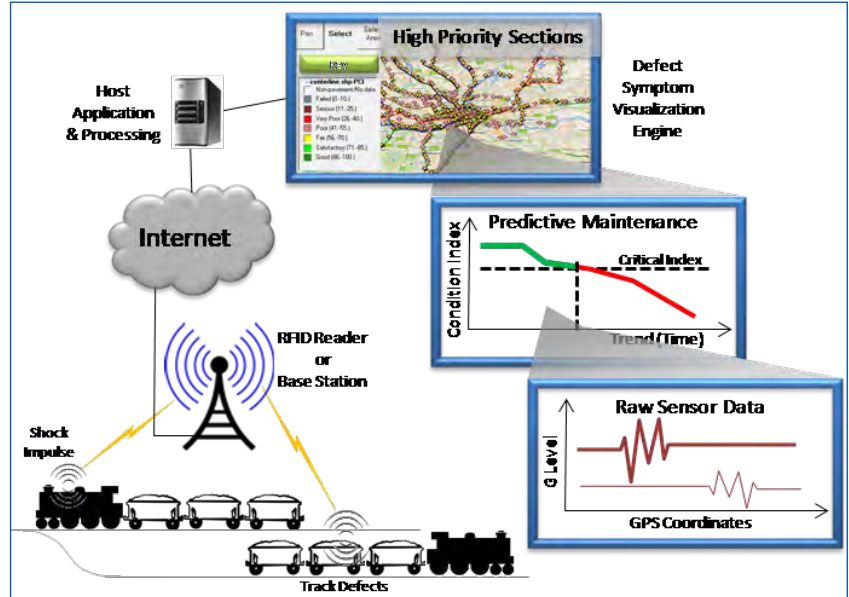# MOUNTAIN-PLAINS CONSORTIUM

MPC 21-445 | P. Lu, R. Bridgelall, D. Tolliver, B. Bhardwaj, and N. Dhingra

TRACK SURFACE
IRREGULARITY POSITION
LOCALIZATION WITH
SMARTPHONE-BASED
SOLUTION

1. Report No. MPC-551	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Track Surface Irregularity Position Localization with Smartphone-Based Solution		5. Report Date November 2021	
		6. Performing Organization Code	
7. Author(s) Pan Lu; Raj Bridgelall, Denver Tolliver, Bhavana Bhardwaj, and Neeraj Dhingra		8. Performing Organization Report No. MPC 21-445	
9. Performing Organization Name and Address North Dakota State University 1340 Administration Ave Fargo, ND 58105		10. Work Unit No.	
		11. Contract or Grant No.	
12. Sponsoring Agency Name and Address Mountain-Plains Consortium North Dakota State University PO Box 6050, Fargo, ND 58108		13. Type of Report and Period Covered Final Report	
		14. Sponsoring Agency No.	
15. Supplementary Notes Supported by a grant from the U.S. DOT, University Transportation Centers Program			

16. Abstract

Tracks are a critical and expensive railroad asset, requiring frequent maintenance. Railroads companies often rely on the accurate localization and identification of the track anomalies that could cause serious damage to infrastructure, environment, and the traveling public. However, the deployed method of inspection and maintenance is expensive, slow, increases the safety risk of workers, and requires track closure. Also, the technical limitations of present methods prevent their network-wide scaling to all railroads. Low-cost GPS receivers and accelerometers aboard regular vehicles offer a promising alternative to monitor all railroads in real time. However, low resolution and low accuracy of GPS receivers and non-uniform sample rates of the inertial sensors produced signal position misalignment and additive signal noise. Subsequently, signal-to-noise ratio (SNR) decreases and the detection error to locate the track irregularity will increase. In particular, false positives and false negatives can increase when SNR decreases. The introduced framework in this research reduces detection error while enhancing the quality of the signals and extracted features. This research demonstrates the potential use of low-cost sensors aboard hi-rail vehicles to monitor automatically and continuously for inertial events caused by irregular track geometry. The study characterizes and validates its accuracy by comparing the estimated positions of detected irregularities with the actual positions of irregularities that the railroad inspector observes. Therefore, railroad agencies that employ developed frameworks and methods will benefit from reliable track and equipment condition situation to make informed decisions, leading to resource optimization.

17. Key Words Accelerometer, Global positioning system, Railroad irregularity localization, Signal filtering, Interpolation, Ensemble average, Smart phones, Regular service vehicle, Sensors, Road Impact Factor, Feature Extraction, Track Geometry, GPS error.	18. Distribution Statement Public distribution		
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of pages 49	22. Price n/a

Form DOT F 1700.7 (8-72)

Track Surface Irregularity Position Localization with Smartphone-Based Solution

Prepared By

Dr. Pan Lu (PI)
Associate Professor
Department of Transportation, Logistics, and Finance
Upper Great Plains Transportation Institute
North Dakota State University
Pan.Lu@ndsu.edu
ORCID:0000-0002-1640-3598

Dr. Raj Bridgelall (Co-PI)
Assistant Professor
Department of Transportation, Logistics, and Finance
Upper Great Plains Transportation Institute
North Dakota State University
Raj.Bridgelall@ndsu.edu
ORCID:0000-0003-3743-6652

Dr. Denver D. Tolliver (Co-PI)
Director
Upper Great Plains Transportation Institute
North Dakota State University
Denver.Tolliver@ndsu.edu
ORCID:0000-0002-8522-9394

Bhavana Bhardwaj
Research Assistant/PhD student
Department of Transportation, Logistics, and Finance
Upper Great Plains Transportation Institute
North Dakota State University
Bhavana.Bhardwaj@ndsu.edu
ORCID: 0000-0002-4379-1565

Neeraj Dhingra
Research Assistant/PhD student
Department of Transportation, Logistics, and Finance
Upper Great Plains Transportation Institute
North Dakota State University
Neeraj.Dhingra@ndsu.edu
ORCID: 0000-0001-9970-7185

November 2021

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

ABSTRACT

Tracks are a critical and expensive railroad asset, requiring frequent maintenance. Railroads companies often rely on the accurate localization and identification of the track anomalies that could cause serious damage to infrastructure, environment, and the traveling public. However, the deployed method of inspection and maintenance is expensive, slow, increases the safety risk of workers, and requires track closure. Also, the technical limitations of present methods prevent their network-wide scaling to all railroads. Low-cost GPS receivers and accelerometers aboard regular vehicles offer a promising alternative to monitor all railroads in real time. However, low resolution and low accuracy of GPS receivers and non-uniform sample rates of the inertial sensors produced signal position misalignment and additive signal noise. Subsequently, signal-to-noise ratio (SNR) decreases and the detection error to locate the track irregularity will increase. In particular, false positives and false negatives can increase when SNR decreases. The introduced framework in this research reduces detection error while enhancing the quality of the signals and extracted features. This research demonstrates the potential use of low-cost sensors aboard hi-rail vehicles to monitor automatically and continuously for inertial events caused by irregular track geometry. The study characterizes and validates its accuracy by comparing the estimated positions of detected irregularities with the actual positions of irregularities that the railroad inspector observes. Therefore, railroad agencies that employ developed frameworks and methods will benefit from reliable track and equipment condition situation to make informed decisions, leading to resource optimization.

