

RAILWAY TRACK CHANGE DETECTION: A DATA-DRIVEN APPROACH TO TRACK  
GEOMETRY AND COMPONENT DEGRADATION

BY

JOSE AUGUSTO VENANCIO DA SILVA RAMOS

THESIS

Submitted in partial fulfillment of the requirements  
for the degree of Master of Science in Civil Engineering  
in the Graduate College of the  
University of Illinois Urbana-Champaign, 2024

Urbana, Illinois

Adviser:

Assistant Professor J. Riley Edwards

**ABSTRACT**

TBD

## ACKNOWLEDGMENTS

First, I would like to thank my parents, Jarmes and Janete, and my brothers, Jose Gustavo and Julio Cesar, for supporting and encouraging me throughout my entire life. Also, my friends in Brazil and at Urbana-Champaign who made these two years a lot easier to come through. Special thanks are due to my present and past roommates Vitor, Arthur, and Matt, and the B118 folks for sharing the days, engaging in the research, events, and trips together.

To the RailTEC team, Riley Edwards, Marcus Dersch, and Arthur Lima, thanks for giving me the opportunity to study and work with this amazing group. Also, thanks to Chris Barkan, Emma, LB, and Carol for all the support during my stay in the group.

This project would never come to fruition without the industry partners: Brad Spencer (CSX), Adam Fell (FEC), David Wenrich, Richard Fox-Ivey, and Antonio Mauricio (Railmetrics), and consultant Gary Carr. Thank you for supporting this research project. Also, thanks to the folks on the field who shared their insights and made our field trips for track inspecting even more enjoyable.

There are no words to express my gratitude to the Rumo Logistica team that introduced me to the railway world and supported and encouraged me to seek a master's degree.

I would like to thank the American Railway Engineering and Maintenance-of-Way (AREMA) for awarding me with the 2024 AREMA Educational Foundation Scholarship and for supporting our AREMA Student Chapter at UIUC, which contributed greatly to my personal and professional development and connected me with several people in the industry.

I would like to express sincere appreciation to the Federal Railroad Administration (FRA) in general and Cam Stuart specifically for providing support for this project, including funding, insights, and industry connections, without which this research would not be possible.

*To my family:*

*Brazilian song TBD*

## TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION.....	1
1.1    Track Geometry and Regulatory Requirements.....	1
1.2    Track Components and Regulatory Requirements.....	7
1.3    Railroad Track Big Data .....	11
1.4    Objectives.....	<u>15</u>
1.5    Thesis Outline .....	15
CHAPTER 2: STOCHASTIC INVESTIGATION OF THE RELATIONSHIP BETWEEN TRACK COMPONENT HEALTH AND TRACK GEOMETRY DATA	
16	
2.1    Introduction .....	16
2.2    Material and Methods.....	17
2.3    Results and Discussion.....	<u>26</u>
2.4    Conclusions .....	<u>36</u>
CHAPTER 3: CROSS-CORRELATION BASED RAILWAY CHANGE DETECTION: A NOVEL APPROACH.....	<u>38</u>
3.1    Introduction .....	<u>38</u>
3.2    Material and Methods.....	<u>43</u>
3.3    Results and Discussion.....	<u>54</u>
3.4    Conclusions .....	92
CHAPTER 4: CONCLUSIONS AND FUTURE WORK .....	95
4.1    Summary of Findings.....	95
4.2    Recommendations and Future Work.....	96
REFERENCES <u>98</u>	

**Deleted: 14**

**Deleted: 27**

**Deleted: 37**

**Deleted: 39**

**Deleted: 39**

**Deleted: 44**

**Deleted: 55**

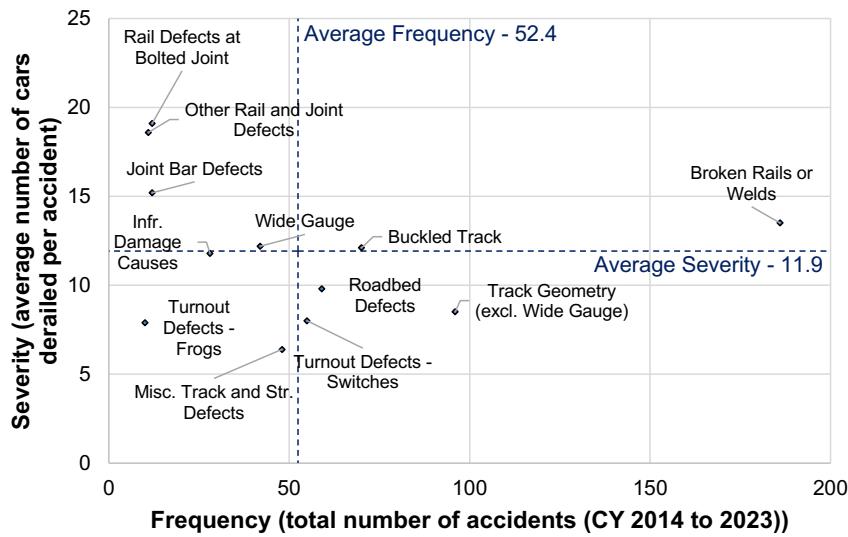
**Deleted: 97**

## CHAPTER 1: INTRODUCTION

### 1.1 Track Geometry and Regulatory Requirements

The railroad track infrastructure and its components are responsible for supporting railroad rolling stock wheel loads and facilitating safe and efficient train operations. In the United States (U.S.), the U.S. Department of Transportation (US DOT) Federal Railroad Administration (FRA) serves as the regulatory body that maintains thresholds that are designed to ensure safe train operations (FRA, 2024). The FRA also maintains a database of reportable accidents. For 2024, FRA reportable accidents were ones in which monetary damages surpassed \$12,000 USD. Accidents that occur frequently and have high severity (i.e., the number of cars derailed) have the highest associated risk, and thereby are high-priority candidates for further investigation. For the 10-year period between January 2014 and December 2023, *broken rails or welds, track geometry defects (excluding wide gauge), and buckled track* were the three most frequent causes of track-related accidents on Class I freight railroad mainlines and sidings ([Figure 1.1](#)).

**Deleted:** Figure 1.1



**Figure 1.1: Frequency and severity of track-related accidents on Class I railroad mainlines and sidings between January 2014 and December 2023 (FRA, 2023).**

To ensure the safety of train operations and compliance with FRA regulations, railroads perform their own internal track inspections via high-rail vehicles as often as twice per week depending on FRA Track Class (FRA, 2024). These classes are defined by FRA and maintained in the FRA Track Safety Standards (FRA, 2024). The FRA track class dictates the maximum allowable speed for both passenger and freight trains (FRA, 2024) ([Table 1.1](#)). Most freight and intercity passenger rail operations in the U.S. are conducted on track that is maintained to FRA track Classes 1-5. The majority of the primary Class I freight corridors in the U.S. are maintained to FRA track Classes 4 and 5.

**Deleted:** Table 1.1

**Formatted:** Font: Not Bold

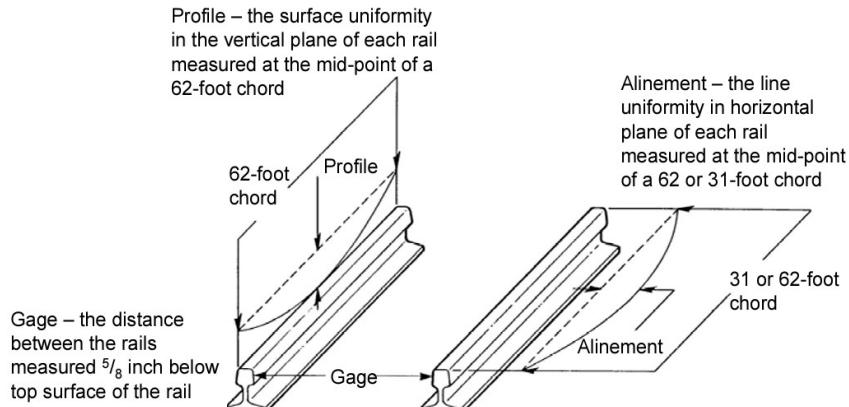
**Table 1.1: FRA Track classes and maximum allowable train speeds (FRA, 2024).**

FRA Track Class	Maximum Allowable Train Speed (mph) [km/h]	
	Freight Trains	Passenger Trains
0 (Exception)	10 [16]	Not Allowed
1	10 [16]	15 [24]
2	25 [40]	30 [48]
3	40 [64]	60 [97]
4	60 [97]	80 [129]
5	80 [129]	90 [145]

Beyond setting different maximum speeds for train operation, FRA track classes also require geometric standards that are commensurate with the operating speed. FRA regulates a variety of track geometry parameters which include gauge, profile, and alignment (or alinement)

(Figure 1.2).

**Deleted:** Figure 1.2



**Figure 1.2: Figure and description of typical track geometry parameters (Federal Railroad Administration, 2018).**

Gauge is defined as the distance between both running rails, and is measured in a plane five-eighths of an inch (15.875 mm) below the top of the rail head (FRA, 2024). Compliance with gauge standards is essential for ensuring good wheel contact and stability of track system and proper gauge is crucial for safe and efficient train operations. The FRA-prescribed limits for gauge vary based on the FRA class of track ([Table 1.2](#)).

**Deleted:** Table 1.2

**Formatted:** Font: Not Bold

**Table 1.2 – Gauge conditions based on the class of track (FRA, 2024).**

Track Class	Gage must be at least --	But not more than --
Excepted	N/A	4'10 1/4" (1480 mm)
1	4'8" (1422 mm)	4'10" (1499 mm)
2 and 3	4'8" (1422 mm)	4'9 3/4" (1467 mm)
4 and 5	4'8" (1422 mm)	4'9 1/2" (1461 mm)

As an example, for FRA Classes 4 and 5, track gauge should not exceed 4'9 1/2" (1460.5 mm) else a temporary speed reduction to FRA Track Class 3 is required (FRA, 2024). Usually, the gauge values recorded and presented by railroads correspond is the deviation from standard track gauge (56.5 in [1435mm]). I will also employ this method of reporting gauge as deviation from standard.

Profile is a measurement of the consistency of the vertical location of rails in the longitudinal direction and is also referred to as “surface.” Profile is commonly calculated using a 62-foot chord (18.89 m) as defined by FRA (2024). Profile represents the deviation from a uniform profile (straight line) on either rail at the mid-ordinate of a chord (e.g., 31 ft (9.45 m) distance for a 62-foot chord (18.89 m)). Each rail is individually assessed using a moving 62-foot (18.89 m) window. Inspections are frequently carried out to compare the geometry measurements to thresholds that are railroad-specific or required by the FRA as shown in [Table 1.3](#).

**Formatted:** Font: Not Bold

[1.3](#). For instance, FRA Class 4 track profile deviations should not exceed 2 inches (50.8 mm) at

**Deleted:** Table 1.3

the midpoint of a 62-foot (18.89 m) chord, else a temporary speed reduction to FRA Track Class 3 is required (FRA, 2024).

**Table 1.3 – Track surface conditions based on class of track (FRA, 2024).**

Track surface	FRA Track Class				
	1	2	3	4	5
The deviation from uniform profile on either rail at the mid-ordinate of a 62-foot chord may not be more than:	3" (76 mm)	2 ¾" (70 mm)	2 ¼" (64 mm)	2" (51 mm)	1 ¼" (32 mm)

Alignment refers to the variation in the curvature of each rail on the track. For tangent track where the intended curvature is zero, alignment is the deviation from zero (FRA, 2024).

On curved track, alignment is the deviation from the desired uniform curvature over a defined distance (FRA, 2024). Inspections are carried out to compare the geometry measurements to thresholds that are railroad-specific or required by the FRA as shown in [Table 1.4](#).

**Deleted:** Table 1.4

**Formatted:** Font: Not Bold

**Table 1.4 – Track alignment thresholds based on FRA Track Class (FRA, 2024).**

Track alignment	FRA Track Class				
	1	2	3	4	5
Tangent: The deviation of the mid-offset from a 62-foot line <sup>1</sup> may not be more than:	5" (127 mm)	3" (76 mm)	1 ¾" (44 mm)	1 ½" (38 mm)	¾" (19 mm)
Curve: The deviation of the mid-offset from a 31-foot chord <sup>2</sup> may not be more than:	NA	NA	1 ¼" (32 mm)	1" (25 mm)	½" (13 mm)
Curve: The deviation of the mid-offset from a 62-foot chord <sup>2</sup> may not be more than:	5" (127 mm)	3" (76 mm)	1 ¾" (44 mm)	1 ½" (38 mm)	5/8" (16 mm)

<sup>1</sup> The ends of the line shall be at points on the gage side of the line rail, five-eighths of an inch below the top of the railhead. Either rail may be used as the line rail, however, the same rail shall be used for the full length of that tangential segment of track.

<sup>2</sup> The ends of the chord shall be at points on the gage side of the outer rail, five-eighths of an inch below the top of the railhead.

For instance, FRA Class 4 track alignment deviation on curved track should not exceed 1.5 inches (50.8 mm) at the mid-offset of a 62-foot (18.89 m) chord, else a temporary speed reduction to FRA Track Class 3 is required (FRA, 2024).

To ensure safe and efficient operations, railroads invest in track inspection and measurement devices that include Autonomous Track Geometry Measurement System (ATGMS) railcars, ultrasound for rail flaw detection, Ground Penetration Radar (GPR), gauge restraint measurement systems (GRMS), and others. Many of these devices send alerts in near real-time if an exception is identified. Moreover, the frequency of track inspections for the primary HAL corridors commonly exceeds what is required by FRA, generating large track condition datasets over time. Since track inspection data come from different inspection platforms, each platform captures more than one channel of data, and data are commonly captured at frequent distance intervals (i.e. foot by foot), they are commonly considered “big data” (Zarembski, 2014; Jamshidi et al., 2017; Zarembski et al., 2017; Ghofrani et al., 2018; Bilheri et al., 2023).

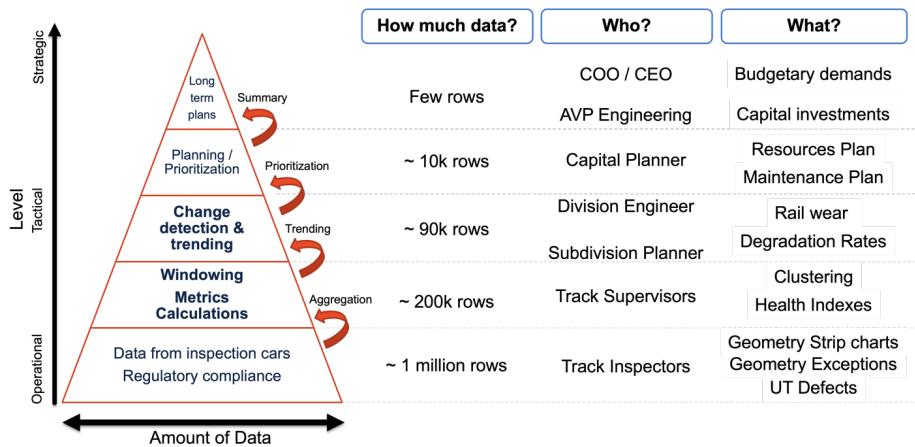
These inspection and condition data are consumed at different processing levels by railroad practitioners and researchers depending on the specific question being answered. Data are commonly used to address challenges related to immediate and objective evaluation of the track all the way to forecasting and predicting when and where geometry thresholds will be exceeded to avoid speed restrictions and optimize maintenance spending (Esveld, 2014).

According to Carvalho (2023), data analysis within the railroad industry is closely tied to the organizational hierarchy of a company. At the operational level, detailed and raw data is essential for handling tasks such as corrective maintenance and managing speed restrictions. At the tactical and strategic levels, the focus shifts towards broader responsibilities like capital

planning and resource allocation across larger areas, including multiple subdivisions, divisions, or even the entire network. At these higher levels, the reliance on raw data decreases, and need for processed data covering larger portions of a rail network increases. The strategic level is primarily focused on long-term investment planning and requires minimal data resolution but emphasizes processed results for trend forecasting. This structured flow of data and its distillation starting from the operational level and culminating at the strategic level can be depicted as a pyramid (Figure 1.3).

**Deleted:** Figure 1.3

**Formatted:** Font: Not Bold



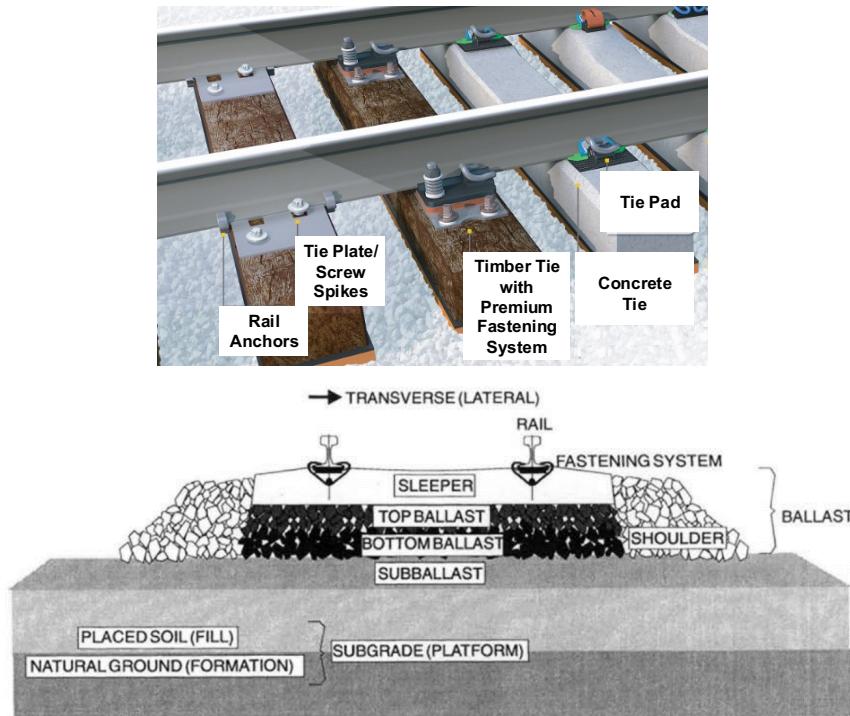
**Figure 1.3: Data pyramid and examples of quantity, users, and type of information (Carvalho, 2023).**

## 1.2 Track Components and Regulatory Requirements

The components that make up the track system are a critical part of the rail infrastructure system. Track components work together as a system to safely and efficiently transmit loads from the rolling stock to the infrastructure, provide a smooth-running surface for the train, and maintain proper geometry to ensure safe train operations. It is critical that such components are in a state

of good repair and - at a minimum - are compliant with FRA Regulations (2024). The most common track components providing stability and safety to train operations are the rails, crossties (sleepers), tie plates, fasteners, and ballast (Figure 1.4).

**Deleted:** Figure 1.4  
**Formatted:** Font: Not Bold



**Figure 1.4: Typical track structure components (top) (adapted from Progress Rail, 2024), and track cross sectional view (bottom) (Selig and Waters, 1994).**

Similar to their regulatory oversight of track geometry parameters, the FRA also maintains and enforces regulations for the presence, condition, and function of track components. In addition to the Track Safety Standards (CFR 213), the FRA also maintains the Track and Rail and Infrastructure Integrity Compliance Manual (Federal Railroad Administration, 2018) to

provide additional commentary and aid in the interpretation of regulations included in CFR 213. Moreover, railways have their own internal maintenance standards and business rules which are designed to be more restrictive than FRA's regulatory requirements (BNSF Railway, 2018; CSX Transportation, 2020; Union Pacific Railroad, 2022). What follows is a summary of the most common types of track components, their function, and the FRA's regulations regarding their placement and condition.

A railway crosstie (also known as a *railroad tie*, *sleeper*, or *tie*) is a rectangular support that rests on the track bed and is used to hold the rails in place (Hay, 1982b). Crossties have traditionally been made from timber, but are increasingly made of concrete, steel, or plastic composites to improve their durability and performance. Crossties serve a critical function in the track system by maintaining the correct gauge (the distance between the rails) and distributing train loads to the ballast at levels that do not crush or deform the ballast layer. They also help absorb vibrations and provide stability to the track structure in all three axes, ensuring safety.

According to the FRA standards (FRA, 2024), crossties must be manufactured from materials that securely fasten rails and comply with requirements for track stability. Each 39-foot (11.88 m) segment of track must have enough crossties to maintain proper gauge, surface, and alignment, with the minimum number determined by track class and curvature, as shown in

[Table 1.5](#)

**Deleted:** Table 1.5  
**Formatted:** Font: Not Bold

**Table 1.5 – Minimum number of crossties in a 39-foot track segment (FRA, 2024).**

Minimum number of crossties	FRA Track Class			
	1	2	3	4 and 5
Tangent track and curved track less than or equal to 2 degrees	5	8	8	12

Tangent track and curved track greater than 2 degrees	6	9	10	14
--	---	---	----	----

At least one non-defective (i.e., healthy, functioning) crosstie must be present at joint locations. Timber crossties must not be excessively split or worn, while concrete crossties must be free from significant deterioration and cracking, ensuring they prevent lateral movement beyond specified limits and maintain the necessary support and fastening strength to uphold track integrity and safety.

According to Esveld (2014), railway fasteners are critical components that secure the rails to the crossties, ensuring track stability under train loads. Fasteners may include spikes, clips, bolts, or screws, depending on the track system design. Their primary role is to hold the rails in place preventing both lateral and longitudinal movement, while allowing for the rail to expand and contract under reasonable thermal loading. Proper fastening system securement is crucial for maintaining track alignment, reducing the risk of misalignment, and minimizing wear on both the track and rolling stock. For timber crossties, cut spikes and screw spikes most common, with elastic systems used in locations with high loading demand. In contrast, elastic fastening systems are always used for concrete crossties, obviating the need for rail anchors.

According to FRA (2024), for Classes 3 through 5 constructed with timber crossties, tie plates must be installed under the running rails on at least 8 out of every 10 consecutive crossties. The track fastening system must maintain the proper track gauge per FRA regulations, and its components should be regularly evaluated to ensure they are performing their design function. For concrete crossties, the combination of crossties, fasteners, and rail anchors must provide effective longitudinal restraint. If fasteners interfere with insulated joints, they can be modified or removed as long as the rail remains supported.

Railway ballast is the layer of aggregate (crushed stone or gravel) placed beneath and around crossties to provide stability to the track structure. The primary functions of ballast are to distribute the loads from passing trains, facilitate drainage, and maintain the track's alignment by holding the crossties in place. The ballast also helps absorb the vibration and shock loads caused by train traffic, which reduces the wear and tear on the track components.

Traditionally, hard, angular aggregate such as granite are preferred due to their durability and resistance to degradation (Hay, 1982a). The angular shape of ballast particles allows them to interlock, providing resistance to crosstie movement and maintaining track geometry. This helps control the longitudinal and lateral forces caused by trains. Over time, the ballast may degrade due to crushing from train traffic, thus requiring tamping (surfacing) and ballast replacement maintenance to restore its effectiveness and reestablish proper geometry (Selig and Waters, 1994).

### **1.3 Railroad Track Big Data**

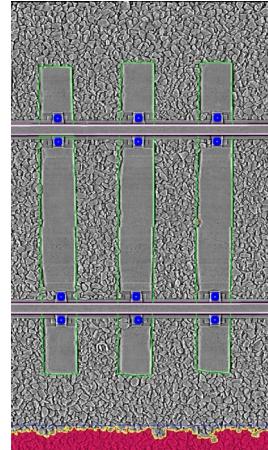
Traditionally, track inspections were conducted manually, requiring inspectors to visually assess and measure track conditions on-site. The primary systems in use were geometry inspection systems, most of which were operated by personnel who evaluated track alignment, gauge, and other critical dimensions in person. Recently, however, advancements in automated inspection technology have significantly reduced the need for manual involvement. Automated and unmanned systems, equipped with sensors and data collection capabilities, now allow for continuous and highly accurate monitoring of track condition, improving both the efficiency and reliability of track inspections.

Some novel and emerging devices have been made available for the railways to inspect the state of components in the last decade. Their primary advantages are better productivity, repeatability, and safety. The system used for collection of component health data described in the research was developed by Railmetrics, Inc. and is called the Laser Rail Inspection System (LRAIL). LRAIL relies on a very high 3D scan resolution, capturing over 100 million points per second (Fox-Ivey et al., 2020a; Fox-Ivey et al., 2020b). In addition to capturing laser scans, it simultaneously records 2D images and then processes the data using deep convolutional neural networks (DCNNs) to identify rail components and their condition. Such attributes include ballast levels, surface fouling, crossties, tie plates, fasteners, insulators, anchors, and joint bars (Harrington et al., 2022; Harrington et al., 2023; Germoglio Barbosa et al., 2024). [Figure 1.5](#) shows a hi-rail mounted LRAIL system and an example output for concrete crosstie track.

