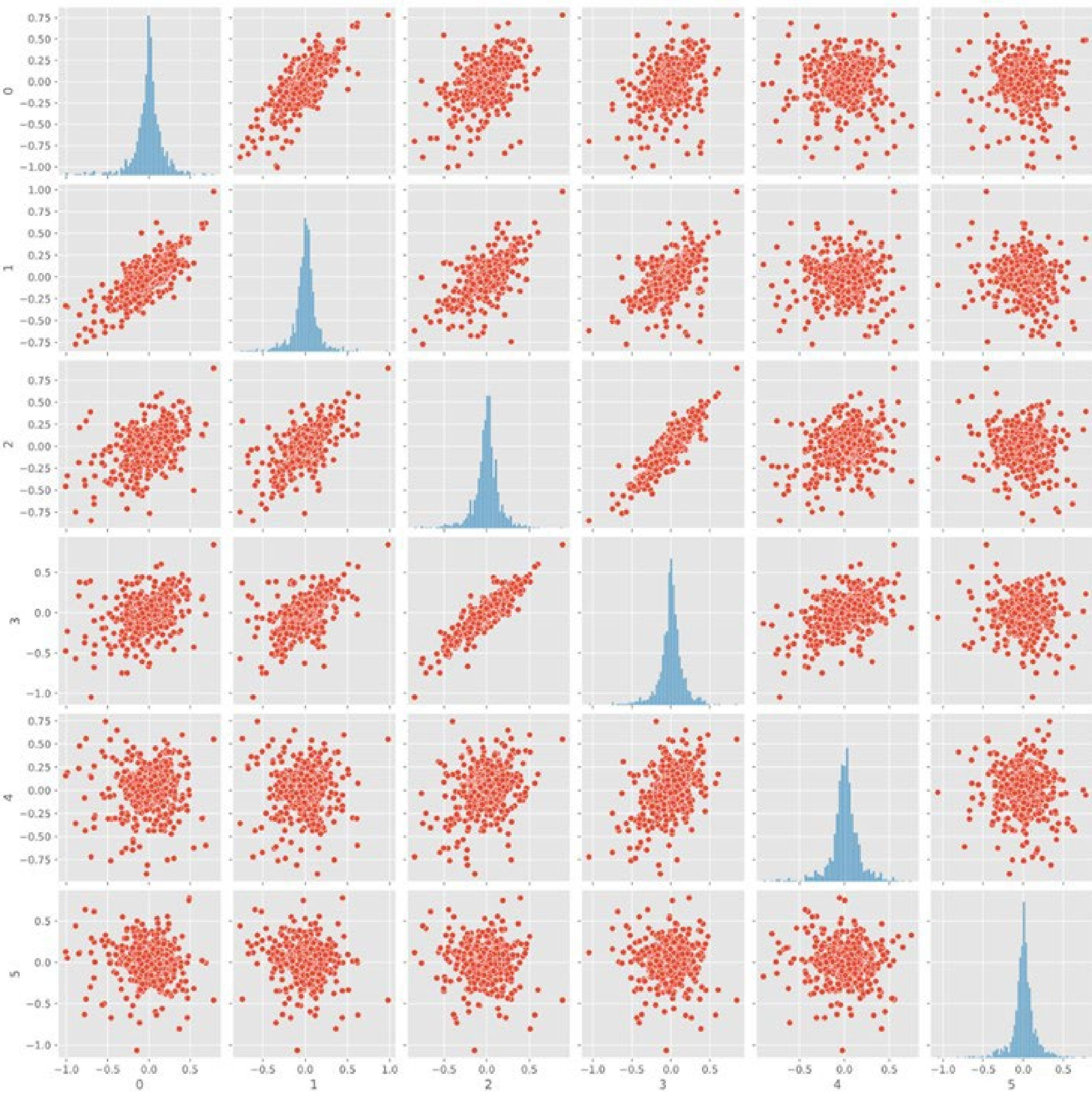


Figure 2.2: Scatter and histogram plot for the first to 6th month

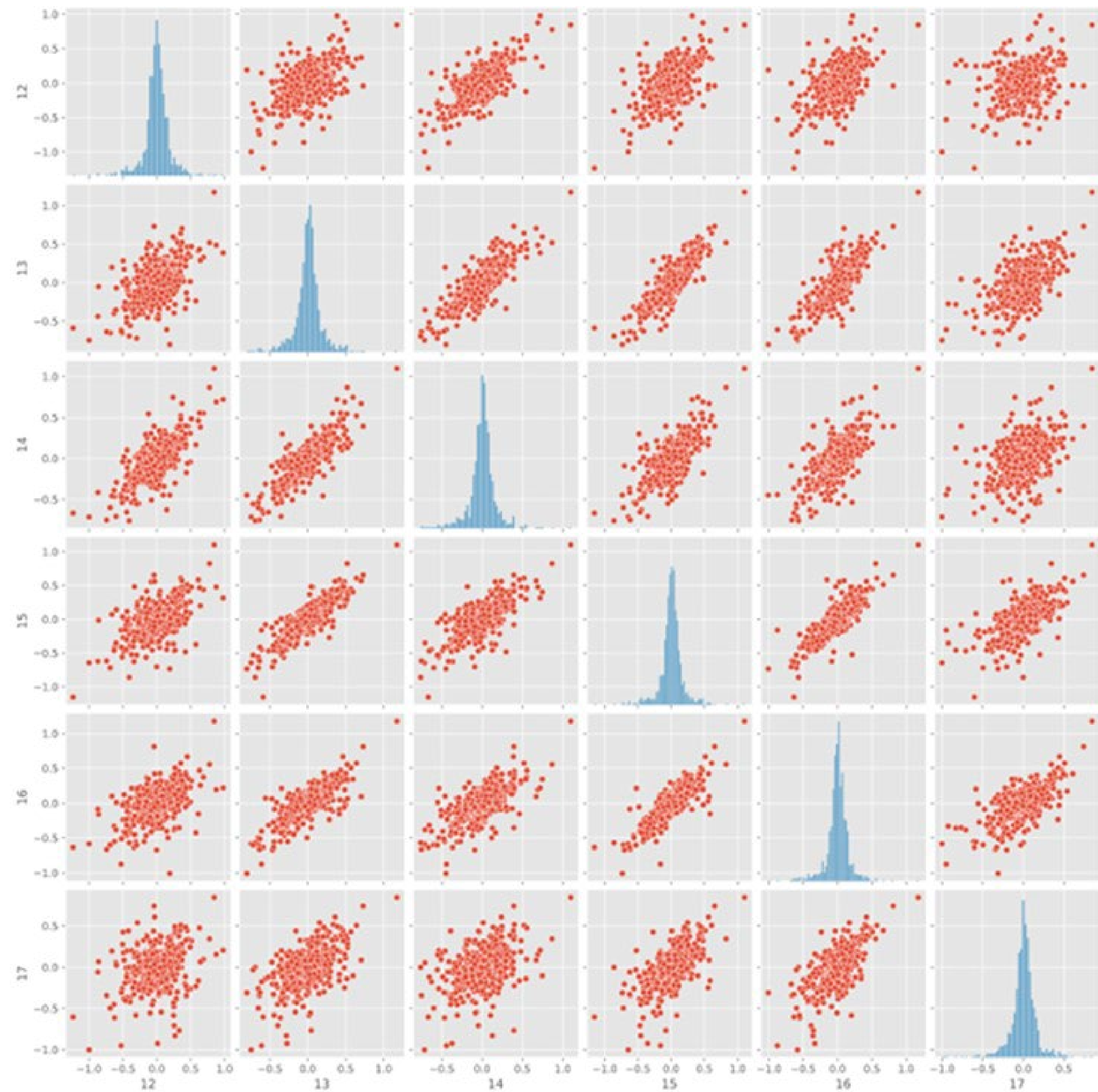


Figure 2.3: Scatter and histogram plot for the 7th to 12th month

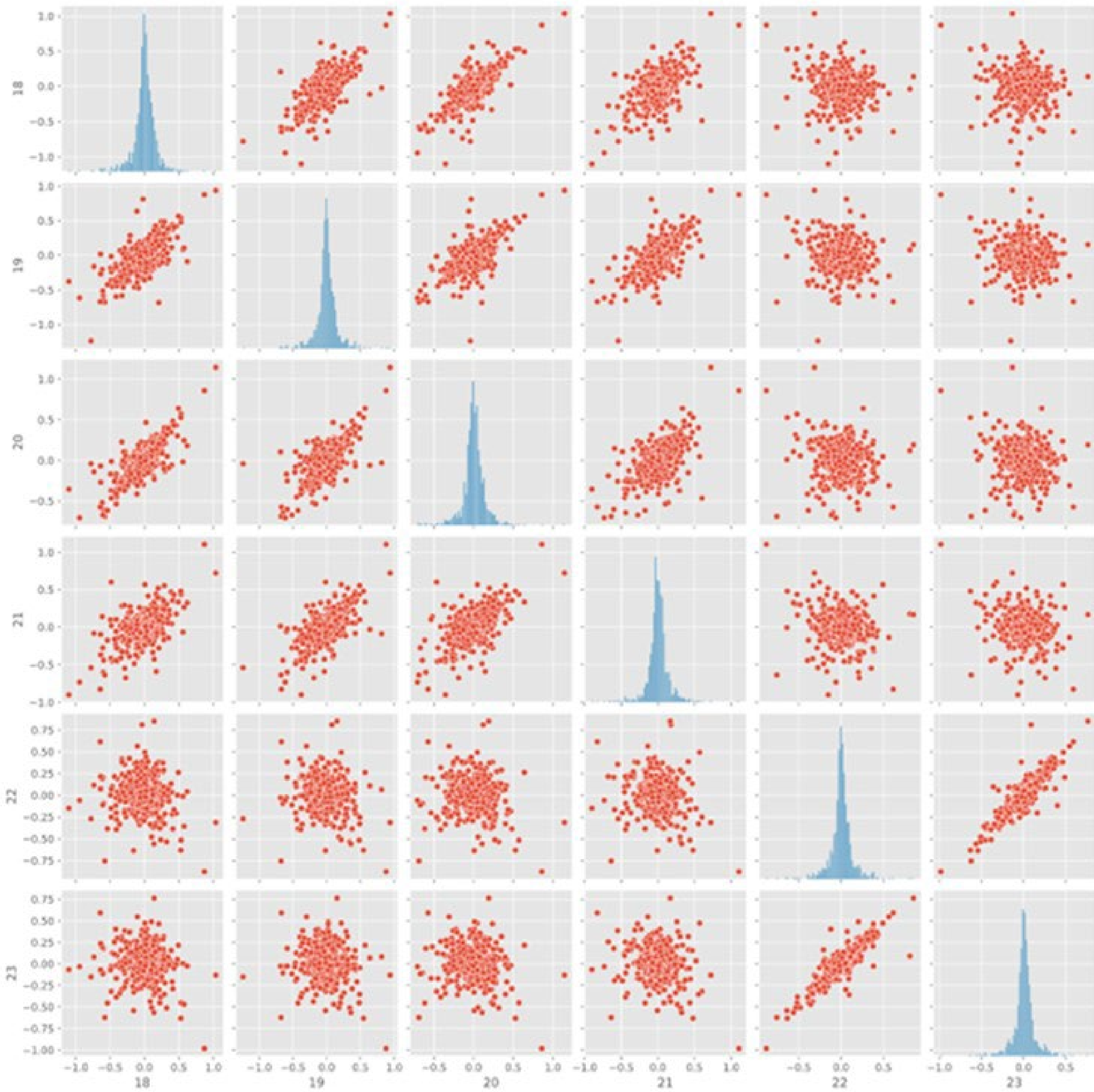


Figure 2.4: Scatter and histogram plot for the 13th to 18th month

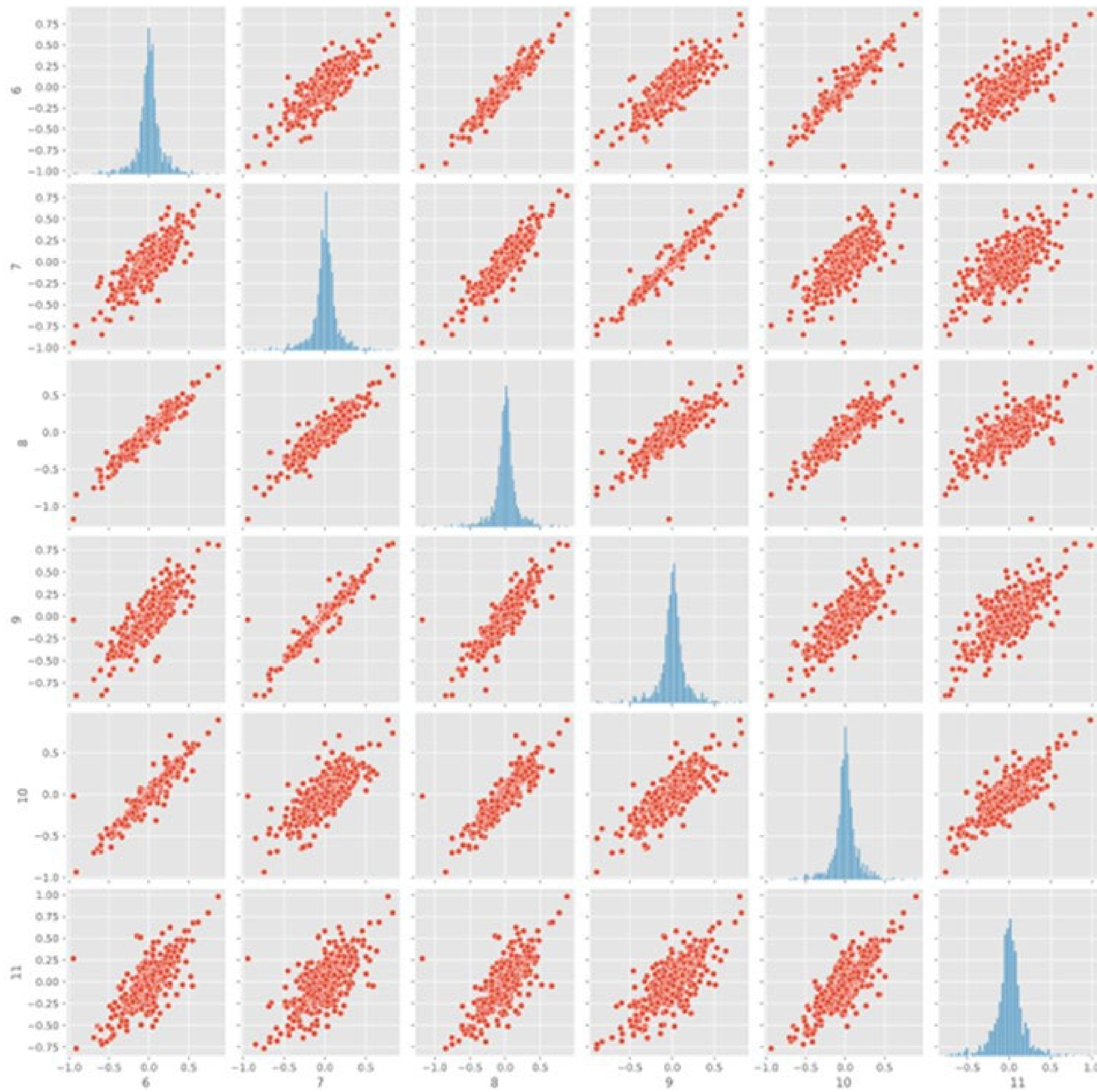


Figure 2.5: Scatter and histogram plot for the 19th to 24th month

2.2.3 Correlation Plot Visualization

A correlation plot is a graphical representation of the correlations between variables in a dataset. This can help us understand the connections between railway variables and identify significant associations using Correlation coefficients, which quantify the strength and direction of relationships between these variables. These coefficients range between -1 and +1. A score of +1 shows a significant positive correlation, which means that when one variable increases, the other tends to increase as well, and vice versa. A score of -1, on the other hand, suggests a high negative correlation, indicating that when one variable increases, the other tends to decrease. A correlation coefficient of 0 suggests no correlation between the variables.

The correlation coefficient is calculated using the variables' covariance and standard deviations of each variable. The equation for the correlation coefficient, denoted as θ, is given as:

$$\theta = \frac{COV\,(x, y)}{\sigma_x \sigma_y}$$

where:

θ = correlation coefficient
$COV\,(x, y)$ = covariance of variables x and y.
σ_x = standard deviation of x.
σ_y = standard deviation of y.

The correlations between variables can be visualized in a correlation plot using scatter plots or heatmaps. Each data point in the scatter plot represents a combination of values from two variables. By looking at the general arrangement of the data points, it is possible to determine the direction and magnitude of the association. A scatter plot implies a positive correlation with an upward trend in the data points, whereas a downward trend implies a negative correlation. In railways, examining these correlations may identify closely associated variables and understand the direction and strength of these relationships. This information could be useful for decision-making, maintenance, optimization, and performance improvement. For example, suppose cross-level and warp have a high positive correlation. In that case, it shows that adjustments to one variable may impact the other, emphasizing the necessity for coordinated maintenance efforts.

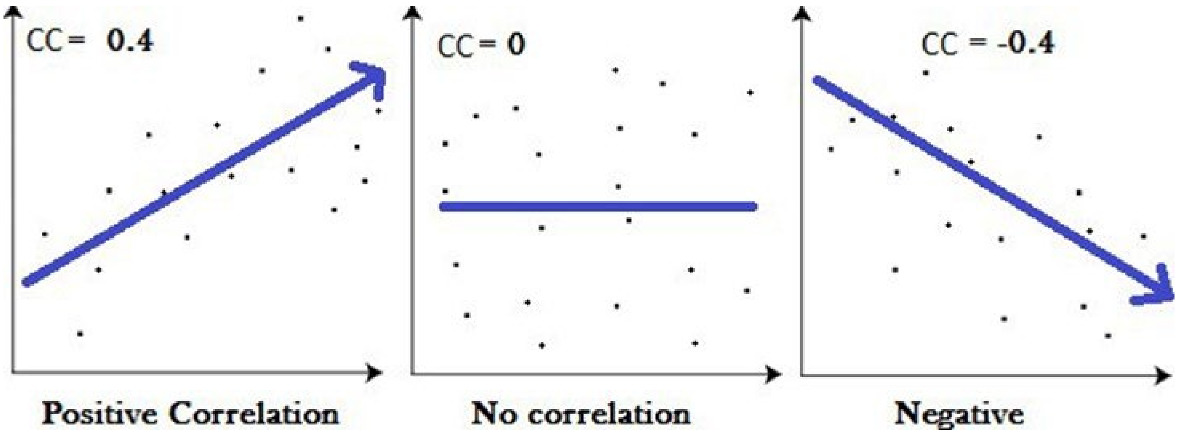


Figure 2.6: correlation coefficient

Figure 2.7 shows a correlation plot for 24 months of railway profile. The Pear- son equation was used to calculate the Correlation coefficient between the variables. The plot shows a significant positive association between each month and the months before it, with correlation values reaching +0.80, which suggests a linear relationship, meaning that an increase in profile value in one month is likely to be accompanied by an increase in the following months.

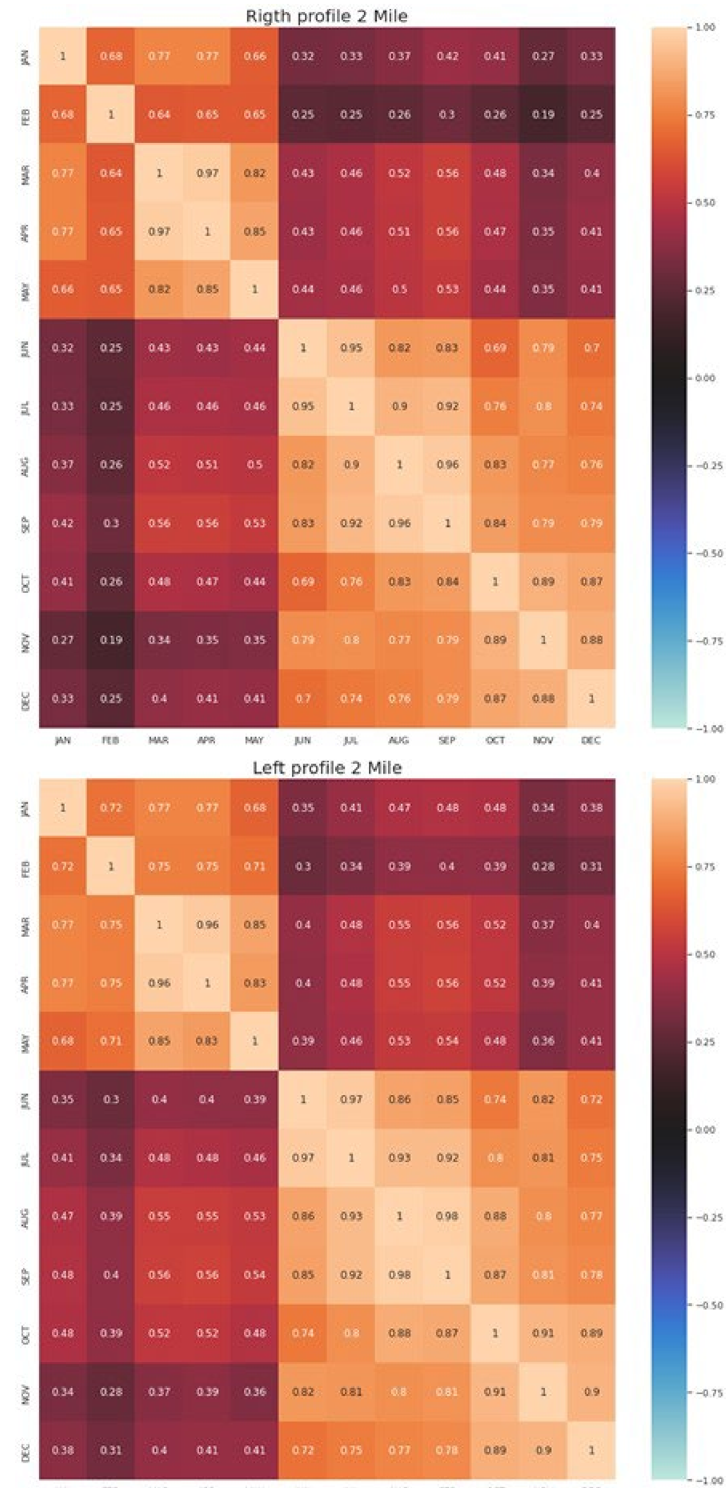


Figure 2.7: Correlation plots for the left and right profile

2.2.4 Box and Whisker Plot

In exploratory data analysis (EDA), box and quantile plots can shed light on the distribution, central tendency, and data variability. They are effective ways to summarize and visually analyze a dataset's characteristics. A box plot, known as a box-and-whisker plot, provides a graphical representation of data distribution, metrics of central tendency, and variability (Peck et al. 2015). The plot is made of a box reflecting the middle 50% of the data, called the interquartile range (IQR), as shown in Figure 2.8. The box's" whiskers" represents the range of the data, excluding any outliers, while the line inside the box shows the median. The box plot shows the outliners, which are points that deviate significantly from the rest of the data, by individual points or asterisks outside the whiskers. Examining the box plot gives an easy way to identify a dataset's spread, skewness, and presence of outliers. Figure 2.9 shows a box plot for 24 months of the profile for the whole track. The plot reveals the spread of the variables. The spread of the variables can be determined by observing the width of the boxes. In this case, nearly all the months have the same spread and similar median and quantile.