EXECUTIVE SUMMARY

Transportation agencies are dependent on the accurate and frequent logging of localized irregularity to prioritized inspection and maintenance needs. However, the existing methods and practices of railway condition monitoring are expansive and relatively slow. The current practices are not only expensive but they decrease rail operations' productivity and increase the safety risk for railway inspection workers. Nonetheless, the inability of proper resources and network capacity of railroads limit the continuous and network-wide monitoring. Sensors such as inertial sensors, accelerometers, gyroscopic sensors, and global position systems (GPS) installed on regular railway vehicles can have the potential to continuously monitor and inspect rail assets to meet the growing needs of safety improvements for reliable and low-cost rail operations.

The non-uniform sample rate of accelerometers and inaccurate position estimates of GPS receivers reduce the accuracy and precision of the signal needed to detect track anomalies. Moreover, these factors are also responsible for signal position misalignment and additive noise. Consequently, it reduces the S/N ratio when it increases the false positive and false negative. Smartphones have become a flexible and convenient tool to develop and test the viability of an automated track monitoring system that uses dedicated sensors onboard regular service trains. Therefore, this research introduced a framework that considers a technique called "ensemble average" that leverages the large number of sensor signals collected from multiple traversals of roadway or railway segment to enhance the signal's quality and reduce the detection error. Subsequently, increases in the SNR and reduces the false positive and false negative rates. Moreover, the signal quality improves continuously with the additionally available data stream.

The developed technique uses distance interpolation, heuristic, and correlation alignment to align the signals across the traversals. The algorithm then extracts the feature from the aligned and filtered signals. The experiments were conducted to demonstrate the reliability and accuracy of the employed method. Tradeoff analysis found that a window size of 5 meters provided a good balance between data reduction, accuracy, precision, and the consistency of anomaly detection while minimizing the potential for false positives and false negatives. Hence, the framework methods with ensemble averaging can be generalized to any sensor data from multiple traversals.

The research also proposed an approach investigating the potential use of low-cost sensors aboard regular vehicles to screen automatically and continuously for the inertial events produced by irregular track geometry. The method employed a road impact factor (RIF) transform to extract features from the sensor signals that are proportional to the amount of track geometry irregularity. The method reduces the data by generating RIF features from the rotational rates. The authors validate the accuracy of the method by comparing the estimated positions of detected irregularities with the actual positions of irregularities called ground truth area (GTA), which the railroad inspector observed. It is found that the estimated position of the irregularity is within a reasonable visual distance of the actual position so that it can be seen during a manual follow-up inspection.

The applied approach suggests that the methods have low computational complexity and are easy to use in practical applications. The railroads will benefit by integrating the computationally simple models into their decision support platform. With the enhanced situational understanding, railroad companies will be able to optimize inspections and maintenance procedures to reduce the frequency of slow orders and track closures while minimizing cost and safety risk.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.   INTRODUCTION

Railroads, comprising freight and passenger transportation operations, are the most prominent and primary mode of transportation in various countries. In the United States, approximately 140,000 rail miles are operated by seven Class I railroads, 22 Class II railroads, and 584 Class III railroads (2021 Infrastructure Report Card – Rail, 2021). During 2018, railroads delivered 1.4 billion tons of freight, a figure which is expected to increase 13% by the end of 2045 (2021 Infrastructure Report Card – Rail, 2021). This traffic increases track usage and exacerbates track geometry irregularities even more rapidly, resulting in frequent increases of the risk of incidents and accidents over time. Such risks could be reduced by keeping safe traffic conditions with regular inspections and timely detection and rectifications of track irregularities. The inspection process of rail tracks is expensive and time consuming, but it is essential because it helps reach the highest safety track geometry standards and enhances vehicle equipment ride quality and safety.

For that reason, new track inspection and maintenance methods are continually introduced (Barbosa, 2016). One such technology is based on intelligent condition monitoring systems (ICMs) to perform cost-effective and dynamic railway track inspections to reduce the potential risks of rail breaks and derailments. Such methods include ultrasonic signaling (Jarmulak, Kerckhoffs, and van't Veen, 1997; Podofillini, Zio, and Vatn, 2006; Campos-Castellanos et al., 2011; Carboni and Cantini, 2012), vision-based inspections (Singh et al., 2006; Kaleli and Akgul, 2009; Resendiz, Hart, and Ahuja, 2013; Feng et al., 2014), ground-penetrating radar (GPR) (Eriksen and Gasgoyne, 2004; Fontul et al., 2016), and wireless sensor-based systems that include fuzzy logic and surveys (Aboelela et al., 2006; Flammini et al., 2010; Sharma et al., 2014; Hodge et al., 2015; Komalan and Chauhan, 2017; Moreu, Kim, and Spencer, 2017).

Recently, the federal government mandated railroad companies to install a new system known as positive train control (PTC) integrated with standard-based wireless communication across the entire network. As explained by Peters and Frittelli ("Positive Train Control (PTC): Overview and Policy Issues," 2018), PTC is an "overlay-type" system in which the sensors, signals, and transponders are installed over existing track along with other wayside infrastructure. Analog radio signals integrate the whole system by communicating between wayside infrastructure, transponders, and trains. The entire system functions when the network operating center sends a one-way communication to a train that passes over a transponder embedded in the track. If the track movement is irregular or has a hindrance for normal functioning, the train receives information through installed equipment about speed restrictions, movements, and work zones, which alter the engineer accordingly. If the operating engineer fails to react on time or accidentally takes remedial action other than what is required, then this system can supersede the train engineer's authority to slow or even to stop the train in case of an emergency.

PTC provides a technological edge to the industry, which is designed to prevent train-to-train collisions and over-speed derailments, limit work zone accidents, and prevent the movement of trains through a switch left in the wrong position. Conversely, the system has some functional boundaries that limit its benefits. Such limitations include that the operating system does not track the real-time train location and can only communicate when a train passes the wayside infrastructure; Wi-Fi is not practical in the PTC, which restricts greater data transferability.

Some of the limitations of PTC have been addressed in a more sophisticated and more expensive variant of the system known as communications-based train control (CBTC). CBTC has a global positioning system (GPS) and other instrumentation installed, which helps track the train location. These instruments also help locate the train's whereabouts even when GPS cannot locate the signal. Like cell phone technology, the advanced infrastructure will communicate continuously using digital signals that provide

greater precision to the whole system ("Positive Train Control (PTC): Overview and Policy Issues," 2018).

Although PTC and CBTC have different capabilities and associated costs, FRA recommends that developing both functionalities simultaneously will be more beneficial and cost effective. Both PTC and CBTC are limited to focusing on improving communication between the system and train engineers to reduce specific train accidents and curbing their severity. But these technologies are disregarding the track-related issues, which are responsible for causing a more significant number of accidents.