**Deleted:** Figure 1.5



a)



b)

**Figure 1.5: Hi-rail based LRAIL equipment (left) and JPEG output example (right).**

With new equipment and improvement in data collection methods, railroads now have very large track condition datasets to explore. Some geometry inspections are carried out using autonomous inspection vehicles at a very high inspection frequency (e.g., multiple passes per week). This frequency of inspection generates a large track condition database. Proper analysis of these data could facilitate more advanced predictive maintenance activities as opposed to corrective maintenance, thereby saving resources and enhancing safety (Sharma et al., 2018; Goodarzi et al., 2022). Implicit in this advancement that is made possible through the capture of big data is the development of a quantitative understanding of railway track systems and component degradation rates. Quantification of these rates is useful for efficiency and effectively managing track maintenance and ensuring safe train operation.

Increases in computational power over the past few decades has allowed researchers to develop, apply, and improve mathematical models that consume large amounts of data (Ghofrani et al., 2018). One example is the use of data science techniques by Liao et al. (2022) to assess track degradation to schedule maintenance, thereby reducing the risk of service failures and saving resources. Zhu et al. (2013) explored vertical profile and alignment irregularities and concluded that both can be modeled as a Gaussian random process. They proposed a power spectral density (PSD) method to approximate the random process and aid in the stochastic modeling of track degradation. Vale and Lurdes (2013) explored a stochastic approach for track geometry degradation in tracks under different vehicle speeds on a Portuguese railway. They found that the Dagum distribution best represents the track degradation process when considering the metric of vertical track profile (surface).

Goodarzi et al. (2023) applied a stochastic approach to investigate the track degradation rate using Ground Penetration Radar (GPR) and Light Detection and Ranging (LiDAR) data.

They used 10 years of geometry inspection data and concluded that higher degradation rates are 10 times more likely to happen at locations with highly fouled ballast as opposed to locations with clean ballast. They also found that high degradation rates are 30% more likely in sections with poor drainage conditions than those with moderate drainage. Alsahli et al. (2019) investigated the relationship between track geometry defect occurrence and timber crosstie condition. They used stochastic and statistical analyses to conclude that certain geometry defects including excess cant, wide gauge, and profile deviations correlate with crosstie condition.

Zarembski et al. (2017) used data science algorithms to investigate the relationship between ballast volume deficit and track geometry exceptions. They developed a logistic regression to predict the probability of occurrence of yellow and red-tag geometry defects using the ballast volume measured by LiDAR data and tonnage information provided by the railroad.

Researchers have also used data science approaches to investigate track component degradation. Soufiane et al. (2022) explored crosstie life cycle based on the support condition of adjacent crossties. They used two different algorithms and concluded that premature degradation and crosstie failure are generally associated with a loss of support condition quality.

One critical step in all the recent developments is the need for proper alignment of data. Palese (2020) described the challenges involving railway data alignment. The authors used dynamic time warping (DTW). This method, described theoretically by Salvador and Chan (2007), was originally developed to align time-series data for applications like movement and speech recognition. It calculates the piecewise distance between two signals, producing a distance matrix. Due to the computational demands of this process, the use of cloud computing was essential. Cross-correlation was also described as a way to evaluate track health data misalignment before and after the alignment correction method was applied (Palese et al., 2020).

## **1.4 Objectives**

The primary objectives of my work are to:

- Investigate the relationship between track component degradation and geometry degradation rates and identify any correlations;
- Explore an alternative and practical way to identify maintenance events on intensive railway track condition datasets.

## **1.5 Thesis Outline**

My thesis is divided into four chapters including this introduction. The following paragraphs provide a brief description of the overall scope of each chapter.

Chapter 2 describes the methodology and results for investigating the relationship between track component degradation and geometry degradation. Track component health indexes are introduced as is the stochastic modeling methodology. I also present model results and discuss their implications.

Chapter 3 describes a novel use of the cross-correlation algorithm to identify maintenance events in large railway infrastructure condition datasets. The methodology is described along with suggested parameters and results are generated using revenue service field data.

Lastly, Chapter 4 summarizes the key findings of my thesis, offers recommendations based on the results, and suggests potential directions and needs for future research.

## CHAPTER 2: STOCHASTIC INVESTIGATION OF THE RELATIONSHIP BETWEEN TRACK COMPONENT HEALTH AND TRACK GEOMETRY DATA

### 2.1 Introduction

While prior research had been conducted on track geometry and component degradation (Vale and M. Lurdes, 2013; Soleimanmeigouni and Ahmadi, 2016; Higgins and Liu, 2018; Wilk and Li, 2021; Goodarzi et al., 2023), most models did not take into consideration the rate of change in the boundary conditions. The objective of this study is to apply a stochastic method to investigate the influence of the degradation rate of ballast section on the degradation rate of track profile.

In this chapter, I will investigate the relationship between geometry and component inspection data using a stochastic approach. The results will be compared across different components using revenue service data from a Class I railroad mainline as the subject test corridor. Section 2.2 presents the data and methods used and introduces the hypotheses and statistical tests that I carried out. Section 2.3 describes the results I obtained and presents descriptive statistics, Spearman correlation coefficients, the overall stochastic approach to data classification and stratification, and a discussion of the results and their impact. Lastly, Section 2.4 presents conclusions and recommendations for future work.

Two datasets were used in this study: track geometry data and track component inspection data. Data were collected over 20 miles (32 km) of track between June 2021 and October 2023 on a major U.S. Class I freight railroad's primary corridor in the southeastern U.S. The area of study is maintained to FRA Track Class 4 and is constructed of timber crossties with a maximum operating speed of 60 mph (97 km/h). Approximately 90 million gross tons (MGT) of freight are

**Commented [JE1]:** I still don't understand what this means in the context of your work. Can you clarify?

**Commented [VDSRJA2R1]:** Is it better?

**Commented [JE3R1]:** No - I am sorry - but I still don't understand the "boundary condition" reference.

transported on this segment of track each year and there are no regularly scheduled passenger train operations. Locations with crossings, turnouts, or bridges were identified and removed from the dataset, since they are known to have operational and maintenance characteristics that differ from open track (Goodarzi et al., 2021).

## 2.2 Material and Methods

### 2.2.1 Data Collection

#### 2.2.1.1 Geometry Inspection Data

Track geometry data were collected using an Autonomous Track Geometry Measurement System (ATGMS) railcar that traverses the corridor in revenue service intermodal trains. The dataset consists of 76 geometry inspection runs collected over two-and-a-half years. Each dataset comprises 42 individual parameters (e.g., gauge, profile, alignment, and degree of curvature) which are geo-located. The specific track geometry parameter used in this research is track profile, which is output on an industry-standard foot-by-foot basis.

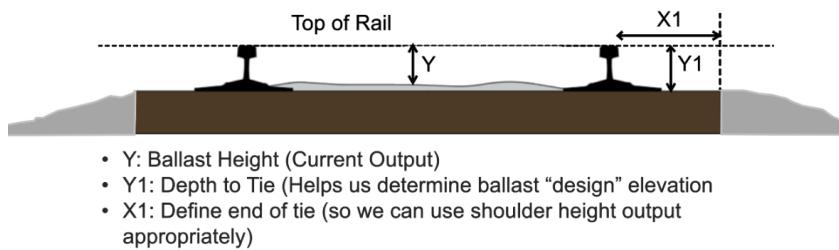
As presented in Section 1.1, profile is a measurement of the consistency of the vertical location of rails. Profile was calculated using a 62-foot chord (18.89 m) as defined by FRA (2024). For FRA Class 4 track, profile deviations should not exceed 2 inches (50.8 mm) at the midpoint of a 62-foot (18.89 m) segment, else a speed reduction (i.e., slow order) to FRA Class 3 is required (FRA, 2024).

#### 2.2.1.2 Component Inspection Data

This study included collection and processing of six LR4IL inspections that were collected using a highway-rail (hi-rail) inspection truck. While many other LR4IL outputs were collected as a

part of each data collection effort, I will focus on the ballast and its profile. For ballast, LR4/L measures the distance from the top of rail to ballast at multiple points that are as close as 1 mm apart. These data are used to generate a 3D point cloud before undergoing additional processing to evaluate whether the measured cross section matches the design cross section ([Figure 2.1](#)).

**Deleted:** Figure 2.1



**Figure 2.1: Ballast level measurement methodology employed by LR4/L.**

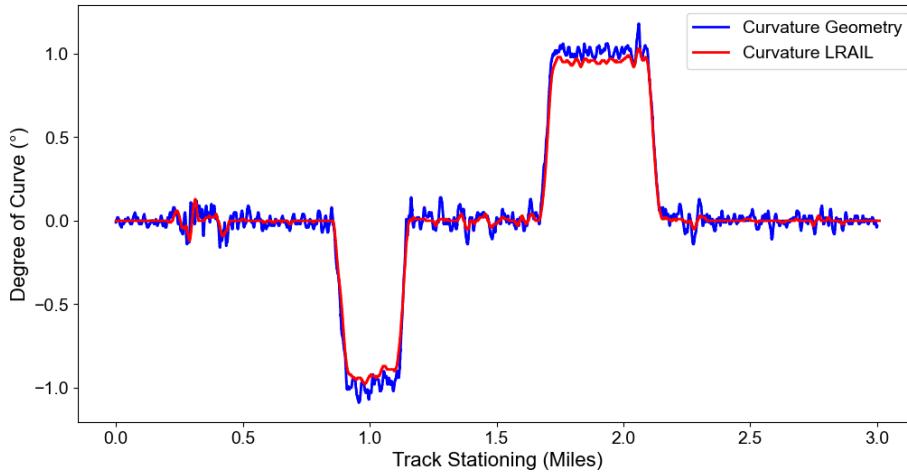
To better consume LR4/L component condition data, and given the lack of available indices, the Track Component Health Index (TCHI) and its sub-indices were developed (Germoglio Barbosa et al., 2024). These sub-indices include the Crosstie Health Index (CHI), Fastener Health Index (FHI), and Ballast Health Index (BHI), and each presents a quantitative measurement of track component condition using a 0-10 scale (Germoglio Barbosa et al., 2024).

Most relevant to my research is the BHI, which compares the ballast height at the crib and shoulder with that of the top of the crosstie and assigns numerical values (Germoglio Barbosa et al., 2024). Since the BHI is reported using a 0-10 scale, compliance and safety thresholds are defined at 6.5 and 2.5, respectively. The BHI uses a 10-meter (33-foot) moving window for all calculations, and equal weighting is assigned to both shoulders and the crib ballast outputs. Thus, lower BHI values may indicate a loss of ballast across the entire section or inadequate ballast in a specific area.

### 2.2.2 Data Alignment

An essential and challenging task in management of condition data for linear assets is data alignment (Palese et al., 2020). Within this dataset, both types of inspections were carried out at different times and using different inspection equipment. The geometry data were aligned prior to data being transferred to our research team, and the milepost and GPS location of those runs were used as reference to align the LR4/L component inspection data. To align the data, a k-dimensional-tree (k-d tree) algorithm was employed to assign milepost to the component data. The k-d tree organizes points in a k-dimensional space for efficient nearest-neighbor searches and recursively partitions points along axes, balancing the tree for optimal performance (Miller et al., 2009). Several searches traverse the tree, pruning branches to effectively locate nearest neighbors. An example of the post-processed data alignment is illustrated below demonstrating a good match between data collected on both tangents and curves ([Figure 2.2](#)). Also, it is shown that each device contains a distinct filter to process the signal based on the present noise.

**Deleted:** Figure 2.2



**Figure 2.2: Example of alignment between LR4IL and geometry data.**

### **2.2.3 Data Aggregation**

Different approaches have been developed, proposed, and used to model track degradation as summarized by Liao et al. (2022). Neuhold et al. (2020) developed an algorithm to identify improvements in track geometry that indicate undocumented tamping actions. The algorithm calculates differences between consecutive standard deviation values for vertical track geometry within a cross section, using sequential measurement runs. When track quality worsens, the following standard deviation value is higher, yielding a positive difference. In contrast, a lower following value suggests an improvement in track quality, signaling a maintenance action. Generally, it is assumed that a linear fit adequately represents the deterioration of the track geometry over time and/or tonnage (Neuhold et al., 2020). In addition, a linear model is easily interpreted and implemented (Jovanovic, 2004; Bilheri et al., 2023). Based on the documented accuracy for modeling track degradation and the ease of implementation for practitioners, I will employ linear regression in this study.

### **2.2.4 Degradation Calculation**

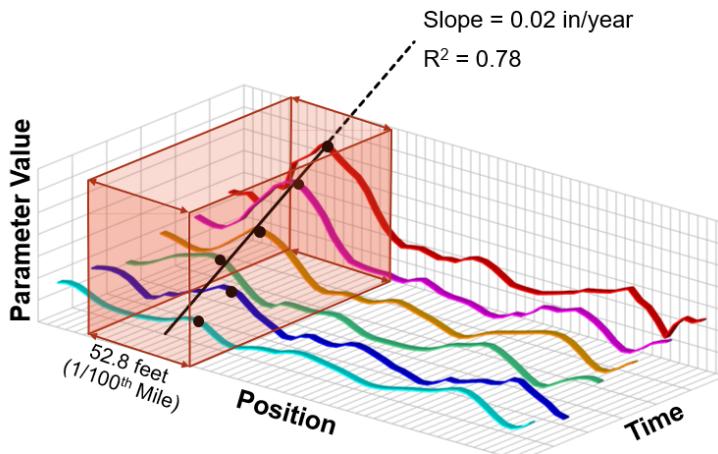
Once the aggregation functions were defined and deployed, it was possible to calculate how each parameter changed over time for each 1/100th of a mile (52.8 ft [16.09 m]) fixed length track section. A linear regression was used as shown in Equation 2.1:

$$Y = ax + b \quad (2.1)$$

Where Y is the dependent variable (i.e., geometry parameter value), a is the slope or degradation rate, x is the independent variable (time in years), and b is the intercept.

A schematic representation of the degradation calculation is presented in [Figure 2.3](#). Each colored curve represents a single inspection run and the parameter's value within the defined window size is combined to plot a linear regression. The degradation rate is then defined as the slope of the regression line (parameter b from Equation 2.1).

**Deleted:** Figure 2.3



**Figure 2.3: Change detection using a linear trend line  
(adapted from Carvalho (2023)).**

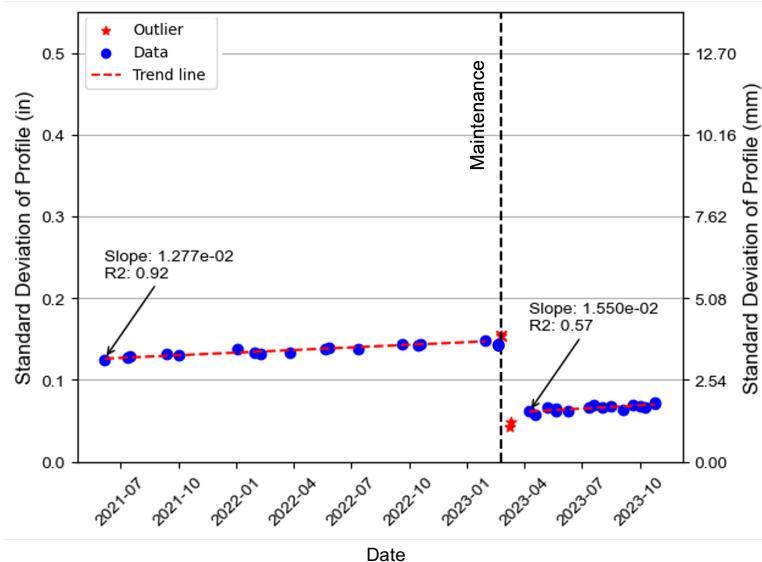
Another essential step in track degradation evaluation is proper identification of track locations that were subject to maintenance. One common metric to quantify track degradation (and restoration) is the standard deviation of profile which is known to be sensitive to maintenance activities. [Figure 2.4](#) provides an example indicating maintenance occurred in late February 2023. To address these stepwise changes in degradation, maintenance events were identified and only the first degradation cycle was included in the analysis (i.e., June 2021 – February 2023 in [Figure 2.4](#)). The first degradation cycle was selected due to more LR4/L inspections being available than were available in the second degradation cycle. Maintenance

**Deleted:** Figure 2.4

**Deleted:** Figure 2.4

identification was performed based on an adaptation of Goodarzi et al.'s (2022) methodology.

Fewer track component measurements were available overall, but this inspection frequency was considered adequate given the expected lower rate of change of these features as compared to track geometry.



**Figure 2.4: Degradation of standard deviation of track geometry profile over time and maintenance identification.**

The maintenance threshold for separating different degradation trends was defined by profile change, which is commonly addressed by ballast dumping and track surfacing (tamping). Ballast dumping directly affects BHI since the BHI is primarily a metric for quantifying ballast level. Sections where BHI suddenly increased just before maintenance were excluded from the BHI slope calculation.

Degradation is represented with a positive slope for geometry features ([Figure 2.4](#)) but a negative slope for component features. This happens since lower values of the geometry parameter indicate a better track, whereas higher values signify worse track, such as a high value on the standard deviation of profile that represents uneven track ([Table 2.1](#)). Conversely, for the BHI, lower values on the 10-point scale indicate poorer track (in terms of ballast profile) while higher values indicate better track.

**Deleted:** Figure 2.4

**Deleted:** Table 2.1

**Formatted:** Font: Not Bold

**Table 2.1 – Geometry and component condition feature names and descriptions.**

Feature	Description
Initial value of SD of profile	The first available value of the standard deviation of the profile (62-foot-chord) at the 1/100th of a mile (16.09 m) at the first degradation cycle in inches or millimetres. Initial value of the feature. The lower the better.
Slope of SD of profile	The degradation rate of the standard deviation of the profile (62-foot-chord) at the 1/100th of a mile (16.09 m) track section at the considered degradation cycle in inches or millimetres per year. The lower the better.
Initial value of BHI	The value of the average Ballast Health Index at the 1/100th of a mile (16.09 m) track section at the first available inspection of the first degradation cycle. Initial value of the feature. The higher the better.
Slope of BHI	The degradation rate of Ballast Health Index at the 1/100th of a mile (16.09 m) track section at the considered degradation cycle in index value per year. The higher the better.

### 2.2.5 Data Analysis

The dataset was divided into curve and tangent sections given these two track conditions are known to perform differently. A boxplot was generated and 25th and 75th percentiles were calculated to provide thresholds for further stratifying the data ([Figure 2.5](#)). To test if the data for curve and tangent track segments is statistically different, the Mann-Whitney U test was used with a significance level of  $\alpha = 0.05$ .

**Deleted:** Figure 2.5

The following null and alternative hypotheses were used:

**H<sub>0</sub>:** The data is statistically equal between the two groups (curve and tangent);

**H<sub>1</sub>:** The data is not statistically equal between the two groups.

To evaluate the relationship between features, the Spearman correlation test was performed and the results were presented as a correlation matrix. The Spearman correlation coefficient is defined by Equation 2.2:

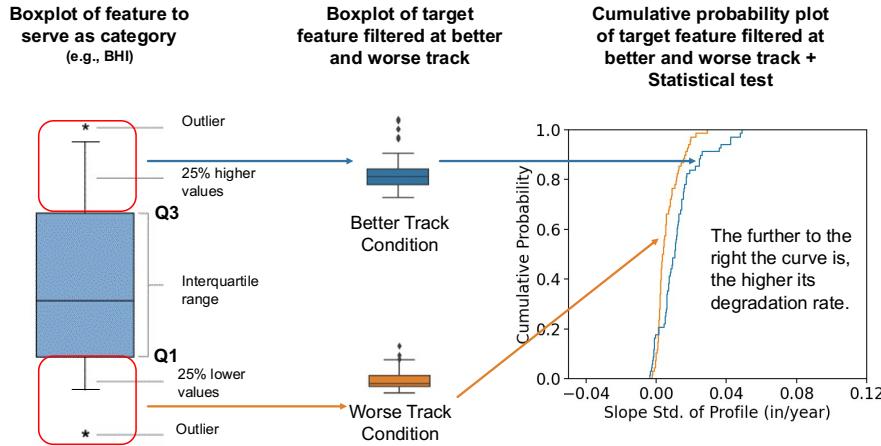
$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)} \quad (2.2)$$

Where  $\rho$  is the Spearman's rank correlation coefficient,  $d_i$  is the difference between the two ranks of each observation, and  $n$  is the number of observations.

Another way to analyze the data is through a stochastic approach. Given the complex set of interactions and the nature of factors influencing railway operations and their associated uncertainties, two-parameter linear comparisons are not always suitable models. Stochastic and probabilistic models offer viable strategies for estimating expected degradation rates by employing probability distributions to represent the overall degradation across parameters (Attoh-Okine, 2017). Usually, categories are defined to evaluate the difference between track segments with different conditions, such as curves and tangents.

In this chapter, I propose a novel method of comparison by analyzing the correlation between change in track geometry and component condition. To conduct this analysis, I convert the component parameter slope into categorical values to determine if the distribution of a geometry parameter's degradation is changing in track sections where component degradation rates are at distinct levels ([Figure 2.5](#)).

**Deleted:** Figure 2.5



**Figure 2.5: Example of the proposed approach for degradation rate comparison.**

To ensure consistency in interpretability for both geometry and component features, two groups were established for track sections demonstrating a "worse condition" or "better condition." The data within the 0-25th and 75-100th percentiles of each feature were used as ranges for the two categories. The "worse condition" category represents track sections with less favorable conditions, such as accelerated degradation rates, higher profile values, or lower BHI values. Conversely, "better condition" track segments have more desirable feature values, such as higher BHI. After defining these categories, the distribution of track geometry slopes in these regions was plotted.

To evaluate if there is a difference between the geometry degradation within the distinct categories, the Mann-Whitney U test was used with a significance level of  $\alpha = 0.05$ . The following null and alternative hypotheses were used:

**H<sub>0</sub>:** The slope (degradation rate) is statistically equal between the two groups (worse or better track condition);

**H<sub>1</sub>:** The slope (degradation rate) is not statistically equal between the two groups.

For the scenarios where the null hypothesis was rejected, the one-sided Mann-Whitney U test was performed to check which track category (worse or better) the geometry slopes are higher. For example, I checked to see if the degradation rate for the profile is lower (less accelerated) at locations with a lower BHI degradation rate. I used the following null and alternative hypotheses for the one-sided Mann-Whitney U test:

**H<sub>0</sub>:** The slope (degradation rate) is statistically equal between the two groups (worse or better track condition);

**H<sub>1</sub>:** The distribution of the degradation rate for the worse condition track group is stochastically greater than the distribution of the degradation rate for the better condition track group.

If the calculated p-value is lower than 0.05, the null hypothesis can be rejected.

## 2.3 Results and Discussion

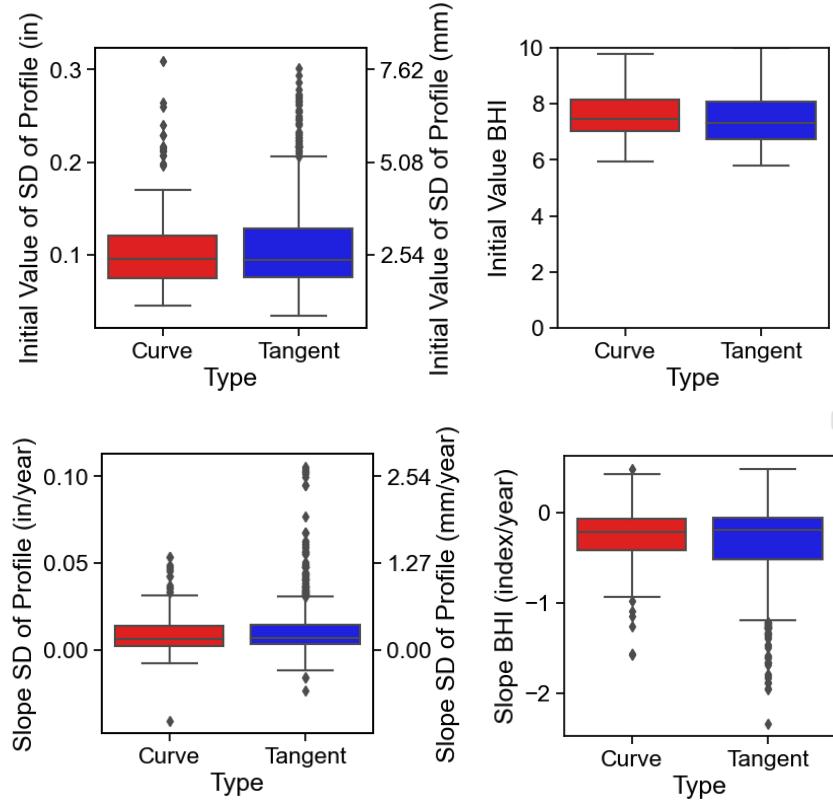
### 2.3.1 Descriptive Statistics

Box plots of the features based on the type of asset ([Figure 2.6](#)) and the 25th and 75th percentiles that serve as condition (worse or better) thresholds are presented ([Table 2.2](#)). The number of samples (individual track sections) for the curve group is 270 while the tangent group is 1,093.