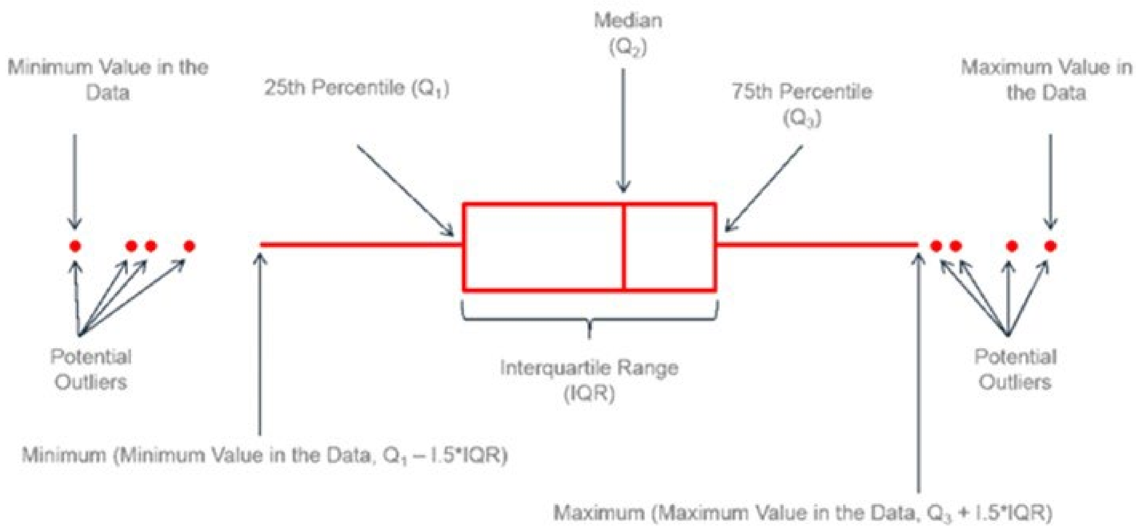


Figure 2.8: Box and Whisker Plot (Parker 2023)

Outliers can skew outcomes and lead to inaccurate prediction model interpretations. However, in railways, outliers in profile indicate the diversity and variability of track conditions and the presence of track defects. Instead of discarding these outliers, including them in the analysis is important. Doing so and considering typical and atypical observations can develop a comprehensive understanding of the track geometry, ensuring all relevant information is incorporated into the analyses and modeling development. In this track, for class 6 and according to the FRA limits, the profile safety limit is 1.0 inch. From figure 2.9, it is evident that months 12 and 21 have the highest number of defects and the smaller number of defects in month 2.

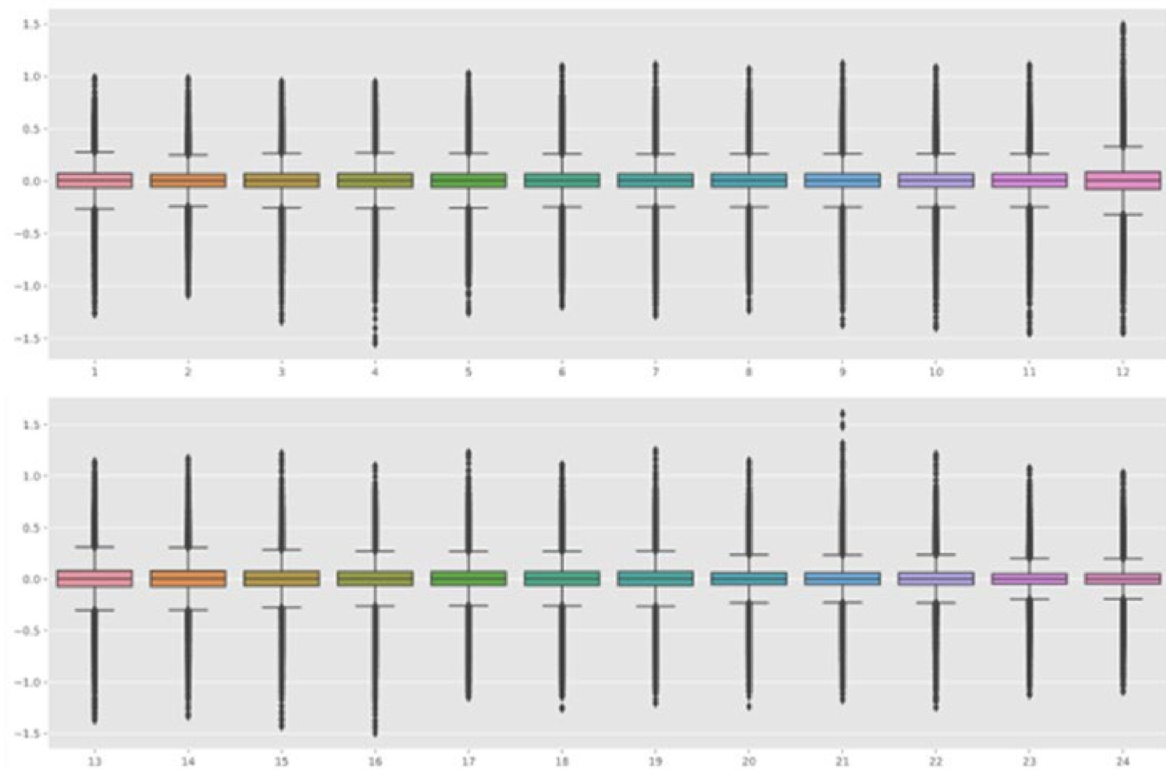


Figure 2.9: Box Plot for the 24 months of the profile

2.2.5 Quantile-Quantile (QQ) Plot

Quantile-Quantile (QQ) plots are useful tools for determining the distribution of a random variable in exploratory data analysis (EDA). QQ plots can demonstrate if the data follows a given distribution, such as Gaussian, uniform, exponential, or Pareto, by comparing the observed quantiles of a variable with the expected quantiles of a theoretical distribution. In the railway's context, QQ plots provide insights into the similarity between the two sets of quantiles and if both sets come from the same distribution. Figure 2.10 shows QQ plots for months no. 1 and 2. However, the points on both plots form a roughly straight line, suggesting that the observed and theoretical quantiles are drawn from the same distribution. The plot reveals a high correlation between the profile with time for most quantiles, with less correlation observed towards the tail ends of the distributions. This behavior is generally observed across all datasets.

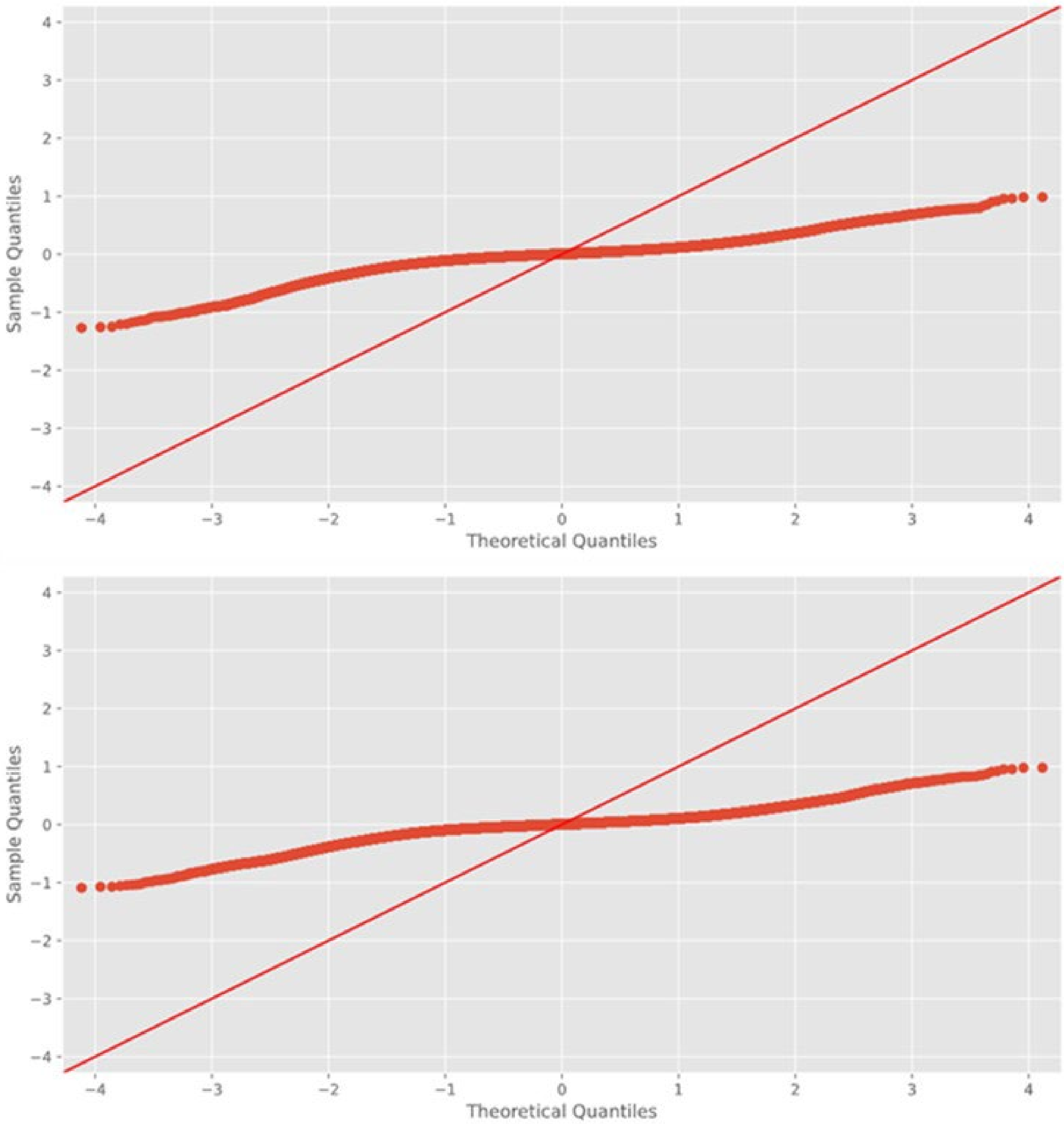


Figure 2.10: Quantile-Quantile (QQ) Plot for the first and second month

2.3 Data preprocessing

2.3.1 Data Alignment

The first analysis of EDA generally focused on the track on a macro basis. However, in the next part, the track will be analyzed on a micro basis. However, alignment for the data is essential.

Aligning the inspection data longitudinally is essential in analyzing railway track inspection data over time. Usually, automated track inspection vehicles gather numerous data parameters to evaluate the state of track components, including track geometry, rail profile, crosstie condition, and ride quality. GPS coordinates and milepost markers align these, allowing for spatial reference. However, aligning the data longitudinally, using only GPS coordinates and milepost markers, with sufficient accuracy for direct comparison, is often challenging. Time series techniques and Traditional methods of comparing data over time rely heavily on averaging techniques which are easily influenced by slight calibration errors or differences in sampling intervals between different measuring systems, resulting in misalignment between data collection points. The measured profile magnitude fluctuates as track components deteriorate, but the overall pattern is preserved. To solve this problem of misalignment, data alignment techniques have been developed, such as RFID tags or automatic location identification systems. However, in track geometry data, the cross-correlation function can be applied (Palese et al. 2020). Cross-correlation evaluates how similar two functions are in relation to a constant shift or lag. The cross-correlation function calculates the lag at which two datasets are most closely aligned by considering the data as a time series e.g. this can be done by selecting one dataset as the reference and lining up each additional dataset with it. Each lag's correlation coefficient is calculated, and the lag with the highest correlation coefficient is considered the best alignment, and the optimal lag is then used to shift the data to align them with the reference. The equation for Cross-correlation is given in the Equation:

$$lag(t) = \frac{\sum_{n=1}^{n}[(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum_{n=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{n=1}^{n}[(y_i - \bar{y})^2}}$$

Firstly, the track is divided into 500 feet segments. Then to align the data, the cross-correlation is calculated for each lag. Cross-correlation is mathematical technique to measure the similarity between two signals or data sets at different time lags. The optimal lag, which reflects the time shift or alignment required to match the respective segments of the track data, is obtained by computing the cross-correlation. Figure 2.11 shows the profile after alignment. This optimal lag guarantees the data segments are properly aligned, minimizing any offsets or misalignments. The term "spikes" refers to certain irregularities or anomalies in the track profile that are visible before the alignment process. These spikes appear to be offset from one another, indicating a lack of alignment between the corresponding segments of the track data. Figure 2.12 show the before and after alignment for 500 feet segment.

In Figure 2.13, the correlation coefficient p(l) for each lag value, l for month 2 correlated with month 1 for segment 6000 to 6500 foot, for the allowable lag range of ±100 samples. The maximum value of 0.98 at a lag of l = -2 = L can be clearly seen. Note that the pre-aligned correlation coefficient, p(l = 0) was 0.95.

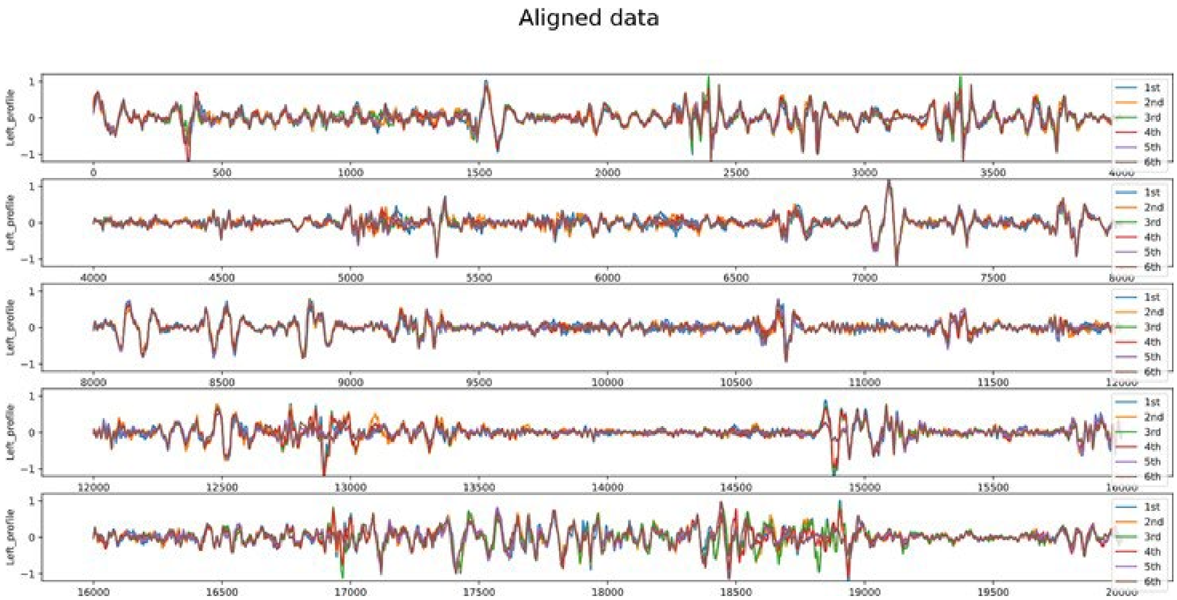
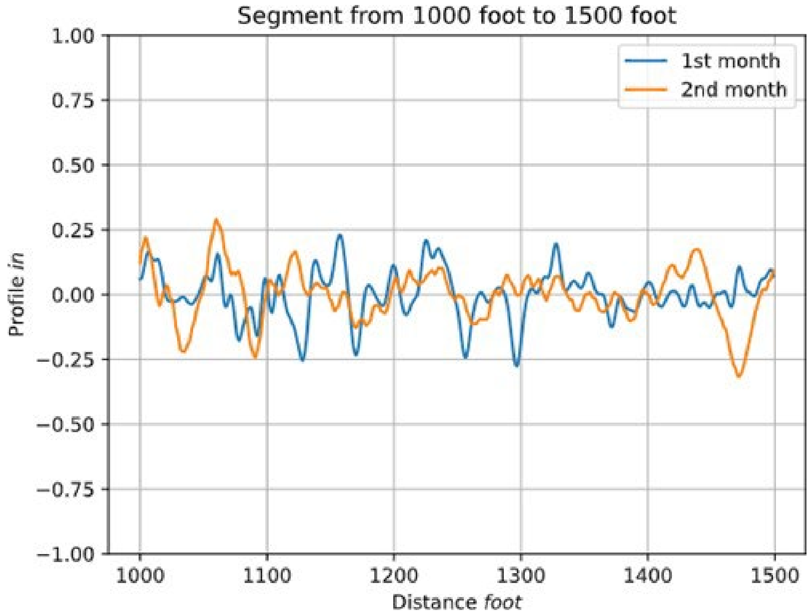
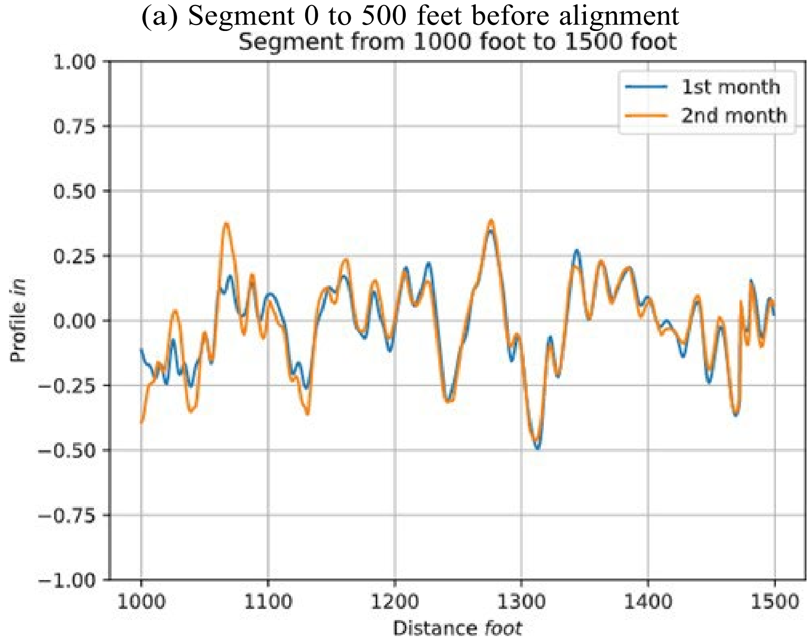


Figure 2.11: Align left profile for the 6 months

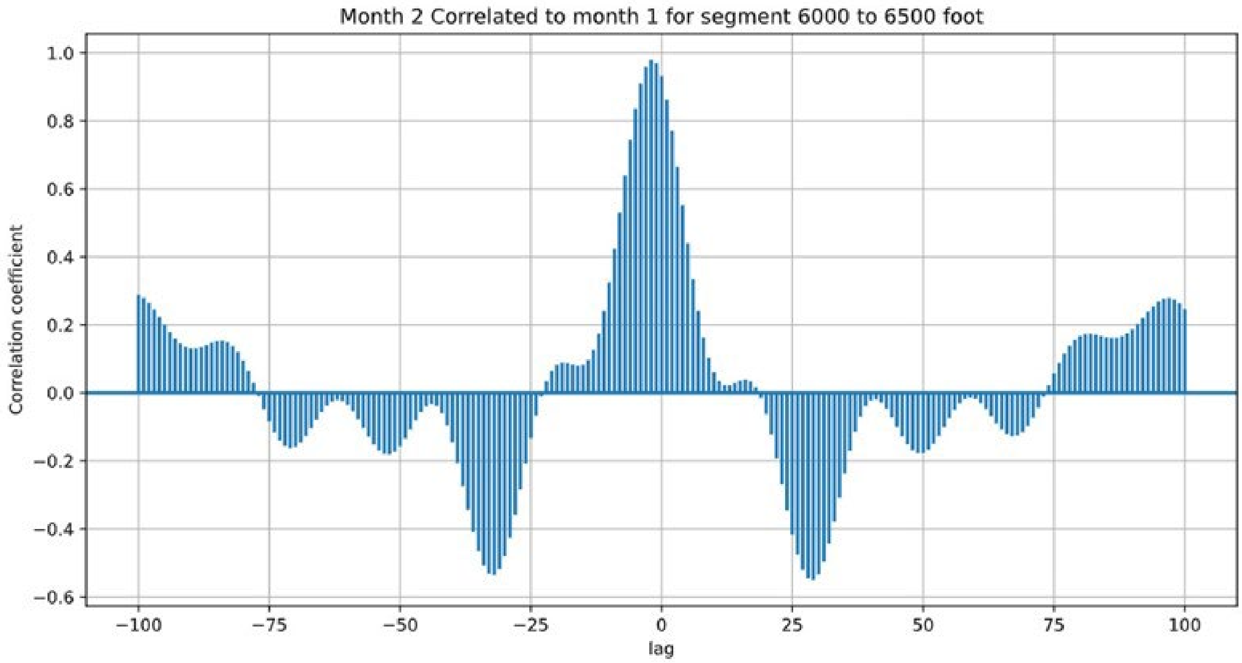


(a) Segment 0 to 500 feet before alignment



(b) Segment 0 to 500 feet after alignment

Figure 2.12: Segment 0 to 500 feet before & after alignment



2.3.2 Previous maintenance

Lack of knowledge regarding the timing and location of past maintenance actions is a common problem when working with railway data. This information is important in data preprocessing and preparation and could bias the model results. During maintenance, the railway infrastructure undergoes restoration and adjustments, which can significantly impact the measured values. This could make the model attempt to establish relationships between variables that may not be present after each maintenance intervention. Therefore, the model might falsely perceive a substantial improvement in the rail condition and try to establish correlations that do not exist.

Figure 2.13 is a heat map showing the degradation of the standard deviation of the profile over the two years for around 100 track sections. The heat map provides a visual representation of the changes in the track's profile variability, with a reduction indicating tamping or renewal actions. By analyzing the heat map, it is possible to identify both the timing and location of maintenance and renewal activities along the track. This information is valuable for effective planning and scheduling of maintenance operations.