To focus on track-related issues, based on PTC and CBTC foundations, this project uses inertial sensors aboard an in-service vehicle or locomotive that presents an opportunity to meet all these expectations and requirements. An in-service vehicle equipped with low-cost sensors may serve as a probe to detect and analyze real-time monitoring data with position information obtained from a GPS. Monitoring information obtained from in-service vehicles may enable the detection of maintenance problems at an early stage, thereby contributing to the renaissance of the railways by making maintenance tasks more efficient (Tsunashima et al., 2015). However, the implementation of low-cost sensor and GPS receivers pose significant challenges and limitations in detecting signals at high accuracy and high precision. First, the non-uniform sample rate of an accelerometer causes a problem in signal detection. Figure 1.1 demonstrates the challenge by showing the sample period distribution for an accelerometer set at maximum sample rate. From the distribution, it is clear the sample periods ranged from 1 to 18 milliseconds, but the primary mode is at approximately 11 milliseconds.
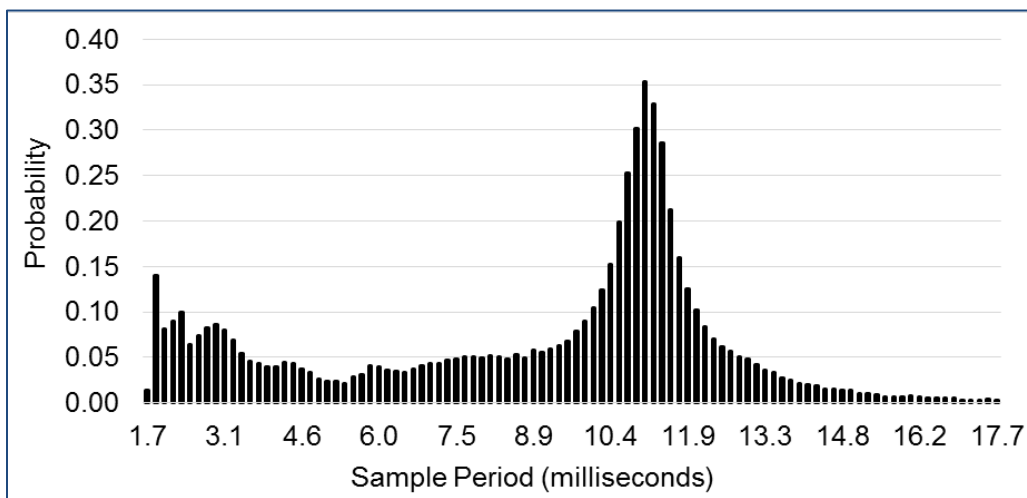
Figure 1.1 Sample period distribution

The non-uniform sampling caused each signal to have a different number of samples, and the distances associated with each sample were also non-uniformly distributed. Even with the uniform sampling, the distances associated with each sample will be misaligned because of the asynchronous spatial updates among traversals. Second, GPS receivers do not always provide reliable and accurate position information due to the following five reasons:

1. Standard low-cost GPS receivers provide position updates approximately once per second (Hunter et al., 2009). Given an inertial sensor that samples at 64 hertz on average, the GPS coordinate will update after every group of 64 inertial samples. Therefore, the system will, on average, tag blocks of 64 inertial samples with the same GPS coordinates. This causes low position resolution and consequently signal misalignment.

2. Position updates in a traversal are spatially asynchronous. That means the GPS updates from an individual traversal will be at different geospatial points along the path. Therefore, some position updates will not tag some signal peaks.
3. The geospatial position error is in two dimensions, causing deviation in position updates from the travel path.
4. The STD of the position estimates from GPS receivers is three to five meters along the travel direction (Hughes, 2004).
5. The non-line-of-sight (non-LOS) condition triggered by clouds, trees, or tunnels could block the reception of GPS signals in some locations, causing non-uniform update rates (Wang, Groves, and Ziebart, 2013).

The main purpose of this research is to develop and evaluate an automated and real-time inspection screening system for railroad tracks monitoring called Railway Autonomous Inspection Localization System (RAILS). The system will locate and characterize the possible track irregularity by analyzing the inertial dynamics of a regular service vehicle. The goal of this research is to develop a method that can utilize the data from very low-cost sensors, such as accelerometers and GPS receivers that are ubiquitous in smartphones. Therefore, the objectives of this research are:
1. Introduce the proposed system architecture and methods of big rail data processing.
2. Develop a technique that reduces the signal detection error and enhance the signal quality by ensemble average collected from multiple traversals of sensors aboard the vehicle. The technique includes distance interpolation and signal position alignment.
3. Determine and measure the appropriate digital filter cutoff frequency that will maximize the signal-to-noise ratio for subsequent feature extraction.
4. Develop a statistical model to enhance the detection and localization accuracy of track irregularities by extracting and combining features of the inertial signals obtained from multiple traversals of a track segment.
5. Investigate and validate the potential use of low-cost sensors aboard a regular service vehicle to monitor automatically and continuously for the inertial events produced by irregular track geometry.

The remainder of this report is organized as follows. Section 2 summarizes the literature review   Section 3 establishes the proposed system's architecture and describes the data collection and processing methods. Section 4 presents the framework that describes the methodology required to enhance the detection and localization of the track irregularity. The framework briefly demonstrates the effectiveness of the proposed method. Section 5 shows the potential use of low-cost sensors aboard the regular vehicle. The introduced approach estimates and validates the position of irregularity by comparing it with GTA overserved during the rail inspection. Section 6 summarizes the concluding remarks, prospects of generalizing the application, and comments on future work.

2.   LITERATURE REVIEW

A literature review describes the condition monitoring in railway applications and the types of railway track irregularities and their measurements. The review also highlighted the current track monitoring and inspection practice. Nonetheless, a review of the literature reveals the gap and scope for the improvement of established monitoring and inspection techniques.

2.1   Condition Monitoring in Railway

A condition monitoring system is generally used to track anomalies or deviations from some predetermined thresholds or a specific divergence from normal conditions by diagnosis data and prognosis for predicting future failure (Alemi, et al., 2017). Over the years, condition monitoring techniques have evolved from measurement-oriented strategies to a computer-based stratagem. Moreover, the development and integration of various algorithms, including signal filtering, artificial intelligence, and big data analytics, provides more resilience to the computer-based condition monitoring because of the possibility of rapid, dynamic, and cost-effective data processing and performing efficient developments (Ngigi, et al., 2012). Due to this rapid advancement, condition monitoring becomes a reliable and effective monitoring system for various industries. Ollier (Ollier, 2006) highlighted and recommended that condition monitoring is and will be crucial for the effective management of extensive rail infrastructure. Because condition monitoring can detect and identify structure or rail infrastructure deterioration, which is otherwise difficult to detect during normal railroad inspections, it also helps raise the alarm even before the hidden deterioration could cause any failure.

2.2   Rail Track Irregularities & Measurements

A railroad structure comprises a rail track structure and sub-structure. This report will include only the track structure faults that categorize into track anomalies and track abnormality. A rail anomaly is a rail surface that deviates from what is a standard or normal rail track geometry. This research only focuses on track abnormality detection.