**Deleted:** Figure 2.6

**Deleted:** Table 2.2

**Formatted:** Font: Not Bold



**Figure 2.6:** Boxplot of the different dataset parameters based on classification by track type: tangent or curve.

**Table 2.2:** Threshold limits for worse and better track conditions for tangent and curved track sections.

Feature	Section Type and Condition			
	Curve: Better Condition	Curve: Worse Condition	Tangent: Better Condition	Tangent: Worse Condition
Initial value of SD of profile (in) [mm]	$\leq 0.075$ [1.899]	$\geq 0.120$ [3.056]	$\leq 0.076$ [1.941]	$\geq 0.128$ [3.259]

Slope of SD of profile (in/year) [mm/year]	$\leq 0.02$ [0.056]	$\geq 0.014$ [0.356]	$\leq 0.003$ [0.081]	$\geq 0.014$ [0.361]
Initial value of BHI	$\geq 8.142$	$\leq 7.016$	$\geq 8.099$	$\leq 6.748$
Slope BHI (index/ year)	$\geq -0.068$	$\leq -0.42$	$\geq -0.062$	$\leq -0.519$

The mean values for the variables show low variation between the two track types.

Additionally, the minimum and maximum values for the variables are close, thus demonstrating a low range and tightly grouped data. The Mann-Whitney U test was performed to evaluate the statistical significance of the difference between the curved and tangent parameters ([Table 2.3](#)).

**Deleted:** Table 2.3

**Formatted:** Font: Not Bold

**Table 2.3 - Mann-Whitney U test results for the difference between tangent and curved track.**

Feature	p-value	Statistical Difference
Initial value SD of profile	$3.63 \times 10^{-1}$	No
Slope SD of profile	$2.09 \times 10^{-1}$	No
Initial value BHI	$8.9 \times 10^{-4}$	Yes
Slope BHI	$8.63 \times 10^{-1}$	No

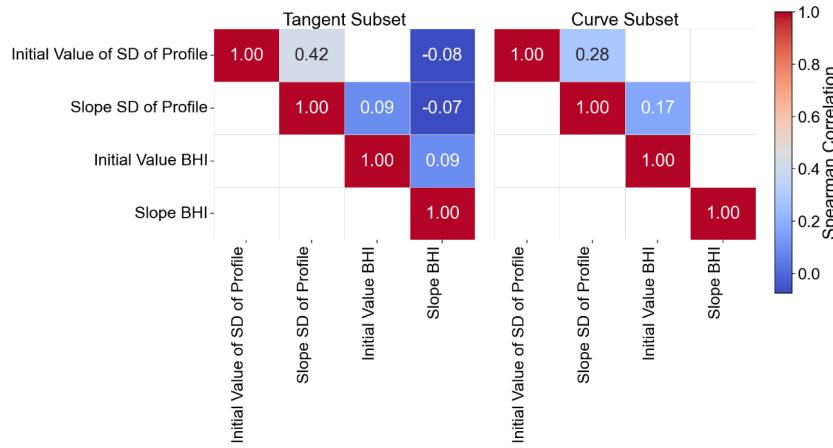
The only feature that presented a p-value below 0.05 was the initial value of BHI. The duration of the degradation cycle may play a role in the degradation rates, where one and a half years could be insufficient to highlight differences in degradation slopes based on curvature on the evaluated parameters.

Given the combined use of initial values and slopes, measurements are segregated based on curve and tangent track for the analysis I describe in the subsequent sections.

### 2.3.2 Spearman Correlation Calculation

The Spearman correlation coefficients between the pairs of parameters on tangents and curves are presented below (Figure 2.7). Only the coefficients presenting statistical significance are shown ( $\alpha=0.05$ ).

**Deleted:** Figure 2.7



**Figure 2.7: Spearman correlation coefficients for different parameters on tangent and curved track segments.**

The analysis of the Spearman correlations reveals distinct relationships for the "Slope SD of Profile (in/year)" across both tangent and curve track locations. In the tangent subset, the "Slope SD of Profile" exhibits moderate positive correlations with the "Initial Value of SD of Profile (0.42)" and weaker positive correlation with "Initial Value BHI (0.09)," indicating that an increase in these initial values is associated with an increase in the slope of the profile's standard deviation. Negative correlation is present with "Slope BHI (-0.07)." In the curve subset, the "Slope SD of Profile" also shows moderate positive correlations with "Initial Value of SD of Profile (0.28)" and weaker positive correlations with "Initial Value BHI (0.17)." These findings highlight the nuanced differences in how the slope of the profile's standard deviation is

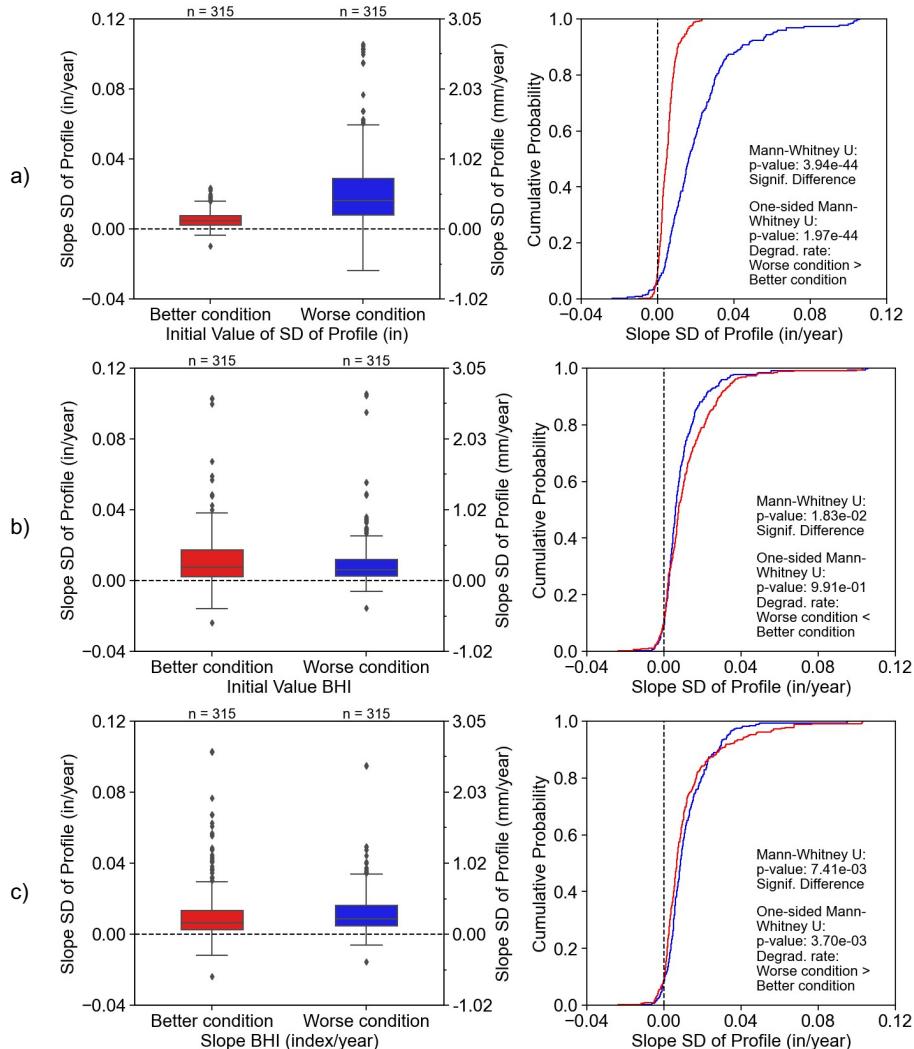
influenced by initial conditions and other parameters in tangent and curve sections.

### **2.3.3 Stochastic Approach**

Since linear regressions do not capture all relationships between parameters, a stochastic method was applied to the data to model the behavior of the track degradation. The parameters considered to influence profile degradation are the initial track geometry profile value, the initial BHI value, and the BHI slope. These features are utilized to categorize track sections into worse and better conditions, with data further partitioned into curved or tangent track segments.

The box plot and the cumulative distribution function (CDF) plots of the slopes of the standard deviation of the profile for tangent sections are presented in [Figure 2.8](#).

**Deleted:** Figure 2.8



**Figure 2.8: Boxplots and CDFs of profile slope categorized per worse and better track condition based on a) Initial value of profile, b) Initial value BHI, and c) BHI slope for tangent track sections.**

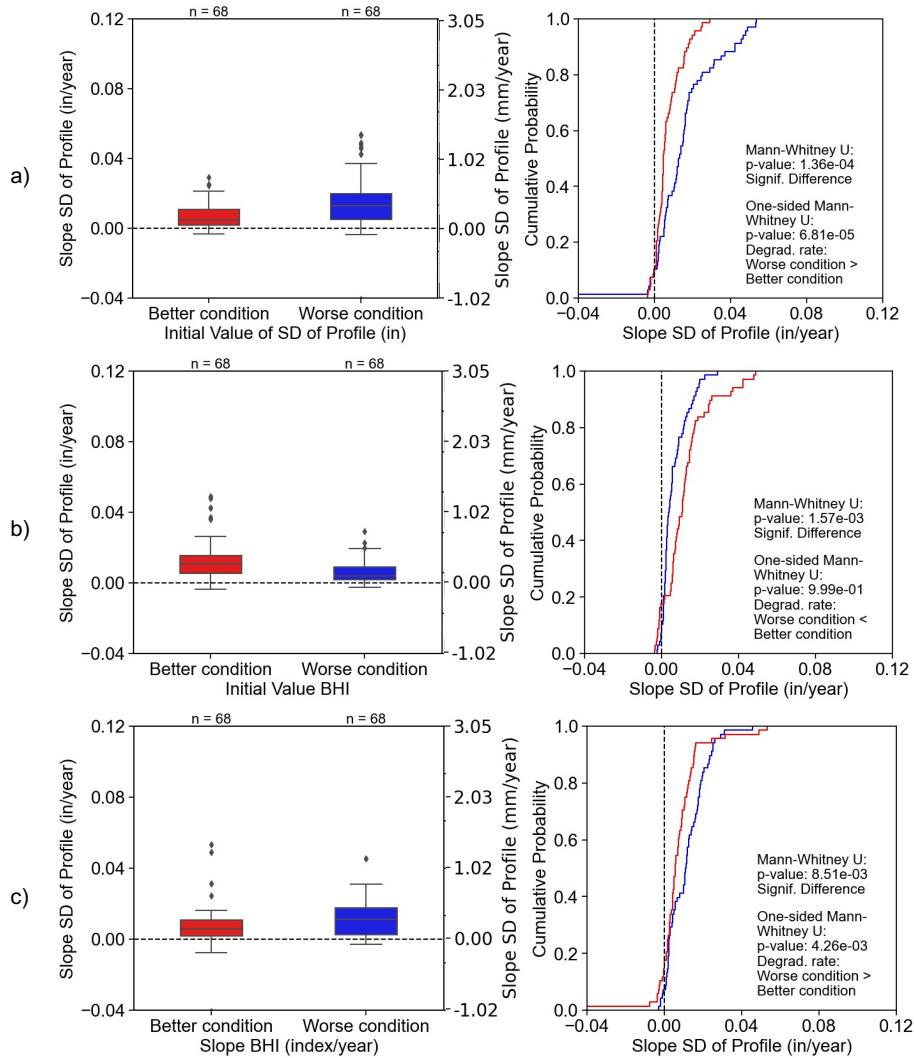
The statistical results presented in [Figure 2.8](#) refer to the statistical difference between the groups and which one presents more accelerated degradation rates. The CDF chart facilitates visual interpretation of the different distributions of values and facilitates the extraction of a degradation rate for a given percentile on the test track. When two curves are present, the most rightward curve represents the subset of track conditions with the higher degradation rate.

**Deleted:** Figure 2.8

The p-values of the Mann-Whitney U tests were lower than 0.05 in all cases, thus, there is sufficient evidence to conclude that the profile degradation rate is different between the groups. Regarding the one-sided Mann-Whitney U tests, the groups categorized as worse track regions based on the initial value of the profile and slope of BHI were also the regions that presented more accelerated profile degradation. On the other hand, regions with a higher initial BHI value showed a higher degradation rate. Further discussion of these results will be presented after the curved track segment results.

The box plots and CDF plots of the slopes of the standard deviation of the profile for curved sections and the results of their statistical tests are presented in [Figure 2.9](#).

**Deleted:** Figure 2.9



**Figure 2.9: Boxplots and CDFs of profile slope categorized per worse and better track condition based on a) Initial value of profile, b) Initial value BHI, and c) BHI slope at curved track sections.**

The p-values from the Mann-Whitney U tests were lower than 0.05 in all cases, indicating a difference between the groups. Furthermore, the groups categorized as worse track regions based on the initial value of profile and slope of BHI also presented more accelerated profile degradation. Conversely, locations categorized as better track conditions by initial BHI presented a higher degradation rate.

The most significant difference was observed for tangent track segments grouped by initial value of the standard deviation of profile ( $p$ -value =  $3.94 \times 10^{-44}$ ). For these sections, at a probability of 90%, the degradation rate would be 0.044 in/year (1.117 mm/year) at worse initial values and 0.011 in/year (0.279 mm/year) at locations exhibiting better track conditions. This represents a 300% increase in degradation rate.

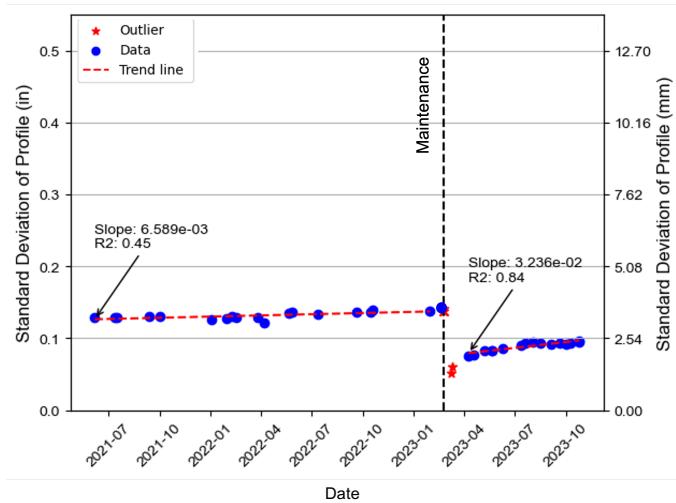
Initially, the impact of the initial profile value might suggest a non-linearity in its degradation, thus implying that the degradation rate is more accelerated in places with initial higher values (worse condition). Prior research has shown that profile degradation typically accelerates exponentially during initial settlement after tamping (Caetano and Teixeira, 2016). One hypothesis could be that locations with higher degradation rates are coincident with regions containing drainage or other substructure deficiencies since they are a potential source of accelerated degradation (Goodarzi et al., 2023). However, locations with worse values for BHI actually presented lower profile degradation rates.

This behavior could be investigated by considering all three features simultaneously. Locations with higher BHI values may have undergone more frequent or more recent maintenance such as surfacing or track raising. Prior research indicates that maintenance activities can significantly alter the degradation rate of profile standard deviation due to factors like ballast breakdown, fines interaction, and drainage (Selig and Waters, 1994; Audley and

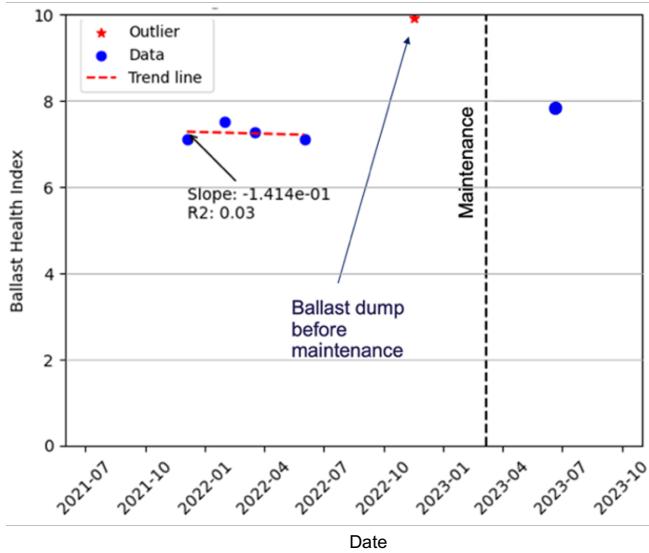
Andrews, 2013; Soleimanmeigouni et al., 2020; Wilk and Li, 2021; Guo et al., 2021). Therefore, this is a likely explanation for the profile degradation results. In that sense, monitoring profile is valuable for proper identification and prioritization of locations requiring surfacing. [Figure 2.10](#)

**Deleted:** Figure 2.10

a)



b)



**Figure 2.10: Example of track maintenance effects on the geometry profile (top) and ballast health index (bottom).**

## 2.4 Conclusions

Over the past several decades, track degradation research has focused on individual track geometry attributes and their degradation over time and tonnage, with little focus on the degradation of components or interactions between component and geometry degradation. To add to the body of research and overcome some of the prior limitations, I presented a stochastic modeling approach to transform continuous features into categories to evaluate their relationship with geometry degradation. The advantage of the stochastic model is the application of a probabilistic approach and its ability to model the inherent variability of field data. This stochastic method, used in conjunction with the Mann-Whitney U test, proved to be more robust for identifying relationships than a traditional correlation matrix. Based on the results from a Class I railroad revenue service test section, the following observations were made:

- Stochastic models proved useful for modeling track geometry degradation and evaluating its change related to track component initial conditions and degradation rates.
- Track profile degradation rates are statistically significantly (more accelerated) in regions with higher (worse) initial track geometry profile. This indicates the importance of monitoring and managing this parameter with respect to surfacing maintenance planning.
- Regions with accelerated BHI degradation are associated with higher profile degradation rates. However, regions with a higher initial value of BHI showed higher profile degradation. Therefore, areas that have more frequent maintenance and maintain a higher BHI may experience increased profile degradation.

Building on these results, future research should analyze the relationship between changes in additional geometry and component features, especially in regions with more variability in crosstie and fastening system types and conditions. Also, applying novel data science and machine learning models to analyze the effect of change by combining several features at once could improve our understanding of this relationship. Modeling track degradation can also provide valuable insights into enhancing safety measures in the railway industry.

Overall, insights from this study can contribute to the development of more effective maintenance scheduling and practices for the rail industry, further improving its safety and state of good repair, and bolstering its role as a reliable mode of transportation for people and goods.

## **CHAPTER 3: CROSS-CORRELATION-BASED RAILWAY CHANGE DETECTION: A NOVEL APPROACH FOR MAINTENANCE DETECTION**

### **3.1 Introduction**

As described in Section 1.3, cross-correlation was found to be an effective tool for aligning data from multiple datasets that were captured across different platforms (Palese et al., 2020). Based on the success of its application to data alignment, and the need to determine which portions of the track infrastructure were the subject of maintenance, I studied cross-correlation's applicability for maintenance detection. By identifying changes in track condition data before and after maintenance activities, cross-correlation can help isolate the sections of track where maintenance has impacted the geometry signal pattern. Detection of maintenance is crucial to degradation trend analysis (Neuhold et al., 2020; Goodarzi et al., 2021) and can aid in the precise evaluation of maintenance effectiveness for interventions such as tamping. These will allow for further optimization of maintenance schedules based on the actual track condition and effectiveness of treatments applied to it. These data open new avenues for improving the current state of practice for track monitoring and quantification of maintenance activities and effectiveness, enhancing both efficiency and safety.

Neuhold et al. (2020) used the standard deviation of the profile calculated over a 200-m (0.12 mi) window to identify non-recorded maintenance activities. The method was based on the difference between consecutive measurements by subtracting the previous value from the following one. Negative values were identified as potential maintenance activities. To improve the method, the authors incorporated outlier detection using the mean absolute deviation (MAD). For each measurement, the algorithm examined the differences between the subject value and its

adjacent measurements. If both differences exceed the MAD and one is negative, the measurement is flagged as an outlier and removed. Next, the process was repeated but using as a threshold the new calculated MAD multiplied by a factor of 0.79. Authors reported the identification of 99% of maintenance activities on a 200-km (124.3 mi) track. The number of false positive results was not presented.

Soleimanmeigouni et al. (2020) developed a data-driven methodology for predicting isolated track geometry defects. They used data from a Swedish railway, collected from January 2015 to July 2018 over 82 km (50.9 mi) of track. Track sections were categorized as good, with a standard deviation of longitudinal level less than 80% of the lower limit for corrective maintenance, and poor, with a standard deviation of longitudinal level larger than 80% of the lower limit for corrective maintenance. For a good track section, a 15% reduction in the standard deviation was considered a tamping intervention. For poor track sections, Equation 3.1 was used to identify undocumented tamping interventions, whenever the tamping ratio satisfies the inequality.

$$TR = \frac{SDLL_a}{SDLL_b} < 0.9 - \frac{0.16}{SDLL_b} \quad (3.1)$$

Where TR is the tamping ratio, SDLL<sub>a</sub> is the SD of longitudinal level after tamping, and SDLL<sub>b</sub> is the SD of longitudinal level before tamping.

Goodarzi et al. (2021) used a method to identify maintenance events based on the SD of profile calculated over a 200-ft (60.9 m) window. Their track geometry data were gathered from two ballasted passenger tracks in the U.S. from 2011 to 2020, totaling 287 miles (461.9 km). Their threshold for maintenance detection was defined by Equation 3.2.

$$\frac{SDP_1 - SDP_2}{SDP_1} \times 100 > 20\% \quad (3.2)$$

Where SDP<sub>1</sub> and SDP<sub>2</sub> are profile standard deviations of two consecutive inspections.

In another publication, Goodarzi et al. (2022) presented a different method to identify previous maintenance activities also based on the standard deviation of the profile of two consecutive runs ( $SD_1$  and  $SD_2$ ). In this research, more conditions were included to flag a maintenance event. First, if  $(SD_2/SD_1) < 0.7$  and  $(SD_2-SD_1) > 0.03$  a maintenance event was flagged. Additionally, if  $(SD_2/SD_1) < 0.9$ ,  $(SD_2-SD_1) > 0.03$ , and the variance of the four last inspection SD values is less than  $4 \times 10^6$  the location was identified as having maintenance. This same methodology was used to identify maintenance events in research focused on understanding the relationship between track geometry degradation rates based on GPR and LiDAR data (Goodarzi et al., 2023).

In the previous research I described above (Soleimanmeigouni et al., 2020; Goodarzi et al., 2021; Goodarzi et al., 2022; Goodarzi et al., 2023), the authors did not present the performance of their maintenance detection models quantitatively. Also, cross-correlation calculation was presented as an intermediate step for data alignment in their papers.

Schatzl et al. (2024) presented research evaluating three approaches for dealing with unrecorded maintenance detection. The first method, namely SEARCH algorithm is a further improvement of the method developed by Fellinger (2020). The second one is based on the Cross-Section- and RANSAC-Based (CRAB) Algorithm (Fischler and Bolles, 1981). The last method is based on the work of Loidolt et al. (2023) and is referred to as the Cumulative Track Geometry-Based Algorithm (CTG). The F1-Score was used as a metric to compare the methods. All the methods used the standard deviation of the longitudinal level ( $SD_i$ ) for each iteration date i.

Developed by Fellinger (2020), the SEARCH algorithm iterates throughout the standard deviation values calculated for each track cross-section over time, including the next available

point at each iteration. Five conditions are verified for each iteration. The first condition is applied when only two points are available, if the  $SD_2$  value is 0.25 mm (0.01 in) or lower than  $SD_1$ , a maintenance event is recorded. Rule 2 analyzes datasets when the iteration achieves 3 to 5 points. A linear regression is calculated using all but the final point to predict its value. If the final point shows an improvement of more than 0.25 mm (0.01 in) over the predicted value, and the next point shows at least a 0.1 mm (0.004 in) improvement from the prediction, it indicates a tamping action before the last point. Rule 3 applies if a tamping action is detected and the first two points in the dataset do not meet any conditions. The first point is excluded if three criteria are met: the quality improvement from  $SD_1$  to point  $SD_2$  is greater than from  $SD_0$  to  $SD_1$ ,  $SD_2$  is lower than  $SD_1$ , and the improvement from  $SD_1$  to  $SD_2$  exceeds 0.05 mm (0.002 in). Rule 4, like Rule 3, requires a detected tamping activity and three points in the dataset. It excludes the second point if three conditions are met: the quality improvement from  $SD_2$  to  $SD_3$  is greater than the improvement from  $SD_0$  of the previous deterioration to  $SD_2$ ,  $SD_3$  shows better quality than  $SD_2$ , and the improvement from  $SD_2$  to  $SD_3$  exceeds 0.05 mm (0.002 in). Rule 5 is the main rule for detecting tamping actions and requires at least four points without previously-detected tamping. A linear regression is calculated using all but the final point to predict its value within a 99.5% confidence interval. If the actual value exceeds the predicted range, an outlier or tamping action is probable. If the following point also surpasses the predicted quality range, a tamping action is confirmed, and all but the final point are removed from the dataset. If the final point is an outlier, it is excluded from future analysis. The improvement applied from the authors (Schatzl et al., 2024) over Fellinger et al. (2020) was including recorded maintenance to the data as well.