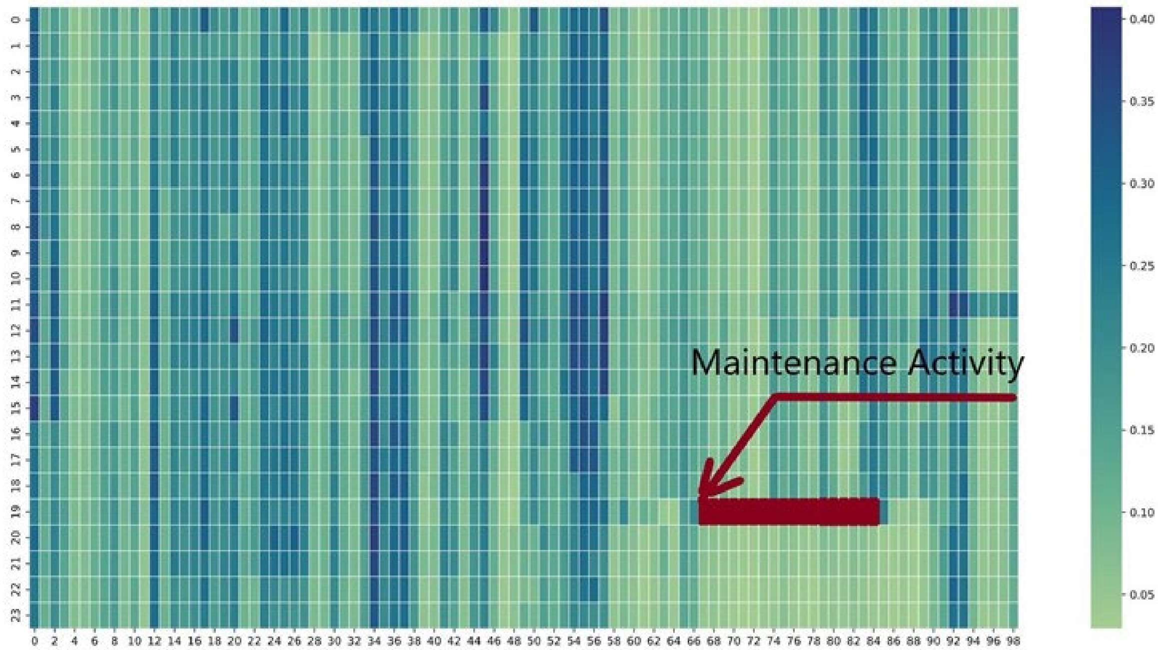


Figure 2.13: Heat map of the evolution of Standard Deviation for the profile over two years

2.4 Summary

During the Exploratory Data Analysis (EDA) phase, our main objective was to comprehensively understand the dataset and uncover any discernible patterns or relationships within the data. To achieve this, we utilized various visualization techniques, such as histograms, box plots, scatter plots, and correlation plots. The histograms visually represent the data distribution, allowing us to identify any outliers or unusual patterns. Box plots, on the other hand, summarized the distribution of the data by displaying its central tendency, variability, and any potential outliers. These visualizations helped us gain insights into the spread and characteristics of the dataset.

We also calculated descriptive statistics, including mean, median, standard deviation, quantiles, and interquartile range. These statistics further quantified different aspects of the data, providing a deeper understanding of its central tendencies and variability. In addition, correlation plots were employed to visualize the relationships and associations between variables in the dataset. By examining the strength and direction of these correlations, we identified potential dependencies and interactions among the variables.

The insights gained from the EDA provided a solid foundation for understanding the properties and characteristics of the dataset. Identifying patterns, relationships, and data distribution enabled us to make informed decisions and guide further analysis in our research. Based on the findings from the EDA, we will now proceed to the next section of our thesis, where we will delve into the modeling and prediction techniques, leveraging the knowledge gained from the EDA to develop accurate and robust predictive models for track profile forecasting.

Chapter 3 FUNCTIONAL NETWORKS

3.1 Introduction

The efficiency and safety of railway operations are crucial for any railway system and to ensure that track maintenance and monitoring are required. Predictive maintenance is an approach that offers a significant advantage by scheduling maintenance activities based on predicted defect occurrences. While some data-driven models exist for predicting track degradation rates using geometry data, the proposed approach aims to predict individual geometry values instead of track degradation for segments and improve prediction accuracy by combining mechanical models with data-driven approaches, specifically employing functional networks, Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks instead of traditional neural networks.

The main goal is to accurately predict future irregularities in railway track using geometry data. To achieve this, we will utilize a machine-learning technique called functional networks. This approach involves predicting values at every foot along the track based on previous measurements and their behavior. The research will be based on monthly railway track geometry data collected over a period of 24 months.

Functional networks are machine-learning techniques utilized to forecast values at every foot along the track based on previous measurements and their behavior. It is a crucial component of this research, focusing on applying functional networks and presenting the obtained results and their limitations. In these efforts, we employ machine learning techniques, specifically functional networks, to improve our under- standing of how these models might contribute accurate predictions of track geometric data, consequently, enable more informed maintenance planning decisions.

Two primary purposes will be highlighted. Firstly, the introduction of functional networks as a machine learning method suitable for predicting track geometry data with a comprehensive explanation of functional networks, their specification, discussion on the rationale behind choosing them, and their relevance to the research objective. Secondly, present and analyze the results obtained from the functional network models. In addition to providing visual representations of the predicted track geometry data, the effectiveness of these networks will be assessed, and a discussion of the limitations and challenges encountered during the application of this class of models. This analysis and evaluation will shed light on potential causes of errors or inaccuracies, thus, showing the boundaries and areas for improvement in the methodology.

3.2 Functional Networks

3.2.1 Overview of Neural Networks

Functional Networks are a machine learning model class that has demonstrated promising results in various applications. These networks are designed to capture the functional equations between input variables and output values. Functional Networks leverage the power of Neural Networks (NNs) to model intricate relationships between input features and output predictions. Neural Networks are one of the sub-fields of artificial intelligence (AI); they are computational models inspired by the structure and functioning of the human brain. By using nonlinear functions on weighted sums of input data, NNs seek to mimic the activity of individual neurons. They comprise interconnected layers that include input, hidden, and output layers, via which the network's information is transmitted.

NNs have gained a lot of attraction and popularity in recent years due to their ability to identify complex patterns from large amounts of data and learn high-level features, leading to deep learning (DNNs). DNNs are neural networks with multiple hidden layers that excel at capturing and representing intricate data relationships. These features make them excellent for complex prediction tasks like track geometry data forecasting. DNNs training involves finding the optimal weights and biases inside the network, whereas inference utilizes the trained model to generate predictions on new, unseen data. During training, the network modifies its weights to reduce the loss, which is the difference between the true and predicted values. The backpropagation algorithm computes the partial derivatives of the gradient using the chain rule of calculus, allowing one to determine how each weight impacts the loss. Techniques for training DNNs include supervised, unsupervised, semi-supervised, and reinforcement learning. DNNs were first developed in the 1940s, and commercial applications started to appear in the 1980s. However, accuracy improvements for DNN-based applications only occurred in the early 2010s. Numerous variables, such as the availability of big training datasets, improvements in computing power, and algorithmic developments, are responsible for this development.

One of drawbacks of DNNs is their need for significant computational resources and massive datasets, typically handled in cloud systems. However, depending on the requirements of the individual application, the inference phase may be carried out either at the edge or in the cloud. For example, performing DNN inference processing closer to the sensor in computer vision applications is preferable to decrease latency and improve real-time capabilities.

3.2.2 FUNCTIONAL NETWORKS

Functional networks (FNs) are a novel generalization of neural networks (NNs) that combine domain knowledge with data in machine learning tasks. While NNs are powerful tools for solving various problems, they have limitations, such as black box models with low interpretability. FNs go beyond being black box models by incorporating domain-specific knowledge, considering functional constraints, leveraging this knowledge to design the network's structure, and using data to estimate the neuron functions (Castillo and Gutiérrez 1998). This makes them suitable for predicting railway geometric data, where understanding the underlying functional relationships and incorporating domain knowledge is crucial. While NNs have fixed internal neuron functions, FNs do not have fixed internal neuron functions, which are learnable from the data. Moreover, the weights of the connections between neurons are absent in FNs, and the effects of these connections are subsumed within the neuron functions. This enables FNs to include functional constraints determined by functional qualities, such as associativity and distributivity.

Functional networks have several important components (Castillo et al. 2012):

- Input layer: In this layer, input data are shown as small black circles (e.g., x, y, and z in Figure 3.1).
- Output layer: This layer contains the output data, Displaying the output data as little circles. (e.g., w in Figure 3.1).
- processing units: These units receive input values from the preceding layer (intermediate or

input units) and evaluate them to generate a set of output values for the next layer (intermediate or output units). Each neuron within the functional network is associated with a neuron function, which can be multivariate and have as many arguments as there are inputs to the neuron. The individual components of the neuron function are called functional cells (e.g., f, g, and h in Figure 3.1).

- intermediate storing units: Intermediate storing units refer to the layers within the functional network that store intermediate information generated by the neuron units. These units are depicted as small circles (e.g., u, v, and w in Figure 3.1).

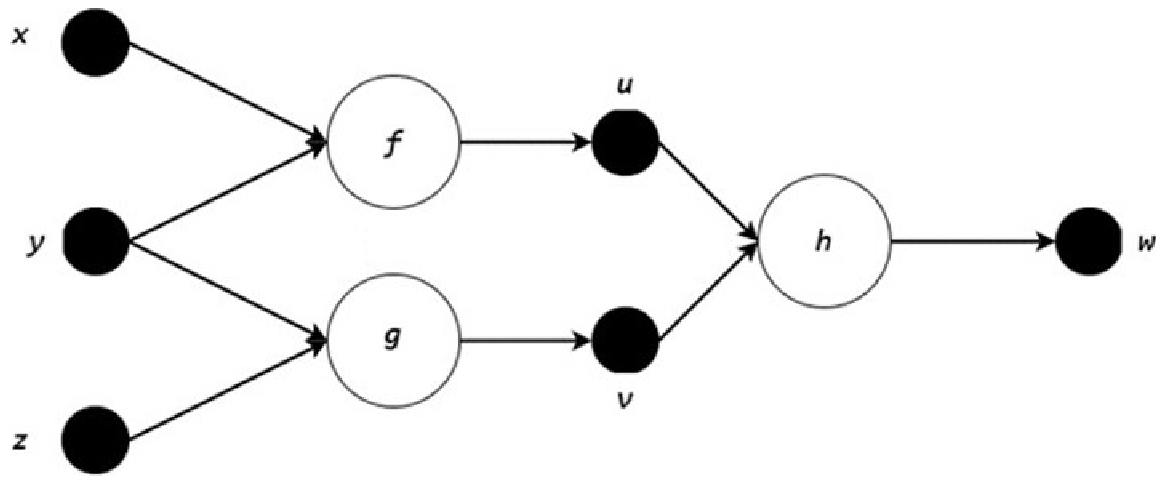


Figure 3.1: a functional network architecture (Castillo et al. 2012)

3.3 Types of FN

3.3.1 The Generalized Associativity Model

The Generalized Associativity Model is a functional network with the property of generalized associativity, in which the output can be computed using two alternative sets of inputs. The model can be made simpler by describing the functions involved in terms of arbitrary continuous and strictly monotonic functions, as in the equation below. Figure 3.2 shows examples of the Generalized Associativity Mode architecture (Castillo and Gutiérrez 1998).

$$F\,[g(x, y), z] = K[x, N(y, z)]$$

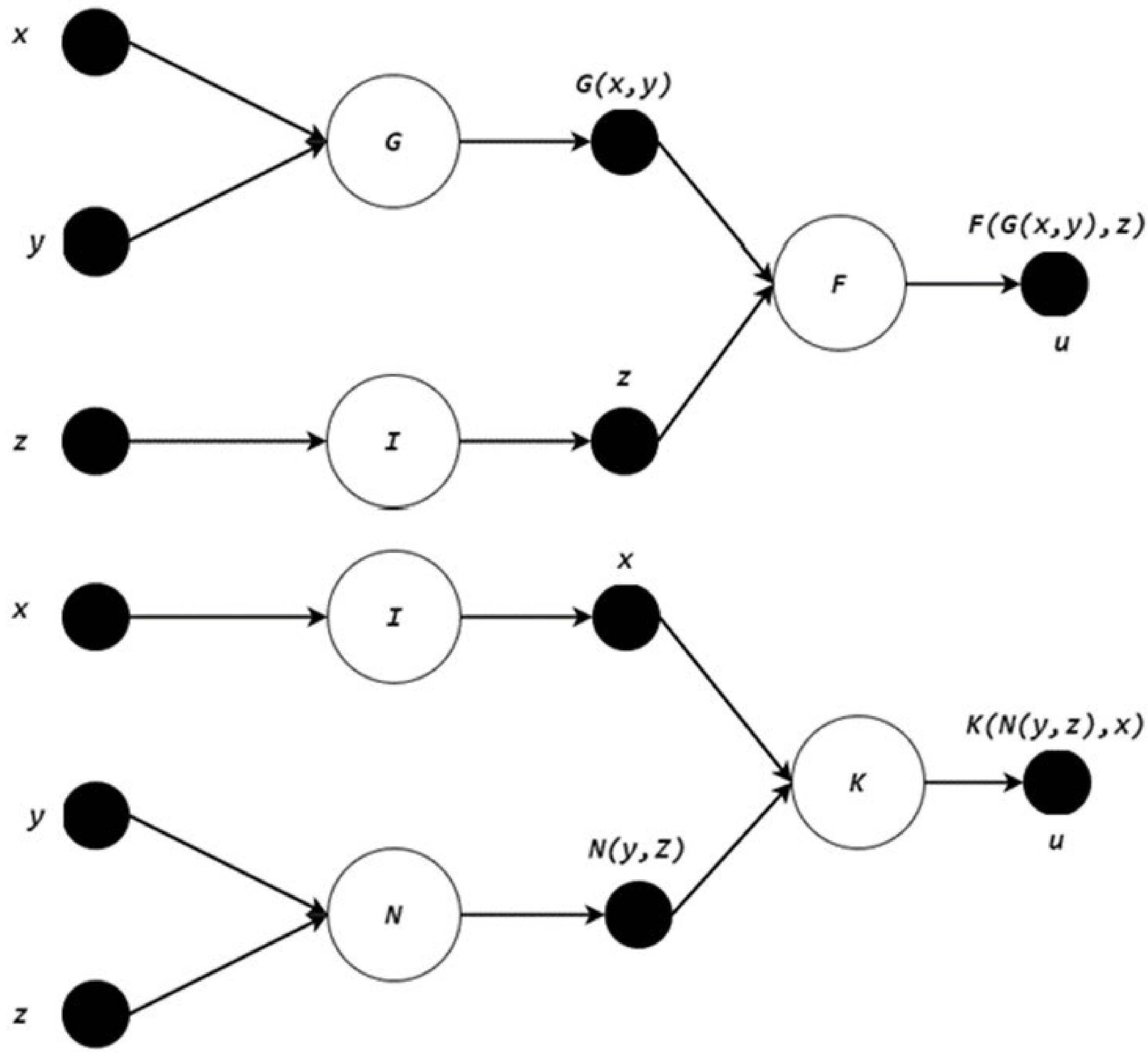


Figure 3.2: Example of the Generalized Associativity Model architectures (Castillo et al. 2012)

3.3.2 The Separable model

Figure 3.3 illustrates the separable model, which is described by the equation below. The model consists of linearly independent functions $f_i(x)$ and $g_i(y)$. In order to streamline the model, it is assumed that the sets of functions are also linearly independent.

$$z = F(x, y) = \sum_{i=1}^{n} f_i(x) g_i(y) = \sum_{j=1}^{m} h_j(x) k_j(y)$$

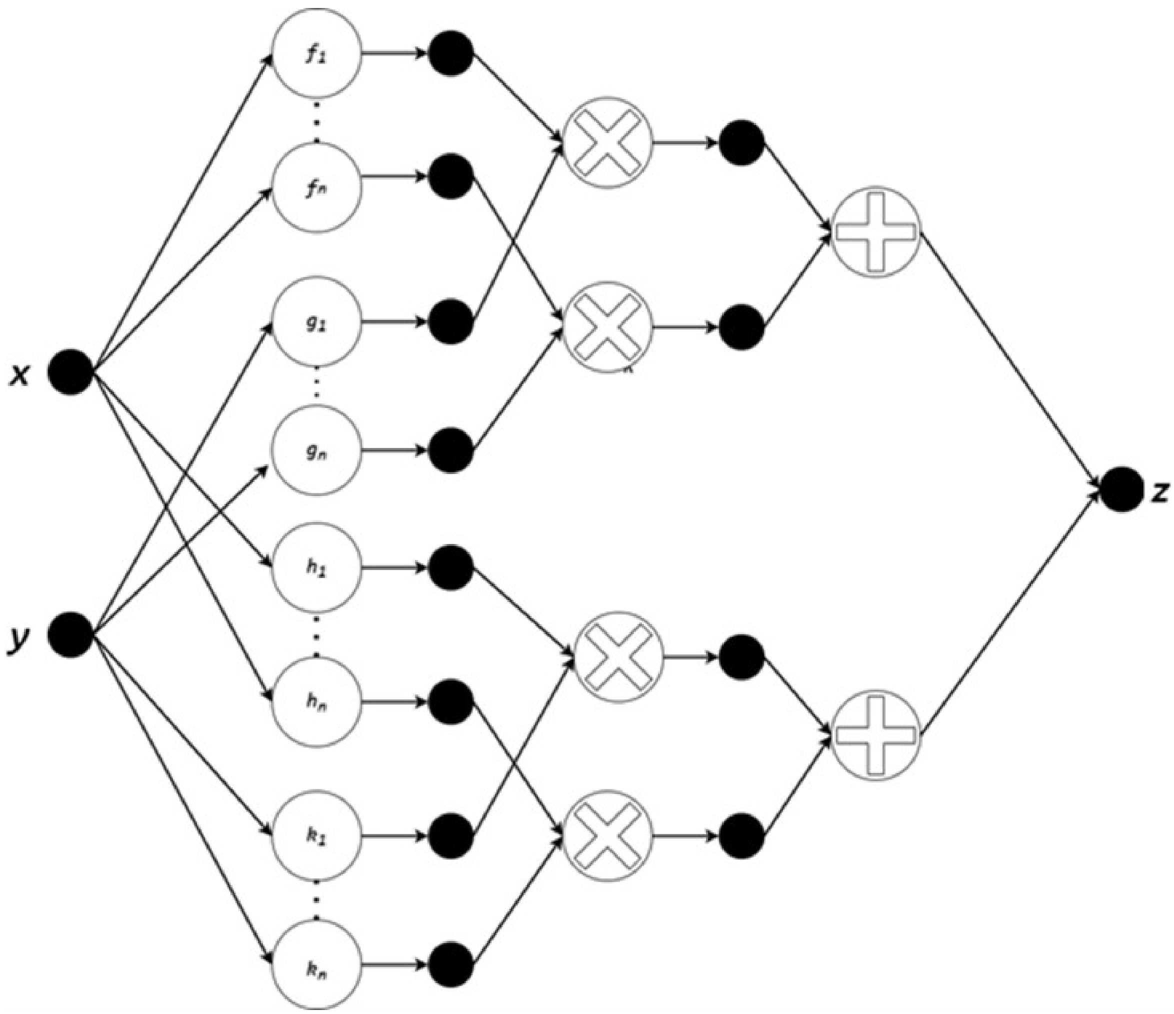


Figure 3.3: Example of the Separable model architectures

3.3.3 The Serial functional model

A sequential network of functional units on one layer is called a serial functional model, described in the below equation. A serial functional network's output can be modeled as a composite of the same function applied more than once, as shown in figure 3.4. The problem can be reduced to the translation functional equation by investigating the scenario where every neuron is the same.

$$y_{n+1} = f_n(f_{n-1}(\ldots f_1(x)\ldots))$$

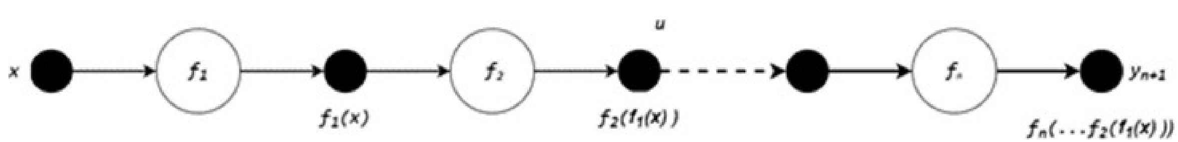


Figure 3.4: The Serial functional model architectures (Castillo et al. 2012)

3.4 FN selection

Traditional techniques like Fourier transforms, and Box-Jenkins models have been extensively used in the fields of prediction and forecasting. With the help of these techniques, future predictions can be made by fitting the data into a model. These models, however, frequently include linear assumptions and might not be able to fully rep- resent the intricate dynamics connected to the stochastic behavior of railway systems. Therefore, there has been a shift away from linear methods in recognition of their limitations. It has become clear that neural networks are effective for modeling non-linear systems. But for this study, we have opted to use a particular class of neural network called functional neural networks. FNs' flexibility in handling various equation families that can be suitable for modeling the complexity of railway geometry data. They can accommodate equation families such as the polynomial family (e.g., $\phi = 1, x, x^2, ..., x^q$), trigonometric functions (e.g., $\phi = 1, sin(x), cos(x), ..., sin(qx), cos(qx)$), and exponential functions (e.g., $\phi = 1, e^x, e^{(-x)}, ..., e^{qx}, e^{(-qx)}$). Unlike sigmoidal functions commonly used in traditional neural networks, these equation families are not limited to a specific range of outputs. This approach offers a framework for forecasting track geometry information that is more intuitive and interpretable by taking advantage of the non-linear modeling natural of FNs while maintaining simplicity and interpretability, which will enable deeper insights into the factors influencing railway behavior and track geometry.

3.4.1 Box-Jenkins model

In 1970, George Box and Gwilym Jenkins developed the Box-Jenkins model, also known as the ARIMA (Autoregressive Integrated Moving Average) model (Box Jenkins 1973). This approach is widely used in various fields, such as economics, finance, and engineering, for analyzing and predicting time series data. The Box-Jenkins model is an essential model in the field of time series analysis and forecasting, which is based on the idea of autoregressive and moving average components, along with differencing, to capture and forecast the patterns and dynamics present in a time series dataset. The autoregressive model, known as AR(p), depicts a process in which the most recent value is expressed as the weighted sum of previous values plus noise. An autoregressive process of order p has the following generic form:

$$x_i = \alpha_1 x_{i-1} + ... + \alpha_p x_{i-p} + \alpha_i$$

The coefficients $\alpha_1, \alpha_2, ..., \alpha_p$ in the Box-Jenkins model are obtained by fitting the data into a linear model using methods like least squares. The model can be non-linear in functional networks, allowing for a more adaptable and expressive representation. Our approach is to use the Box-Jenkins model with functional neural networks by applying the associative model to autoregressive; the general form will be:

$$x_i = f(x_{i-1}) + ... + f_p(x_{i-p}) + \alpha_i$$

Here, f_1 and f_2 represent the activation functions applied to x_{i-1} and x_{i-2}, respectively. These activation functions introduce non-linearity to the model, enabling it to capture more complex dynamics in the railway geometric data. The challenge is to find a suitable Finding equation family that could fit the track geometric data and find a good equation that describes the relationship between the current value (x_i) and the previous values ($x(i-1)$ and $x(i-2)$). A polynomial family of equations, $\phi = 1, x, x^2, ..., x^q$, is considered suitable for the activation functions f_1 and f_2, to ensure flexibility and handle varied patterns in the data. Figure 3.5. shows the model structure.

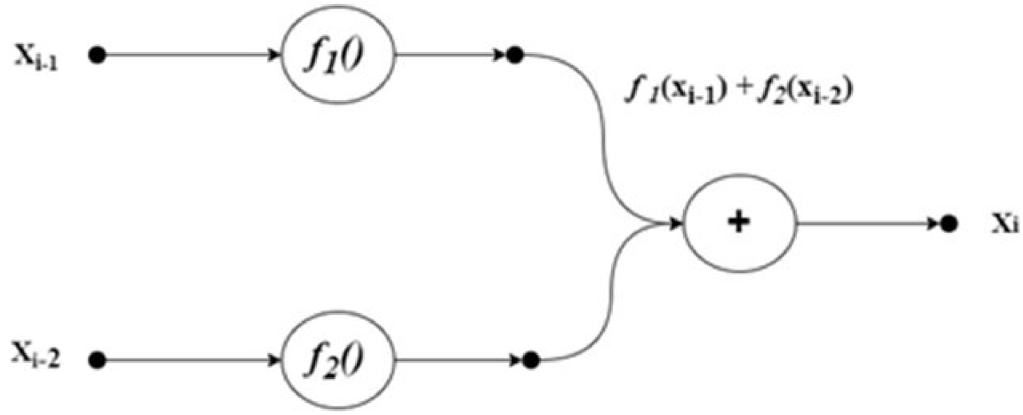


Figure 3.5: The Box-Jenkins functional network architectures where:
X_i = the current value of the profile
X_{i-1} = the $i-1$ value of the profile
X_{i-2} = the $i-2$ value of the profile
f_1 = the activation function for X_{i-1}
f_2 = the activation function for X_{i-2}

3.5 Methodology

3.5.1 Data preprocessing

Several measures were taken to ensure the data was high-quality and suitable for the functional network before it was ready for model training. Subsequent actions were carried out in the data preprocessing stage.