Rail track geometry consists of several parameters that describe each rail's position or the track centerline occupied in space. Track geometry parameters can be specified or grouped by projecting into various planes they reside in. There are horizontal plane parameters, longitudinal vertical plane parameters, transverse vertical plane parameters, and track plane parameters. The main parameters of the track geometry include profile (longitudinal vertical plane), alignment (horizontal plane), cross-level and warp (transverse vertical plane), and gage (track plane) (Fateh, 2000; Sadeghi, 2010). Table 2.1 summarizes the types of track geometry parameters, detection tools, and their measurements with corrective action.

Table 2.1 Summary of Rail Track Surface Anomalies and Their Measurements

Rail Track Surface Anomalies	Detection Tools	Measurements	Corrective action
Gauge (56 ½ inches)	Gauge measurer	56 inches ≤ gauge ≤ 58 inches	Adding spikes or tie
Alignment (62 feet)	String	> 5 inches	
Cross level	Track level	> 3 ½ inch on curves > 2 ½ inch on normal	Tamping fresh ballast
Profile (62 feet)	String	> 3 inches	Tamping fresh ballast
Warp	Track level	> 3 inches	Tamping fresh ballast
Cant	Track geometry recording trolley	< 5 inches	
Twist	Track geometry recording trolley	1 inch in 62 feet	

2.3   Current Track Monitoring and Inspection Practice

In the railroad industry, a number of track monitoring techniques and practices are used to maintain a safe railroad network (FRA, *Track Inspection Time Study*, 2008). They are summarized in the following sections.

2.3.1   Visual Inspection

Visual inspection is the most common and traditional technique used in the railroad industry. It is conducted either on foot or in a hi-rail vehicle. Hi-rail is a dual-mode vehicle that can operate on both roads and tracks. Hi-rail trucks have a set of flanged steel wheels that, when deployed, allow the vehicle to travel on railways. In visual inspection, the railroad can deploy a single inspector or multiple inspectors depending upon the requirement of the track inspection. In the case of multiple tracks, two inspectors may conduct the inspection together using the hi-rail vehicle so that each inspector can focus and address all expected track defects on their respective track. This method is very slow, laborious, and requires track closure to search for possible track defects. Other track defects are subtler and require vigilance and proper training to be detected.

2.3.2   Gage Restraint Measurement System (GRMS)

GRMS is a technique that measures the track structure's ability to maintain its gage under constant vehicle applied gage-spreading load. GRMS vehicles include a third axle that applies a constant lateral and vertical load on each of the rails. The gage strength is quantified by comparing the loaded and unloaded gage measurements depending on the applied load. Therefore, it will allow railroads to focus their tie and fastener replacement programs effectively.

2.3.3   Track Geometry Measurement System (TGMS)

TGMS technique has been widely used over the past decades to measure track geometry. To implement TGMS, most railroads have specially equipped railcars called "track geometry cars." Railroads usually deploy their cars one to three times per year over a section of track. These cars can record track geometry measurements foot by foot, making them faster than manual measurements of track geometry data.

2.3.4 Ground Penetrating Radar (GPR) System

Ground-penetrating radar (GPR) systems are not as extensively used by the railroad industry as TGMS and GRMS. These systems use electromagnetic wave technology to capture information on different layers of the track structure. GPR systems are also capable of identifying moisture content, subgrade discontinuities, and other anomalies at various depths in the track structure, which could result in poor support as the track settles; as a result, GPR allows for the detection of subsurface track structure issues before they are established on the surface in the form of poor track geometry.

2.3.5 Trends and Other Development of Railroad Condition Monitoring Techniques

Nowadays, using in-service vehicles for the condition monitoring of railroads is prevailing and gaining greater popularity worldwide. This technique is generally referred to as vehicle/track interaction where accelerometers, gyros, and a GPS are often used to detect track surface irregularities, vehicle speed and location, and vehicle dynamic behavior. Moreover, the accelerometer measurements can be used to quantify the ride quality (FRA, *Track Inspection Time Study*, 2008). Such a technique can be considered as a reasonable approach for monitoring the condition of the track during regular operation hours (Tsunashima et al., 2015).

Lederman et al. (Lederman et al., 2017) presented a data fusion approach for data-driven track monitoring with in-service trains. They mounted two uniaxial accelerometers vertically on two trains and placed a BU-353 GPS receiver under the roof of the first train. Real et al. (Real et al., 2011) developed a technique to estimate track profile and relate it to track conditions. Weston et al. (Weston et al., 2015) adopted a sensor mounted on a rail bogie to monitor lateral and vertical irregularities by double integration of acceleration signals and high pass filtering. Kim et al. (Kim et al., 2009) introduced a high-speed track irregularity measurement system, installed it into an HSR-350x locomotive, and analyzed the test run result on a high-speed line. The Korean high-speed train (HSR-350x) is composed of seven cars, two power cars, two motorized cars, and three trailer cars. The system measures longitudinal level, cross-level, gauge, alignment, and twist irregularity. A GPS and an encoder attached to the axle also measured train speed. The result shows excellent performance of the measurement system.

Many railroad companies have also adopted machine-vision inspection technologies to detect anomalies in the track structure. This type of system uses automated algorithms for the extraction of designated features from high-resolution 2D images or 3D light detection and ranging (LIDAR) data. Santur, Karaköse, and Akın used 3D laser cameras to provide high accuracy rates in real time and found that the system was prone to false positives. The cost of these applications is high because of the need for high accuracy of detection at high speeds (Santur, Karaköse, and Akın, 2016). Likewise, Alippi et al. (Alippi et al., 2000) represent an innovative approach to track profile measurements. A real-time image-processing-based technique was adopted to reconstruct and measure the profile by analyzing a laser-scanned CCD-camera image. A light source and high-resolution cameras found an increase in false positives due to stains from oil and dust particles.

All these studies and research demonstrated a successful assessment of irregular track geometry. However, they all focused on the use of complex and expensive sensors and cameras, which result in high maintenance costs, greater energy requirements, and low robustness. Therefore, research gaps are 1) the impacts of the non-uniform sample rate of accelerometer sensor and GPS error of the GPS receiver on the detection of irregularity; 2) the effects of combining the signals from multiple train traversals at different speeds to detect track irregularities throughout the network.