The CRAB algorithm process starts by using known locations where maintenance has been recorded and splitting data into different degradation cycles the authors refer to as “rooms” (Schatzl et al., 2024). If no maintenance record is available, the whole period is considered as one room. For each room, every possible combination of two measurement points represented by the SD of the longitudinal level is selected iteratively and a linear model is fitted to them to define an interval (1/3 of the SD value). All data points in the room within the defined interval are recorded, and after all points are iterated, the set of data more frequently identified as contiguous is defined as a degradation branch. Measurements within the same period of the defined degradation branch that fall outside of the interval are considered as outliers and discarded. The following points, outside of the defined branch period, are submitted to a second iteration using the same process to define other potential degradation branches. As a second step, three measurement points are employed by different combinations across the room, with first and third points defining the straight line. The vertical distance between the second data point and the defined regression line is calculated, and if the value is between half and double the whole room’s standard deviation calculated using the data below the linear fit line, maintenance was executed prior to the second data point. If the second point is below the lower limit, the data point is considered an outlier and discarded.

The CTG algorithm calculates the cumulative sum of the square roots of the longitudinal level. The calculated index is called the Cumulative Index (CI) and the length of the section can be selected arbitrarily. The gradient of cumulated curves reflects track geometry quality, where steeper gradients indicate lower quality. Differences between two curves may signal track degradation or recent maintenance. To identify maintenance events, CI values are subtracted to create a difference signal (DCI) which is smoothed over a 100-meter (328.2 ft) length to reduce

noise. Based on the analysis, three scenarios are possible. In Case 1, no maintenance occurs, thereby leading to rapid track deterioration and presenting a positive gradient in the DCI signal due to increased longitudinal level amplitudes. In Case 2, maintenance is performed, thus reducing the longitudinal level and resulting in a lower CI gradient after maintenance and a negative DCI gradient in the affected area. Case 3 indicates track condition following a track renewal where new components minimize deterioration, leading to almost identical and flatter CI gradients with the DCI showing a low but positive gradient (Schatzl et al., 2024).

In this chapter, I will investigate the use of cross-correlation as a maintenance detection tool. I will apply the approach using revenue service data from a Class I railroad mainline as the subject test corridor. I will rely on two datasets for this study: track geometry inspection data and track component inspection data. Data were collected over 100 miles (160 km) of track between June 2021 and July 2024 on a Class I freight railroad's primary corridor in the southeastern U.S. The track is maintained to FRA Track Class 4 with a maximum freight train operating speed of 60 mph (97 km/h) and is constructed primarily of timber crossties. A few locations with concrete crossties are present. Approximately 90 MGT of freight are transported on this segment of track each year.

### 3.2 Material and Methods

#### 3.2.1 Data Collection, Storage, and Alignment

In this chapter, I followed the same data collection, storage, and alignment methods previously described and used in Sections 2.2.1 and 2.2.2. The data used on the cross-correlation calculation is the average of the left and right profile values measured on a mid-chord offset of 62-ft (18.9 m).

**Commented [JV4]:** @Edwards, John Riley they were not removed, in here, all single track was used

### 3.2.2 Data Analysis

#### 3.2.2.1 Cross-correlation Calculation

A cross-correlation function (CCF) is a statistical measure that assesses the similarity between two time series as a function of the time-lag applied to one of them. A CCF helps to identify the degree of correlation between the two series at various shifts. Mathematically, for the time series  $X(t)$  and  $Y(t)$ , the cross-correlation coefficient  $p(k)$  at lag  $k$  is given by Equation 3.3.

$$p(k) = \frac{\sum_{t=1}^N (X_t - \mu_X)(Y_{t+k} - \mu_Y)}{N\sigma_X\sigma_Y} \quad (3.3)$$

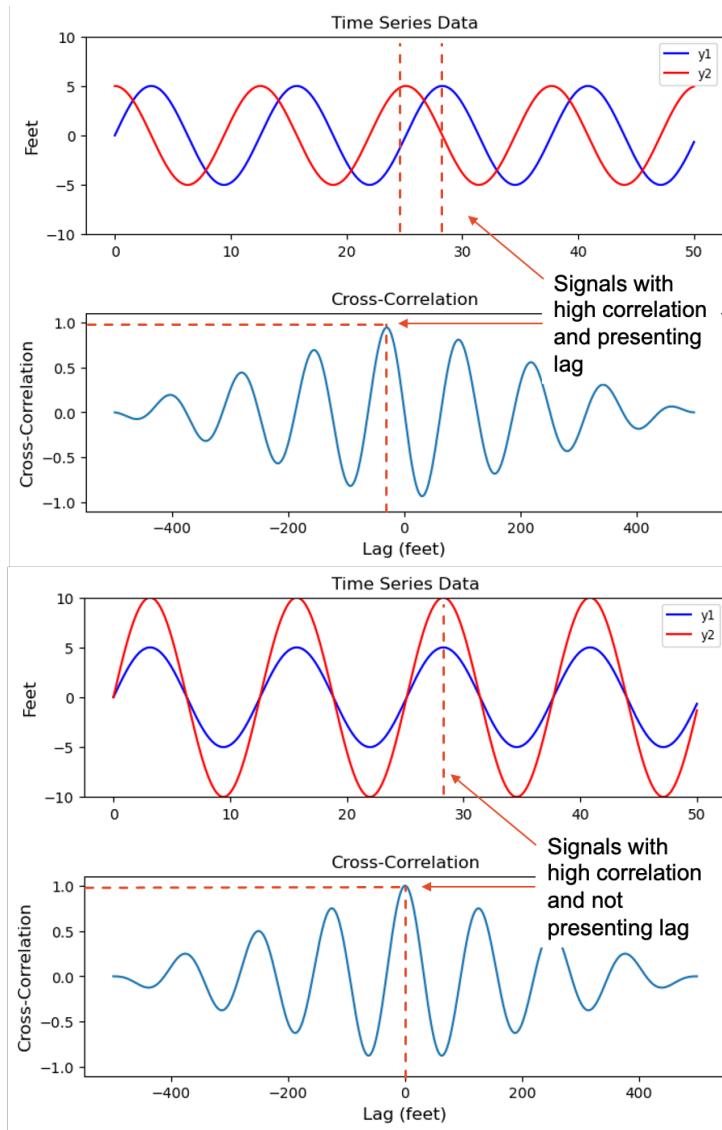
Where  $N$  represents the number of observations,  $\mu_X$  and  $\mu_Y$  are the means of  $X$  and  $Y$ , and  $\sigma_X$  and  $\sigma_Y$  are their standard deviations, respectively.

Lag refers to the amount one time series is shifted relative to the other. A positive lag indicates that the second time series is shifted forward, while a negative lag means it is shifted backward. The lag between two series can be calculated as the lag  $k$  associated with the maximum CCF coefficient.

The CCF coefficient provides insight into the lead-lag relationship between the two series. If the coefficient at a specific lag  $k$  is high, it suggests that the value of  $X(t)$  is highly correlated with  $Y_{t+k}$ , which implies that changes in  $X$  may precede changes in  $Y$ , or vice versa.

When the coefficient is close to +1 or -1, it indicates a strong linear relationship at that lag that is either positive or negative, respectively. A value close to zero suggests little or no linear relationship at that specific shift. [Figure 3.1](#) presents examples of different cross-correlation and lags between two time-series signals. We can notice that the presence of lag in the first example and the change in amplitude in the second example did not decrease the CCF coefficient.

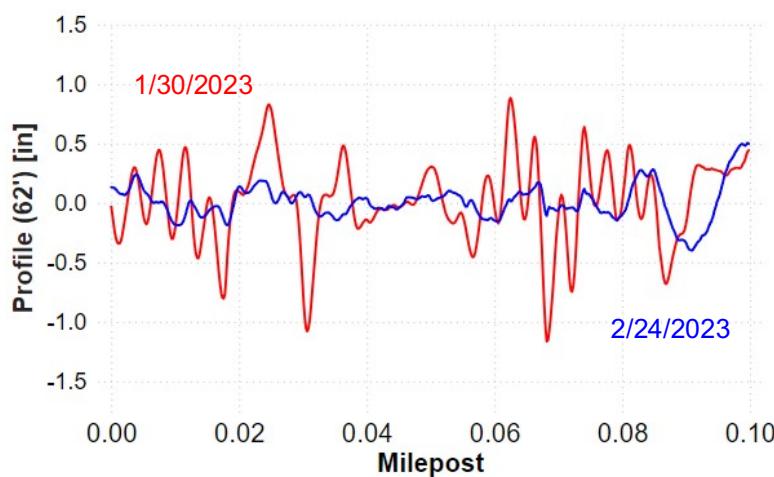
**Deleted:** Figure 3.1



**Figure 3.1: Example of CCF evaluation on two signals.**

In the case of continuous track measurements, such as gauge and profile, changes can be identified by selecting a window length with enough data to highlight the signal changes. Different approaches have been used to model track degradation, as summarized by Liao et al. (2022). Researchers presented that after a tamping event, there is a logarithmic behavior on track degradation however, it might be replaced by a more accelerated linear degradation followed by a shallower one (Xu et al., 2012; Neuhold et al., 2020). For computing optimization, not all available inspections are being used in the present research, but a sample per month, in that case, the non-linear behavior of the curve is likely to not be captured. Thus, it is assumed that a linear fit is a fair approximation to represent the deterioration of the track geometry over time and/or tonnage, as supported by previous research (Neuhold et al., 2020). Thus, natural track degradation over time and tonnage is not expected to disrupt the signal identity, only increasing its magnitude (Palese et al., 2020) and keeping a high CCF coefficient. However, human-induced intervention or an extreme weather event (e.g., tornado or floods) will disrupt the signal identity and present a low CCF coefficient ([Figure 3.2](#)).

**Deleted:** Figure 3.2



**Figure 3.2: Section of track where the profile measurement signal changed after a potential maintenance activity.**

A key element in the proposed approach is the window size, which should reflect the length of a typical maintenance activity (e.g., surfacing) and include relevant parameters (e.g., track profile) and include enough relevant data to accurately capture the signal change.

According to Andrews (2013), there are different ways to correct track alignment when it falls outside acceptable tolerances including mechanized tamping or localized manual tamping. When ballast is no longer functional, more comprehensive measures such as shoulder cleaning and undercutting are necessary. Machine-based maintenance is faster, more accurate, and less labor-intensive than manual options. Tamping machines are particularly effective, using vibrating arms to pack ballast under sleepers, aligning rails, and creating level tracks. Large capitalized workgroups usually cover many miles per day (CSX Transportation, 2020; Union Pacific Railroad, 2022). Conversely, manual interventions can cover shorter lengths of track, to address local exceptions. Ballast dumping often covers short distances of track where there is a deficit, on the order of 0.1 mile (160.9 m).

In addition to window length, the *type* of window is an important parameter to establish. When modeling track degradation, windows can be defined as either *moving* or *fixed (discrete)* (Offenbacher et al., 2020). A moving window adjusts its position incrementally across the dataset, capturing overlapping sections of data as it moves. This approach allows for the detection of degradation patterns that might emerge at the boundaries of adjacent windows. In contrast, a fixed window remains static, analyzing each defined segment separately without overlap. Moving windows improve the likelihood of identifying events located at the boundary between two windows. For instance, Neuhold et al. (2020) used 100 m (328 ft) sliding windows

calculated every 5 m (16.4 ft). On the other hand, FRA TQI uses a 528 ft (160.9 m) window size (Sung, 2005).

Moving windows require a step value to determine how much the window shifts with each iteration. Setting a very small step, such as 1 foot (0.3 m), would be computationally intensive and likely unnecessary. For example, with a 528-foot (160.9 m) window, a 1-foot (0.3 m) step would change less than 0.5% of the dataset at each iteration. This would have minimal effect on the CCF coefficient. The optimal step size could be determined using actual data, balancing the resolution of change identification with computational efficiency to determine a window size that can capture meaningful changes.

Fixed windows evaluate each segment independently and may be more suitable for applications where segments are distinct and non-overlapping by nature. In this research, I will use a fixed window approach to facilitate the assessment of degradation cycles within track sections containing fixed assets known to contribute disproportionately to maintenance costs, such as turnouts, bridges, and crossings. As an example, Goodarzi et al. (2021) used a 200-ft (60.9 m) fixed window and removed windows containing fixed assets. Fixed segmentation allows for targeted analysis of critical areas where degradation patterns may differ due to structural influences and differences in the track's condition and loading demands.

Sections where data recording was not performed during certain months can prove challenging when developing track degradation models, as inconsistent subset lengths may lead to misleading interpretations. To address this problem, segments with less than 85% coverage of the defined window length were excluded from the dataset. This filtering ensures that only sections with sufficient data representation are included in the analysis, thereby reducing the risk of biased or inaccurate degradation assessments. Furthermore, in this case of my research,

several inspections are available each month and only the first available inspection for each section is kept. In case the first inspection did not cover the whole section, the following inspection was used to complete the dataset.

As mentioned earlier, the main use of cross-correlation on linear assets is the evaluation of the lag between runs. In that sense, the equation quantifies the maximum lag between each window for consecutive runs. It is unlikely that a high lag is present since a specialized third-party company previously aligned the data. Based on this knowledge and prior experience with the geometry dataset, the evaluated lag was limited to  $\pm 50$  ft ( $\pm 15.2$  m), which is approximately 10% of the window length. The lag calculation will also allow us to check the correlation between the calculated cross-correlation and lag.

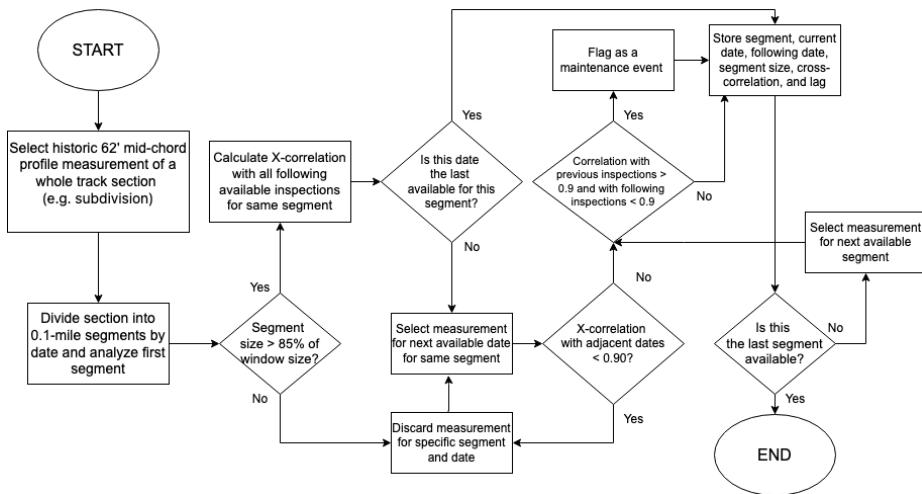
The CCF coefficient calculation process starts with the algorithm selecting a subset of data with the defined window size for the first available date. Then, this subset will be compared with the equivalent subset on all following and previous inspections with the cross-correlation between them recorded. This process is then repeated for the following track windows. Lag and window size are also recorded. Microsoft PowerBI was used to display the results but different data management tools, such as Python, Tableau, and Excel can also be used.

Neuhold et al. (2020) emphasized the importance of data quality when assessing track measurement data over time, as unfiltered data can lead to inaccurate conclusions. To resolve this, the authors enhanced their algorithm to identify and remove outliers using the MAD values (Neuhold et al., 2020). Goodarzi et al. (2021) used Cross Power Spectral Density (CPSD) to evaluate the similarity between two consecutive signals and consequently discard unreliable data.

In my research, the cross-correlation coefficient value was used to identify outliers. Since the natural degradation of track is not believed to drastically change between two consecutive

inspections (Palese et al., 2020), events of low correlation with respect to adjacent measurements for the same location on different dates were removed from the dataset. [Figure 3.3](#) illustrates the maintenance detection approach I am developing and applying in this thesis. The values of thresholds presented will be explained on the following sections.

**Deleted:** Figure 3.3



**Figure 3.3: Flowchart of the proposed approach for maintenance detection.**

### 3.2.2.2 *Ground-truth Verification*

The next step was defining a threshold that would properly identify locations with a high likelihood of maintenance. To do so, descriptive statistics were calculated, and the value considered as the lower limit of CCF coefficient was defined as the threshold. The lower limit was calculated based on the Equation 3.4.

$$\text{Lower Limit} = Q1 - 1.5 * \text{IQR} \quad (3.4)$$

Where Q1 is the value of the first quartile and IQR is the interquartile range, or difference between the first and third quartiles of the data.

Different methods can be used to ground truth a model. I employed two approaches to verify if maintenance occurred. First, I checked the signal difference between the two different dates where the change was shown to occur. I also checked the previous and following dates to ensure there was an effective change in the track's state that was kept forward – as opposed to a signal quality issue for a singular inspection run that would be corrected for subsequent runs. To verify this, a heatmap of correlation coefficients with all pairs of inspection dates for each section was developed.

The second method I used was checking the degradation of the track segment based on the SD of the track profile measurement. The use of the SD as an aggregation function is a common way to summarize a track quality indicator and has seen widespread use by previous researchers (Andrews, 2013; Neuhold et al., 2020; Goodarzi et al., 2021; Goodarzi et al., 2022). Using this method, dates where tamping maintenance occurs are highlighted by a marked decrease in this parameter. The SD of the profile was calculated over the fixed windows of 528 ft (160.9 m) using Equation 3.5.

$$SD = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}} \quad (3.5)$$

Where  $x_i$  is the current value,  $\mu$  is the mean, and  $n$  is the number of values.

True positives (TP) and false positives (FP) were considered to assess the model's performance. In cases where there is no clear change in the linear degradation of the SD of profile, but the CCF coefficient calculated is lower than the adopted threshold, this event will be classified as a FP. Events where a clear change in the degradation is present, but a high CCF coefficient is generated will be classified as a false negative (FN). The metrics used to summarize the model performance were precision, recall, and F1 score, as defined in Equations 3.6, 3.7, and 3.8, respectively.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (3.6)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (3.7)$$

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3.8)$$

Precision is a metric that gives the proportion of TPs to the number of total positives the model predicts. Recall is a metric that measures the model's ability to correctly identify all relevant instances within the dataset, calculated as the ratio of TPs to the sum of TPs and FNs. It is particularly useful in cases where missing relevant instances is costly. The F1 score is the harmonic mean of precision and recall, providing a balanced metric that reflects both the accuracy of positive predictions and the model's completeness in identifying relevant instances. This score is especially valuable when precision and recall are both critical and should be appropriately balanced. I did not evaluate true negative (TN) values because they represent all the other instances and do not add to the interpretation of the model in this context.

In addition to the CCF coefficient threshold defined by the lower limit, intermediate values were defined to evaluate the optimal value to be adopted. The F1 score was the metric used to choose the optimal CCF coefficient threshold.

### **3.2.2.3 Linear Degradation Change**

Finally, I used linear modeling to assess the effectiveness of the method in defining different degradation cycles. I calculated the linear degradation for geometry parameters considering both the entire time frame as well as splitting it based on the maintenance detection.

To split the degradation cycles, the CCF coefficient value associated with the best F1 score was used. The degradation slopes for the dataset were calculated following the methodology

described in Section 2.2.4. For track profile, the aggregation function used was the SD of the values present at each date for each 0.1-mile (160.9 m) section. Four different parameters were used in the evaluation of the slopes: slope value (i.e., degradation rate), coefficient of determination ( $R^2$ ), root mean square error (RMSE) and mean of absolute percentage error (MAPE). Only degradation cycles with at least three measurements were considered in this analysis since using two or fewer points would decrease the reliability of the results.

$R^2$  was chosen as it is one of the most common metrics for evaluating the goodness of fit in regression analysis. However, slopes with minimal variation can yield a low  $R^2$  value, leading to misinterpretations of the model's fit. In such cases, alternative metrics or additional contextual interpretation may be necessary. The  $R^2$  value can be calculated using Equation 3.9.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.9)$$

Where  $\hat{y}_i$  is the point on the regression line;  $\bar{y}$  is the mean value; and  $y_i$  is the measuring point.

RMSE is a useful parameter for assessing the magnitude of prediction errors. Its main advantage lies in its intuitive interpretation as an average error in the model's units, making it easy to understand, interpret, and compare across models. However, RMSE is highly sensitive to low values, especially those close to zero, which can disproportionately influence the overall error measure. RMSE values can be calculated using Equation 3.10.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (3.10)$$

Where  $\hat{y}_i$  is the point on the regression line;  $y_i$  is the measuring point, and  $n$  the number of samples.

MAPE is an effective metric for evaluating model accuracy by quantifying prediction errors as a percentage of the actual values. Its key strength lies in its straightforward interpretation, allowing for easy comparison across different models and datasets, regardless of their scale. However, MAPE tends to be problematic with low actual values, where even small deviations in predictions can result in inflated error percentages, skewing the overall assessment. Equation 3.11 illustrates how MAPE is calculated.

$$MAPE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n y_i} \times 100 \quad (3.11)$$

Where  $\hat{y}_i$  is the point on the regression line;  $y_i$  is the measuring point, and  $n$  the number of samples.

Finally, the slope—or degradation rate—of the fitted equation is included because the natural behavior of track profile standard deviation is expected to increase over time due to degradation. It is defined by the  $a$  coefficient in Equation 2.1. A negative slope, suggests an unexpected trend, potentially indicating data anomalies or maintenance, and thus should be disregarded from degradation analysis.

The combined use of these parameters and the degradation plots helps to interpret the benefits of the change detection method for track degradation evaluation. Moreover, it can bring insights about thresholds that could be adopted based on these metrics to discard poor fitting data.

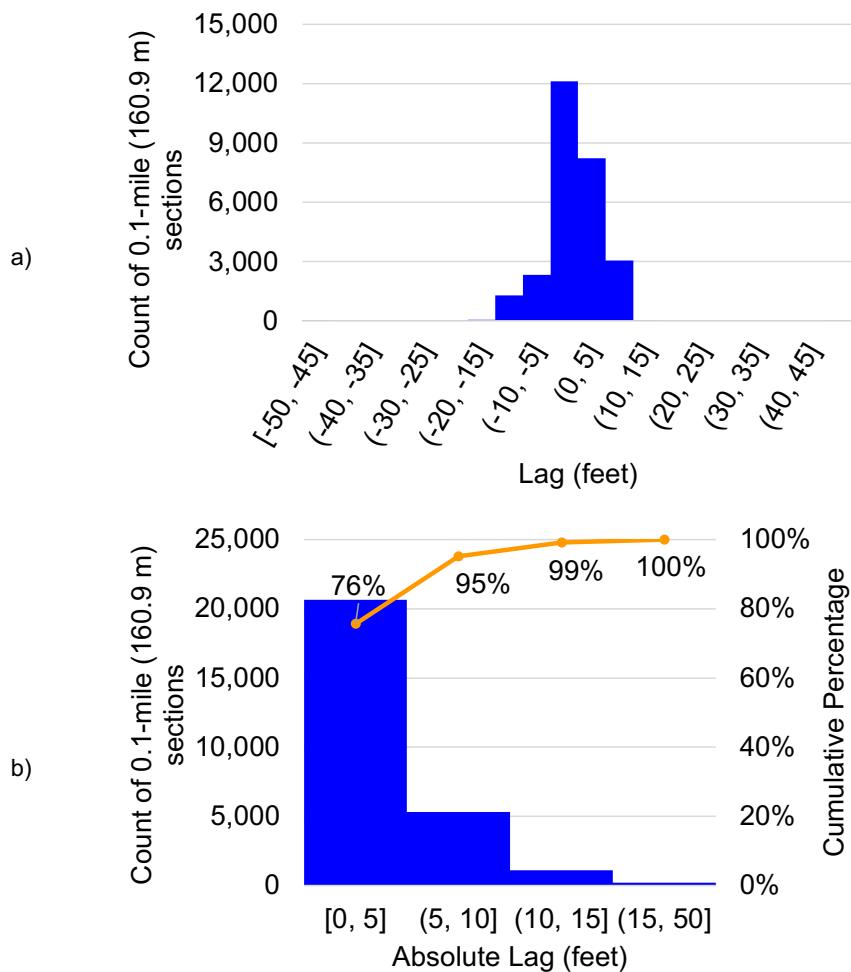
### 3.3 Results and Discussion

#### 3.3.1 Cross-correlation and Lag Results

To date, and as described previously, the primary use of cross-correlation calculation for linear assets is for data alignment. For the current dataset, the maximum lag was limited to 50 ft (15.24

m) which represents approximately 10% of each window's length. [Figure 3.4](#) presents the distribution of the lag associated with the maximum cross-correlation for each window with its consecutive inspected date and Figure 3.4 (separate them) shows the cumulative distribution of the absolute value of lag for the same values.