Data cleaning: The model's performance may be negatively impacted by missing values or noise in the raw data. To preserve the integrity and quality of the data, procedures such as imputation of missing values, alignment, and removing tamping data.

Normalization: To avoid bias or the dominance of features during model train- ing, it is crucial to scale the input data to a standard scale. Standard normalizing methods, including min-max scaling and z-score normalization, were used to ensure that all features contribute equally to learning.

Data Splitting: Training and testing sets of the prepared data were created. The split was set at an 80:20 ratio, where 80% of the data was used to train the functional network, and the remaining 20% was held back for performance testing. This division aids in evaluating the model's generalizability to new data and reducing the risk of overfitting.

Adjusting Time Steps: the size of the data after data preparation was large than 300,000. To handle the large dataset, the time step was adjusted by considering a specific time interval, such as one month, as a single time step. This adjustment allowed for manageable chunks of data to be fed into the model while still capturing the temporal behavior of the track profile over multiple tamping cycles. Inspections within the same time step were ignored to simplify the analysis and focus on the overall behavior of the track profile. After these preprocessing steps, the data was ready for model training functional neural network model.

3.5.2 Model Training

The specific functional network architecture used for predicting railway geometric data is based on the functional network shown in figure 3.6 and figure 3.5. For each output value xi, it takes two delayed values of the profile as input, namely x_{i-1} and x_{i-2}, which allow the model to capture the dependencies and patterns between the current profile value and its two preceding values. For the neuron functions, the quadratic polynomial functional family: x, x^2, is used. The quadratic polynomial function was found to be the best suitable for the profile data and enables the model to capture non-linear relationships between the input values and the output profile. The specific architecture can be represented as follows:

$$x_i = f1(x_{i-1}) + f_2(x_{i-2}) + \alpha_i$$

3.5.3 Training Process

The training process involves improving the functional network to minimize the difference between the predicted profile values and the ground truth values in the training data. A suitable loss function, mean squared error (MSE), was selected and used to quantify the discrepancy between the predicted and actual values in the training data. The training process aims to minimize this loss function by adjusting network parameters.

The model was trained using the prepared data, and the current profile value (x_i) was predicted using the delayed input values (x_{i-2} and x_{i-1}). The training procedure modified the functional network's parameters based on the estimated loss function. The objective was to find the functional network's ideal configuration that can accurately capture the relationships between the delayed input values and the output profile values.

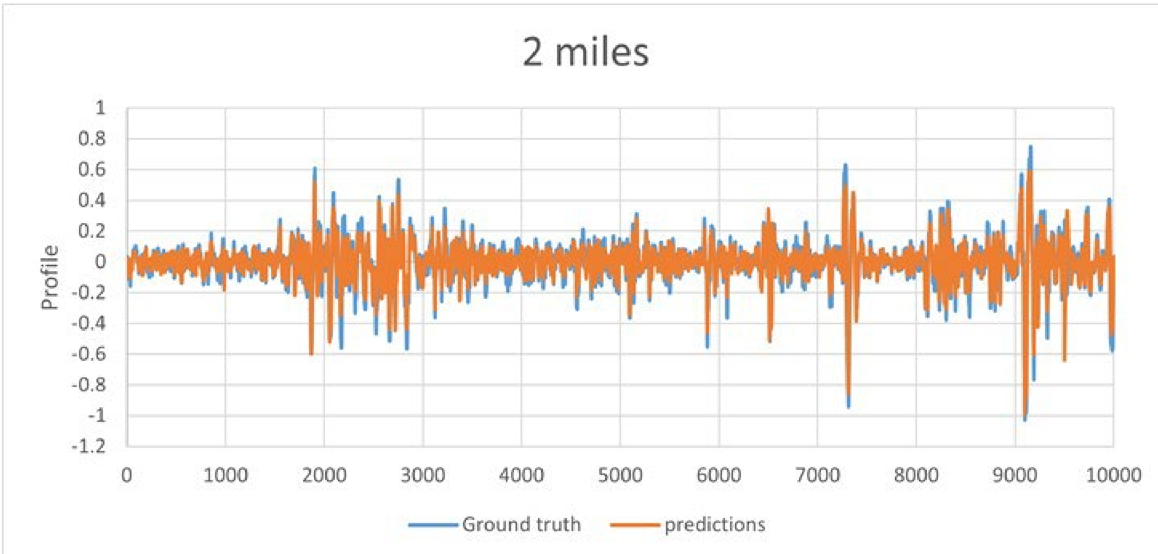
3.6 Results and Analysis

3.6.1 Metrics Used for Evaluation

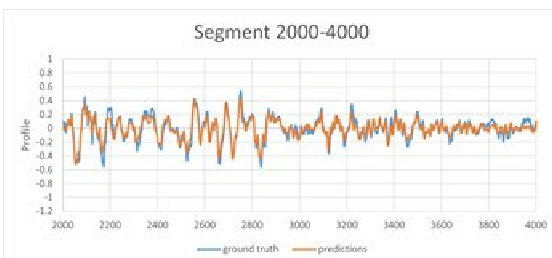
The mean squared error (MSE) metric was used to evaluate the performance of the predictive models using railroad geometry data. MSE calculates the average squared difference between the railroad profile predictions and actual values. Since there is less variance between the expected and actual profiles, a lower MSE indicates higher predictive ability The modeling was conducted using Python in Google Co- lab. Google Colab is a cloud-based development environment that provides a Jupyter notebook interface, allowing users to write and execute Python code collaboratively.
The functional network model generated the following equation for predicting the railway profile is as follows:

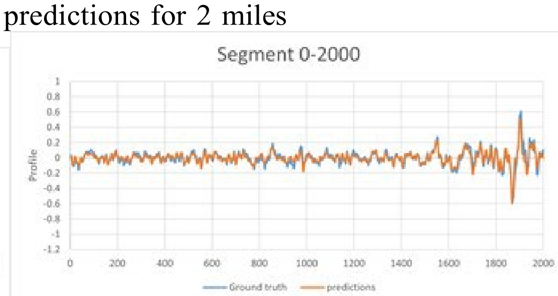
$$x_i = 0.00006 + 0.5706x_{i-1} - 0.0483x^2_{i-1} + 0.2535x_{i-2} + 0.0431x^2_{i-2}$$

To evaluate the performance of the functional network models, a comparison was made between the actual profiles and the predicted profiles. This assessment is depicted in Figures 3.6, 3.7, and 3.8. The models were utilized to forecast the profile data for a period of three months into the future.



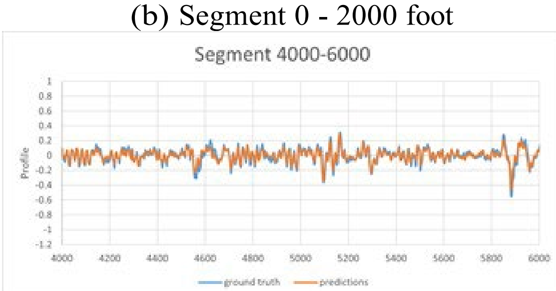
(a) The 1st month's predictions for 2 miles

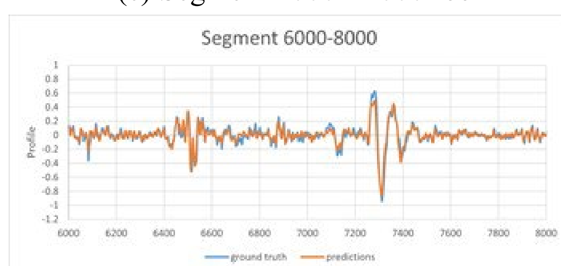




(b) Segment 0 - 2000 foot

(c) Segment 2000 - 4000 foot

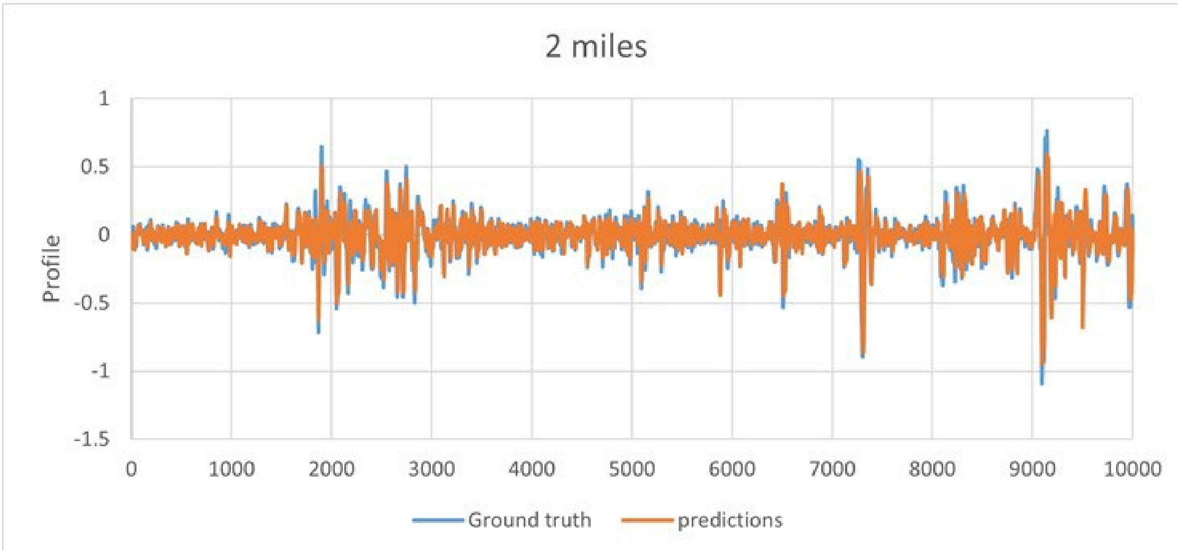




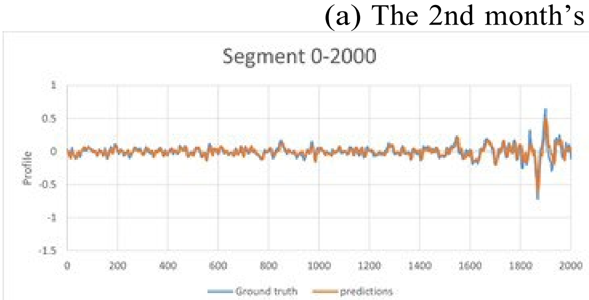
(d) Segment 4000 - 6000 foot

(e) Segment 6000 - 8000 foot

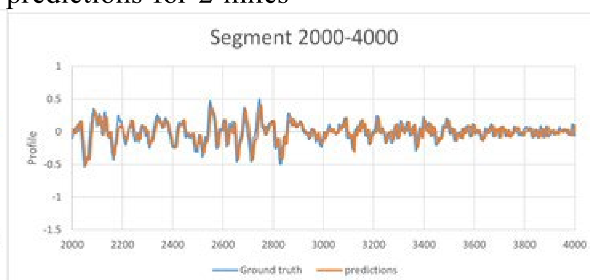
Figure 3.6: The 1st month's predictions



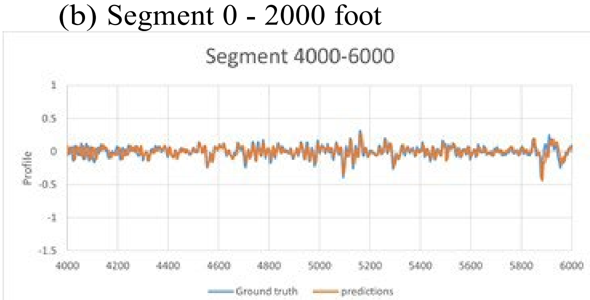
(a) The 2nd month's predictions for 2 miles



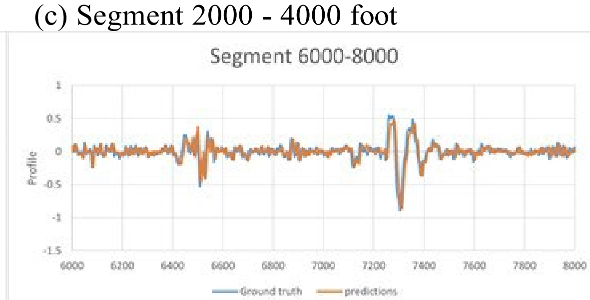
(b) Segment 0 - 2000 foot



(c) Segment 2000 - 4000 foot

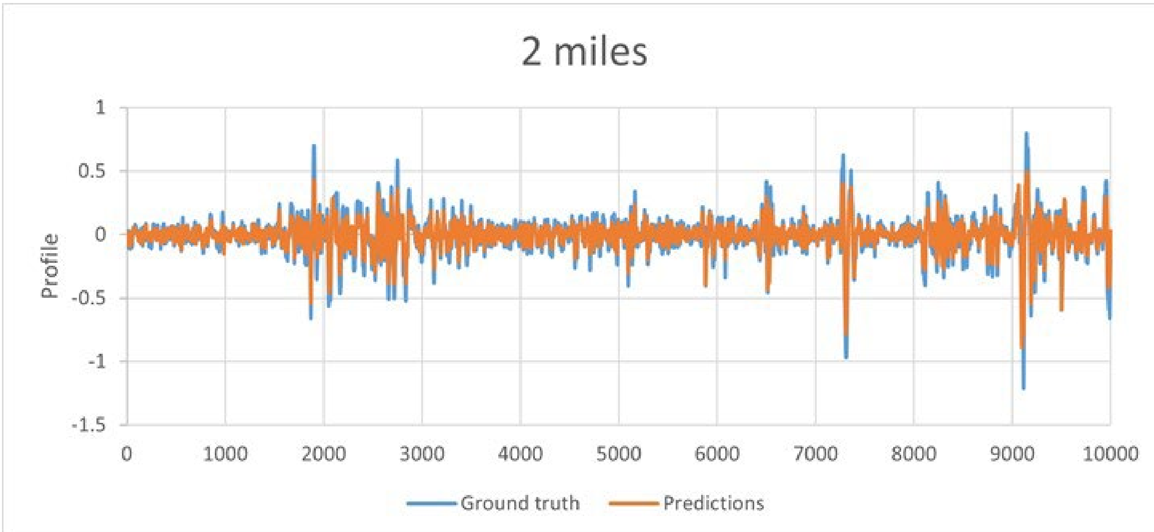


(d) Segment 4000 - 6000 foot

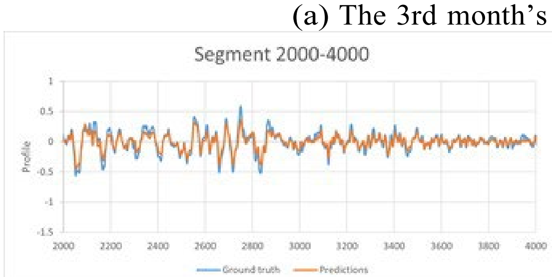


(e) Segment 6000 - 8000 foot

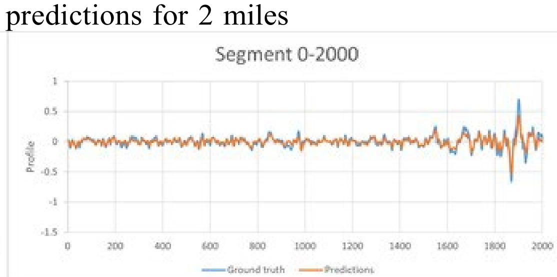
Figure 3.7: The 2nd month's predictions



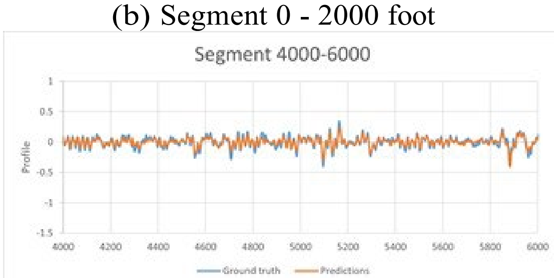
(a) The 3rd month's predictions for 2 miles



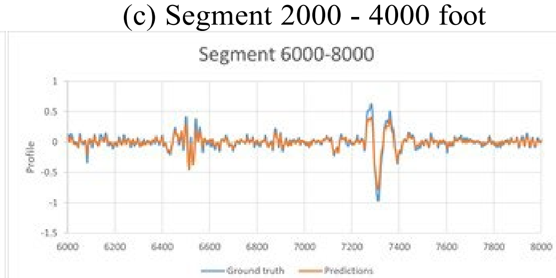
(b) Segment 0 - 2000 foot



(c) Segment 2000 - 4000 foot



(d) Segment 4000 - 6000 foot



(e) Segment 6000 - 8000 foot

Figure 3.8: The 3rd month's predictions

3.6.2 Discussion of Findings

Figures [3.6- 3.8] show the ground truth (expected) profiles and the corresponding prediction results for the first, second, and third months. The orange line in these figures represents the predictions made by the Functional network (FN), while the blue line represents the actual data. Additionally, each figure includes subfigures that provide a closer look at specific data segments.

To evaluate the accuracy of the predictions, the test was conducted using two specific months as input for the model. These two months served as the historical data for the FN model. The output of the model is the predicted values for the first month in the future; after that. These predicted values are compared with the actual values presented in the Figures [3.6- 3.8]. Figure 3.6 displays the actual values, predictions for the first month, and subfigures showing magnified data parts. These subfigures enable a more in-depth analysis of the FN's predictions about the actual values. Moving on to Figure 3.7, it shows the ground truth and predictions for the second month. Here, the model utilizes the predictions from the first month as input to forecast the second month. Lastly, Figure 3.8 displays the ground truth and predictions for the third month. In this case, the FN employs the predictions from both the first and second months as input to generate a forecast for the third month. By incorporating multiple previous predictions, the model seeks to predict multiple months in the future using only two months of historical data as input.

However, the FN model is considered a one-time-step forecast model. The ML one-time step forecast model relies on machine learning algorithms that learn from historical data patterns to make predictions for a single time step ahead. These algorithms analyze the input data, such as historical values, and identify relevant features and patterns contributing to the prediction. The model is trained using a specific technique, in this case, a Functional network, to learn the underlying relationships and dependencies in the data. The goal is to generate accurate predictions for a single time step ahead, such as forecasting the value for the next month given the data from the previous months.

The performance of the Functional Network (FN) in forecasting exhibits a discernible pattern. Initially, in the first month of forecasting, the FN demonstrates strong performance, particularly when predicting small values. However, as the fore- casting horizon extends further into the future, the accuracy of the FN's predictions diminishes. This behavior can be attributed to the inherent limitations of the FN as a one-time-step prediction model.

Notably, the Mean Squared Errors (MES) for the first, second, and third months are 0.001542, 0.0046, and 0.0036, respectively. These values indicate the average squared difference between the actual and predicted profiles. Furthermore, the maximum absolute errors for the first, second, and third months are 0.382081, 0.528598, and 0.611619, respectively. These metrics provide insights into the magnitude of the largest discrepancies between the actual and predicted values.

Taken together, these findings suggest that while the FN performs well in short- term forecasting, its accuracy diminishes as the time horizon extends further into the future. These limitations highlight the challenges associated with using a one-time step prediction model and emphasize the need for further investigation and potentially exploring alternative modeling approaches to improve long-term forecasting performance.

The FN's functionality is constrained to predicting only the immediate next step based solely on the current input. It lacks the capability to incorporate information beyond this immediate future. As a result, the FN encounters challenges when attempting to forecast further into the future, leading to reduced accuracy. The model's limitations prevent it from effectively capturing long-term patterns or trends that may exist in the data. The decrease in prediction accuracy over time can be understood by considering the accumulation of errors. As each subsequent prediction relies on the previous prediction, any errors or inaccuracies in the initial forecasts are carried forward and compounded with each subsequent step. This cumulative effect amplifies the discrepancy between the predicted values and the actual data.

3.7 Concluding Remarks

In conclusion, the FN exhibited satisfactory performance in the short term, especially for small-value predictions. However, as the prediction horizon extended, the accuracy of the FN declined due to its one-time step prediction nature—consequently, errors accumulated over time, resulting in reduced accuracy. Notably, the FN faced challenges in accurately predicting peak profiles, which are crucial for effective railway maintenance and operational planning. Accurate forecasting is vital in optimizing maintenance schedules, minimizing downtime, and ensuring the seamless operation of railway systems.

The findings of this study underscore the significance of selecting appropriate models for railway maintenance and operational forecasting. While the FN demonstrated effectiveness in short-term predictions, it proved inadequate for longer time horizons. Furthermore, the accurate prediction of peak values in the profiles is paramount for railway maintenance. Peaks often indicate critical locations requiring maintenance attention. By precisely forecasting these peaks, operators can allocate resources and schedule maintenance activities accordingly, minimizing disruptions and ensuring passenger safety. However, the FN's inability to predict peaks effectively highlights the need for more advanced models to capture and anticipate such fluctuations accurately.

Furthermore, the study emphasizes the importance of considering the time horizon when selecting forecasting models for railway maintenance and operation. Short- term predictions may be adequate for immediate decision-making, such as scheduling routine maintenance tasks. However, long-term forecasts are essential for strategic planning and resource allocation. Therefore, there is a need to adopt models capable of capturing long-term patterns and trends. The limitations of the FN make it in- adequate for accurately predicting future events beyond a single time step. However, RNNs and LSTM offer solutions to better capture dependencies over extended periods. By incorporating these models, railway maintenance and operation planners can obtain more accurate forecasts, enabling them to address potential issues and allocate resources effectively and proactively. They are designed to handle sequential data and can capture dependencies over time. Developing models that can effectively predict future values in railway maintenance and operation profiles is possible.

Chapter 4 LONG SHORT-TERM MEMORY NETWORKS

4.1 Introduction

Through the study, it was learned that accurate prediction of track geometry data is crucial for ensuring safe and efficient railway operations. In recent years, machine learning methods have been increasingly utilized to track and predict track geometry data, aiming to capture the intricate relationships between various external factors and forecast future track geometry values.

To overcome the challenges and limitations observed in the forecasting performance of the Functional Network (FN) model, further research could explore alter- native modeling approaches. For instance, incorporating more sophisticated machine learning techniques, such as recurrent neural networks (RNNs) or long short-term memory (LSTM) networks, may help capture temporal dependencies and improve forecasting accuracy. Additionally, additional external factors and domain-specific features in the modeling process could enhance the model's ability to capture the complexities of track geometry data.

Long Short-Term Memory (LSTM) networks are one machine learning approach that has demonstrated significant promise in time-series prediction challenges. Recur- rent neural networks (RNNs) with LSTMs are made to handle input sequences with variable lengths and long-term dependencies. Because LSTM networks can store and retrieve data for longer periods than conventional models, they are ideally suited for time-series prediction tasks with high temporal dependence.

4.1 RNNs and LSTM

RNNs, are a type of neural network that can process sequential data by retaining information from previous time steps. They have connections between nodes that allow information to flow in cycles, enabling them to capture temporal dependencies. RNNs are therefore ideally suited for problems requiring sequential data, such as time-series prediction, speech recognition, and natural language processing.