3.   SYSTEM SETUP

3.1   Architecture

Figure 3.1 provides the overall architectural overview of the proposed approach. The smartphone aboard the in-service vehicle, or in any other locomotive sense, records multi-directional inertial, velocity, and geospatial position data during ordinary commercial operations. The measured data are transferred over a communication channel using a secure standard protocol and uploaded to the remote server. The remote algorithms compress and process three-dimensional linear acceleration, angular acceleration, and geospatial data to identify and localize the signature of track irregularity. Railroad inspectors can comprehend and visualize reliable condition information by using color-coded TRQ values onto maps of the rail routes via the defect symptom visualization engine. Figure 3.2 shows the block diagram of the approach. Railroad agencies can use such a tool to optimize inspection and maintenance procedures to reduce the frequency of slow orders and track closure while minimizing the cost and safety risk. It can also help asset managers maintain reliable track and equipment condition information to make informed decisions, leading to resource optimization.
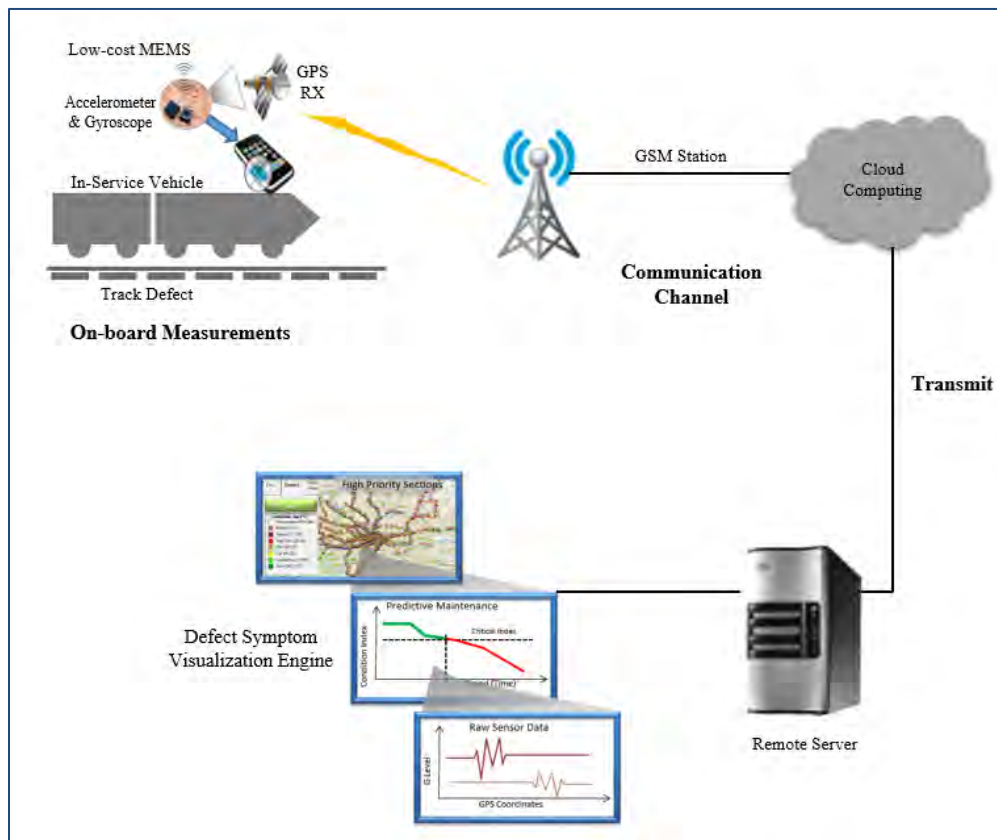
Figure 3.1 Architectural overview of the rail track condition monitoring system
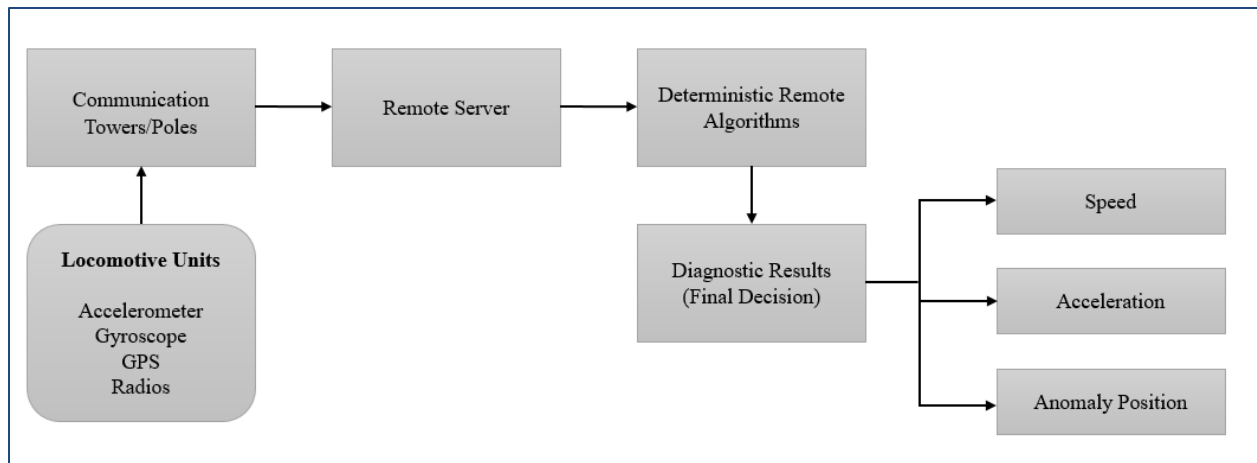
Figure 3.2 Block diagram of the rail track condition monitoring system

3.2 Data Collection and Processing

Smartphones provide all the capability to develop and test the viability of an automated track monitoring system that uses a dedicated sensor onboard regular locomotives (Lane et al., 2010; Bridgelall and Tolliver, 2020). In lieu of the absence of low-cost and low-power commercial sensors, this study uses a smartphone application "Railroad Infrastructure and Vehicle Evaluation Technology" (RIVET), which is capable of autonomously collecting and uploading data from hi-rail vehicles, geometry cars, locomotives, and end-of-train cars where power is available.

3.2.1 Data Collection Setup

The study uses Harsco Rail's LD515 HY-RAIL equipped with Harsco model 1515 rail gear with steel wheels. The device used for data collection was Google Pixel. The smartphone location setup considered the different kinetic energy responses, power supply resources, and the GPS signal strength. Based on the potential consideration, three android phones are placed securely in a hi-rail vehicle shown in Figure 3.3. The first phone is installed on the dash and the other two phones placed in the rear interior of the vehicle under a seat. Therefore, the device recorded the intensity of roughness that the seated driver and passenger would have experienced. All three phones can access the power and collect the data from multiple traversals over the same segments.