**Deleted:** Figure 3.4

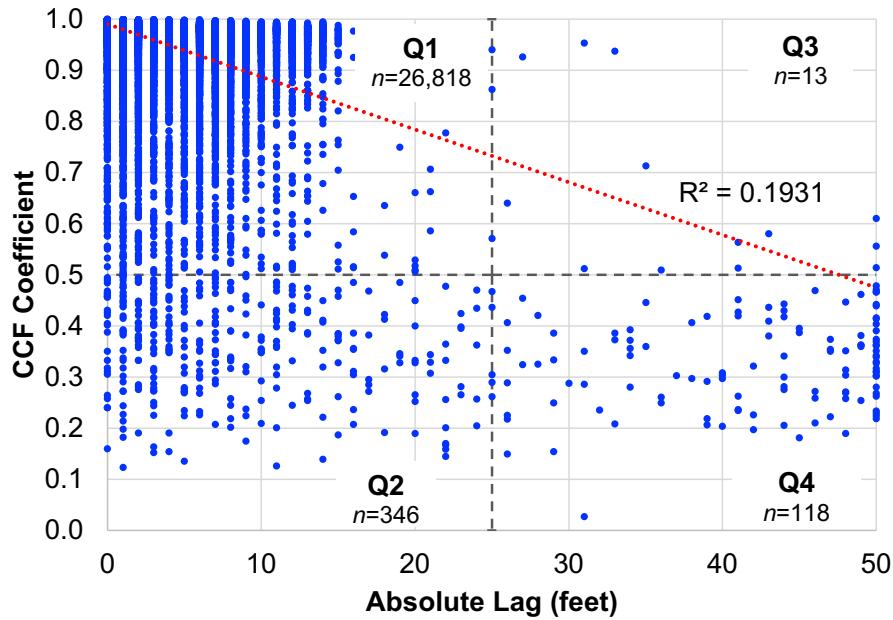


**Figure 3.4: (a) Histogram of calculated lag and b) cumulative distribution of absolute lag.**

Almost all (95%) of the lag between consecutive dates for each section is lower than 10 ft (3.05 m). Given the window length of 528 ft (160.9 m) this represents only 3% of its size and provides confidence in the alignment of the data and ability to apply the cross-correlation methodology.

[Figure 3.5](#) presents the correlation between the absolute lag value and CCF coefficient. To aid in understanding different scenarios depicted within the figure, four quadrants were drawn to represent high cross-correlation coefficient and low lag (Q1), low cross-correlation coefficient and low lag (Q2), high cross-correlation coefficient and high lag (Q3), and low cross-correlation coefficient and high lag (Q4). The choice of the quadrants' limits is based on the half of the values scale and was arbitrary for visualization and data exploration purposes. Most likely, a cross-correlation coefficient of 0.6 cannot be considered high in this application.

**Deleted:** Figure 3.5

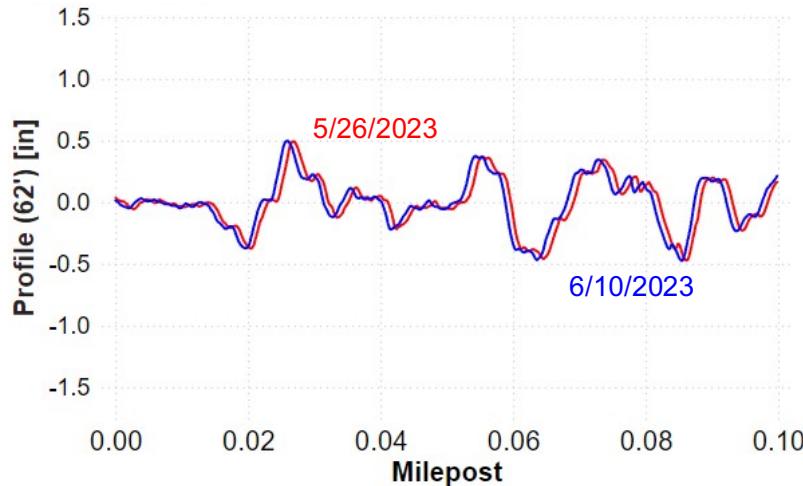


**Figure 3.5: Scatter plot of the absolute lag x cross-correlation coefficient.**

Figure 3.4 shows a concentration of samples (98.3%) in the upper-left corner of the chart (Q1) indicating lower absolute lag and high cross-correlation coefficient. On the other hand, there are comparatively few track windows (0.8%) in the bottom right corner of the chart that demonstrate high lag and low cross-correlation coefficient. A linear correlation between absolute lag and cross-correlation showed a low positive correlation ( $R^2 = 0.19$ ) thus indicating most samples present both low lag and high cross-correlation. It is worth mentioning that the maximum lag for this project was limited to 50 ft (15.24 m) given the window size and prior observation of data alignment accuracy. Thus, samples with this exact value may not present a realistic value for lag and likely contribute to lowering the  $R^2$  value.

The following are examples of the actual data and its CCF coefficient and lag results that fall within each of the four quadrants. [Figure 3.6](#) presents the profile for two consecutive dates resulting in a lag of 6 ft (1.83 m) and a CCF coefficient of 0.98 representing data in Q1.

**Deleted:** Figure 3.6



**Figure 3.6: Example of track profile measurements showing low lag and high cross-correlation (Q1 of [Figure 3.5](#)).**

**Deleted:** Figure 3.5

Both profile data series present the same signature, resulting in a high CCF coefficient. Lag is also visually observable in the time series data. To provide further insight into the dataset, a heatmap of the cross-correlation for all pairs of data at this segment is presented ([Figure 3.7](#)).

**Deleted:** Figure 3.7

**Formatted:** Check spelling and grammar

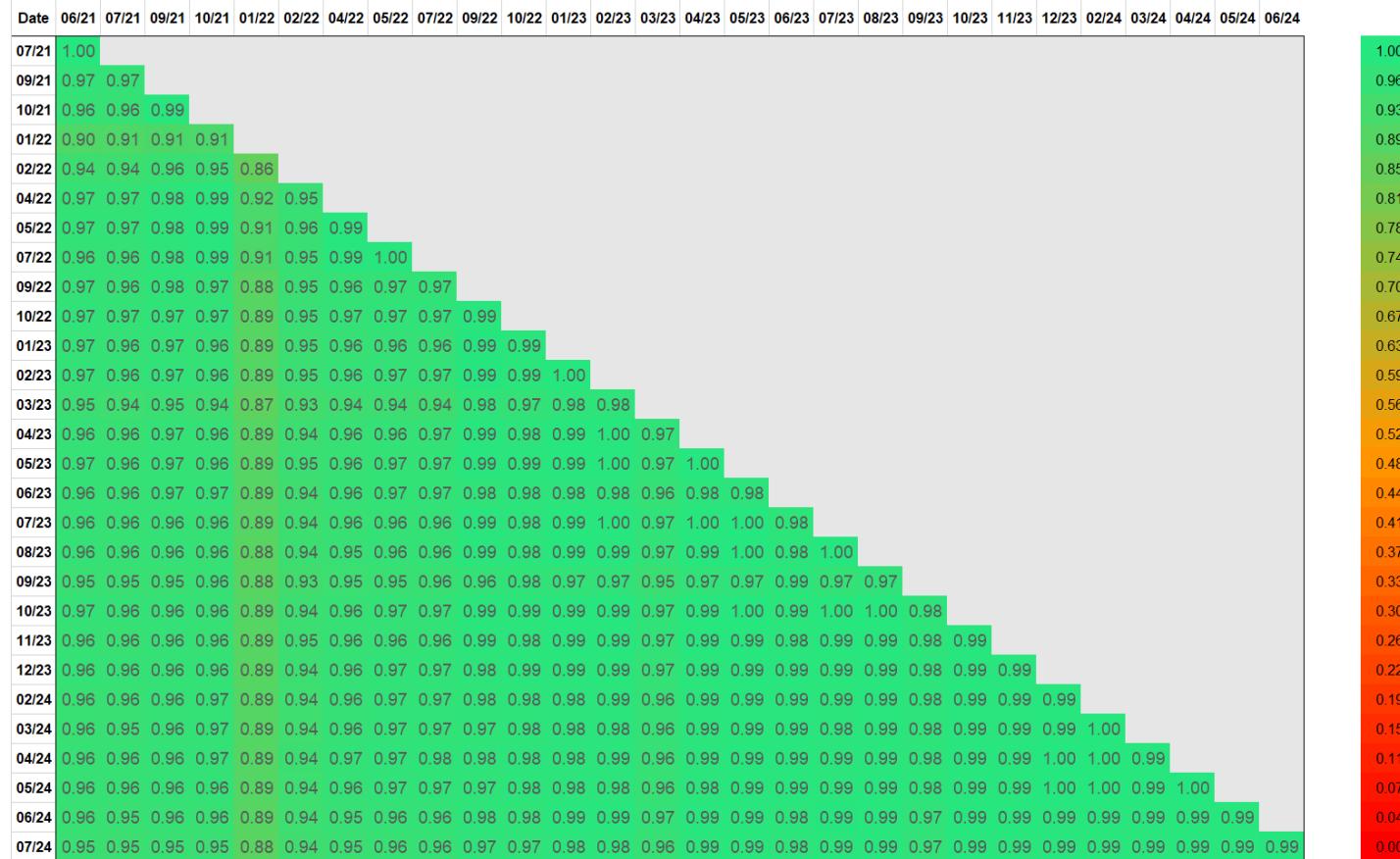


Figure 3.7: Cross-correlation heatmap (Q1 of Figure 3.5).

One key observation when reviewing data in the heatmap ([Figure 3.7](#)) is that the high cross-correlation coefficient is not only between the consecutive dates (05/23 and 06/23) under study but that it is present between all pairs of inspections. For instance, the cross-correlation between the profile measured in June 2021 and July 2024 is 0.95. This finding would indicate that the change in track condition is due to natural degradation given the data show a very similar signature run-over-run (Palese et al., 2020). In other words, the natural degradation of the profile might not strongly influence the cross-correlation result.

**Deleted:** Figure 3.7

The results as mentioned earlier indicate the possibility of identifying maintenance interventions even with gaps in data collection given the cross-correlation coefficient was kept high after three years. It is also likely that the parameter used for change detection and its expected degradation rate will play a significant role in the application of cross-correlation to this task. For instance, Palese (2020) reported the cross-correlation between successive rail wear measurements in six years obtaining values between 0.69 and 0.82. The authors did not mention any potential maintenance events that could have impacted the lower correlation value. [Figure 3.8](#) shows that no profile-adjustment maintenance occurred at this location given the high goodness of fit for the linear trend line for profile degradation. Practically, segments with only pairs of inspections presenting a high coefficient and a low lag can be used for linear track degradation analysis without having to split the data into degradation cycles.

**Deleted:** Figure 3.8

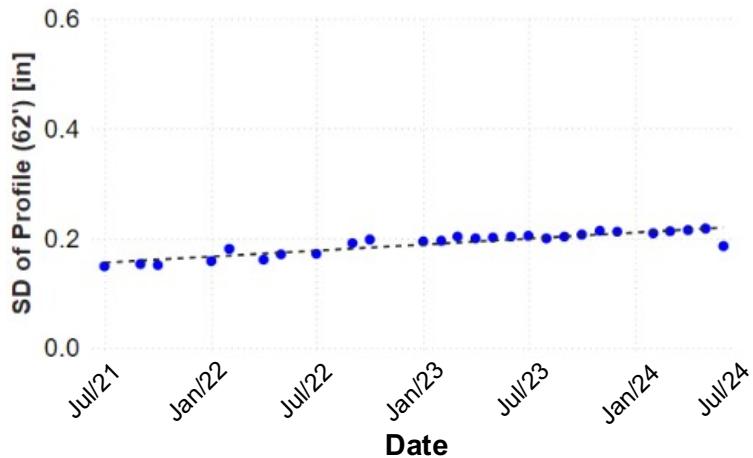


Figure 3.8: SD of profile representing track degradation over time (Q1 of [Figure 3.5](#)).

**Deleted:** Figure 3.5

[Figure 3.9](#) presents the measured profile on two consecutive dates for a track section in Q2, resulting in a lag of -7 ft (-2.13 m) and a cross-correlation coefficient of 0.29.

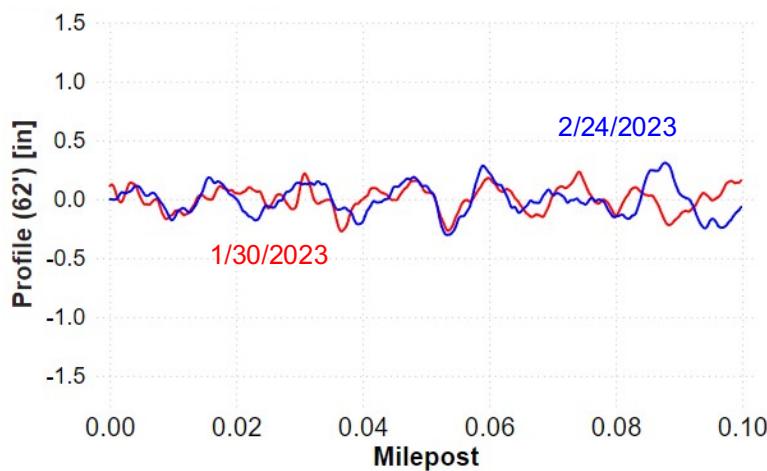


Figure 3.9: Example of track profile measurements showing low lag and low cross-correlation (Q2 of [Figure 3.5](#)).

**Deleted:** Figure 3.5

A review of cross-correlation results also reveals that some locations within the window have similar peaks and valleys but reveal distortions upon further review of the data. To provide additional insight into the data and to better visualize the output of cross-correlation results, I generated a heatmap of the cross-correlation calculation among all pairs of dates for this track segment ([Figure 3.10](#)).

[Figure 3.10](#) presents a red area at the intersection of the data before (columns) and after (rows) February 2023. This indicates that – after February 2023 - the measurements at this section of the track showed a lower correlation regarding the measurements before that date. The green areas related to the cross-correlation values after February 2023 are also remarkable. These suggest that the change in February 2023 was not an outlier and that there was a legitimate change in the profile geometry signal that continued beyond that date.

**Deleted:** Figure 3.10

**Deleted:** Figure 3.10

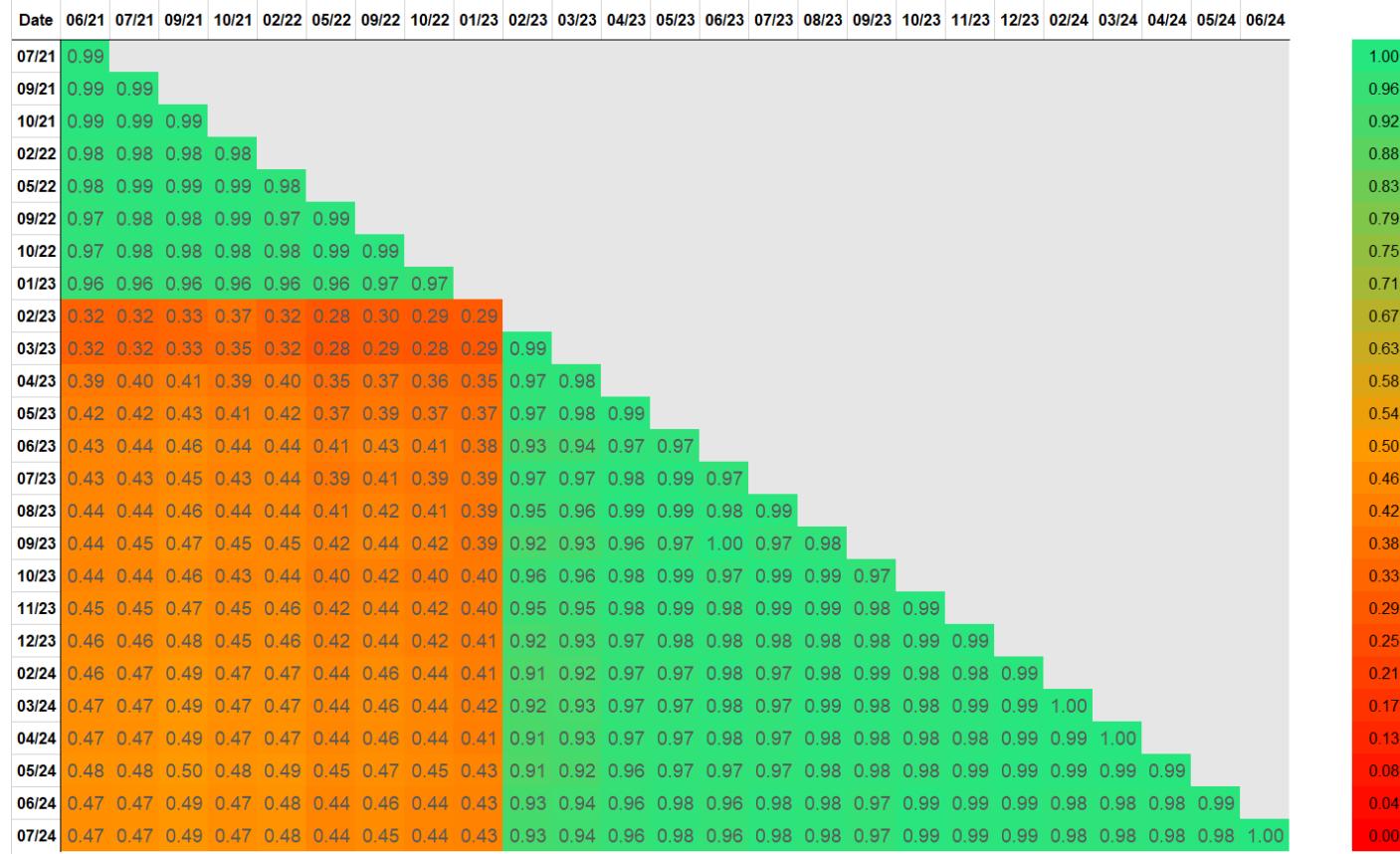
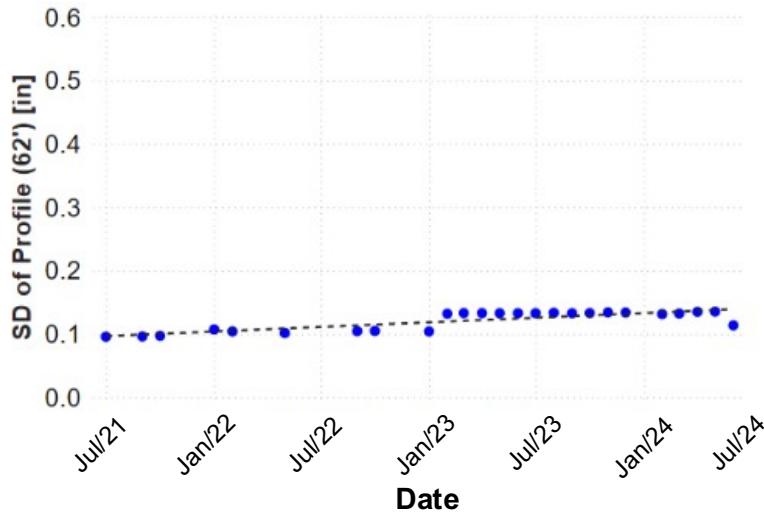


Figure 3.10: Cross-correlation heatmap (Q2 of Figure 3.5).

This change in condition is also observable in [Figure 3.11](#). We can notice that the theoretical linear degradation built with the whole dataset does not look ideal. There is a visible shift in the linear trend before and after February 2023

**Deleted:** Figure 3.11



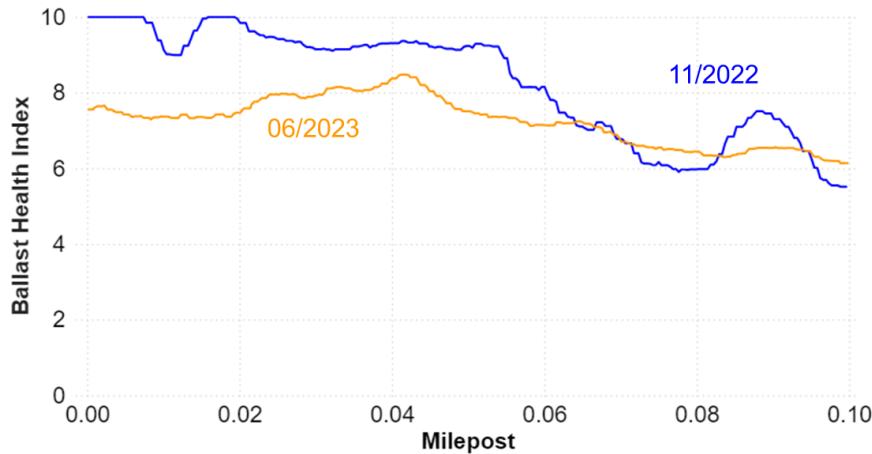
**Figure 3.11: SD of profile representing track degradation over time (Q2 of [Figure 3.5](#)).**

**Deleted:** Figure 3.5

What is unusual with respect to this example is the shift upwards of the SD. For a tamping maintenance event, it is expected that the track profile will be improved (made more even) thus resulting in a lower standard deviation of profile. In the section, the track got slightly worse in terms of profile SD (i.e., track profile became rougher). This could be the result of a different source of maintenance or a naturally-occurring disruptive event. Changes in the measurement device or calibration, as well as wheel slipping, might also be a potential cause.

The BHI can also be used to help illustrate this track segment ([Figure 3.16](#)). The two LRA/L measurements closest to the date of the change event are in November 2022 and June 2023.

**Deleted:** Figure 3.16



**Figure 3.12: BHI measurement at track segment example presented (Q2 of [Figure 3.5](#)).**

**Deleted:** Figure 3.5

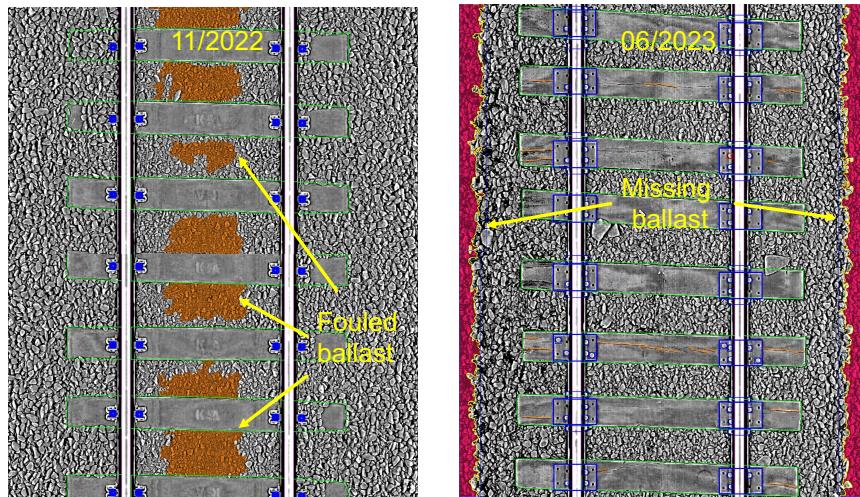
There is lower ballast level present in both inspections at the end of the segment, from milepost 0.07 to 0.1. Furthermore, from November to June there was an overall decrease in the BHI, which might be associated with the deterioration of the SD of profile. A track intervention other than tamping likely disrupted the ballast.

Even though the data collection with LRA/L had a time gap of seven months, making it hard to determine the precise date by this device only, the images ([Figure 3.13](#)) bring the hypothesis that the maintenance event was a replacement from concrete to timber crossties. The presence of concrete crossties' debris in the righthand image highlights the occurrence of such intervention. Two other important facts that can be drawn from the images are the fouled ballast/mud spot present before the maintenance and the lack on the shoulder ballast present on

**Deleted:** Figure 3.13

the post-maintenance. Currently, the fouled ballast is not considered in the BHI calculation, but the shoulder level is, explaining the decrease in its quality presented in [Figure 3.16](#).