Due to the vanishing gradient issue, traditional RNNs cannot accurately capture long-term relationships in sequences. During the training process of an RNN, gradients are calculated to update the network weights, allowing the model to learn from the data. However, as the gradients propagate through time, they can become extremely small. This occurs due to the repeated multiplication of gradient values, which can result in a vanishingly small gradient. When the gradients become too small, they no longer effectively convey information about the input data and previous time steps to update the network weights. Consequently, the RNN struggles to capture long-term dependencies and cannot effectively model relationships that span distant time steps. To overcome this challenge, LSTM networks were introduced. The Long Short-Term Memory (LSTM) networks are a Recurrent Neural Network (RNN) type created to address the issue of disappearing gradients in standard RNNs. Hochreiter and Schmidhuber (1997) presented LSTM networks as a solution to the vanishing gradient problem. The LSTM design incorporates memory cells and gating mechanisms that allow for the selective storage and retrieval of information over long periods. The memory cell retains essential information from previous time steps, enabling the net- work to recognize long-term dependencies in input sequences. The gating mechanisms, consisting of input, output, and forget gates, regulate the flow of information into and out of the memory cell.

To overcome this challenge, LSTM networks were introduced. The Long Short- Term Memory (LSTM) networks are a Recurrent Neural Network (RNN) type created to address the issue of disappearing gradients in standard RNNs. Hochreiter and Schmidhuber (1997) presented LSTM networks as a solution to the vanishing gradient problem. The LSTM design incorporates memory cells and gating mechanisms that allow for the selective storage and retrieval of information over long periods. The memory cell retains essential information from previous time steps, enabling the network to recognize long-term dependencies in input sequences. The gating mechanisms, consisting of input, output, and forget gates, regulate the flow of information into and out of the memory cell.

An LSTM cell has three gating mechanisms: an input gate, an output gate, and a forget gate, as shown in figure 4.1.

- The input gate determines which information should be stored in the cell.
- The forget gate regulates which information in the cell is maintained or forgotten.
- The output gate governs which information from the cell should be outputted.

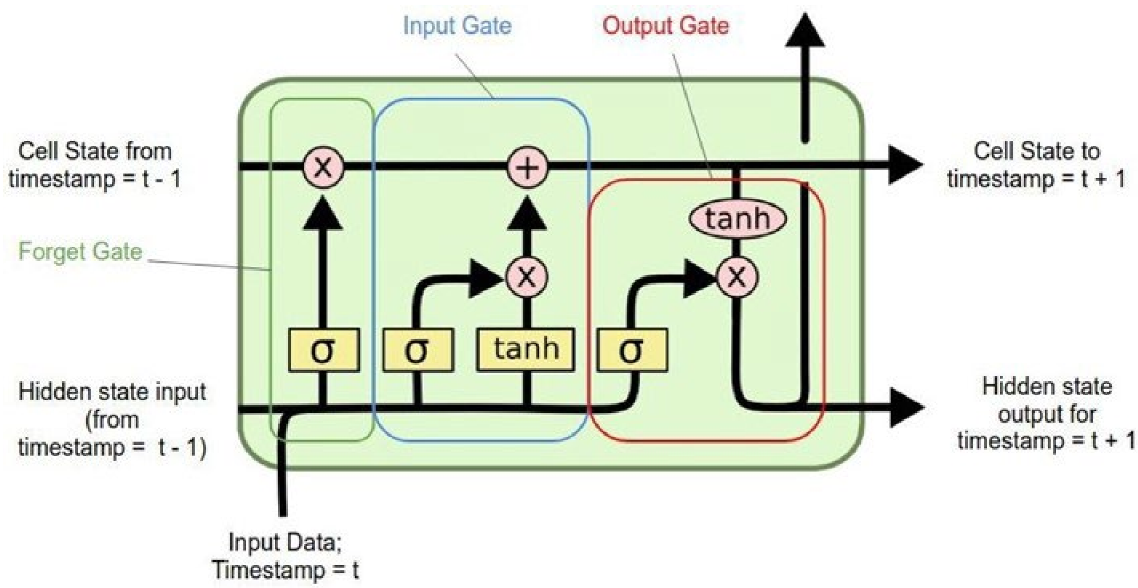


Figure 4.1: Long Short-Term Memory cell

The LSTM network can selectively store and retrieve information as needed by managing the flow of information into and out of the cell, making it well-suited for time-series prediction applications with a high degree of temporal dependence.

LSTM networks are particularly effective in predictive modeling tasks when the input sequences involve long-term dependencies and variable lengths. This is because LSTM networks can handle variable-length sequences and store information over long periods, enabling them to recognize intricate patterns in the input data.

In the context of time-series prediction, LSTM networks have shown promise in capturing intricate interactions between variables and making accurate forecasts. By learning the correlations between input features and their corresponding future values, LSTM networks can predict outcomes based on historical patterns. This is particularly useful in railway maintenance and operation, where variables like track geometry parameters, weather conditions, train speed, and train weight can impact the future state of the track. LSTM networks can effectively capture these dependencies and make accurate predictions, maximizing track maintenance and enhancing railway safety.

4.2 Methodology

4.2.1 Data preparation

Historical track geometry data must be collected and processed to train an LSTM model for predicting track geometry parameters. This data may comprise numerous factors, such as alignment and profile, collected for 24 months. The model initially focused on predicting only the profile data.

After the data has been aligned, preprocessing is required to make it appropriate for training the LSTM model. Scaling and normalizing are often performed during this preprocessing stage. Scaling is used to ensure that all of the input features are on the same scale. The input data becomes more consistent after scaling, and outlier effects are reduced with the help of normalization. Finally, the data can be split into training and testing sets. We split the data into training and test sets to prepare the dataset for training and testing. We selected the first 75% of the data for training and the remaining 25% as the test set. This allowed us to evaluate the LSTM network's performance on data it had not been trained on. The training set is used to train the LSTM model. The testing set evaluates the model's performance on unseen data and measures its accuracy and predictive power. The profile data shown in Figure 4.2 shows the profile for 500 feet for six months. As you can see from the figure, the data was dynamic, highly nonlinear, and contained high noise. This made it a challenging dataset for predictive modeling.

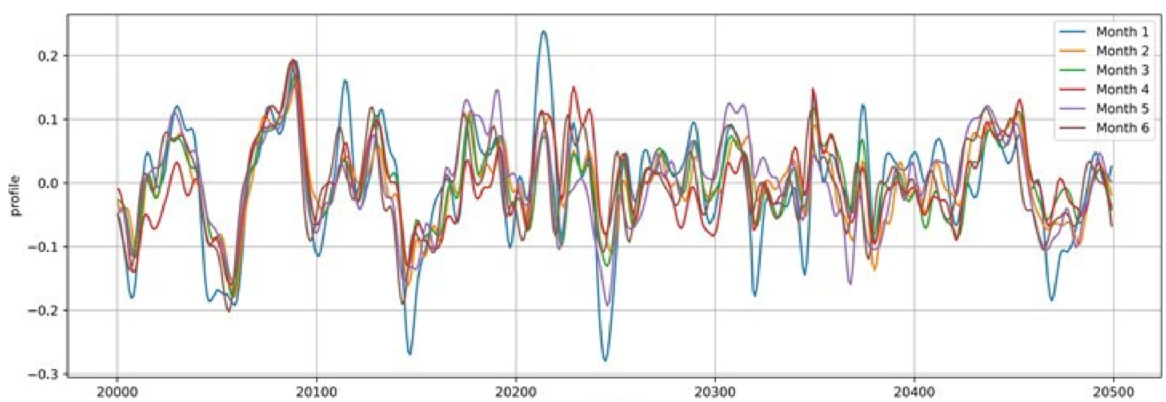


Figure 4.2: The profile for six months from 20000 foot to 20500 foot

In this study, we tested several configurations for the LSTM network, but two showed promising results. The first configuration used six months of input data and predicted the next six months of output data. In comparison, the second configuration used 12 months of input data and predicted six months of output data.

The first configuration, with six months of input data and six months of output data, was designed to test the performance of the LSTM network over a shorter time horizon. This configuration may be more suitable for applications where short-term predictions are needed, such as predicting track geometry data for the next few months. The second configuration, with 12 months of input data and six months of output data, was designed to test the performance of the LSTM network over a longer time horizon. This configuration may be more suitable for applications where longer-term predictions are needed, such as predicting track geometry data for the following year.

The LSTM network was trained using the same training and testing datasets in both configurations. We compared the performance of the two configurations to evaluate the effectiveness of using longer or shorter time horizons for the LSTM network. By testing multiple configurations, we were able to gain insights into the performance of the LSTM network and determine which configuration is more appropriate for different applications. This allowed us to identify the strengths and weaknesses of the LSTM network and optimize its performance for track geometry data prediction.

4.2.2 Model Training

The size of the network layer and the number of neurons in each layer are not always strictly determined when developing neural network models. In the described approach, the parameters of each model layer were determined through multiple experimental adjustments. Specifically, the model consisted of two layers of Bidirectional LSTM followed by a single LSTM layer. The models were supervised and trained using the Adam algorithm, which optimizes a predefined objective function to obtain the optimal model parameters. The modeling for this study was conducted using Python in the Visual Studio Code (VS Code) integrated development environment (IDE). VS Code provides a user-friendly and customizable environment for writing and executing Python code. In the first configuration, each time step consists of 6 months, so the input shape of the network is (6, 1). The output shape of the network is also (6, 1). This means that the network predicts six values for each input for the next time step. In the second configuration, each time step contains 12 months, so the input shape of the network is (12, 1). The output shape of the network remains (6, 1), meaning that for each input time step, the network still predicts six values for the next time step. After training the network, it is evaluated using the test data. Only the data for the current time step must be provided as input to the network during evaluation. The network then predicts the values for the next time step. The goal is to minimize the mean square error (MSE) between the predicted values and the actual observations of the next time step. A smaller MSE indicates better prediction accuracy, which signifies a closer match between the predicted values and the ground truth observations.

The hyperparameters of the models, such as the number of layers and batch size, were determined empirically through experiments. During the training process, we utilized data pairs consisting of the real current time step data and the corresponding next time step data to train the network. Overall, our approach allowed us to tailor the architecture of the neural network to the specific requirements of our study and optimize its performance through experimentation and fine-tuning of the model parameters.

During training, the Adam optimization algorithm was used to optimize a pre- determined objective function and obtain the optimal model parameters. The hyper-parameters, such as the number of layers and batch size, were all obtained through empirical experiments.

4.2.3 Metrics Used for Evaluation

After training the network, we evaluated its performance using the test data. During the evaluation, we provided the data of the current time step to the network, and the network output predicted values for the next time step. We aimed to minimize the root mean square error between the predicted values and the actual observations of the next time step.

When evaluating the performance of our LSTM models, we used two metrics: Absolute Value Error (MAE) and Mean Square Error (MSE). These metrics commonly measure the difference between predicted values and actual observations.

MAE is a simple metric that calculates the average absolute difference between the predicted values and the actual observations. Its calculation formula is as follows:

$$MAE = \frac{1}{N} \sum |y_{pred} - y_{true}|$$

Where y_{pred} is the predicted value, y_{true} is the actual value, and n is the number of data points. MSE is another popular metric considering the squared differences between the predicted and actual values. Its calculation formula is as follows:

$$MSE = \frac{1}{n} \sum (y_{pred} - y_{true})^2$$

where y_{pred} and y_{true} are defined the same as for MAE, and n is the number of data points.

An MSE value of 0 indicates that the predicted values perfectly match the actual values. Lower MAE and MSE values generally indicate better model performance predicting the target variable. By comparing the MAE and MSE values of our LSTM models, we can determine which configuration performs better in predicting the track geometry data.

4.3 Results and Analysis

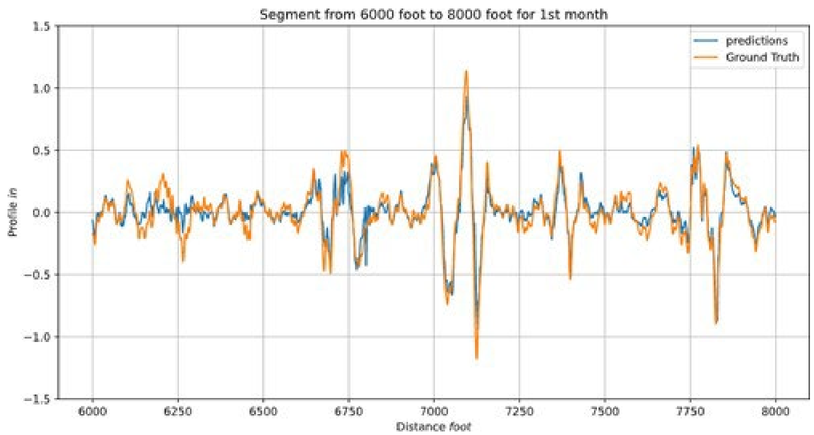
4.3.1 Results

The left profile ground truth (expected) and prediction results for the first and sixth months, for the 6-month and 12-month configurations, are shown in Figures 4.4 and 4.5. The blue line in the figures represents the predictions made by the LSTM network, while the red line represents the actual data. Moreover, Figure 4.3 provides a closer examination of small segments comprising 2000 feet. The purpose of focusing on these smaller segments is to assess the localized effect of the model and gain insights into its performance on a finer scale. By zooming in on smaller segments, it becomes possible to evaluate the model's accuracy in capturing the detailed variations and irregularities present in the track geometry.

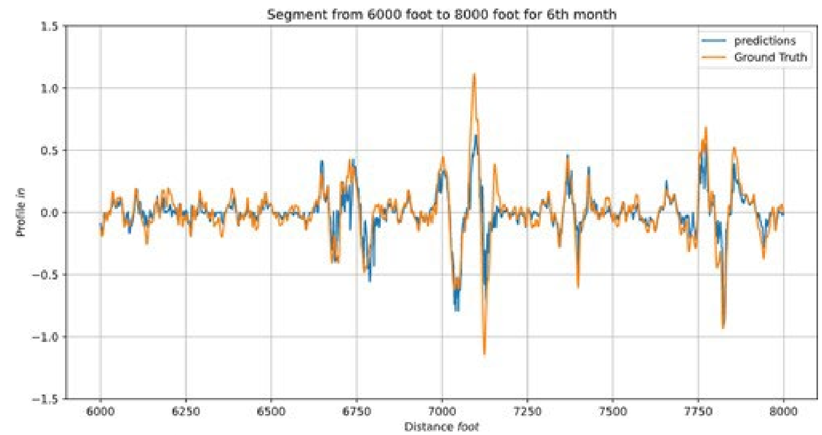
The results show that both LSTM configurations performed well in predicting the left profile of track geometry. However, the 12-month configuration outperformed the 6-month configuration. This is due to the LSTM network's ability to accurately describe and capture the dynamic nature of time series data, which is difficult to predict using traditional statistical methods, enhancing its prediction accuracy. The increased input length from 6 to 12 months provides more information to the LSTM network, allowing it to identify hidden correlations and learn long-term dependencies between time steps. Additionally, supervised learning enables it to learn from the past and predict the future.

Figure 4.3 shows 2000 feet segments of the track profile, which provides information into the successes of the LSTM model in accurately predicting track geometry. This segment showcases the alignment between the predicted profile and the real pro- file, indicating the model's ability to capture the underlying trends and fluctuations in the data. The predicted profile closely follows the ups and downs of the actual profile, capturing the overall trend and fine-grained fluctuations present in the data. The model successfully identifies the subtle changes in the track geometry, accurately forecasting the corresponding profile values. This alignment between the predicted and actual profiles suggests a high level of accuracy and reliability in the LSTM model's predictions.

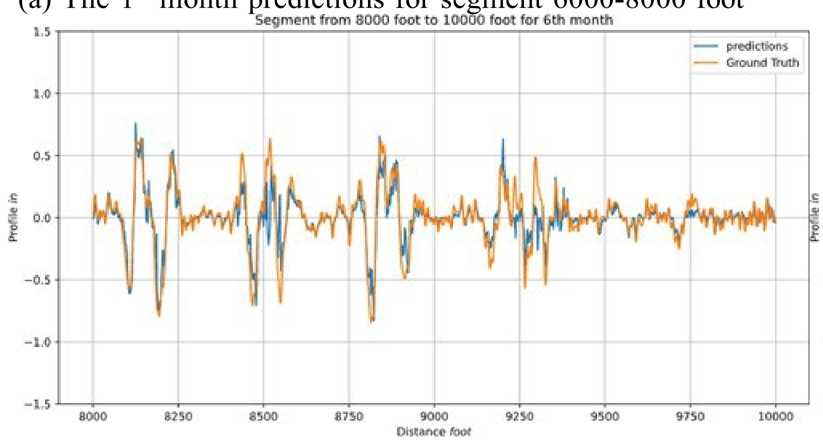
Accurate Representation of Trends and Fluctuations The LSTM model effectively captures the trends and fluctuations in the track geometry data within the examined segments. It accurately predicts both the upward and downward movements, mirroring the real profile changes. This level of accuracy is crucial for maintaining the integrity and safety of railway operations, as it ensures the proper assessment and management of track geometry deviations. While acknowledging the areas of underperformance, it is crucial to recognize the successes achieved by the LSTM model in accurately predicting and aligning with the real profile within specific segments. These successes demonstrate the model's capacity to capture trends and fluctuations in track geometry data, ultimately contributing to safe and efficient railway operations. Building upon these achievements, further advancements can be made to enhance the model's performance and address the remaining challenges in profile prediction.



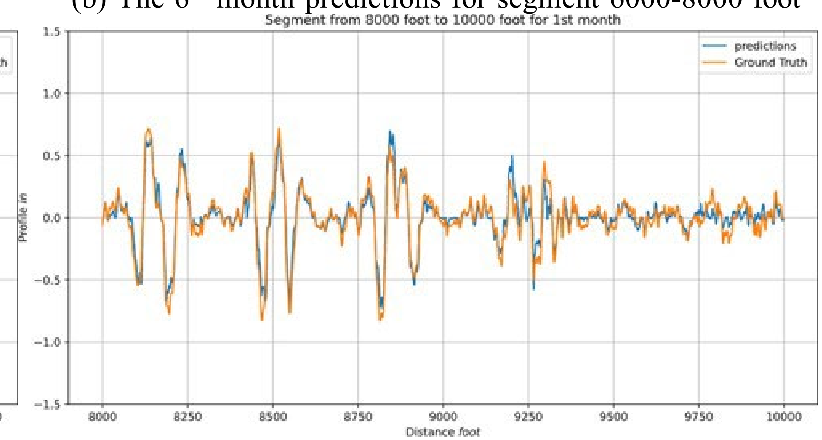
(a) The 1^{st} month predictions for segment 6000-8000 foot



(b) The 6^{th} month predictions for segment 6000-8000 foot

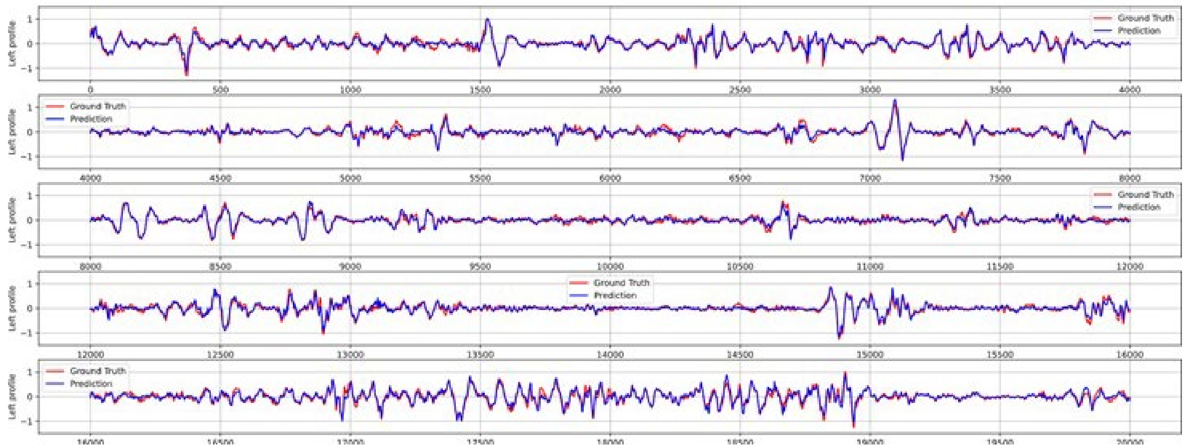


(c) The 1^{st} month predictions for segment 8000-1000 foot

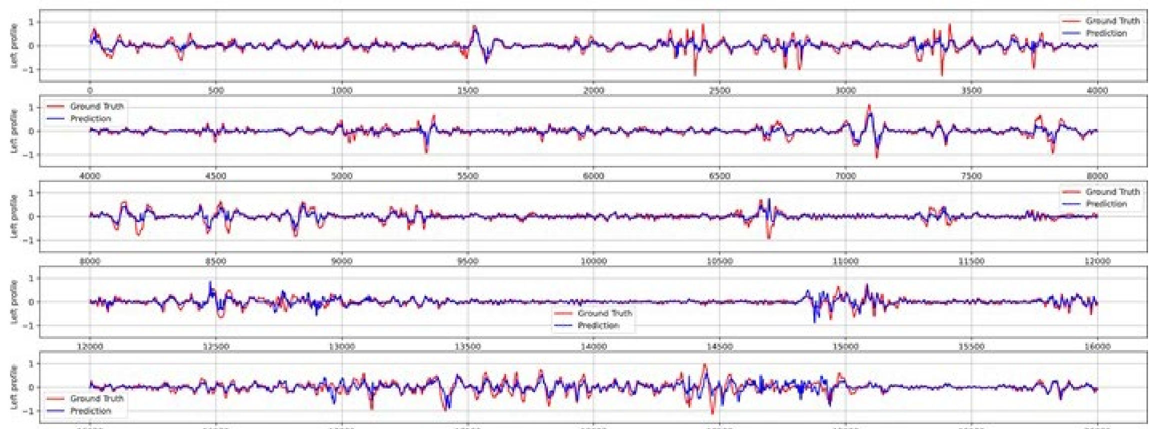


(d) The 6^{th} month predictions for segment 8000-1000 foot

Figure 4.3: The profile Prediction for segments 6000-1000

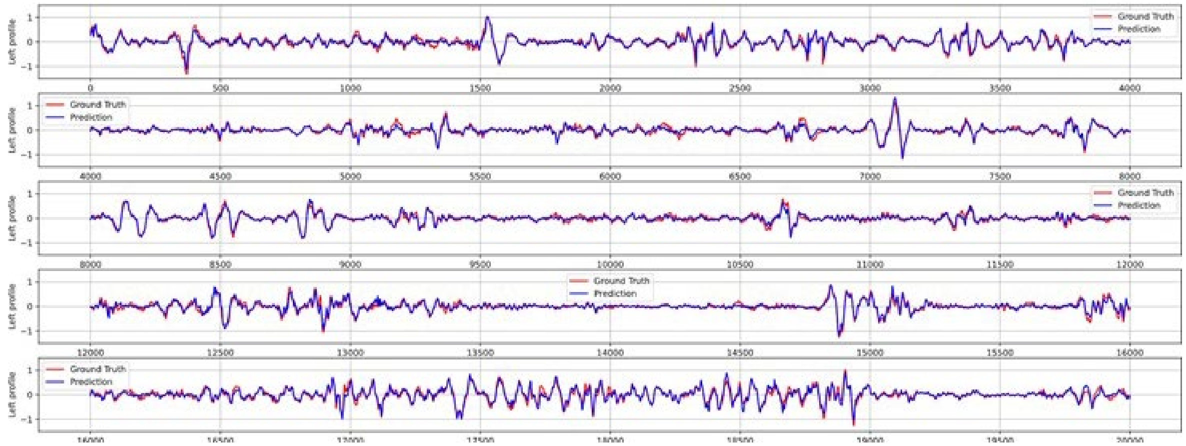


(a) The 1^{st} month predictions

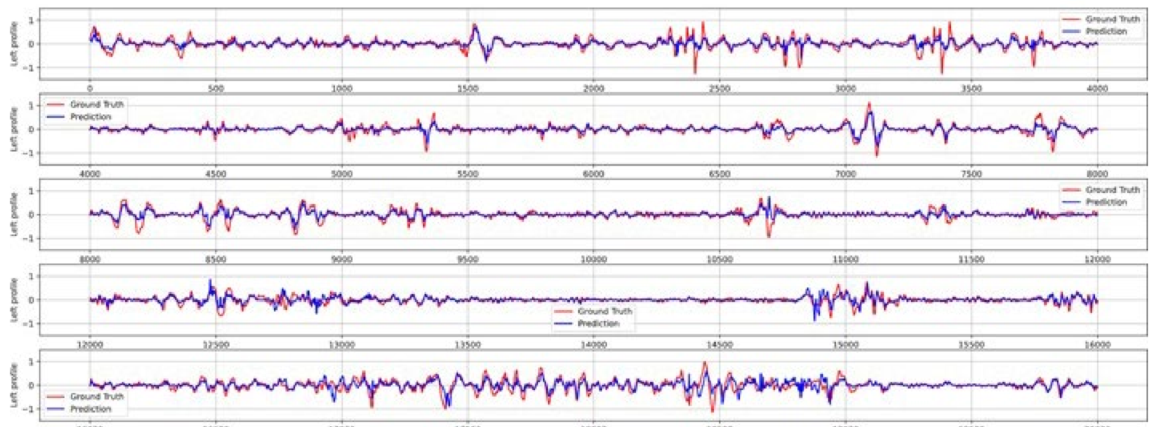


(b) The 6^{th} month prediction

Figure 4.4: The 6 month LSTM configuration prediction



(a) The 1^{st} month predictions



(b) The 6^{th} month prediction

Figure 4.5 The 12 month LSTM configuration Prediction

4.3.2 Analysis

Figure 4.6 and Table 4.1 show the mean square error (MSE) and absolute mean error (MAE) for each month, providing information on the accuracy of our model's predictions over time. The figure clearly illustrates an increasing trend in both MSE and MAE as the duration of the predictions progresses. This observation suggests that the model's accuracy diminishes over the study period. In the first month, the MSE is remarkably low at 0.0073 in^2, indicating a high level of accuracy. Similarly, the MAE for the same month is measured at 0.0542 *in*, further supporting the model's accurate predictions. These low error values signify a close alignment between the model's projected values and the actual data point.