The primary areas of interest are locations where known geometry issues such as warp, profile, and alignment exist. Basically, measurements recorded by the inspector are compared to measurements predicted by the automated in-vehicle system.
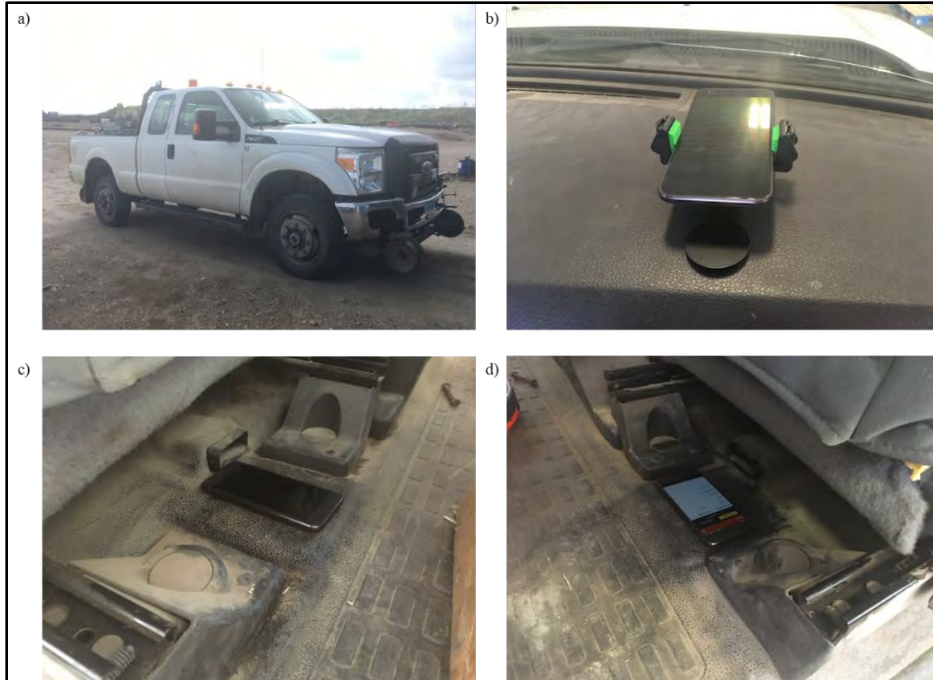
Figure 3.3 a) Hi-rail vehicle used b) Phone 1 installed on dashboard
c) Phone 2 installed under driver seat d) Phone 3 installed under passenger seat

Before the project commences, a data confidentiality pledge will be signed by UGPTI to protect the confidentiality of the data collected on the Northern Plains Railroad's property. The data will be used for research purposes only and will not be shared with the public. Specific location information will also be kept confidential.

3.2.2   Data Collection Tool

The data collection app is called RIVET, which was developed as part of a Mountain Plains Consortium project at North Dakota State University, Fargo, ND, USA. The app currently serves only as a data collection device. Figure 3.4 is a screen shot of the RIVET app. The accelerometer of the smartphone was set to sample at a rate of 400 Hz and the GPS receiver updates location in approximately 1 Hz. The app logged inertial and geospatial position data from multiple traversals of railroads. The app generates the output files in comma-separated format with following fields shown in Table 3.1.
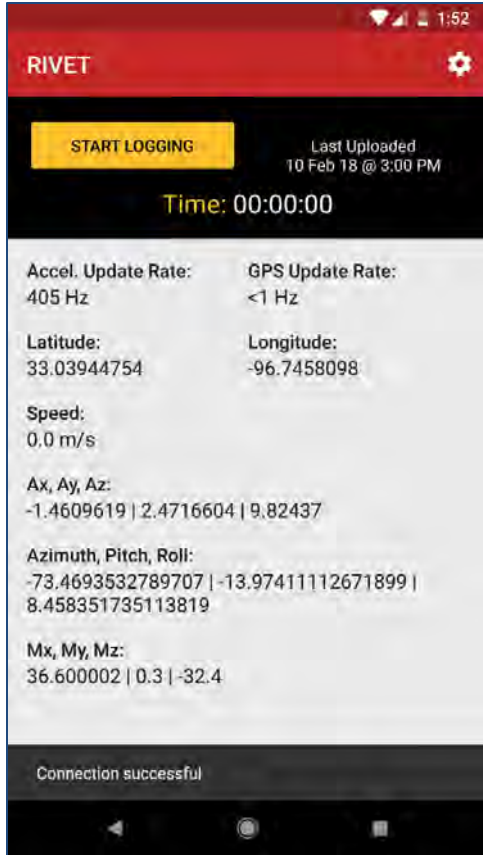
Figure 3.4 Screen shot of the RIVET app

Table 3.1 Format of Smartphone Application Data Log

DateTime	Lat	Lon	Speed	Ax	Ay	Az	Azimuth	Pitch	Roll	Rx	Ry	Rz
1.55E+12	48.17365	-96.2332	10.47	-1.06	-0.73	10.04	-45.9	4.79	8.38	-0.01	0.02	-6.15E-04
1.55E+12	48.17365	-96.2332	10.47	-1.06	-0.73	10.04	-45.9	4.79	8.38	-0.01	0.02	-6.15E-04
1.55E+12	48.17365	-96.2332	10.47	-1.06	-0.73	10.04	-45.9	4.79	8.38	-0.01	0.02	-6.15E-04
1.55E+12	48.17365	-96.2332	10.47	-1.06	-0.73	10.04	-45.9	4.79	8.38	-0.01	0.02	-6.15E-04
1.55E+12	48.17365	-96.2332	10.47	-1.06	-0.73	10.04	-45.9	4.79	8.38	-0.01	0.02	-6.15E-04
1.55E+12	48.17365	-96.2332	10.47	-0.85	0.04	9.561	-45.9	4.79	8.38	-0.01	0.02	-6.15E-04

The first row contains a header with labels for each column of data collected from inertial and GPS sensors. Each data row updates with each sample of the accelerometer data. The DateTime variable is in epoch time format for milliseconds. The GPS data are latitude (Lat) and longitude (Lon) in decimal degree format. The ground speed (Speed) is in units of m s-1. The Ax, Ay, and Az, signals are the linear accelerations in m s-2 along lateral, longitudinal, and vertical directions, respectively. The Rx, Ry, and Rz signals are the angular rotations in degrees-per-second around the X, Y, and Z axis, respectively. The integrated gyroscope produces the "Azimuth," "Pitch," and "Roll" for the sensor orientation angles in degrees. Azimuth is angle around the z-axis in degrees; Pitch is angle around the x-axis in degrees; Roll is angle around the y axis in degrees. Smartphone orientation is positive in the clockwise direction as shown in Figure 3.5.

10

Figure 3.5 Orientation of the mobile device with three angular quantities

3.2.3 Data Processing

Once data are collected through the app, the logged files start uploading automatically to a server. If in the case of any pending logged file, the researcher can tap "stop logging," which uploads any queued file to the server. The app utilizes any of its available wireless connections to communicate with a server. The universal resource locator (URL) for the server is entered in the app setup screen. After the raw data files become available on the server, the pre-processing data algorithm combines all the transmitted data into a big data frame and sorts it by timestamps. After sorting the data, the algorithm removes the bad GPS data, converts linear acceleration Ax, Ay, and Az from ms^{-2} to g-force units by dividing each value by 9.80665 m s-2 and converts rad s^{-1} units into degree s^{-1} units by multiplying by $180/\pi$, before calculating the road impact factor (RIF). After pre-processing, a post-processing algorithm generates the RIF indices using equation 1 and 2. Figure 3.6 shows the computational flow chart with detailed procedure for the RIF indices calculation.