**Deleted:** Figure 3.16

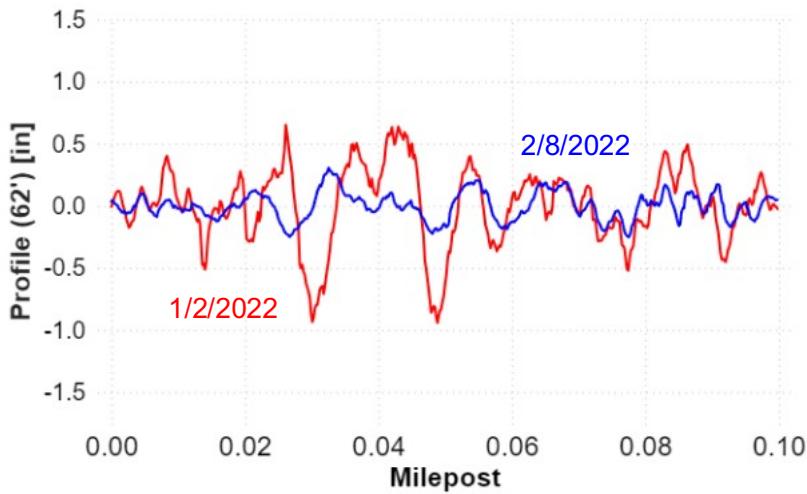


**Figure 3.13: Crosstie replacement identified on track segment example presented (Q2 of [Figure 3.5](#)).**

**Deleted:** Figure 3.5

Another example of a track segment in Q2 is presented in [Figure 3.14](#). In this segment, the obtained lag was 15 feet (4.6 m), and a cross-correlation coefficient of 0.37 between January and February of 2022.

**Deleted:** Figure 3.14



**Figure 3.14: Example of track profile measurements showing low lag and low cross-correlation (Q2 of [Figure 3.5](#)).**

**Deleted:** Figure 3.5

In this example, we can notice a certain similarity between the signals, with a smoothness in February. This implies that track segments in Q2 might also be associated with maintenance activities that decreased track roughness but kept an overall similar profile signature. It is worth it to mention that there is a crossing present at this track segment. In this segment, the SD of profile presented a decrease, as expected on a potential tamping maintenance ([Figure 3.15](#)).

**Deleted:** Figure 3.15

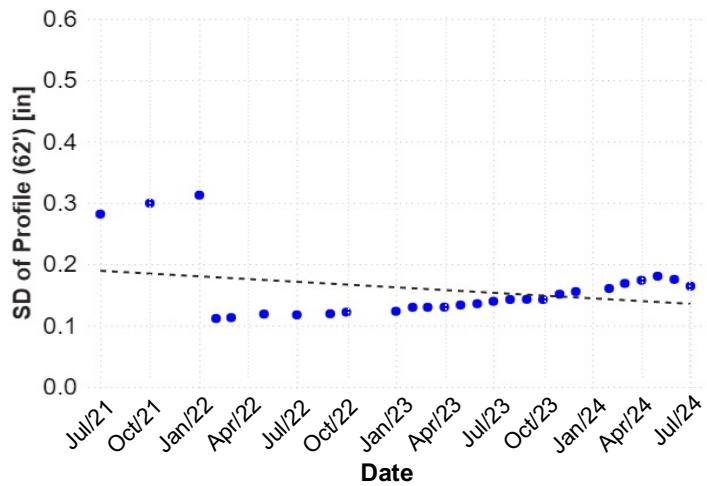
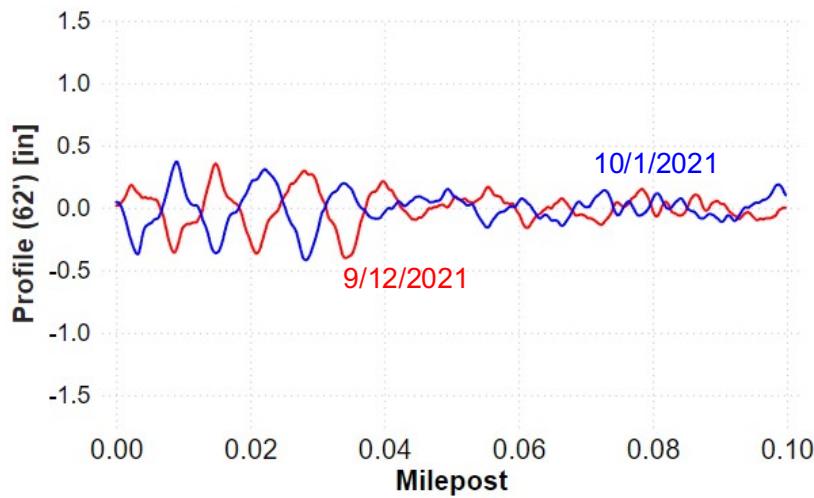


Figure 3.15: SD of profile representing track degradation over time (Q2 of [Figure 3.5](#)).

**Deleted:** Figure 3.5

Q3 represents track segments with a high cross-correlation coefficient and a high lag and is the quadrant with the fewest number of samples. [Figure 3.16](#) presents an example of the profile measured at two different dates for a track section within Q3.

**Deleted:** Figure 3.16



**Figure 3.16:** Example of track profile measurements showing high lag and high cross-correlation (Q3 of [Figure 3.5](#)).

**Deleted:** Figure 3.5

Measurements performed in September and October 2021 are very similar (CCF coefficient of 0.95) but there is visually observable lag between them (31 ft [9.45 m]). This section would be a candidate for a further alignment refining process. [Figure 3.17](#) presents the cross-correlation heatmap over time.

**Deleted:** Figure 3.17

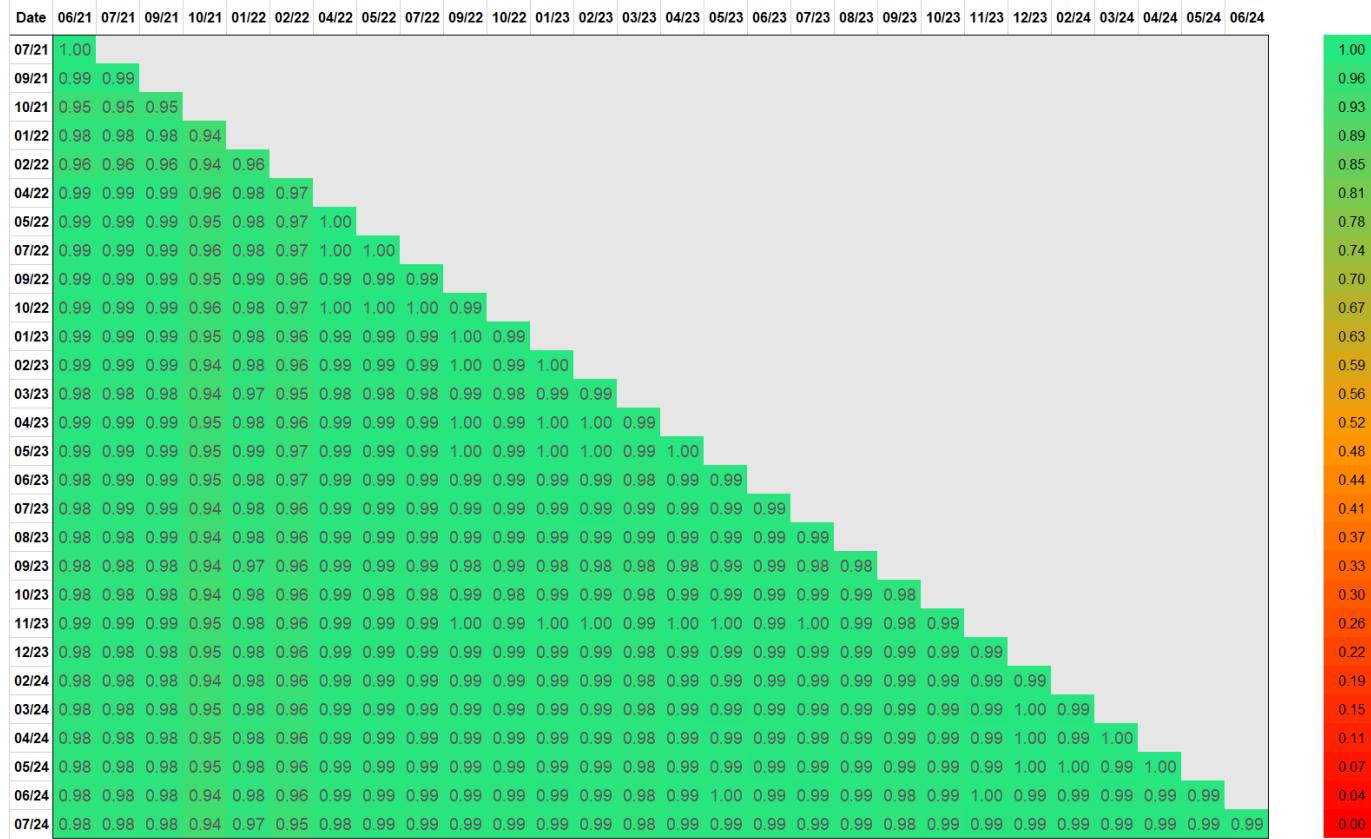


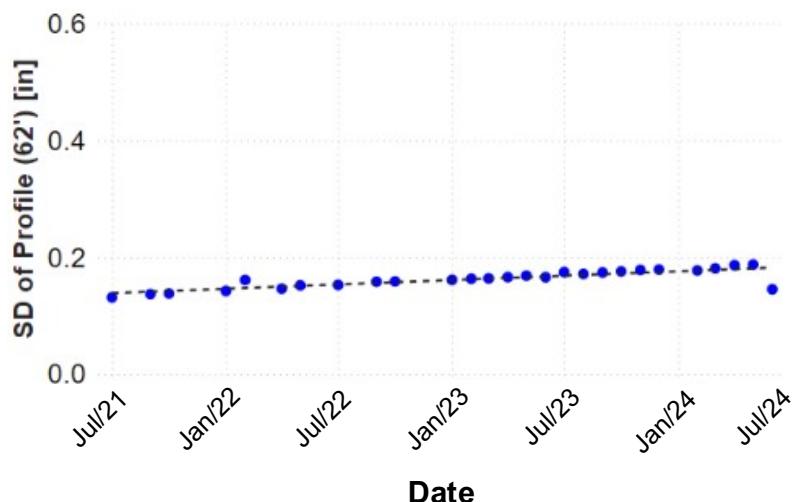
Figure 3.17: Cross-correlation heatmap (Q3 of Figure 3.5).

[Figure 3.17](#) presents that the track location has high cross-correlation throughout the entire inspection period and not just between localized sets of consecutive dates. For instance, the CCF coefficient between the profile measured in July 2021 and in July 2024 is 0.98. This means that within the three-year project duration, the signal of profile was not significantly altered. This behavior shows that even with gaps in the data collection, the cross-correlation can likely capture maintenance events. [Figure 3.18](#) further supports the prior discussion about the heatmap ([Figure 3.17](#)) and shows that no signs of maintenance during the study as the linear trend line provides a good fit for the degradation of the standard deviation of profile for the section.

**Deleted:** Figure 3.17

**Deleted:** Figure 3.18

**Deleted:** Figure 3.17



**Figure 3.18: SD of profile representing track degradation over time (Q3 of [Figure 3.5](#)).**

**Deleted:** Figure 3.5

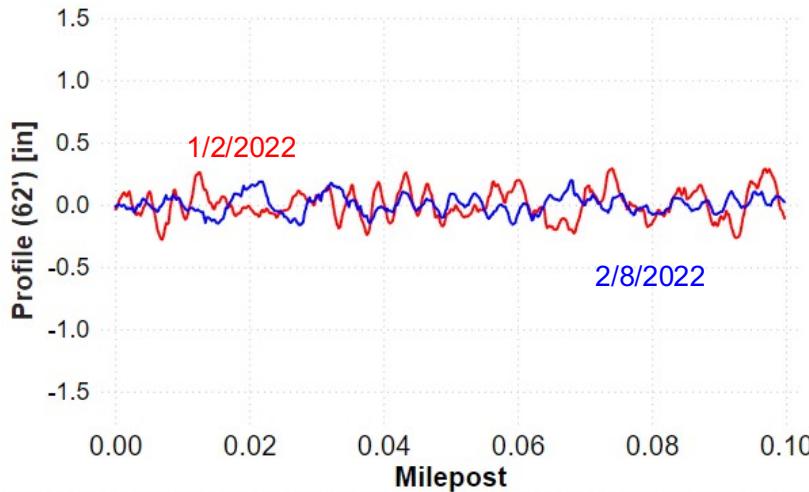
In this case, the window size likely played an important role in the cross-correlation not being impacted by the present lag. The measured lag of 31 ft (9.45 m) represents 5.8% of the window size. In case of a smaller window or higher lag, there would be lower percentual

overlap between the sections, thereby generating results that are inaccurate. Once realigned, data in this quadrant will likely be recategorized into Q1.

Finally, Q4 presents data from track sections with a high lag (41 ft [12.5 m]) and low CCF coefficient (0.26). These are sections that are most likely to represent a maintenance event.

[Figure 3.19](#) presents profile measurements for an example section in Q4.

**Deleted:** Figure 3.19

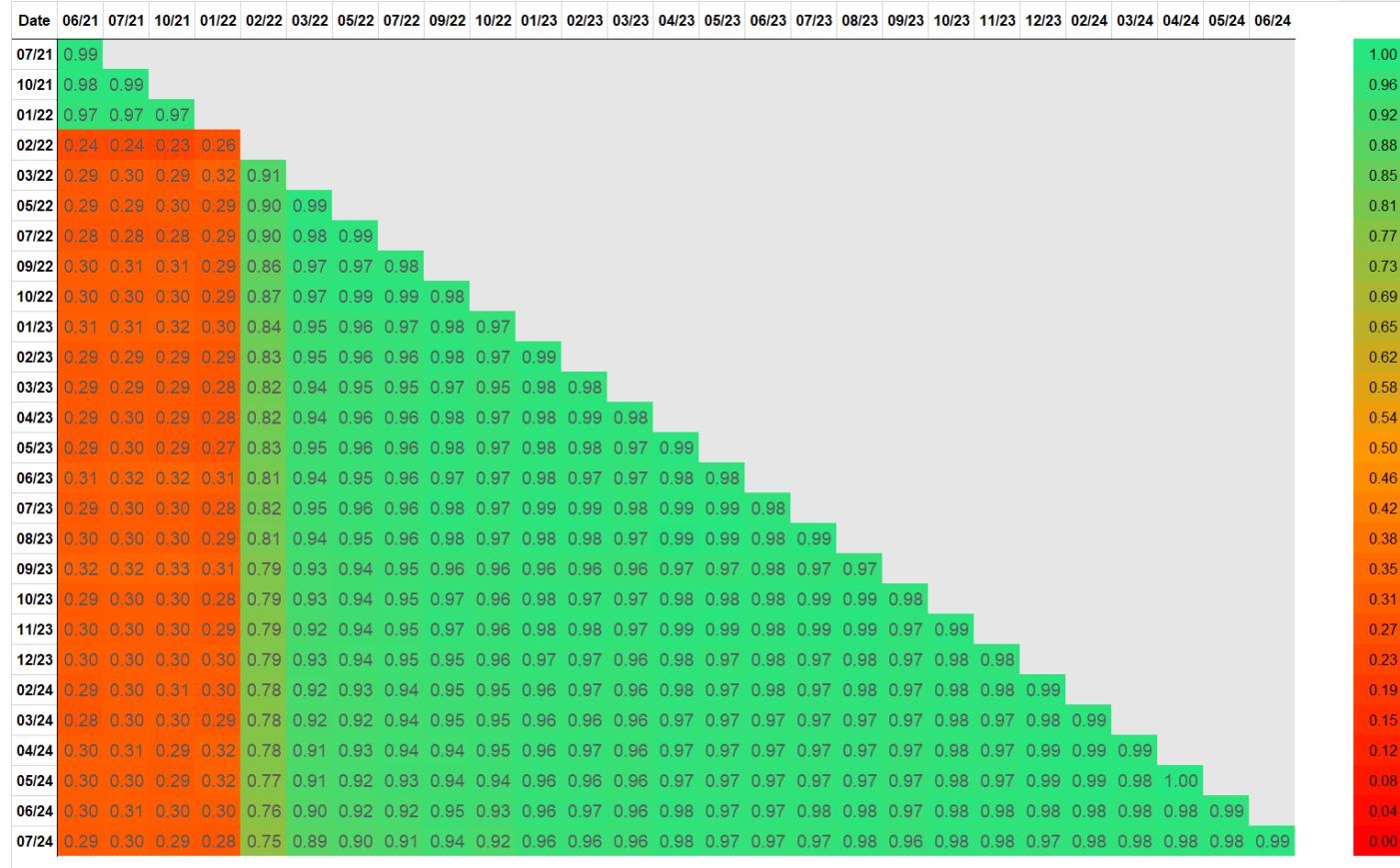


**Figure 3.19: Example of track profile measurements showing high lag and low cross-correlation (Q4 of [Figure 3.5](#)).**

**Deleted:** Figure 3.5

In this section, there is a change in the signal when comparing data from January and February 2022. [Figure 3.20](#) presents the cross-correlation heatmap for this track segment.

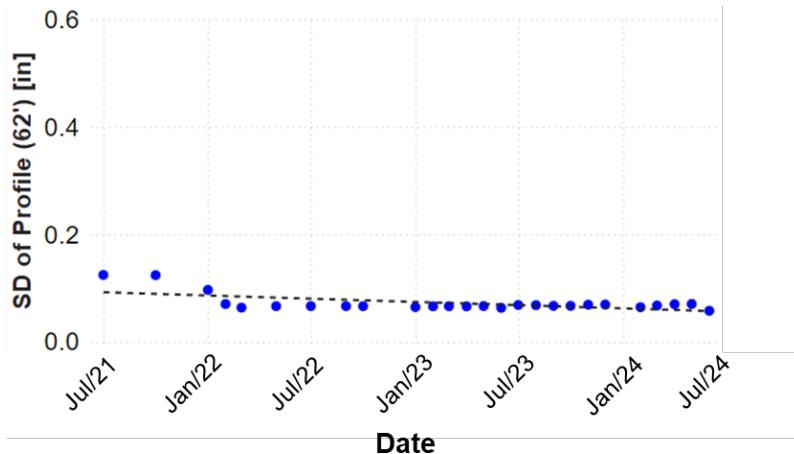
**Deleted:** Figure 3.20



**Figure 3.20: Cross-correlation heatmap (Q4 of Figure 3.5).**

We can observe that the change was permanent (subsequent cross-correlation was high), up until the whole analyzed period. [Figure 3.21](#) illustrates the change in the SD of profile for the same section over time.

**Deleted:** Figure 3.21



**Figure 3.21: SD of profile representing track degradation over time (Q4 of [Figure 3.5](#)).**

**Deleted:** Figure 3.5

Again, visual observation reveals that a change in the SD of profile signal occurred thereby indicating that maintenance was conducted.

Based on the examples presented in this section, the interpretation of events is different in each quadrant. Pairs of consecutive inspections on a track section that falls into Q1 are likely to represent natural degradation of the track, while sections in Q3 are likely to represent natural degradation with a tolerable level of misalignment between runs. Improve alignment could be addressed through further processing and the data would likely be reclassified into Q1. This highlights that events with a high lag are not necessarily associated with a maintenance event, making the CCF coefficient a better metric for it. Segments of the track containing only data in the upper level (high CCF coefficient) of Q1 and Q3 can be used for degradation modeling

without the requirement of splitting degradation cycles. Q4 is the most important quadrant with respect to maintenance detection, where a low CCF coefficient is associated with a high lag value. In Q4, the change in profile affected the signal enough to lower the CCF coefficient and make it difficult for the algorithm to find an optimal match of the lag value. Furthermore, the evaluation of different quadrants reveals that high lag alone cannot be used for maintenance detection, and low CCF coefficient is a more powerful indicator of it. Data in Q2 are more complex to interpret and can be related to a change by a maintenance event other than tamping, measurement/calibration changes, or tamping maintenance that changed the signal, but it kept somehow similar to the previous one.

Another interesting scenario is when the inspection of a certain date presents a low cross-correlation with all the other dates. Based on the previous results, it is unlikely that the natural degradation of the track causes a sudden drop in the CCF coefficient. Additionally, in the case of a drop due to an exceptional event, it is expected that the new condition would remain constant thereby resulting in a high cross-correlation with the following inspections. For instance, for a certain track section, the inspections of January and June 2022 presented low

coefficient with all other dates as can be shown on the cross-correlation heatmap ([Figure 3.22](#)).

**Deleted:** Figure 3.22

**Formatted:** Check spelling and grammar

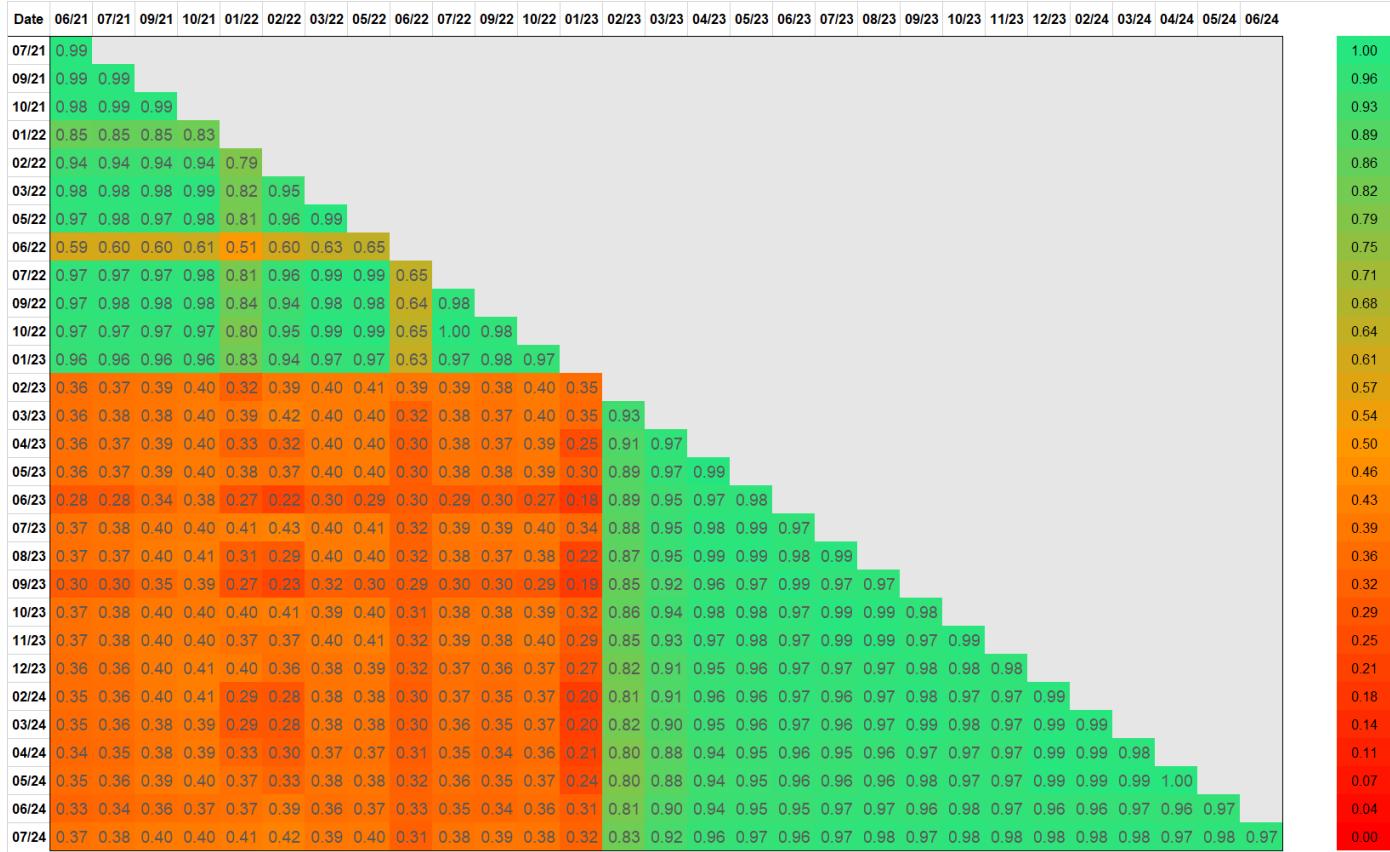
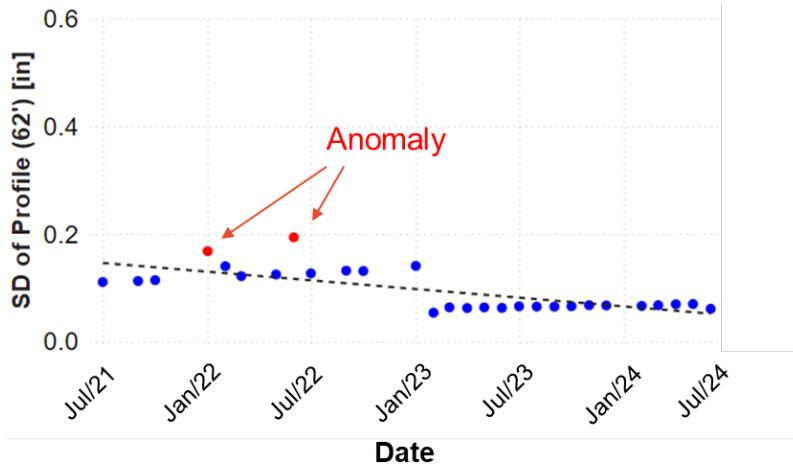


Figure 3.22: Cross-correlation heatmap highlighting anomaly events.