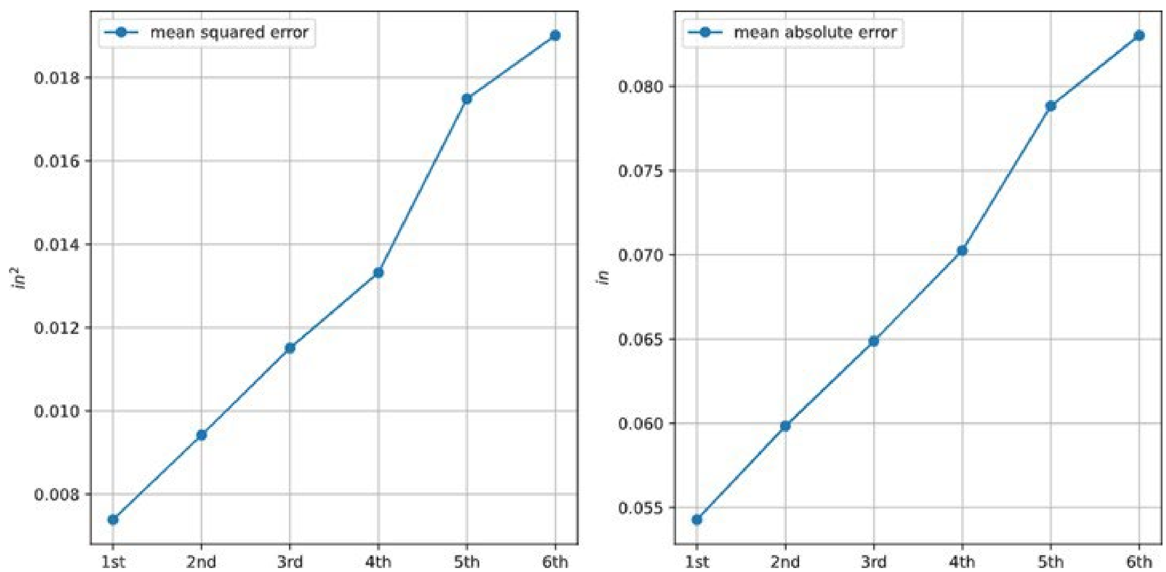


Figure 4.6: The mean square and absolute mean errors

The MSE increases to 0.0190 in^2, while the MAE reaches 0.0830 *in* in the sixth month, representing a 53% increase in error compared to the first month. Although these values indicate a decrease in accuracy, it is important to provide further con- text. To assess the accuracy of these errors, it is valuable to compare them to relevant benchmarks. For example, the profile defects limit set by the Federal Railroad Ad- ministration (FRA 2023) for class 6 is 1.0 in. In the sixth month, the MAE of 0.083 *in* corresponds to 8.3% of the profile defect threshold. This comparison helps us understand that the observed errors remain within an acceptable range and do not deviate significantly from the expected values.

The overall trend depicted in Figure 4.6 underscores the importance of continuously tracking and evaluating the model's performance over time. The increasing errors may indicate potential limitations or changes in the underlying data patterns that the model finds challenging to capture accurately. To ensure the model's prediction capabilities, it is crucial to consider the dynamic nature of the data and regularly assess and improve the model.

Additionally, the analysis of short segments in Figure 4.3 provides further in- sights into the model's performance. By examining the effect of the model on smaller segments of 2000 feet, we can gain a deeper understanding of its strengths and limitations in capturing track irregularities. This analysis can help identify specific areas where the model excels or struggles, enabling targeted improvements and refinements in the predictive capabilities.

Table 4.1: The mean square and absolute mean errors

	mse	mae
1	0.0073	0.0542
2	0.0094	0.0598
3	0.0115	0.0648
4	0.0133	0.0702
5	0.0174	0.0788
6	0.0190	0.0830

4.3.3 Segment-Specific Underperformance Analysis

We examined the predicted profile and compared it with the actual profile to evaluate the performance of the LSTM model. Upon closer examination, we identified some areas where the model underperformed. These segments exhibited notable changes in profile, characterized by both gradual improvements and sudden increases. The complex nature of these variations posed challenges for the LSTM model in accurately capturing the changes.

Segment A, in figure 4.7. Spanning from 2350 feet to 2450 feet, it displayed relatively small rate changes in the first three months. However, starting with the fourth month, the rate of change increased dramatically, with each successive month roughly doubling the preceding month's change. This rapid and substantial variation in the profile presented considerable difficulty for the LSTM model to accurately predict the profile for this segment.

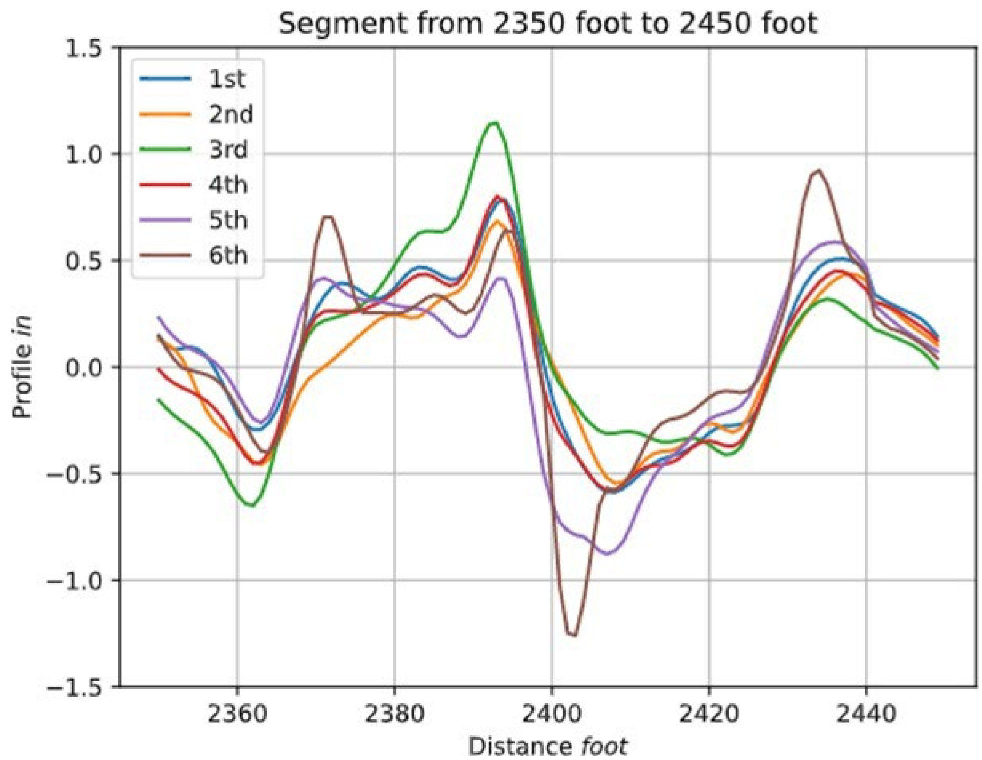


Figure 4.7: Segment A

Comparably, Segment 2, in figure 4.8, which spans 3350 to 3450 feet, showed very moderate rate fluctuations over the first four months. However, the rate of change significantly increased in the fifth and sixth months, with the monthly change approximately doubling compared to the prior months.

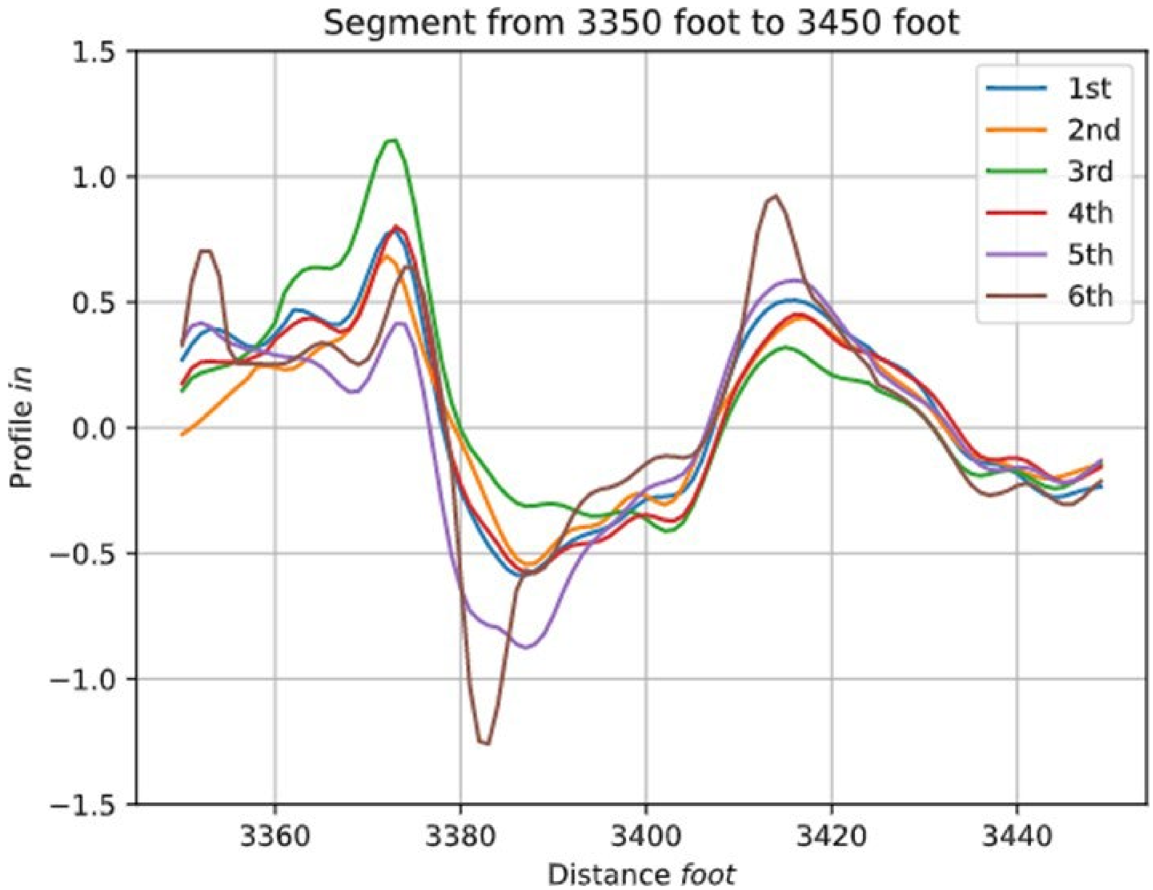
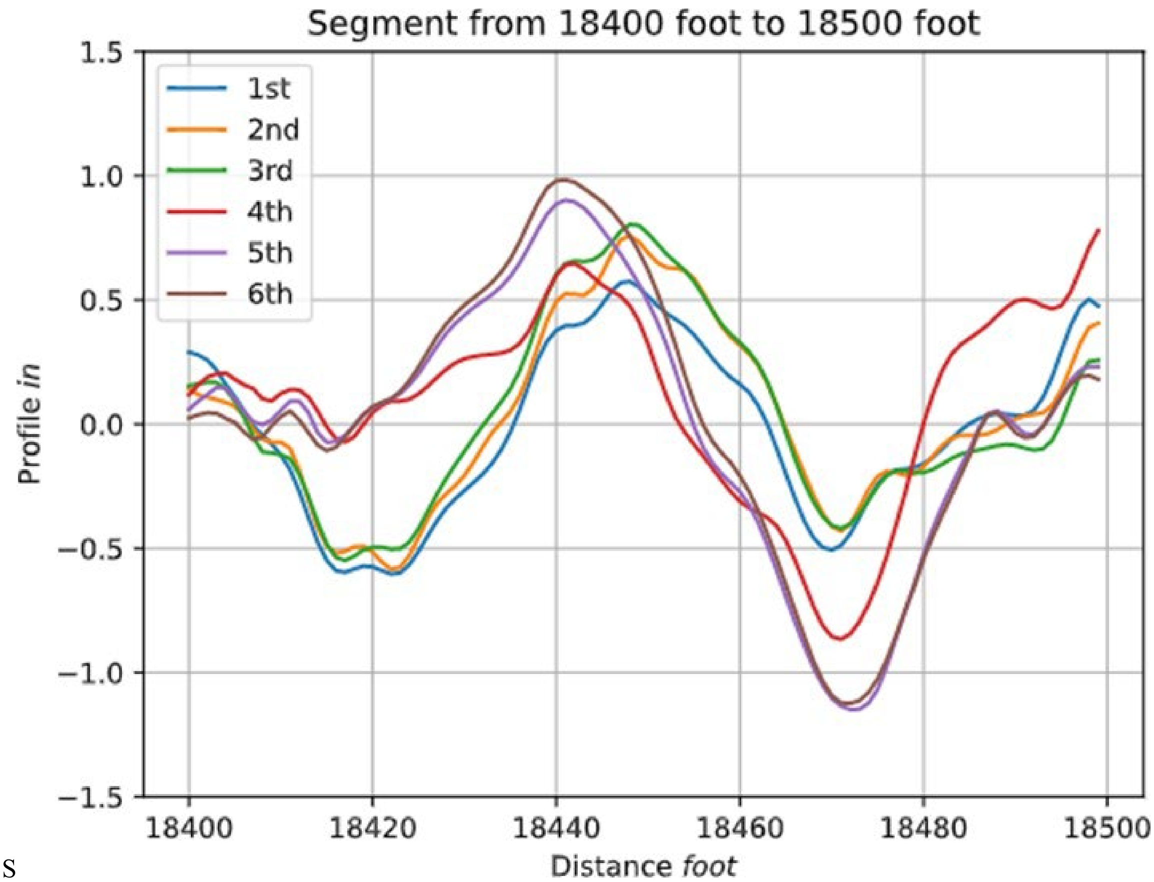


Figure 4.8: Segment B

Moving on to Segment C in Figure 4.9, which is in the range of 18400 to 18500 feet, we can see that the rate variations seen over the first three months were relatively consistent. The rate of change did, however, rise substantially starting in the fourth month. The fourth, fifth, and sixth months experienced a considerable acceleration in the rate of change, with each subsequent month nearly doubling the change compared to the previous month.



S

Figure 4.9: Segment C

These observations and identified segments imply that the LSTM model encountered challenges in predicting the track geometric data accurately when faced with sudden and substantial variations in the profile data. The model's ability to capture and adapt to these rapid and substantial changes is hindered, leading to its underperformance in these specific segments.

To address these challenges, further research and improvements could be made to improve the LSTM model's ability to handle sudden and severe changes in track geometry data. These improvements could include refining the model architecture, incorporating additional features or contextual information, or external factors that may contribute to the sudden changes in the profile.

While the LSTM model performed well in predicting track geometry data, it had problems capturing and forecasting rapid and major alterations in profile data in some sections. These results emphasize the significance of comprehending the model's advantages and disadvantages and offer suggestions for potential future improvements and adjustments to raise the model's performance in difficult situations.

4.4 Development of Multivariable LSTM for Track Geometric Data Pre- diction

This section focuses on developing a multivariable LSTM (Long Short-Term Memory) model to predict and track geometric data. The goal is to leverage the capabilities of multivariable LSTM modeling to predict the four key aspects of the space curve: left profile space curve, right profile space curve, left alignment space curve, and right alignment space curve. By considering these factors collectively, we want to capture their complex relationships and dependencies, allowing us to understand the overall track situation thoroughly. By developing a robust multivariable LSTM model, we can improve the accuracy of track geometry predictions, allowing railway operators to proactively address potential issues and maintain high levels of safety and reliability. This section will delve into the multivariable LSTM model development process, covering the data selection and preparation methods and addressing the model architecture. Through this analysis, we aim to gain insights into the strengths and limitations of the developed model in capturing the complexities of track geometry variations.

4.4.1 Data Selection and Preprocessing

We carefully selected and preprocessed the necessary data inputs to develop the multivariable LSTM model for track geometric data prediction. The dataset used for developing the model contains four fundamental components: left profile space curve, right profile space curve, left alignment space curve, and right alignment space curve. Together, these components give a complete picture of the track geometry and enable the prediction of other measurement channels of the track. The selected data encompasses 12 months, representing a sufficient historical period to capture relevant patterns and trends in track geometry variations. This duration allows the model to learn from past data and make accurate predictions for the subsequent six months. To ensure compatibility with the LSTM model, several preprocessing steps were applied to the data.

One of the key preprocessing techniques employed was normalization or feature scaling. This is because these components do not have the same range. By scaling the data, we bring all the input variables to a common scale, which prevents any variable from dominating the learning process and helps in achieving stable and consistent training and allows the model to converge effectively. By selecting relevant variables and applying appropriate preprocessing techniques, we ensured that the model receives ac- curate and consistent inputs, enabling it to learn and predict track geometry variations effectively.

4.4.2 Model Architecture

The developed multivariable LSTM architecture is an effective tool for predicting track geometry data by leveraging the temporal dependencies and relationships between multiple input features to make accurate forecasts. This section overviews the model architecture and highlights the key design choices to optimize its performance. The multivariable LSTM model's input structure is designed to support the four fundamental space curve channels. Each feature is treated as a separate input, allowing the model to capture each track geometry component's unique properties and variations. This enables a comprehensive analysis of the spatial and alignment aspects and leads to more accurate predictions. In terms of temporal information, the model considers a 12-month sequence length for historical data, and the model uses the last 12 months of data to forecast the upcoming six months. The model can capture long-term trends and patterns in track geometry data by incorporating this comprehensive temporal context, resulting in more accurate predictions.

The LSTM model consists of multiple layers, each contributing to the learning and prediction process. In the initial layers, the input data is combined and analyzed collectively to extract relevant features and capture the overall patterns in the data. This initial analysis helps the model to identify common trends and dependencies shared across different features before proceeding to the feature-specific layers. Following the initial layers, the output of the collective analysis is fed into four separate layers, each dedicated to a specific feature (left profile space curve, right profile space curve, left alignment space curve, and right alignment space curve). These layers allow the model to focus on the unique characteristics and variations of each feature, further enhancing the model's ability to capture the complexities of the track geometry data. The model Architecture is shown in figure 4.10.

The layers in the LSTM model are equipped with adjustable parameters, such as the number of hidden units, dropout rates, and activation functions, which were optimized during the model training process. The specific values of these parameters may vary depending on the complexity of the track geometry data and the desired trade-off between model complexity and generalization capability. The model architecture was designed and fine-tuned through an iterative process, considering the performance metrics such as mean square error (MSE) and mean absolute error (MAE) on a validation dataset. By continuously evaluating and refining the model architecture, we aimed to create an optimal configuration that effectively captures the temporal dependencies and accurately predicts the track geometric data.