The RIF denoted "RIF$_{Gt}$ "a measure of intensity that is proportional to resultant acceleration rate along the x- and z-axis" such that

$$RIF_{Gt} \;=\; \sqrt{\frac{1}{L}\sum_{n=0}^{N-1}|Gt_n\,v_n|^2\,\delta t} \tag{1}$$

where RIFGt is the average g-force magnitude experienced per unit of distance L traveled. A speed sensor produces the instantaneous traversal speed v_n at sample n for N total samples (Bridgelall et al., 2016). The sample period is δt. Gt is a resultant acceleration signal sample in units of g-force such that

$$Gt = \sqrt{(G_x)^2 + (G_z)^2} \tag{2}$$

where Gx and Gz are acceleration signals about lateral and vertical directions, respectively. Vertical accelerations are produced when a locomotive traverses a broken rail or ties on railroad tracks. Lateral acceleration acts transversely to a vehicle's direction of travel. Therefore, it could indicate the presence of irregular track geometry.
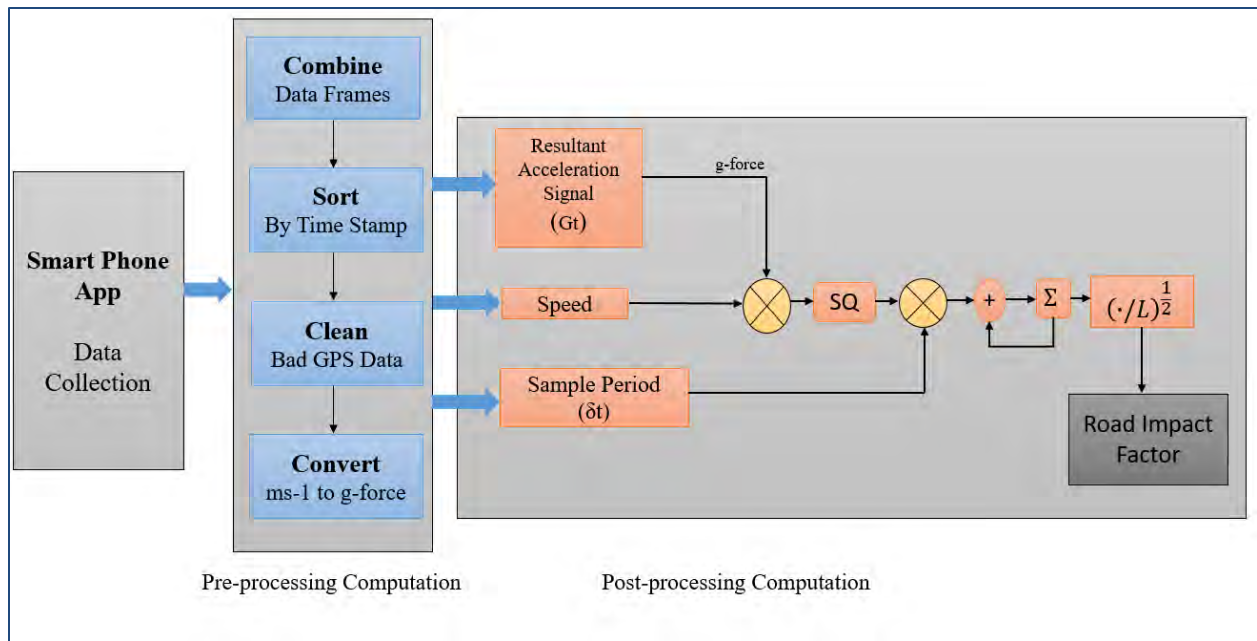
Figure 3.6 Detailed procedure of RIF calculation

Figure 3.7 shows RIF values follow a normal distribution. Based on the distribution, a color-coding scheme and percentile thresholds are defined to visualize the peak inertial event (PIE) shown in Figure 3.7 and Table 3.2. PIEs are the peak accelerometer signals as g-force values are produced by isolated roughness.
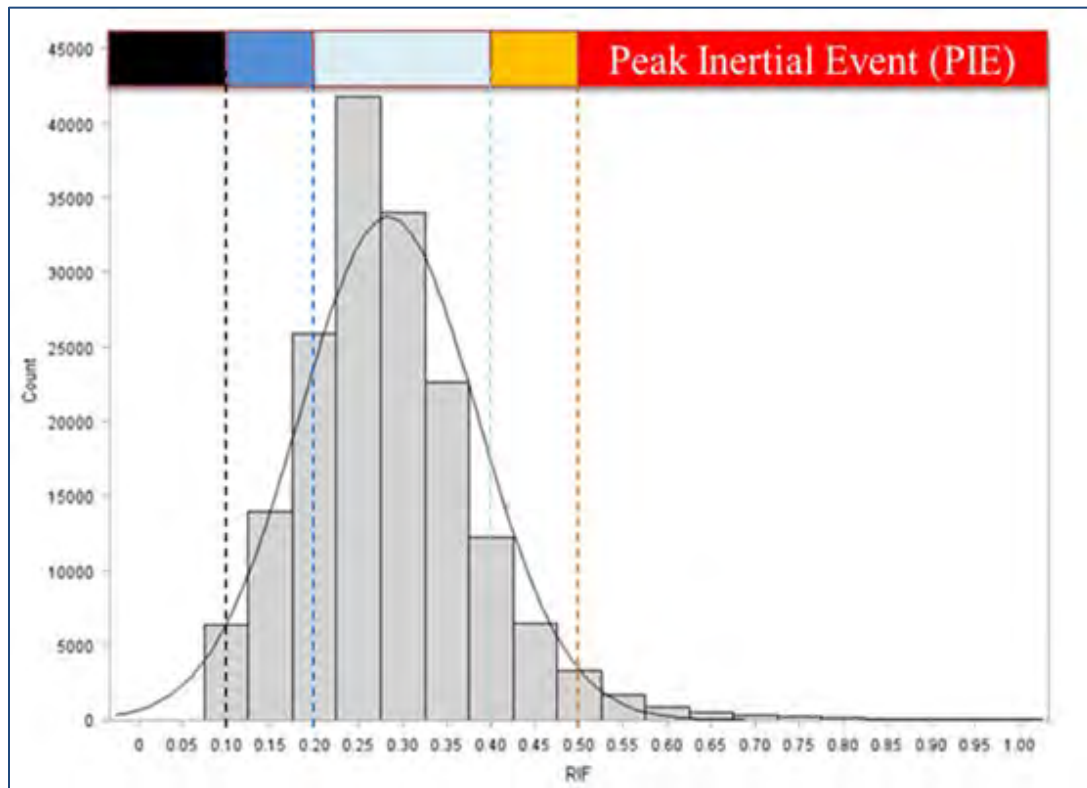
Figure 3.7 Fixed color-coding scheme based on normally distributed RIF values.