The standard deviation of the profile for the mentioned dates is out of the trend line on the degradation cycle ([Figure 3.23](#)). We can observe that the two red dots do not follow the trend.

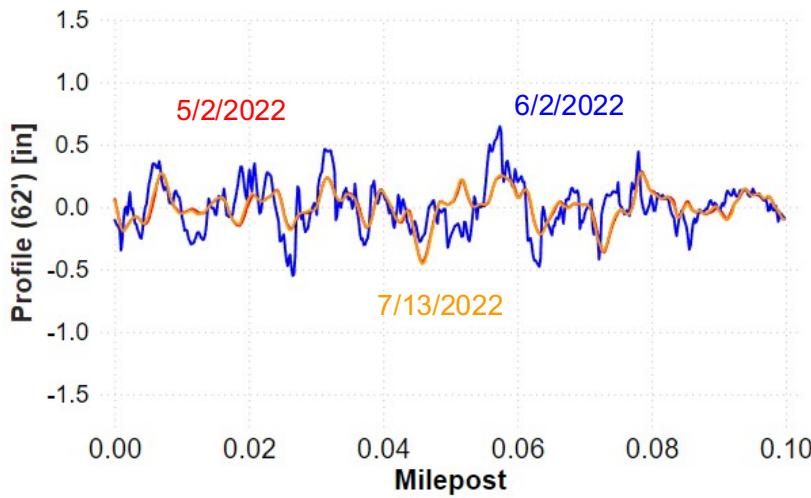
**Deleted:** Figure 3.23



**Figure 3.23: SD of profile representing track degradation highlighting anomalies.**

Retaining this sample in the dataset would impact the performance of the linear degradation and lead to inaccurate conclusions. Considering the raw profile data ([Figure 3.24](#)), it is possible to see that the data from May 2022 and July 2022 match, while June 2022 presents a different signal. The scenario is similar regarding January measurement with adjacent inspections.

**Deleted:** Figure 3.24



**Figure 3.24:** Example of profile measurements with low cross-correlation with previous and following measured data.

Based on this, an anomaly filter was applied to identify and remove data from inspections that presented CCF coefficient lower than 0.90 with the adjacent measurements over time for the same section. Only 3% of the sections were discarded from further analysis, which should not impact the degradation analyses and provides further confidence in the overall cross-correlation approach to maintenance detection. [Table 3.1](#) presents the descriptive statistics of the CCF coefficient for each track segment with its consecutive inspected available date.

**Deleted:** Table 3.1  
**Formatted:** Font: Not Bold

**Table 3.1 – Descriptive statistics of cross-correlation coefficient values.**

Parameter	Value
Count	27,303
Mean	0.953
Standard Error	0.001
Median (Q2)	0.982
Mode	0.986
Standard Deviation	0.100
Sample Variance	0.010

Kurtosis	25.100
Skewness	-4.750
Range	0.973
Minimum	0.027
Maximum	1.000
Q1	0.965
Q3	0.990
IQR	0.025

Results show that the CCF coefficients between track profile measurements are consistently high, with a mean of 0.953 and a median of 0.982, indicating strong overall similarity over consecutive inspections. The low standard deviation of 0.1001 and an IQR of 0.025 suggest that most values are tightly clustered near the mean, reflecting low variability. Additionally, the negative skewness (-4.750) and high kurtosis (25.100) indicate a distribution with a long left tail, driven by a small number of lower cross-correlation values, while the maximum value of 1.00 suggests no change in the signal for some segments.

Considering an outlier value as  $Q1 - 1.5 \times IQR$ , values of cross-correlation lower than 0.927 will be classified into this category. This value will therefore be used as the initial threshold for maintenance detection, which represents 3,226 (11.8%) pairs of dates for track segments being a potential maintenance event. Referring to the previous discussions, 0.927 would be a better value for defining the limit between quadrants with high and low CCF coefficients.

### 3.3.2 Linear change and ground-truth

Using both geometry and 3D laser scanning data, the data were compared to ground truth ([Table 3.2](#)).

The heatmaps were also used to check for permanent changes supporting a track

intervention. The threshold for maintenance used was 0.927. True positives indicate events

**Formatted:** Font: Not Bold

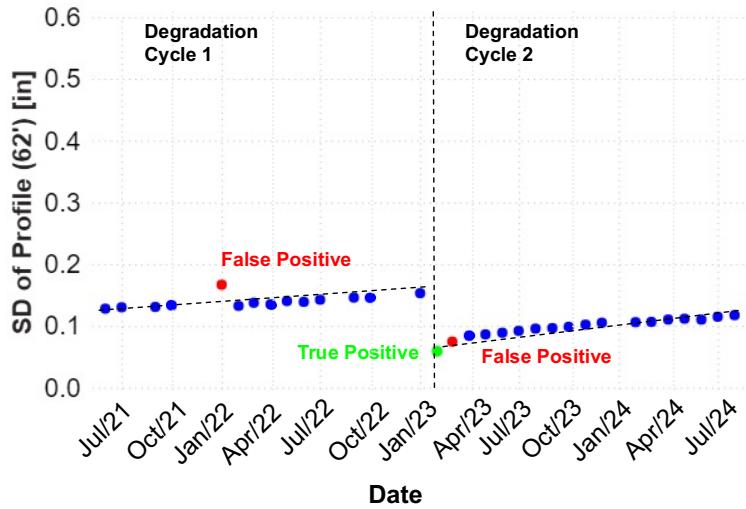
**Deleted:** Table 3.2

flagged as maintenance and subsequently confirmed via ground truthing while false positives indicate events flagged as maintenance and not confirmed ([Figure 3.25](#)). False negatives represent maintenance events with a CCF coefficient higher than the threshold, thus not flagged. Most of the identified false positives were associated with a measurement with a CCF coefficient with the adjacent measurements slightly over the anomaly threshold, measurements right after the true positives, potentially related to a transition from the ballast consolidation to a linear trend or an excessive natural degradation that highly increased the peaks on data. Lower cross-correlation coefficients were also presented to evaluate which one would represent the best F1 score performance.

**Deleted:** Figure 3.25

**Table 3.2 – Performance of the model as a function of different cross-correlation coefficient thresholds.**

Cross-correlation threshold	True Positives	False Positives	False Negative	Precision	Recall	F1 Score
0.927	904	514	38	0.64	0.96	0.77
0.900	850	136	94	0.86	0.90	0.88
0.850	767	67	177	0.92	0.81	0.86
0.800	686	46	258	0.94	0.73	0.82
0.700	518	26	426	0.95	0.55	0.70
0.650	461	15	483	0.97	0.49	0.65



**Figure 3.25: SD of profile representing track degradation highlighting true positives and false positives.**

Precision reflects the accuracy of the model's positive predictions, indicating the proportion of true positives among all positive predictions. As the cross-correlation threshold decreases, precision progressively improves. At a threshold of 0.927, precision starts at 0.64, rising steadily to 0.97 at 0.65. This trend suggests that while lowering the threshold allows for a broader detection of variances, it also reduces the number of false positives, leading to more reliable positive predictions. The model becomes more selective, but the flagged instances are increasingly accurate, demonstrating a balance between detecting fewer instances and ensuring high prediction accuracy.

Recall is the ability of the model to capture true positives among all maintenance events. Results illustrate that as the threshold decreases, its ability to capture true positives generally declines, from 0.96 at a threshold of 0.927 to 0.49 at a threshold of 0.65. This trend indicates

that as the threshold becomes more lenient, the model detects fewer positive instances, potentially due to a wider allowance for variance between track profiles. This suggests that as the threshold decreases, the instances flagged as positive are increasingly accurate, albeit fewer in number.

The F1 Score is calculated by Equation 3.8 and is a metric that quantifies the balance of precision and recall. The F1 Score reaches its highest point of 0.8 at a threshold of 0.9, indicating that this threshold offers the most balanced performance. Above this threshold, recall is relatively high, but the lower precision dilutes the F1 Score. Conversely, as the threshold drops to 0.7 or lower, precision slowly continues to improve, but the sharp decline in recall negatively affects the F1 Score. This pattern implies that a threshold of 0.9 may offer an optimal balance between capturing true positives and minimizing false positives, making it a practical choice for applications where both accuracy and recall are prioritized. Additionally, since the performance between the threshold of 0.85 and 0.9 is close, one can choose one or another based on the precision and recall performance depending on each specific case.

To serve as a reference, the performance of the models obtained by Schatzl et al. (2024) is presented in [Table 3.3](#).

**Deleted:** Table 3.3  
**Formatted:** Font: Not Bold

**Table 3.3 – Performance of the models evaluated by Schatzl et al. (2024) and the cross-correlation-based model.**

Model	Precision	Recall	F1 Score
Cross-correlation-based (0.9 threshold)	0.86	0.90	0.88
SEARCH (Schatzl et al., 2024)	0.57	0.49	0.53
CRAB (Schatzl et al., 2024)	0.87	0.67	0.76
CTG-Based (Schatzl et al., 2024)	0.82	0.68	0.74

It is possible to observe that the model using cross-correlation was able to present a satisfactory performance when compared to the other models. It is important to emphasize that [Table 3.3](#) is to be used cautiously since Schatzl et al. (2024) had access to recorded maintenance events to perform the ground-truth verification whereas in this present research, maintenance records were not available. In addition, different characteristics of track and standard deviation of profile were used. In this context, a reasonable conclusion is that the presented model has the potential to perform as well as state-of-the-art models.

A heatmap was generated to improve the visualization of the potential maintenance events from a system perspective ([Figure 3.26](#)). Each row represents a 0.1-mile (160.9 m) segment of track while each column represents one inspection date. The value in the cell is the calculated CCF coefficient between the current and previous date. Heat map colors were defined with a gradient considering red for a 0 and white at a 0.9 CCF coefficient (adopted maintenance threshold).

**Deleted:** Table 3.3

**Formatted:** Font: Not Bold

**Deleted:** Figure 3.26

	2022-02	2022-03	2022-04	2022-05	2022-06	2022-07	2022-09	2022-10	2023-01	2023-02	2023-03	2023-04	2023-05	2023-06	2023-07	2023-08	2023-09	2023-10
0.92	0.99	0.79	0.98	0.96	0.98	0.97	0.99	1.00	0.58	0.98	0.88	0.97	0.98	0.99	0.98	0.99	0.98	0.98
0.91	0.96	0.97	0.98	0.98	0.76	0.98	0.97	0.99	0.99	0.52	0.99	0.98	0.99	1.00	0.99	0.99	0.99	0.99
0.31	0.91	0.99	0.98	0.98	0.80	0.98	0.94	0.99	0.99	0.42	1.00	0.98	1.00	1.00	0.99	0.99	0.99	0.99
0.97	0.99	0.99	0.99	0.99	0.99	0.96	0.99	0.99	0.39	1.00	0.98	0.99	1.00	0.99	0.99	0.99	0.99	0.99
0.96	0.99	0.81	0.97	0.99	0.65	0.99	0.94	0.98	0.93	0.55	0.99	0.97	0.99	0.99	0.99	0.99	0.99	0.99
0.98	0.98	0.89	0.98	0.99	0.87	0.97	0.93	0.98	0.93	0.70	0.99	0.96	0.99	0.99	0.98	0.99	0.99	0.99
0.91	0.99	0.99	0.99	0.98	0.99	0.92	0.99	0.99	0.53	0.98	0.96	0.99	0.99	0.99	0.99	0.99	0.99	0.99
0.85	0.98	0.99	0.99	1.00	0.95	0.98	0.92	0.99	0.99	0.35	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99
0.74	0.97	0.99	0.97	1.00	0.78	0.98	0.93	0.99	0.99	0.54	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99
0.97	0.98	1.00	0.99	0.99	0.99	0.84	0.99	0.99	0.60	0.99	0.98	0.99	0.99	0.99	0.98	0.99	0.99	0.99
0.95	0.97	0.99	0.99	0.99	0.99	0.98	0.88	0.99	0.99	0.43	0.99	0.97	0.98	0.99	0.99	0.99	0.99	0.99
0.97	0.99	0.99	0.99	0.99	0.99	0.96	0.98	0.99	0.76	1.00	0.99	0.99	1.00	0.99	0.99	0.99	0.99	0.99
0.96	0.98	0.99	0.99	0.99	0.99	0.96	0.99	0.99	0.38	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
0.98	0.98	0.99	0.99	0.99	0.99	0.93	0.99	0.99	0.65	0.99	0.95	0.99	0.99	0.99	0.99	0.99	0.99	0.99
0.98	0.99	0.99	0.99	1.00	0.99	1.00	0.98	0.99	0.99	0.34	0.98	0.96	0.99	0.99	0.99	0.99	0.99	0.99
0.99	0.99	0.99	1.00	1.00	1.00	0.99	0.98	0.99	0.99	0.46	0.99	0.97	0.99	0.99	0.99	0.99	0.99	0.99
0.98	0.99	1.00	0.99	1.00	1.00	0.99	1.00	0.99	0.99	0.34	1.00	0.98	0.99	1.00	0.99	1.00	0.99	0.99
0.97	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.58	0.99	0.97	0.98	0.99	0.98	0.98	0.98	0.98
0.96	0.96	0.98	0.97	0.99	0.99	0.99	0.98	0.99	0.97	0.75	0.98	0.97	0.99	0.99	0.99	0.99	0.99	0.99
0.94	0.97	0.99	0.98	0.99	0.99	0.99	0.98	0.99	0.98	0.83	0.98	0.96	0.99	0.99	0.98	0.98	0.98	0.98
0.97	0.98	0.99	0.98	1.00	0.99	0.99	0.96	0.99	0.98	0.80	0.99	0.97	0.99	0.99	0.98	0.99	0.99	0.99
0.96	0.99	0.99	0.98	0.99	0.99	0.94	0.97	0.98	0.76	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.99
0.98	0.99	0.99	1.00	1.00	1.00	0.97	0.98	0.98	0.86	0.99	0.96	0.99	0.99	0.99	0.99	0.99	0.99	0.99
0.78	0.94	0.99	0.98	0.99	0.96	0.99	0.94	0.99	0.31	0.91	0.98	0.93	0.98	0.99	0.98	0.99	0.99	0.99
0.98	0.99	0.99	0.99	0.99	1.00	0.99	0.99	0.99	0.44	0.91	0.99	0.98	0.98	0.99	0.99	0.98	0.99	0.99
0.96	0.98	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.19	0.96	0.98	0.92	0.98	0.98	0.98	0.99	0.99	0.99
0.97	0.99	0.99	0.99	1.00	0.99	1.00	0.99	0.99	0.99	0.55	0.96	0.97	0.93	0.99	0.98	0.98	0.99	0.99
0.93	0.96	0.99	0.99	1.00	0.99	0.99	0.97	0.99	0.27	0.92	0.99	0.96	0.98	0.99	0.98	0.98	0.98	0.98
0.98	0.98	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.52	0.95	0.98	0.94	0.98	0.99	0.99	0.99	0.99	0.99
0.94	0.96	0.98	0.99	0.99	0.99	0.96	0.99	0.42	0.96	0.99	0.97	0.99	0.99	0.99	0.99	0.99	0.99	0.99
0.96	0.99	0.98	0.99	1.00	0.99	0.99	0.97	0.99	0.46	0.97	0.99	0.97	0.99	0.99	0.99	0.99	0.99	0.99
0.91	0.99	1.00	0.99	1.00	1.00	0.98	0.99	0.55	0.96	0.95	0.96	0.98	0.98	0.98	0.99	0.99	0.99	0.99
0.94	0.99	0.99	1.00	0.99	0.99	0.97	0.99	0.38	0.94	0.97	0.92	0.99	0.99	0.99	0.98	0.99	0.99	0.99
0.97	0.98	0.98	1.00	0.99	0.99	0.94	0.99	0.56	0.96	0.99	0.95	0.99	0.99	0.99	0.99	0.98	0.99	0.99
0.96	0.99	0.99	0.99	1.00	0.98	0.98	0.99	1.00	0.51	0.98	1.00	0.98	0.99	1.00	0.99	0.99	1.00	0.99
0.99	0.99	0.99	1.00	1.00	1.00	0.99	0.97	0.31	0.98	1.00	0.98	1.00	1.00	0.99	1.00	0.99	1.00	0.99

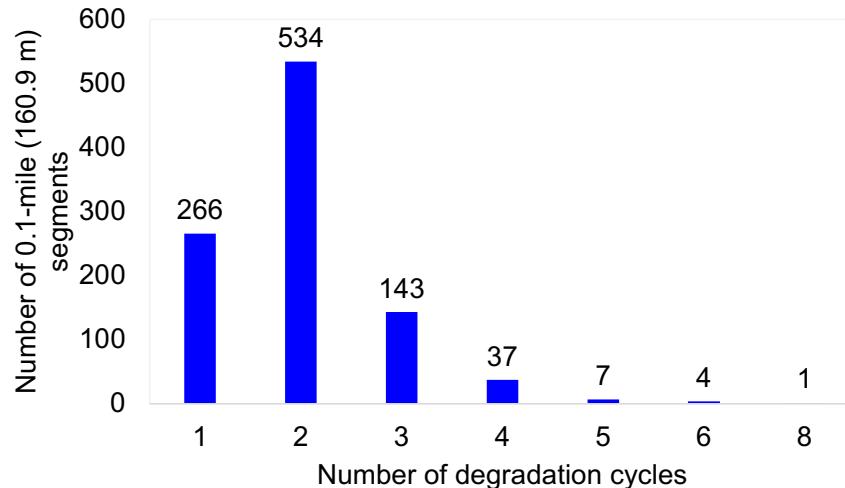
**Figure 3.26: Heat map of maintenance detection with red spots highlighting low cross-correlation with a high likelihood of maintenance activity.**

It is possible to observe continuous maintenance events throughout the sections for a same month. This is plausible with the main tamping activities where long sections of track are tamped at a time. The heatmap also allows us to note the progress of the maintenance crew over time, as noted from March/2023 to April/2023.

To check the effectiveness of the model at identifying maintenance events and facilitating splitting of degradation cycles, the change in the linear trends split based on the potential

maintenance events was evaluated. [Figure 3.27](#) summarizes the number of degradation cycles using a cross-correlation threshold of 0.9.

**Deleted:** Figure 3.27

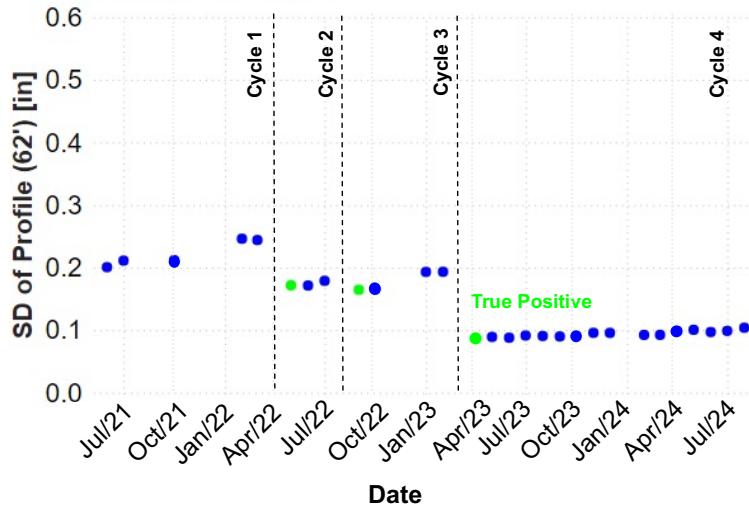


**Figure 3.27: Number of maintenance events per track segment**

Using a cross-correlation threshold of 0.9, the model identified varying numbers of maintenance events for each of the 0.1-mile (160.9 m) segments, with most segments experiencing up to three degradation cycles. As shown, 266 segments had no recorded maintenance events, while 534 segments had at least one event, thus two degradation cycles, suggesting that maintenance interventions were more common in some areas than others. A smaller subset of segments (143) had three degradation cycles, and fewer than 50 segments experienced more than four cycles, with only two segments reaching as many as six or eight. This distribution indicates that while most segments require minimal maintenance, a select few areas might be subjected to recurring maintenance, potentially highlighting locations with persistent and underlying track issues. On the other hand, the track segments with four or more

degradation cycles might be associated with false positive events. [Figure 3.28](#) presents an example of a track section with four identified degradation cycles.

**Deleted:** Figure 3.28



**Figure 3.28: SD of profile representing track degradation on a section identified with 4 degradation cycles.**

In this track segment, the CCF coefficients obtained for the maintenance events were respectively 0.89, 0.85, and 0.43. The first (CCF coeff.: 0.89) and second (CCF coeff.: 0.85) events are associated with a likely spot maintenance over 0.01 miles (16 m) at the beginning of the segment. On the other hand, the third event (CCF coeff.: 0.43) presents a clear smoothening of the profile signal over the whole segment. In that sense, it appears that the lower the coefficient, the more confidence in the occurrence of a change on the track. This effect is also illustrated by the increase in the precision of the model as the threshold decreases ([Table 3.2](#)). From a system perspective, the heatmap ([Figure 3.26](#)) can be a valuable tool to identify places

**Deleted:** Table 3.2

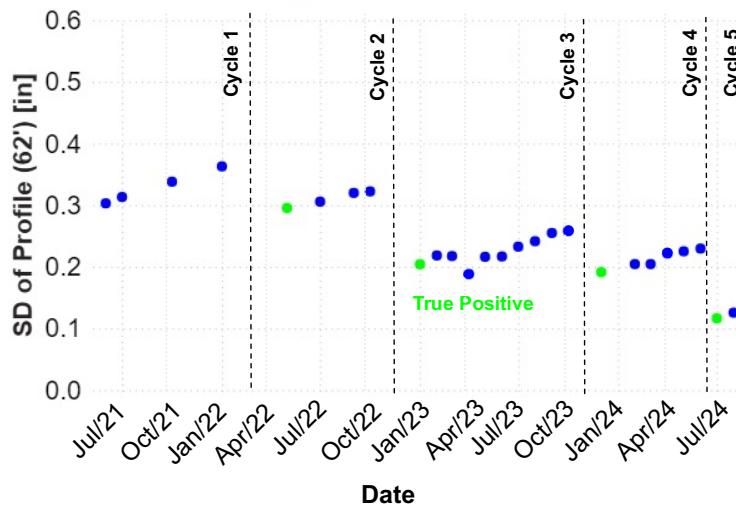
**Formatted:** Font: Not Bold

**Deleted:** Figure 3.26

with more likelihood of being true positives and spot places close to the threshold for review and potential disregard as maintenance events.

[Figure 3.29](#) presents an example of track section with five degradation cycles identified.

**Deleted:** Figure 3.29



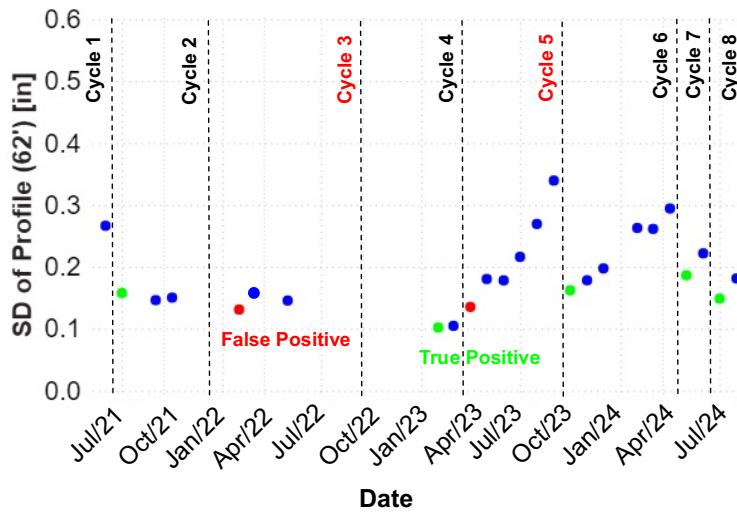
**Figure 3.29: SD of profile representing track degradation on a section identified with 5 degradation cycles.**

In this track segment, the coefficients obtained for the maintenance events were respectively 0.75, 0.6, 0.81, and 0.66. The first (CCF coeff.: 0.75) and third (CCF coeff.: 0.81) events are associated with a likely spot maintenance over 0.02 miles (32.2 m) at the middle of the segment while the second event (CCF coeff.: 0.6) is associated with a likely spot maintenance over 0.03 miles (48.3 m) at the beginning of the segment. The last event (CCF coeff.: 0.66) represents a likely maintenance event over 0.05 miles (80.5 m). The inspection from April 2023 visually looks like a potential maintenance event, although, after inspecting the raw profile data, there is no evidence to believe a maintenance event happened. The CCF

coefficient obtained was 0.97 and spike measurements were present as well as a small lag. It is important to point out that there is a turnout region at this track segment.