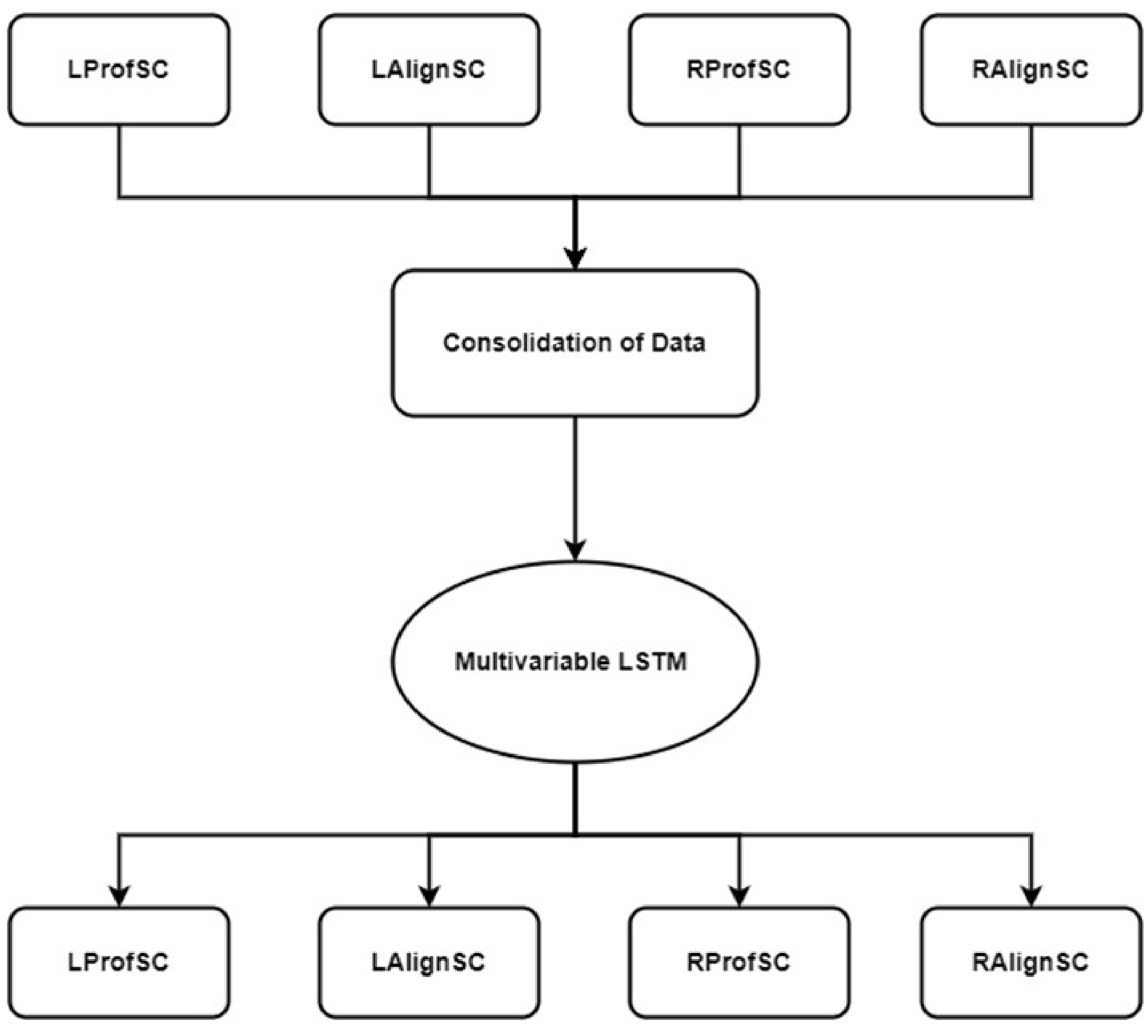


Figure 4.10: Multivariable LSTM

Overall, the multivariable LSTM model architecture used in this study leverages the strengths of LSTM networks in handling sequential data and incorporates specific design choices to accommodate the unique characteristics of track geometric data. By considering each feature separately, capturing long-term dependencies, and optimizing the model parameters, we aimed to develop a robust and accurate prediction model for the track geometric data.

4.4.3 Training and Validation

The multivariable LSTM model was trained using a carefully designed process to ensure its effectiveness in predicting track geometric data. This section outlines the key steps involved in training the model and evaluating its performance through validation. The dataset was divided into training and validation sets to assess the model's ability to generalize to unseen data. A common approach is to use a certain percentage of the data for training (e.g., 80%) and the remaining percentage for validation (e.g., 20%). This division allows the model to learn patterns and relationships from the training data and evaluate its performance on data that it has not been exposed to during training. An optimization algorithm was used during the training process to iteratively change the model's parameters and reduce the prediction error. Stochastic gradient descent (SGD) or one of its derivatives, like Adam, is a frequently used method. These optimization techniques update the weights and biases of the model using gradient information to identify the ideal setting that minimizes the loss function. Hyper- parameters associated with the optimization algorithm, such as the learning rate and momentum, were carefully tuned to ensure efficient convergence and prevent overfitting or underfitting. The learning rate determines the step size at each iteration, while the momentum controls the contribution of the previous parameter updates. The hyper- parameter tuning process involved experimentation and validation to find the optimal values for the specific track geometric data.

To quantify the accuracy of the model's predictions during training, a loss function was employed. The choice of the loss function depends on the specific task and data characteristics. For track geometric data prediction, common loss functions include mean squared error (MSE) and mean absolute error (MAE). The MSE measures the average squared difference between the predicted and actual values, while the MAE computes the average absolute difference. These loss functions allow the model to optimize its parameters by minimizing the discrepancy between predicted and actual track geometric data.

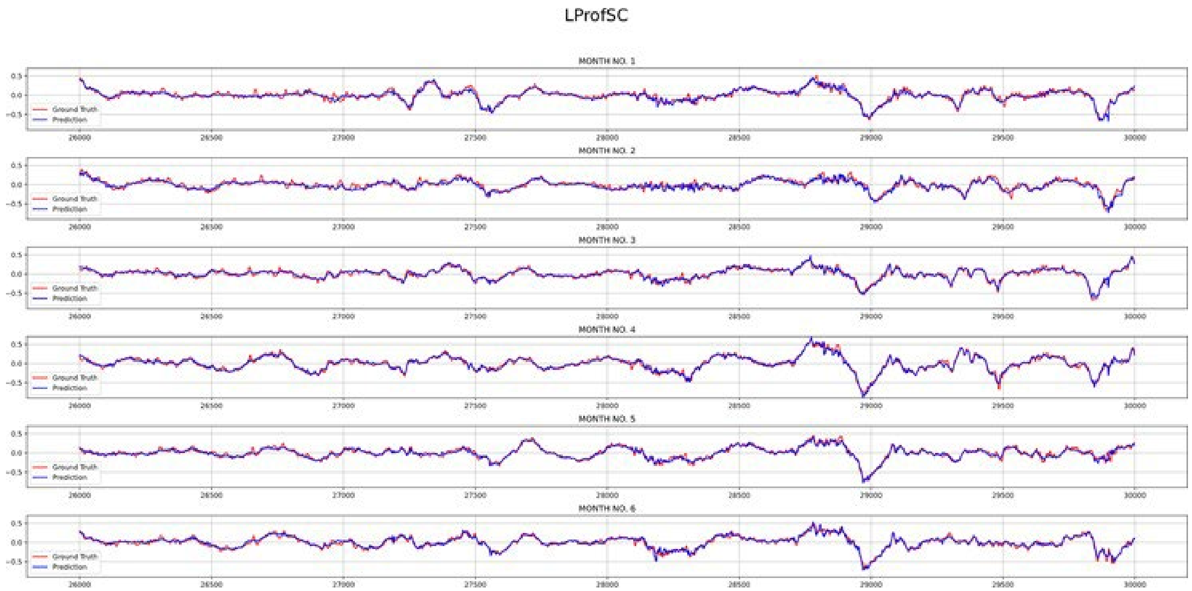
4.4.4 Results and Analysis

The performance of the multivariable LSTM model for the prediction of track geometric data was evaluated, and the results are presented in this section. The model's accuracy, precision, and overall predictive capability were assessed to determine its effectiveness in capturing the complexities of track geometry variations. In Figures 4.11a,4.11b,4.12a, and 4.12b 6 and 7, the multivariable LSTM model's performance results for the prediction of space curve data are shown. Over a 6-month period, these figures represent the ground truth (expected) and predicted outcomes for the left profile space curve, right profile space curve left alignment space curve, and right alignment space curve. The blue line represents the predictions provided by the multivariable LSTM network, while the red line represents the actual data. The projected curve closely follows the trend of the real data, albeit with a smoother shape. The actual data, on the other hand, shows more pronounced fluctuations. This difference can be attributed to the model's ability to capture the underlying trends and patterns in the data while smoothing out some of the inherent noise and variability.

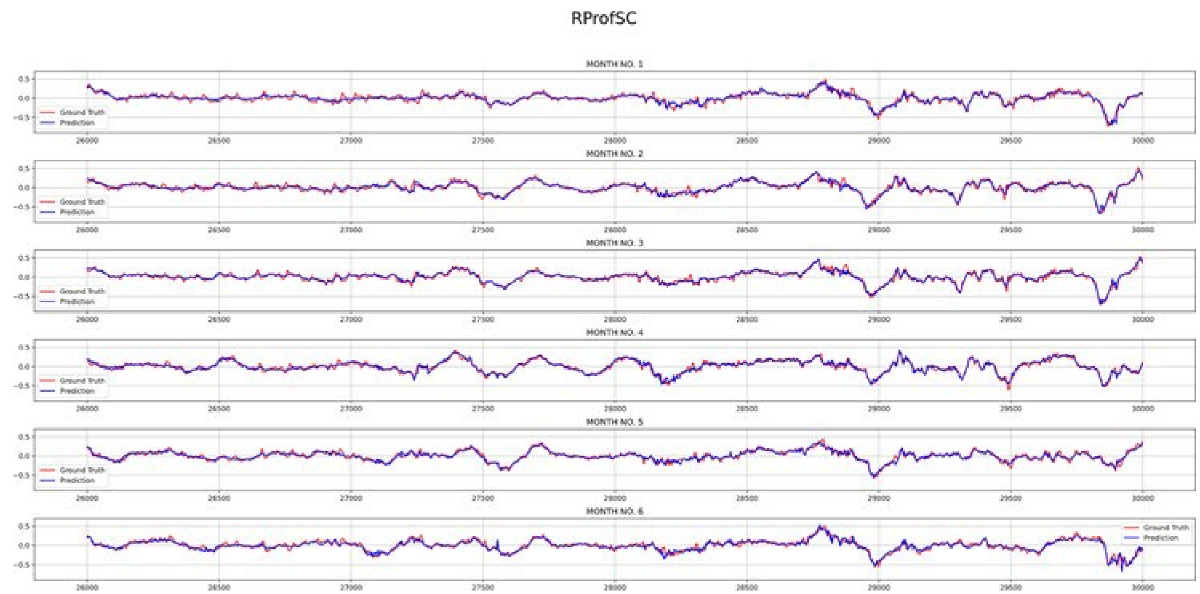
Overall, the multivariable LSTM model performed well in predicting space curve data and successfully captured the general trends and variances, which aligned with the expected patterns. The predictions were especially accurate in capturing the overall trend, which is critical for optimizing maintenance planning and ensuring safe railway operations. However, the model's performance varied across different channels. This discrepancy could be attributed to the fact that each channel was trained separately in the second part of the mode, resulting in predictions accuracy and precision were not constant across all channels, which suggests that certain features or characteristics of the track geometry may be more challenging to capture accurately compared to others.

Certain months outperformed others, showing the presence of specific problems or complexities in the data that influenced the model's performance. This variability in accuracy across the 6-month prediction period could be related to environmental changes, maintenance efforts, or other external variables that influence track geometry variances. When examining the individual channels and MSE and MAE, which are shown in Figures 4.11 and 4.12, it was observed that the left profile space curve (LProfSC) and left alignment space curve (LAlignSC) consistently outperformed the right profile space curve (RProfSC) and right alignment space curve (RAlignSC). These channels exhibited higher accuracy and more reliable predictions compared to their counterparts. Additionally, it is important to highlight the worst-case scenario, which occurred in the fifth month for the right alignment space curve (RAlignSC) channel, which exhibited a higher level of error compared to other months and channels.

In summary, the multivariable LSTM model predicted track geometry data with a fair level of accuracy. Inconsistencies in accuracy were noted across channels and months, with the left profile space curve and left alignment space curve channels per- forming better overall. The mean square error suggested that there was still some error in the forecasts, albeit a tiny amount. The discovery of the worst-case scenario for the right alignment space curve channel in the fifth month underscores the need for additional analysis and improvement in addressing the issues unique to this channel.

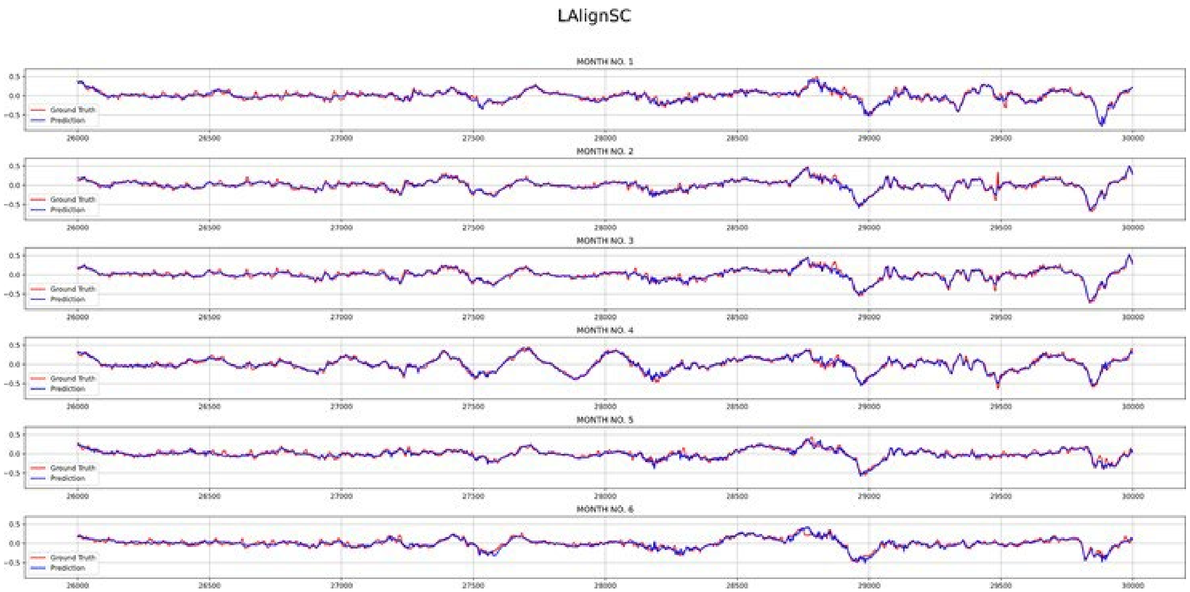


(a) Left profile space curve

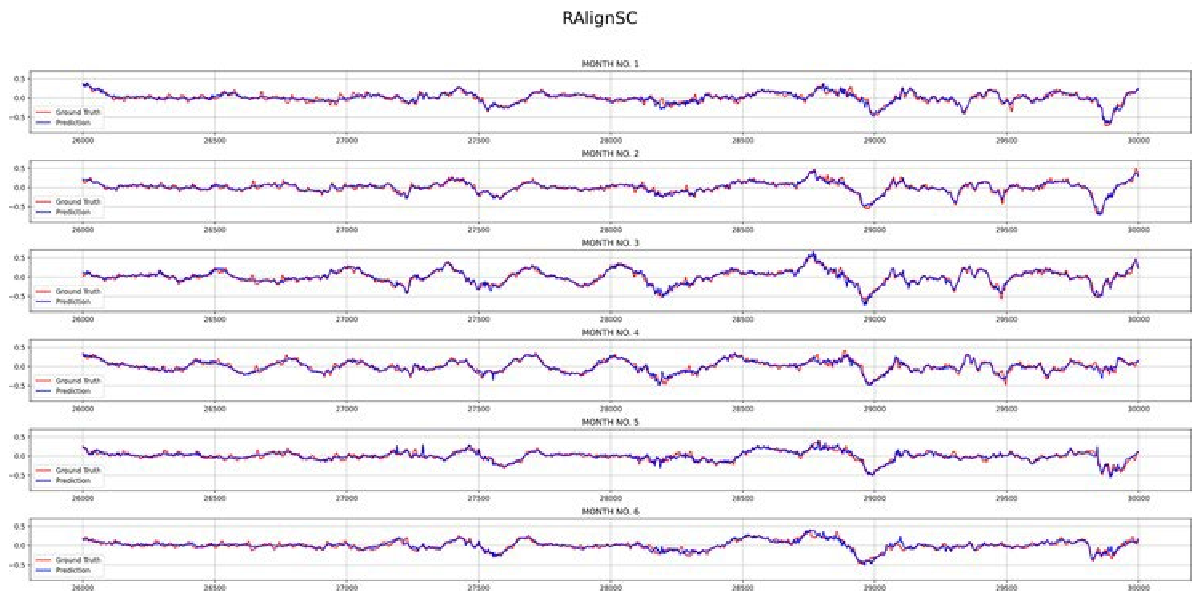


(b) Right profile space curve

Figure 4.11: Left & Right profile space curve

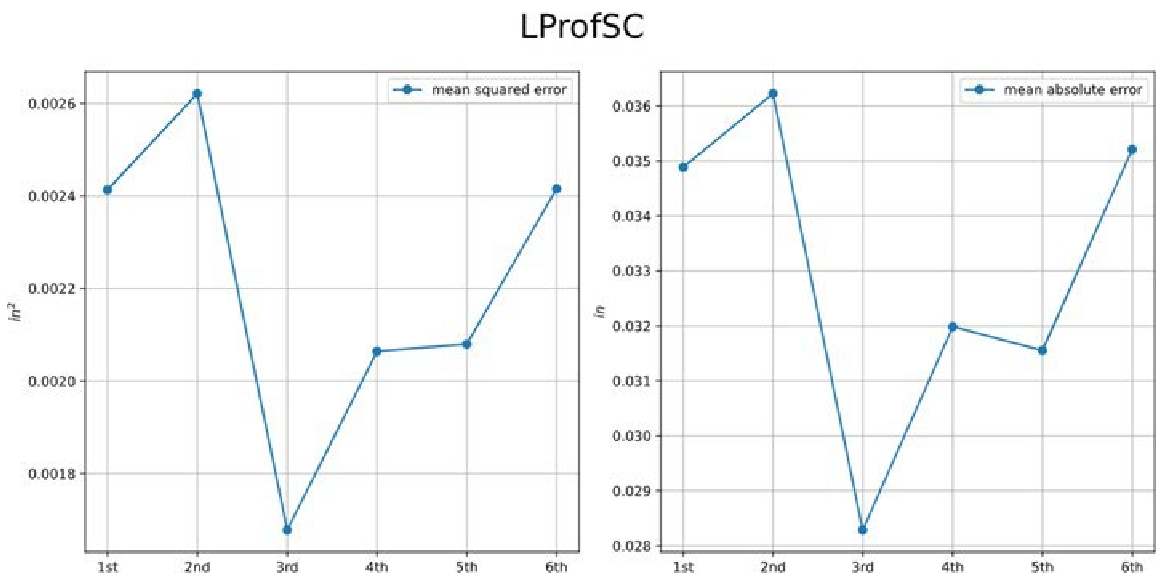


(a) Left alignment space curve.

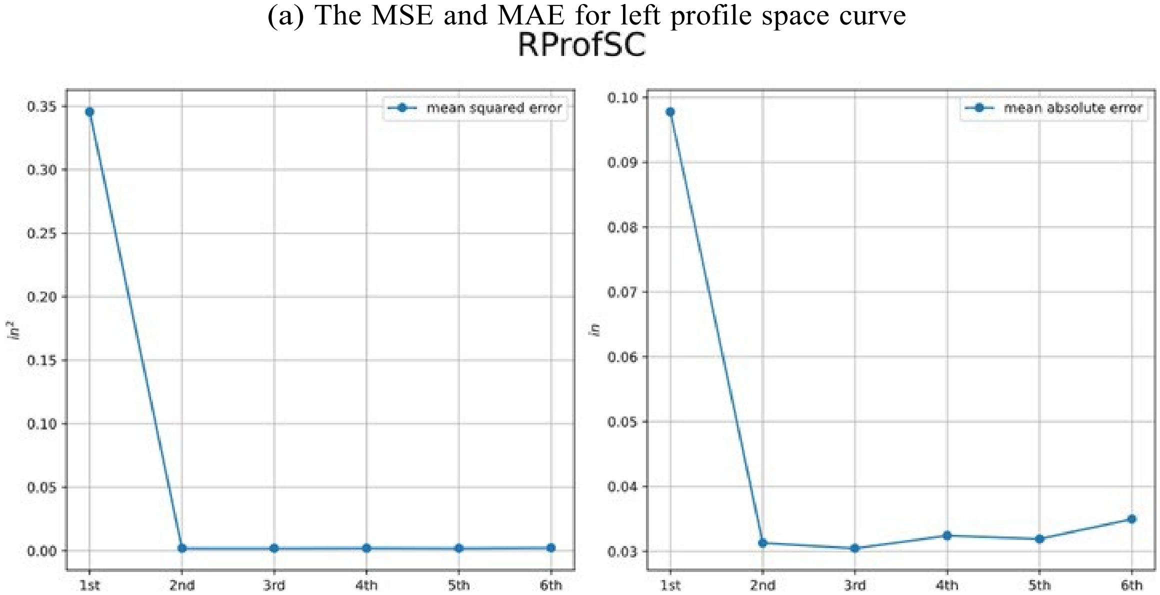


(b) Right alignment space curve

Figure 4.12: Left & Right alignment space curve

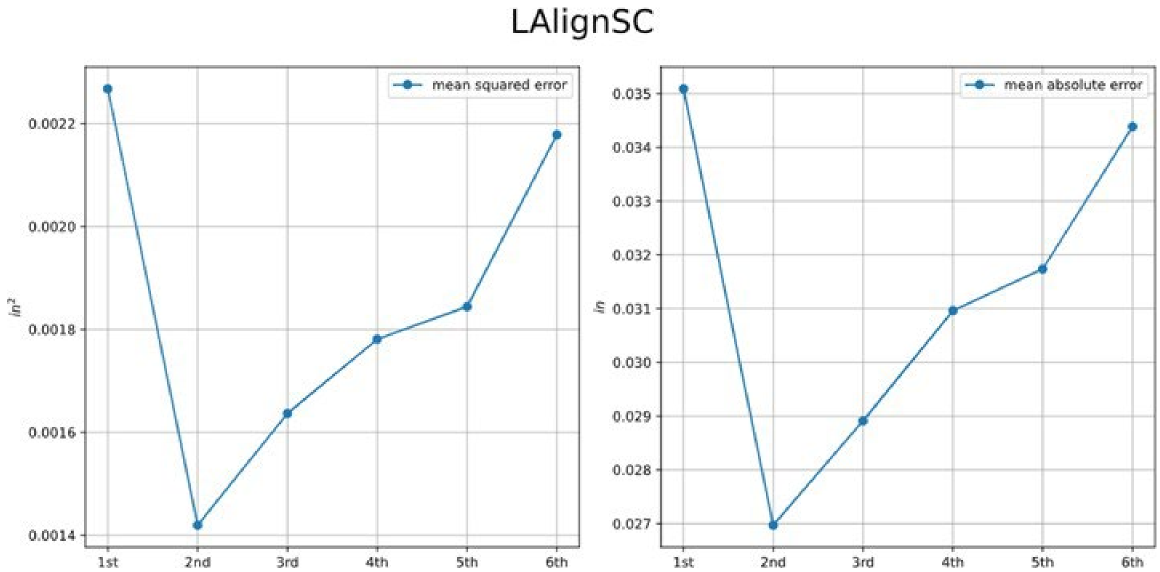


(a) The MSE and MAE for left profile space curve

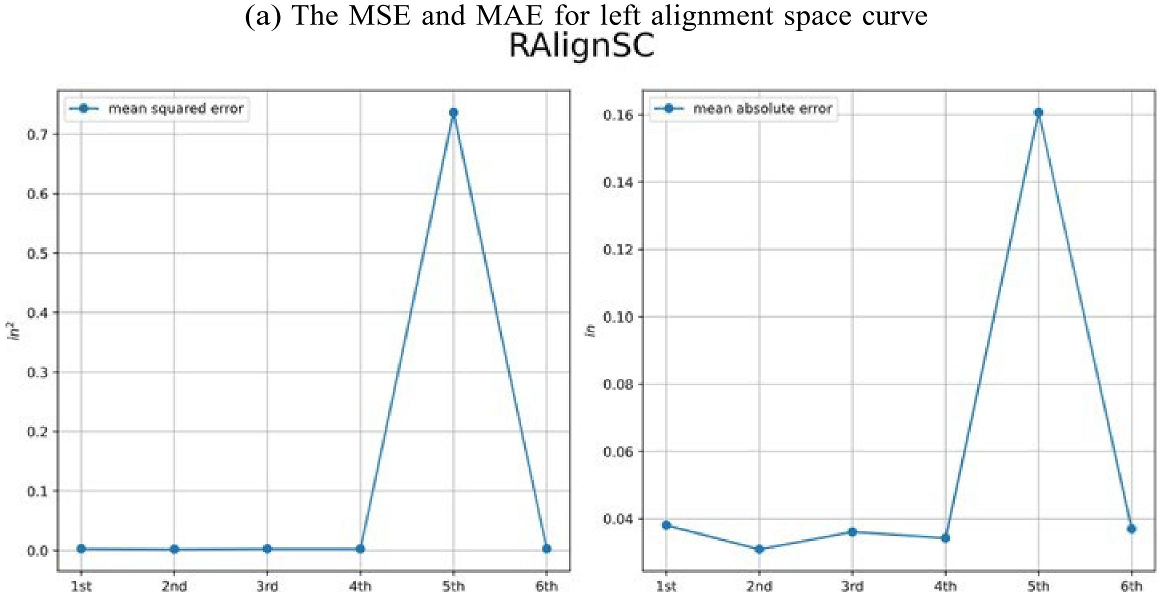


(b) The MSE and MAE for right profile space curve

Figure 4.13: The MSE and MAE for Left & Right profile space curve.



(a) The MSE and MAE for left alignment space curve



(b) The MSE and MAE for right alignment space curve

Figure 4.14: The MSE and MAE for Left & Right alignment space curve

4.4.5 Importance of Space Curve Data

Generally, Traditionally, geometry deviations were measured manually using hand-held devices. To produce a straight-line reference, one conventional approach entailed pulling a string while keeping it close to the rail. The distance between this thread and the rail was then measured. However, in modern practices, automated and continuous measuring systems, such as track geometry cars, are commonly used to obtain accurate and efficient measurements. This is the reason for the space curve data is essential from space. The curve, we can calculate other deviations such as profile, alignment, cross-level, etc.

The mid-chord offset (MCO) is a term used to describe the midpoint offset from a chord when measuring geometry errors along that chord. Specific chord lengths are used in the United States based on the kind of traffic. Chord lengths of 31 ft and 62 ft are often used for freight and passenger, respectively, although a more considerable chord length of 124 ft is used for higher-speed passenger trains. Notably, the 62-ft chord was first chosen because of its practical property: the MCO matches the degree of bend in inches. The 31-foot chord was designed to monitor track characteristics over a shorter distance, but the 124-foot chord was designed to define geometry problems over a longer wavelength for high-speed passenger services.

It's essential to recognize that the series of MCO values obtained do not fully represent the true rail profile or alignment shape. The chords act as mechanical filters, distorting the view of the rail shape. This distortion arises from the fact that the chord length influences the offset measurement, and the ends of the chord are not on a level datum but instead traverse rough track conditions.

We can obtain all these measurements with different chord lengths and purposes from the space curve data. Space curve measurements play a vital role in railway track engineering. By quantifying geometry deviations from the space curve data using reference chords, engineers can assess and monitor the condition of the tracks. They lead to improved safety, reliability, and performance of railway systems. These measurements are a foundation for track maintenance, design optimization, and long- term track management strategies.

Chapter 5 CONCLUDING REMARKS

In this chapter, we will revisit the main research question and objectives ad- dressed in this thesis and highlight the significance of predicting track geometry data for effective maintenance planning activities in railway systems. Our goal throughout this research was to develop machine learning models that could reliably predict the geometric properties of railway tracks, allowing for preventative maintenance planning and assuring the safety and efficiency of railway operations.

The primary research question driving this thesis was:” How can machine learning methods be leveraged to predict and track geometric data and assist in planning maintenance activities for railway systems?” To address the research question, we set the following objectives:

- Investigate the limitations of existing data science models: The first objective of this research was to explore the shortcomings of traditional data science models in accurately predicting track geometric data. We recognized that complex physical processes involved in railway systems require the integration of physics-based models with machine-learning techniques.

- Develop machine learning models for predicting track geometric data: Our second objective was to develop novel machine learning models that could effectively forecast the geometric characteristics of railway tracks. By leveraging the power of data-driven models, we aimed to improve the accuracy and efficiency of track geometry prediction.

- Improve maintenance planning activities: The third goal of this study was to show the importance of forecasting track geometry data for maintenance planning operations. Maintenance personnel may proactively identify maintenance needs, optimize maintenance schedules and assure the safe and reliable operation of the rail network by properly anticipating the future conditions of railway tracks.

The accurate prediction of track geometric data is important for planning maintenance activities in railway systems. Here are some key reasons why this research area is significant:

- Proactive Maintenance: By forecasting track geometry data, maintenance personnel can take a proactive rather than reactive approach to maintenance. Early detection of possible problems enables timely maintenance interventions, avoiding costly repairs and minimizing train service delays.

- Improving Safety: Railway safety is of the utmost importance. Predicting track geometry data allows for the identification of potential safety issues such as track alignment, cross-level, gauge, or twist abnormalities. Maintenance interventions performed on time and based on accurate estimates can decrease safety issues and reduce the likelihood of accidents.

- Optimal Resource Allocation: Efficient maintenance planning necessitates optimal resource allocation, which includes manpower, equipment, and supplies. Accurate track geometry projections enable maintenance teams to precisely deploy resources, optimizing usage and eliminating wasteful costs.

- Cost Reduction: Predictive maintenance based on precise track geometry data estimates can greatly lower maintenance expenses. Maintenance interventions can be carried out in a systematic and cost-effective manner by addressing possible issues before they increase, preventing emergency repairs or costly system failures.

- Increased Operational Efficiency: Well-kept tracks contribute to increased operational efficiency. Accurate maintenance projections allow railways to organize their maintenance schedules in such a way that disruptions to train services are minimized, ensuring smooth operations and customer satisfaction are maintained.

5.1 Research Summary

The research conducted in this study involved employing various methodologies across different chapters to investigate and predict railway geometric data. Chapter 3 (Data and Exploratory Data Analysis) focused on the data collection and preliminary exploration of the track geometry data. The data source comprised 24 monthly inspection runs for a specific location, with various geometric characteristics of the track. The data were consolidated from multiple geometry files into a single dataset, to be analyzed. The Exploratory Data Analysis aimed to understand the data properties and unveil any patterns or relationships present within the dataset. Visualizations, such as histograms, box plots, scatter plots, and correlation plots, were employed to identify outliers, patterns, and relationships among the variables. Descriptive statis- tics were calculated to further characterize the data, including measures such as mean, median, standard deviation, quantiles, and interquartile range. These statistics quantified different features of the data and provided insights into its spread and central tendencies. To visualize the correlations between variables, correlation plots were employed. These plots allow for the understanding of associations and the strength of relationships within the dataset. Box and whisker plots were used to summarize and visualize the distribution, central tendency, and variability of the data. Furthermore, Quantile-Quantile (QQ) plots were employed to assess if the data followed a specific distribution, aiding in understanding the distribution of the profile data.

Chapter 4 introduced the application of functional network (FN) models for predicting railway geometric data. These models incorporated domain knowledge and data to estimate neuron functions, capturing the complex dynamics of railway behavior. Data preprocessing techniques were employed, and the mean squared error (MSE) metric was used for model evaluation. The FN model outperformed a neural network model in terms of predictive accuracy, attributed to the integration of domain knowledge and the use of quadratic polynomial functional families to estimate neuron functions. The FN model's simplicity and interpretability made it a more efficient option, providing valuable insights into functional relationships. However, the model's accuracy relied on the quality and representativeness of the training data, emphasizing the need for diverse and unbiased data.

The study highlighted the importance of selecting appropriate models for railway maintenance and operational forecasting. While the FN model was effective for short-term predictions, accurate prediction of peak values and longer time horizons proved challenging. Accurate peak prediction is crucial for identifying critical maintenance locations and allocating resources appropriately. Therefore, the need for advanced models that can accurately capture and anticipate fluctuations is emphasized, along with considering the time horizon when choosing forecasting models for different decision-making purposes.

Chapter 5 introduced an LSTM model for predicting track geometric data, leveraging its ability to capture the dynamics and hidden correlations within time series data. Two configurations of the

LSTM model were tested, focusing on shorter (6-month) and longer (12-month) time horizons. The model demonstrated improved prediction accuracy overall but encountered difficulties in accurately predicting sudden and significant variations in the profile data.

Furthermore, Chapter 5 explored applying a multivariable LSTM model for predicting space curve data, specifically examining variables such as left profile, right profile, left alignment, and right alignment. This approach effectively captured temporal dependencies and interdependencies between different track geometry parameters, facilitating more accurate predictions for maintenance planning and safe railway operations. The flexibility of the LSTM model allowed it to adapt to varying time steps and capture long-term dependencies in the data. However, the model exhibited variations in accuracy across different channels and months, mainly when predicting sudden variations in the profile data. The model's performance was influenced by factors such as the availability and quality of historical data and any changes or disruptions in the track conditions, underscoring the importance of continuous monitoring, data updates, and model recalibration.

Overall, the research encompassed data collection, exploratory data analysis, functional network modeling, and LSTM modeling. Each chapter contributed unique insights into understanding railway geometric data and predicting its behavior. The FN model demonstrated simplicity and interpretability, while the LSTM model captured temporal dependencies effectively. The LSTM (Long Short-Term Memory) model used in track analysis is capable of capturing 90% of the track with a Mean Absolute Error (MAE) of less than 0.15 inch. However, challenges remain in accurately predicting peak values and sudden variations. The findings underscore the need for advanced models and continuous monitoring to ensure the ongoing accuracy and reliability of predictions for railway maintenance and operational decision-making.

5.2 Challenges

The research on railway maintenance and operational forecasting also highlights several challenges that need to be addressed in future studies:

Prediction Horizon: One of the key challenges is accurately predicting events and conditions beyond a short-term horizon. Forecasting models often face difficulties in maintaining accuracy as the prediction horizon extends.

Peak Profile Prediction: Accurately predicting peak values in railway-measured data remains a significant challenge. Peaks indicate critical locations that require maintenance attention, and precise forecasting of these peaks is crucial for effective resource allocation and minimizing disruptions.

Data Availability and Quality: The availability and quality of data pose challenges in railway maintenance and operational forecasting. Obtaining comprehensive and accurate data, including historical maintenance records, operational data, and ex- ternal factors, can be complex.
External Factors and Uncertainty: Incorporating external factors, such as weather conditions, maintenance schedules, and passenger demand, adds complexity to forecasting models. These factors introduce uncertainties that impact railway maintenance and operation.

Model Selection and Evaluation: Selecting the most appropriate forecasting model for railway maintenance and operational forecasting is a challenge. There is a wide range of models and methodologies available, each with its strengths and limitations. Comparative evaluations and benchmarking studies are needed to identify the most suitable models for different forecasting scenarios and provide guidelines for model selection.

Implementation and Integration: Deploying forecasting models in real-world railway systems can be challenging. Integration with existing infrastructure, data acquisition, and operational constraints require careful consideration.

Interdisciplinary Collaboration: Railway maintenance and operational forecasting require collaboration between different disciplines, including transportation engineering, data science, and operations research. Bridging the knowledge gaps and fostering interdisciplinary collaboration can be challenging but is essential for developing comprehensive and effective forecasting models.

Addressing these challenges through further research and development will contribute to the improvement of railway maintenance planning, operational efficiency, and overall system reliability. Overcoming these challenges will enhance the accuracy and usability of forecasting models, enabling better decision-making and resource allocation for the maintenance and operation of railway systems.

5.3 Future Research

Future research in the field of railway maintenance and operational forecasting can focus on the following areas:

Advanced Modeling Techniques: Further exploration of advanced modeling techniques can contribute to improved forecasting accuracy. This includes investigating the effectiveness of transformer-based models, which have succeeded in other domains like natural language processing and image recognition. Assessing the applicability of trans- formers for capturing long-range dependencies and enhancing forecasting accuracy in railway maintenance and operation would be valuable.

Hybrid Modeling Approaches: Combining multiple forecasting models or methodologies can potentially enhance the overall performance of railway maintenance and operational forecasting. Future studies can explore the development of hybrid models that leverage the strengths of different approaches, such as combining traditional time series models with machine learning algorithms or combining deterministic and stochastic modeling techniques.

Integration of External Factors: Considering external factors that impact rail- way maintenance and operation is crucial for accurate forecasting. Future research can focus on developing methodologies to effectively integrate variables like weather conditions, maintenance schedules, and passenger demand into forecasting models. This integration can enable more robust and reliable predictions, allowing for proactive decision-making and resource allocation.

Uncertainty Quantification: Railway maintenance and operational forecasting inherently involve uncertainties. Future studies can explore methodologies to quantify and incorporate uncertainty into forecasting models. Techniques such as probabilistic forecasting and ensemble modeling can provide a more comprehensive understanding of the range of possible outcomes, aiding in risk assessment and decision-making.

Real-Time and Dynamic Forecasting: Developing real-time and dynamic fore- casting models can support adaptive maintenance and operational planning. Future research can explore techniques that can update forecasts in real-time based on in- coming data and dynamically adjust maintenance schedules and resource allocation accordingly. This can enable more agile and responsive decision-making in rapidly changing operational environments.

Integration of Maintenance Decision Support Systems: Integrating forecasting models into maintenance decision support systems can enhance their practical utility. Future research can focus on developing integrated platforms that combine forecasting models with optimization algorithms, data visualization, and decision support tools. This integration can provide maintenance and operation planners with comprehensive tools for efficient decision-making and resource allocation.

Case Studies and Validation: Conducting case studies and validating forecasting models using real-world data and scenarios is essential. Future research should emphasize the application of forecasting models in practical railway maintenance and operation settings to assess their performance and practicality. Validation studies can provide insights into the strengths, limitations, and areas of improvement for different forecasting models.

Sustainability and Resilience Considerations: Future research can incorporate sustainability and resilience considerations into railway maintenance and operational forecasting. This involves analyzing the environmental impact of maintenance activities, optimizing resource allocation for energy efficiency, and developing forecasting models that support resilient decision-making in disruptive events or emergencies.

By addressing these future research areas, researchers can advance the railway maintenance and operational forecasting field, leading to more accurate and reliable predictions, optimized maintenance planning, and improved operational efficiency in railway systems.

REFERENCES

1. "Train Accidents by Cause FRA." Accessed: Jun. 28, 2023. [Online]. Available: https://railroads.dot.gov/accident-and-incident-reporting/train-accident-reports/train-accidents-cause
2. Sheeran, R., M. Kay, and A. Jenner, "A Short Guide to Network Rail," Natl. Audit Off., Jul. 2015, [Online]. Available: https://www.nao.org.uk/wp-content/uploads/2015/08/Network-rail-short-guide1.pdf
3. López-Pita, A. P.F. Teixeira, C. Casas, A. Bachiller, and P. A. Ferreira, "Maintenance Costs of High-Speed Lines in Europe State of the Art," Transp. Res. Rec. J. Transp. Res. Board, vol. 2043, no. 1, pp. 13–19, Jan. 2008, doi: 10.3141/2043-02.
4. Budai-Balke, G., Operations Research Models for Scheduling Railway Infrastructure Maintenance. Rozenberg Publishers, 2009.
5. Jardine, K.S., D. Lin, and D. Banjevic, "A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance," Mech. Syst. Signal Process., vol. 20, no. 7, pp. 1483–1510, Oct. 2006, doi: 10.1016/j.ymssp.2005.09.012.
6. Jezzini, A., M. Ayache, L. Elkhansa, B. Makki, and M. Zein, "Effects of Predictive Maintenance (Pdm), Proactive Maintenance (Pom) &Amp; Preventive Maintenance(PM) on Minimizing the Faults In Medical Instruments," in 2013 2nd International Conference on Advances in Biomedical Engineering, Tripoli, Lebanon: IEEE, Sep. 2013, pp. 53–56. doi: 10.1109/ICABME.2013.6648845.
7. Marino, F., A. Distante, P. L. Mazzeo, and E. Stella, "A Real-Time Visual Inspection System for Railway Maintenance: Automatic Hexagonal-Headed Bolts Detection," IEEE Trans. Syst. Man Cybern. Part C Appl. Rev., vol. 37, no. 3, pp. 418–428, May 2007, doi: 10.1109/TSMCC.2007.893278.
8. Mohammadi, R., Q. He, F. Ghofrani, A. Pathak, and A. Aref, "Exploring the Impact of Foot-By-Foot Track Geometry on the Occurrence Of Rail Defects," Transp. Res. Part C Emerg. Technol., vol. 102, pp. 153–172, May 2019, doi: 10.1016/j.trc.2019.03.004.
9. Li, Z. and Q. He, "Prediction of Railcar Remaining Useful Life by Multiple Data Source Fusion," IEEE Trans. Intell. Transp. Syst., vol. 16, no. 4, pp. 2226–2235, Aug. 2015, doi: 10.1109/TITS.2015.2400424.
10. Xie, J., J. Huang, C. Zeng, S.-H. Jiang, and N. Podlich, "Systematic Literature Review on Data-Driven Models for Predictive Maintenance of Railway Track: Implications in Geotechnical Engineering," Geosciences, vol. 10, no. 11, p. 425, Oct. 2020, doi: 10.3390/geosciences10110425.
11. Karpatne, A., A. Gowtham, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar, "Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data," IEEE Trans. Knowl. Data Eng., vol. 29, no. 10, pp. 2318–2331, Oct. 2017, doi: 10.1109/TKDE.2017.2720168.
12. Wadud, Z., D. MacKenzie, and P. Leiby, "Help or Hindrance? The Travel, Energy and Carbon Impacts of Highly Automated Vehicles," Transp. Res. Part Policy Pract., vol. 86, pp. 1–18, Apr. 2016, doi: 10.1016/j.tra.2015.12.001.
13. Xie, G., A. Shangguan, R. Fei, W. Ji, W. Ma, and X. Hei, "Motion Trajectory Prediction Based On a CNN-LSTM Sequential Model," Sci. China Inf. Sci., vol. 63, no. 11, p. 212207, Nov. 2020, doi: 10.1007/s11432-019-2761-y.
14. Zhang, Z., X. Che, and Y. Song, "An Improved Convolutional Neural Network for Convenient Rail Damage Detection," Front. Energy Res., vol. 10, p. 1007188, Sep. 2022, doi: 10.3389/fenrg.2022.1007188.

15. Zauner, G., T. Mueller, A. Theiss, M. Buerger, and F. Auer, “Application of Semantic Segmentation for an Autonomous Rail Tamping Assistance System,” Electron. Imaging, vol. 31, no. 7, pp. 462-1-462–6, Jan. 2019, doi: 10.2352/ISSN.2470-1173.2019.7.IRIACV-462.
16. H. Li, B. Qian, D. Parikh, and A. Hampapur, “Alarm Prediction in Large-Scale Sensor Networks 2014; A Case Study In Railroad,” in 2013 IEEE International Conference on Big Data, Silicon Valley, CA, USA: IEEE, Oct. 2013, pp. 7–14. doi: 10.1109/BigData.2013.6691771.
17. Tsui, K.L., N. Chen, Q. Zhou, Y. Hai, and W. Wang, “Prognostics and Health Management: A Review on Data Driven Approaches,” Math. Probl. Eng., vol. 2015, pp. 1–17, 2015, doi: 10.1155/2015/793161.
18. Lederman, G., S. Chen, J. Garrett, J. Kovačević, H. Y. Noh, and J. Bielak, “Track-Monitoring from the Dynamic Response of an Operational Train,” Mech. Syst. Signal Process., vol. 87, pp. 1–16, Mar. 2017, doi: 10.1016/j.ymssp.2016.06.041.
19. Jiang, Y., H. Wang, G. Tian, Q. Yi, J. Zhao, and K. Zhen, “Fast Classification for Rail Defect Depths Using a Hybrid Intelligent Method,” Optik, vol. 180, pp. 455–468, Feb. 2019, doi: 10.1016/j.ijleo.2018.11.053.
20. El-Sibaie, M. and Y.-J. Zhang, “Objective Track Quality Indices,” Transp. Res. Rec. J. Transp. Res. Board, vol. 1863, no. 1, pp. 81–87, Jan. 2004, doi: 10.3141/1863-11.
21. Sadeghi, J. and H. Askarinejad, “Development of improved railway track degradation models,” Struct. Infrastruct. Eng., vol. 6, no. 6, pp. 675–688, Dec. 2010, doi: 10.1080/15732470801902436.
22. Nadarajah, N., A. Shamdani, G. Hardie, W.K.Chiu, and H. Widyastuti, “Prediction of Railway Vehicles’ Dynamic Behavior with Machine Learning Algorithms,” Electron. J. Struct. Eng., vol. 18, no. 1, pp. 38–46, Jan. 2018, doi: 10.56748/ejse.182271.
23. Zhang, G.P., “Time series forecasting using a hybrid ARIMA and neural network model,” Neurocomputing, vol. 50, pp. 159–175, Jan. 2003, doi: 10.1016/S0925-2312(01)00702-0.
24. Heidarysafa, M., K. Kowsari, L. Barnes, and D. Brown, “Analysis of Railway Accidents’ Narratives Using Deep Learning,” in 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL: IEEE, Dec. 2018, pp. 1446–1453. doi: 10.1109/ICMLA.2018.00235.
25. Andrade, A.R., and P. F. Teixeira, “A Bayesian model to assess rail track geometry degradation through its life-cycle,” Res. Transp. Econ., vol. 36, no. 1, pp. 1–8, Sep. 2012, doi: 10.1016/j.retrec.2012.03.011.
26. Hu, C. and X. Liu, “Modeling Track Geometry Degradation Using Support Vector Machine Technique,” in 2016 Joint Rail Conference, Columbia, South Carolina, USA: American Society of Mechanical Engineers, Apr. 2016, p. V001T01A011. doi: 10.1115/JRC2016-5739.
27. Peck, R., C. Olsen, and J. L. Devore, Introduction to statistics and data analysis. Cengage Learning, 2015.
28. Parker, M., “Box Plot with Minitab → Lean Sigma Corporation,” Lean Six Sigma Corporation. Dec. 2015. Accessed: Jul. 14, 2023. [Online]. Available: https://www.leansigmacorporation.com/box-plot-with-minitab/
29. Palese, J.W., A. M. Zarembski, and N. O. Attoh-Okine, “Methods for aligning near-continuous railway track inspection data,” Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit, vol. 234, no. 7, pp. 709–721, Aug. 2020, doi: 10.1177/0954409719860718.
30. Castillo, E. and J. M. Gutiérrez, “Nonlinear time series modeling and prediction using functional networks. Extracting information masked by chaos,” Phys. Lett. A, vol. 244, no. 1–3, pp. 71–84, Jul. 1998, doi: 10.1016/S0375-9601(98)00312-0.
31. Castillo, E., A. Cobo, J. A. Gutierrez, and R. E. Pruneda, Functional networks with

applications: a neural-based paradigm, vol. 473. Springer Science & Business Media, 2012.
32. Box G.E.P. and G. M. Jenkins, “Some Comments on a Paper by Chatfield and Prothero and on A Review by Kendall,” J. R. Stat. Soc. Ser. Gen., vol. 136, no. 3, p. 337, 1973, doi: 10.2307/2344995.
33. Hochreiter S. and J. Schmidhuber, “Long Short-Term Memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

ACKNOWLEDGEMENTS

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

ABOUT THE AUTHOR

Mohammed Ahmed

Mohammed Ahmed was a graduate student when he conducted research on this project in the Civil Engineering Department at the University of Delaware. He obtained his Bachelor degree from the University of Khartoum and his Master's degree from the University of Delaware, respectively.

Joseph W. Palese, MCE, PE

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

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

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

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

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

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