The degradation in the segment with eight cycles is presented in [Figure 3.30](#).

**Deleted:** Figure 3.30



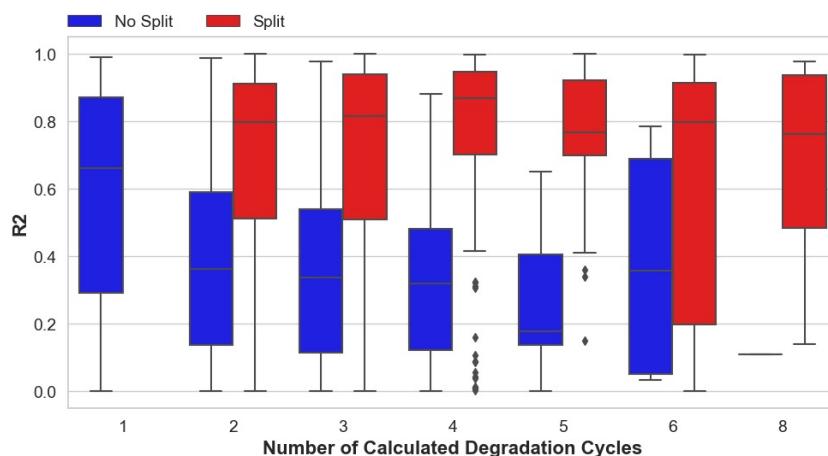
**Figure 3.30: SD of profile representing track degradation on a section identified with 8 degradation cycles.**

In this track segment, the coefficients obtained for the maintenance events were respectively 0.89, 0.87, 0.47, 0.81, 0.88, 0.80, 0.81. The first event (CCF coeff.: 0.89) is associated with a likely spot maintenance at the middle of the section over 0.02 miles (32.2 m). The second (CCF coeff.: 0.87) and fifth event (CCF coeff.: 0.88) were considered an FP, which is likely to be caused by the present lag in the measurement. Cycle 3 was supposed to be part of cycle 2 as cycle 5 was supposed to be part of cycle 4. The third event (CCF coeff.: 0.47) presents a smoothening on the profile signal over about 0.05 miles (80.5 m). This is the lower coefficient present for this segment, and it also presents the biggest time gap between

measurements, making the maintenance date imprecise to define, as somewhere between May 2022 and January 2023. The last three events (CCF coeff.: 0.88, 0.80, and 0.81) are all associated with a decrease in the magnitude of the profile values over 0.02 miles (32.2 m), similar to the first one. This might be associated with a track spot with recurrent issues.

Since the model has limitations, some metrics can be used to evaluate if a degradation line is a good fit to be used or should be discarded. After splitting degradation cycles based on the cross-correlation threshold, the  $R^2$ , slope, RMSE, and MAPE were calculated considering the entire length and the newly partitioned degradation cycles. Only cycles with at least three points were used since the use of two or fewer points would not be very reliable. In addition, it is important to mention that the split subset contains more samples since each whole degradation results in more split degradation cycles. [Figure 3.31](#) presents the boxplot of the calculated  $R^2$ .

**Deleted:** Figure 3.31



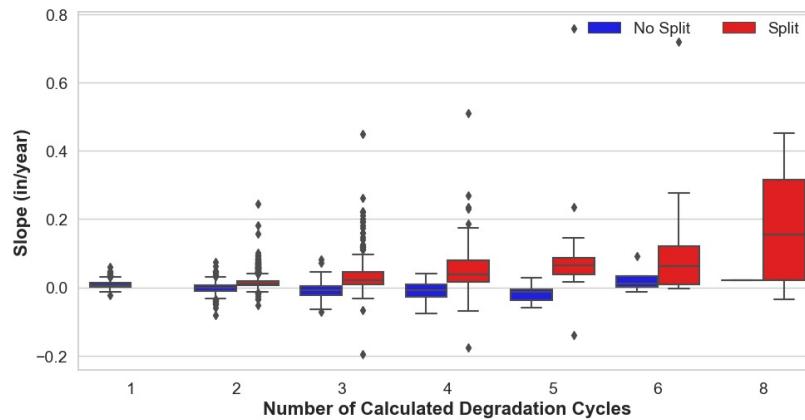
**Figure 3.31: Boxplot of  $R^2$  values based on identified degradation cycles.**

Regarding the  $R^2$ , it is noticeable the difference between split and no split subset. The values are more elevated on the split subset, indicating that the newly generated degradation

lines are overall more accurate to represent the linear fit model. If an  $R^2$  threshold is established to define data quality, using the proposed method to define maintenance events would result in more data available for analysis. An  $R^2$  value lower than 0.6 could be a reasonable threshold for discarding inappropriate data, however, no threshold clearly divides split and no split data. Something to be taken into consideration is that  $R^2$  is sensitive to shallow slopes, resulting in a low value for well-fitting shallow degradation rates.

[Figure 3.32](#), presents the results of the slope of values based on the number of degradation cycles calculated and without splitting degradation cycles.

**Deleted:** Figure 3.32

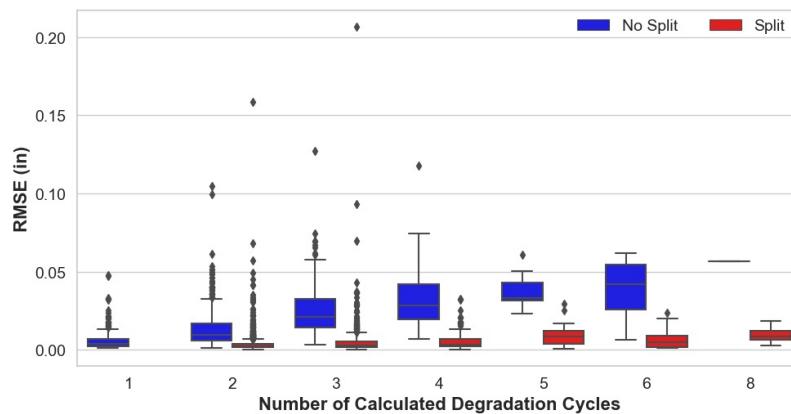


**Figure 3.32: Boxplot of slope values based on identified degradation cycles.**

When analyzing slope values related to the degradation rate of SD of the profile, a common approach to ensure data quality is to exclude trend lines with negative slopes. Naturally, track roughness tends to increase with usage, making a positive slope expected. Although this filtering may result in some data being excluded, it reduces data loss compared to scenarios without any filtering.

[Figure 3.33](#) presents the results obtained from the RMSE calculation.

**Deleted:** Figure 3.33

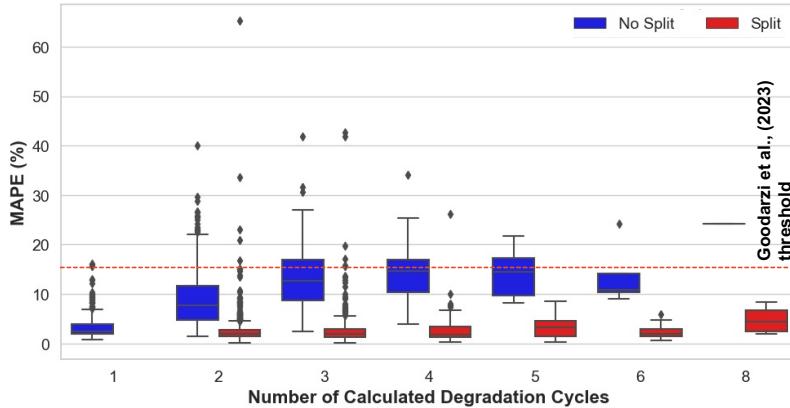


**Figure 3.33: Boxplot of RMSE values based on identified degradation cycles.**

RMSE measures how much the data deviates from the trend line, with lower values indicating better model performance. For the unsplit dataset, increased degradation cycles correspond to poorer performance, reflected in higher RMSE values. Conversely, splitting the cycles results in lower RMSE values, indicating that the data aligns more closely with the degradation trend. Unlike  $R^2$ , RMSE is sensitive to the magnitude of the data, making it crucial to set a threshold for reliability. In this context, a threshold of 0.02 is reasonable for filtering out unreliable data.

[Figure 3.34](#) presents the results obtained from the MAPE calculation.

**Deleted:** Figure 3.34



**Figure 3.34: Boxplot of MAPE values based on identified degradation cycles.**

MAPE measures how much the data deviates from the trend line, with lower values indicating better alignment. In the unsplit dataset, performance declines as degradation cycles increase, resulting in higher MAPE values. Conversely, splitting the cycles improves MAPE, showing closer alignment with the degradation trend. Based on the quality threshold suggested by Goodarzi et al. (2023) linear fits with a MAPE exceeding 15% (indicated by the red dashed line) should be discarded. The unsplit dataset would have more discarded data, whereas most linear fits in the split dataset fall below this threshold. However, MAPE and the 15% threshold may not be optimal for assessing linear fit quality in this study, as a significant number of unsplit samples still meet the threshold. Reducing the threshold to 10% could provide a more reliable assessment if MAPE remains the chosen metric.

### 3.4 Conclusions

By segmenting the data into defined degradation cycles, the analysis generates more consistent trends in track profile deterioration, improving the quality of insights compared to aggregated

analysis across the entire dataset. The benefit of using distinct degradation cycles lies in capturing the natural progression of track wear and highlighting maintenance impacts more precisely. For example, an analysis of the standard deviation in track profile across cycles indicates realistic trends: a persistent increase or stabilization over time rather than an unrealistic negative slope. This aligns with the expected behavior, as degradation in track profile is cumulative, and improvements are only expected to be a result of maintenance interventions as opposed to natural track degradation occurring through the accumulation of train tonnage.

The proposed methodology using the CCF coefficient as a parameter for maintenance detection presented to be a viable alternative, with a performance comparable to state-of-the-art methodologies.

In the case of split cycles, slopes were recalculated within each cycle. This approach shifted many slopes toward positive values, providing a more realistic portrayal of track deterioration. The segmentation also increased the amount of usable data, offering a finer resolution in detecting degradation trends. Nevertheless, even with this refined approach, a portion of the slopes remained negative, suggesting that certain segments or data inconsistencies still impact the model.

The conclusions are similar when evaluating the  $R^2$  values. Across all scenarios, adopting degradation cycles based on cross-correlation improved the  $R^2$  values, indicating improved model fit. Additionally, the RMSE decreased significantly, demonstrating that the calculated linear trends more accurately represent the actual degradation patterns and enhance the model's reliability. The same occurred with the MAPE value, with the processed data presenting a better performance regarding this metric. Potential values to be used as filters based on these metrics were also proposed based on the analysis.

The results consistently show enhanced trends and higher and predictive accuracy across all evaluated scenarios when using this approach. By isolating degradation cycles, the model identifies shifts in track quality more accurately, supporting more timely and targeted maintenance actions. This approach not only avoids misinterpretations—such as the unrealistic expectation of naturally improving profiles—but also enables a more nuanced understanding of degradation processes providing valuable data to inform maintenance planning and resource allocation.

## CHAPTER 4: CONCLUSIONS AND FUTURE WORK

This thesis examines the field of railroad track geometry and component health data by discussing common data types, describing typical railroad applications of this data, and introducing a framework for identifying trends and drawing actionable insights about track and component lifecycles.

Chapter 1 reviews current federal regulations governing track inspections in the U.S., along with an overview of track inspection practices and recent advancements in railroad big data management. Chapter 2 investigates the relationship between track geometry and component data, introducing a stochastic approach to facilitate comparisons of varying degradation rates. Finally, Chapter 3 outlines a framework for identifying track maintenance events on the dataset, providing a foundational step for accurate degradation analysis.

### 4.1 Summary of Findings

#### 4.1.1 *Stochastic investigation of the relationship between track component health and track geometry data (Chapter 2)*

Stochastic models proved effective in modeling track geometry degradation, offering insights into the relationship between initial track conditions and degradation rates. Regions with poorer initial track geometry profiles showed statistically significant, accelerated degradation, underscoring the importance of managing these areas for surfacing maintenance. Additionally, regions with faster BHI degradation correlated with higher profile degradation rates, particularly in areas with higher initial BHI values. This suggests that areas requiring more frequent maintenance, and thus a higher BHI, may also experience increased rates of profile degradation.

#### **4.1.2 Cross-correlation based railway change detection: a novel approach (Chapter 3)**

Segmenting data into defined degradation cycles enhances the clarity and consistency of track profile deterioration trends, yielding more reliable insights than analyzing the entire dataset as a whole. This cycle-based approach captures the natural progression of track wear and highlights maintenance impacts with greater precision. Cross-correlation analysis also appears to be a valid method within this framework, effectively identifying consistent degradation patterns across cycles. For instance, analyzing standard deviation trends across cycles reflects expected patterns, such as a steady increase or stabilization in degradation, rather than suggesting an unrealistic negative slope. This aligns with the cumulative nature of track wear, where improvements occur only after maintenance interventions. By isolating these cycles, the model improves trend visibility and predictive accuracy, enabling timely, targeted maintenance actions and supporting informed planning and resource allocation.

## **4.2 Recommendations and Future Work**

For future research, it is recommended to incorporate multivariable analysis and machine learning techniques to enhance the accuracy of identifying maintenance spots. Integrating component data, such as track geometry and track components, with cross-correlation values could improve the classification of maintenance needs into spot or production maintenance. Additionally, future studies should evaluate the impact of open track versus fixed structures on degradation patterns, as these environments may exhibit distinct deterioration behaviors. Assessing a component health index at locations with higher degradation cycles could provide

valuable insights into component longevity and predict failure risks, enabling more targeted and effective maintenance strategies.

## REFERENCES

- Alsaifi, A., A.M. Zaremski, J.W. Palese and T.L. Euston. 2019. Investigation of the Correlation Between Track Geometry Defect Occurrence and Wood Tie Condition. *Transportation Infrastructure Geotechnology*, 6(3): 226–244 DOI: 10.1007/s40515-019-00077-8.
- Andrews, J. 2013. A modelling approach to railway track asset management. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 227(1): 56–73.
- Attoh-Okine, N.O. 2017. *Big Data And Differential Privacy*. John Wiley & Sons.
- Audley, M. and J. Andrews. 2013. The effects of tamping on railway track geometry degradation. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 227(4): 376–391.
- Bilheri, A., A. Merheb and I. Aragona. 2023. Big data in railroads: the use of unmanned inspection vehicles. In: *Proceedings Of The 12th International Heavy Haul Association Conference 2023 (IHHA RIO 2023)*, Rio de Janeiro, Brazil.
- BNSF Railway, E.S. 2018. *Guidelines For Industry Track Projects*. BNSF Railway Company. Kansas City, KS.
- Caetano, L. and P. Teixeira. 2016. Predictive Maintenance Model for Ballast Tamping. *Journal of Transportation Engineering*, 142: 04016006.
- Carvalho, A.M.B.T. 2023. *Development Of Relationships And Trending Methodologies For Railroad Track Component And Geometry Data*. M.S. Thesis, University of Illinois at Urbana-Champaign, Urbana, IL.
- CSX Transportation. 2020. *Engineering Department Maintenance Of Way Field Manual*. CSX Transportation, Jacksonville, Florida.
- Esveld, C. 2014. *Modern Railway Track (2014)*, Digital ed. MRT-Productions, La Couronne, Nouvelle-Aquitaine, France.
- Federal Railroad Administration. 2018. 2018 Track and Rail and Infrastructure Integrity Compliance Manual, In: U.S. Department of Transportation, Federal Railroad Administration. pp.155.
- Federal Railroad Administration. 2023. Accident Data as reported by Railroads. URL [https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on\\_the\\_fly\\_download.aspx](https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on_the_fly_download.aspx). Accessed 2024-05-06.
- Federal Railroad Administration (FRA). 2024. Part 213 - Track Safety Standards, In: *Title 49 of the Code of Federal Regulations*. US Department of Transportation, Federal Railroad Administration, Washington, DC, USA.
- Fellinger, Michael. 2020. Sustainable Asset Management for Turnouts.
- Fischler, M.A. and R.C. Bolles. 1981. Random sample consensus.

- Fox-Ivey, R., Thanh Nguyen and John Laurent. 2020a. *Laser Triangulation For Track Change And Defect Detection*. Federal Railroad Administration, Office of Research, Development and Technology, DOT/FRA/ORD-20/08. Washington, D.C., USA.
- Fox-Ivey, R., Thanh Nguyen and John Laurent. 2020b. *Extended Field Trials Of LRAIL For Automated Track Change Detection*. Federal Railroad Administration, Office of Research, Development and Technology, DOT/FRA/ORD-20/14. Washington, D.C., USA.
- Germoglio Barbosa, I., A.D.O. Lima, J.R. Edwards and M.S. Dersch. 2024. Development of track component health indices using image-based railway track inspection data. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*,: 09544097231224825 DOI: 10.1177/09544097231224825.
- Ghofrani, F., Q. He, R.M.P. Goverde and X. Liu. 2018. Recent applications of big data analytics in railway transportation systems: A survey. *Transportation Research Part C: Emerging Technologies*, 90: 226–246 DOI: 10.1016/j.trc.2018.03.010.
- Goodarzi, S., H.F. Kashani, S. Chrismer and C.L. Ho. 2021. Using large datasets for finding the correlation between the rate of track settlement and changes in geometry indices. *Transportation Geotechnics*, 31: 100665 DOI: 10.1016/j.trgeo.2021.100665.
- Goodarzi, S., H.F. Kashani, J. Oke and C.L. Ho. 2022. Data-driven methods to predict track degradation: A case study. *Construction and Building Materials*, 344: 128166 DOI: 10.1016/j.conbuildmat.2022.128166.
- Goodarzi, S., H.F. Kashani, A. Saeedi, J. Oke and C.L. Ho. 2023. Stochastic analysis for estimating track geometry degradation rates based on GPR and LiDAR data. *Construction and Building Materials*, 369: 130591 DOI: 10.1016/j.conbuildmat.2023.130591.
- Guo, Y., V. Markine and G. Jing. 2021. Review of ballast track tamping: Mechanism, challenges and solutions. *Construction and Building Materials*, 300: 123940 DOI: 10.1016/j.conbuildmat.2021.123940.
- Harrington, R., J.R. Edwards, A. de O. Lima, M.S. Dersch, R. Fox-Ivey, J. Laurent and T. Nguyen. 2022. Use of Deep Convolutional Neural Networks (DCNNs) and Change Detection Technology for Railway Track Inspections. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*,: 1–9.
- Harrington, R.M., A. de O. Lima, J.R. Edwards, M.S. Dersch, R. Fox-Ivey, J. Laurent and T. Nguyen. 2023. *Automated Track Change Detection Technology For Enhanced Railroad Safety Assessment*. 0704–0188. Washington DC, USA.
- Hay, W.W. 1982a. *Railroad Engineering*, Second ed. John Wiley & Sons, New York, NY, USA.
- Hay, W.W. 1982b. Chapter 22: Crossties, In: *Railroad Engineering*. John Wiley & Sons, Hoboken, NJ, USA. pp.436–468.
- Higgins, C. and X. Liu. 2018. Modeling of track geometry degradation and decisions on safety and maintenance: A literature review and possible future research directions. *Proceedings of the*

*Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 232(5): 1385–1397  
DOI: 10.1177/0954409717721870.

Jamshidi, A., S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, R. Dollevoet, Z. Li and B. De Schutter. 2017. A Big Data Analysis Approach for Rail Failure Risk Assessment. *Risk Analysis*, 37(8): 1495–1507 DOI: 10.1111/risa.12836.

Jovanovic, S. 2004. Railway Track Quality Assessment and Related Decision Making. In: *2004 IEEE International Conference On Systems, Man And Cybernetics*, The Hague, Netherlands, pp. 5038–5043.

Liao, Y., L. Han, H. Wang and H. Zhang. 2022. Prediction Models for Railway Track Geometry Degradation Using Machine Learning Methods: A Review. *Sensors*, 22(19): 7275 DOI: 10.3390/s22197275.

Loidolt, M., S. Marschnig and A. Berghold. 2023. Track geometry quality assessments for turnouts. *Transportation Engineering*, 12: 100170 DOI: 10.1016/j.treng.2023.100170.

Miller, F.P., A.F. Vandome and J. McBrewster. 2009. *Kd-Tree*. VDM Publishing.

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

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

Palese, J.W., A.M. Zaremski and N. Attoh-Okine. 2020. Methods for Aligning Near-Continuous Railway Track Inspection Data. *Journal of Rail and Rapid Transit*, 234(7): 709–721.

Progress Rail. 2024. Track Components. URL <https://www.progressrail.com/en/Segments/Infrastructure/Trackwork.html>. Accessed 2024-10-28.

Salvador, S. and P. Chan. 2007. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5): 561–580.

Schatzl, J., F. Gerhold, M. Loidolt and S. Marschnig. 2024. Enhancing Predictive Maintenance Through Detection of Unrecorded Track Work. *Infrastructures*, 9(11): 204 DOI: 10.3390/infrastructures9110204.

Selig, E.T. and J.M. Waters. 1994. *Track Geotechnology And Substructure Management*. T. Telford ; American Society of Civil Engineers, Publications Sales Dept. [distributor], London : New York.

Sharma, S., Y. Cui, Q. He, R. Mohammadi and Z. Li. 2018. Data-driven optimization of railway maintenance for track geometry. *Transportation Research Part C: Emerging Technologies*, 90: 34–58 DOI: 10.1016/j.trc.2018.02.019.

- Soleimanmeigouni, I. and A. Ahmadi. 2016. A Survey on Track Geometry Degradation Modelling. In: *Current Trends In Reliability, Availability, Maintainability And Safety*, Springer International Publishing, Cham, pp. 3–12.
- Soleimanmeigouni, I., A. Ahmadi, A. Nissen and X. Xiao. 2020. Prediction of railway track geometry defects: a case study. *Structure and Infrastructure Engineering*, 16(7): 987–1001 DOI: 10.1080/15732479.2019.1679193.
- Soufiane, K., A.M. Zarembski and J.W. Palese. 2022. Effect of Adjacent Support Condition on Premature Wood Crosstie Failure. *Transportation Infrastructure Geotechnology*, 9(3): 302–320.
- Sung, L. 2005. *Development Of Objective Track Quality Indices*. FRA, RR05-01.
- Union Pacific Railroad. 2022. *Engineering Track Maintenance Field Handbook (2022)*, February 2022 ed. Omaha, NE, USA.
- Vale, C. and S. M. Lurdes. 2013. Stochastic model for the geometrical rail track degradation process in the Portuguese railway Northern Line. *Reliability Engineering & System Safety*, 116: 91–98 DOI: 10.1016/j.ress.2013.02.010.
- Wilk, S. and D. Li. 2021. *Influence Of Ballast Condition On Track Geometry Degradation*. TTCI, Technology Digest, TD21-031. Pueblo, CO, USA.
- Xu, P., R. Liu, F. Wang, Q. Sun and H. (Harry) Teng. 2012. A Novel Description Method for Track Irregularity Evolution. *International Journal of Computational Intelligence Systems*.
- Zarembski, A.M. 2014. Some examples of big data in railroad engineering. In: *2014 IEEE International Conference On Big Data (Big Data)*, pp. 96–102.
- Zarembski, A.M., J.W. Palese and T.L. Euston. 2017. Correlating Ballast Volume Deficit with the Development of Track Geometry Exceptions Utilizing Data Science Algorithm. *Transportation Infrastructure Geotechnology*, 4(2–3): 37–51 DOI: 10.1007/s40515-017-0039-5.
- Zhu, M., X. Cheng, L. Miao, X. Sun and S. Wang. 2013. Advanced Stochastic Modeling of Railway Track Irregularities. *Advances in Mechanical Engineering*, 5: 401637 DOI: 10.1155/2013/401637.