Table 3.2 Color Coding Scheme for PIEs Visualization

Color	Percentile
Black	Lower 2.5-percentile
Blue	Within the lower 1 to 2 standard deviation from the mean
Light Blue	Within 1 standard deviation of the mean (68%)
Orange	Within the upper 1 to 2 standard deviation from the mean
Red	Upper 2.5-percentile

4. FRAMEWORK: ENHANCE THE DETECTION AND LOCALIZATION OF TRACK IRREGULARITY

This section presents a general framework to develop methods that enhance the detection and localization of rail track irregularity. The authors employed data used in the literature (Bridgelall et al., 2016) to demonstrate the effectiveness of the proposed approach.

4.1 General Framework of the Application

The existing literature categorized onboard condition monitoring sensors as tachometers, inertial sensors, and satellite-based positioning systems (Otegui et al., 2017). An automated track geometry monitoring system requires accurate localization. The global navigation satellite systems, such as the Global Positioning System (GPS), that collect and record the locations of detected track anomalies have been used for decades. However, low resolution and low accuracy of GPS receivers and non-uniform sample rates of the inertial sensors result in signal position misalignment and additive signal noise. Consequently, these factors reduced the signal-to-noise ratio (SNR) needed to reliably detect and localize track irregularities. In particular, false positives and false negatives can increase when SNR decreases. Also, it poses significant challenges in signal processing, feature extraction, and signal combination.

Methods for estimating the position updates from the inertial sensors (accelerometers and gyroscopes) and magnetometers (Bistrov, 2012) are explored. Conversely, sensor bias can cause drift in the velocity, position, and attitude information. The bias term affects the velocity and attitude linearly with time, and the position quadratically (Sukkarieh, Nebot, and Durrant-Whyte, 1999). Therefore, distance interpolation between GPS positions is necessary to increase the position resolution. Moreover, to enhance the signal quality, the technique leverages the large volume of sensor signals from multiple traversals of a road or rail segment by ensemble average. The ensemble averaging will boost the information and reduce the noise in signals. A significant benefit of the ensemble averaging approach is that both precision and accuracy increase with the number of traversals. Then again, ensemble averaging requires position repeatable signals. Consequently, this research dominates these challenges by integrating techniques of distance interpolation, heuristic and correlation alignment. Nonetheless, appropriate noise filtering is necessary to maximize the SNR of each signal prior to feature extraction. This section introduced a general framework that reduces detection error and enhances the quality of the signals and features to get the ensemble averaging (EA) through heuristic alignment, correlation alignment, and signal filtering that improves S/N ratio and minimizes the false positives and false negatives.

4.2 Distance Interpolation

The developed approach position aligns the inertial signals by replacing the GPS position tags with higher resolution distance tags and identifies the starting position along the traversal path. Interpolation can boost the distance resolution using the instantaneous speed and sample periods obtained from the onboard sensors. Table 4.1 shows the format of the data collected using iPhones® retrofit onto the dashboard of a sedan.

Table 4.1 Format of Data Sample

Time	Gz	Speed	Lat	Lon
44.142	-1.057	9.586	45.263	-93.711
46.768	-1.216	9.586	45.263	-93.711
50.260	-1.087	9.586	45.263	-93.711
62.927	-0.854	9.586	45.263	-93.711
73.909	-0.912	9.586	45.263	-93.711
86.754	-0.942	9.586	45.263	-93.711
95.669	-1.001	9.586	45.263	-93.711
110.365	-1.022	9.586	45.263	-93.711
118.253	-1.096	9.586	45.263	-93.711
128.695	-1.013	9.586	45.263	-93.711

The dataset contains geospatially tagged inertial signals from 53 traversals of a road segment that contained a single isolated bump. Traversing the bump produced a consistently large PIE that is identical to the inertial signal produced when a railroad car traversed a verified track irregularity near a rail-grade crossing. The time column is the sample period in milliseconds. The inertial signal "Gz" is linear accelerations in the vertical directions with values in units of g-force. The ground speed (GSpeed) is in units of meters-per-second (m s-1). The GPS receiver produced latitude (Lat) and longitude (Lon) in decimal format. The GPS receiver sampled at approximately one hertz. It means that the coordinates remained unchanged for GPS blocks of inertial samples. Figure 4.1 illustrates non-uniform distribution of accelerometer updates and the position of the GPS update.
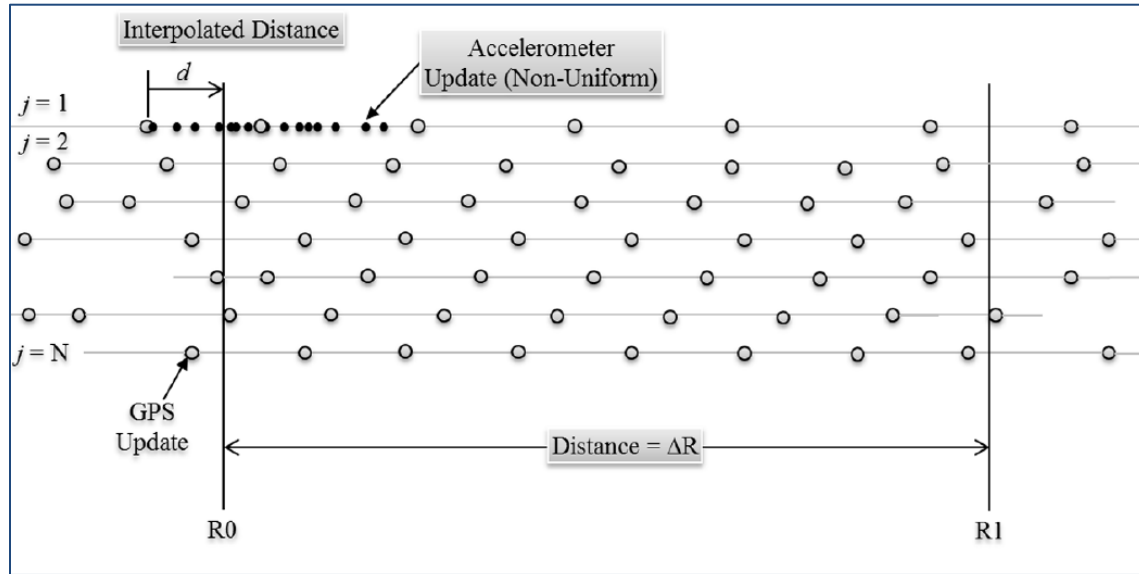
Figure 4.1 Non-uniform distribution of accelerometer signals and the position of the GPS update

The interpolation uses the speed and sample intervals in the dataset to accumulate distances from a reference point. The distance for the first reference point d_0 is set to zero and subsequent samples are located at the accumulated distance

$$d_n = d_{n-1} + v_n \times \Delta\tau_n \tag{3}$$

15

where n is the sample number, v_n is the instantaneous speed for sample n, and $\Delta\tau_n$ is associated the sample period. The high-resolution differential time interpolation produces in a high-resolution distance signal. It allows the ensemble averaging to compute signal means within small, fixed distance windows along the traversal path for a vehicle traveling at any speed and sampling at any rate. The applied strategy is compatible with traditional linear referencing techniques that specify distances from a reference position such as a milepost or a landmark to improve accuracy and precision (Curtin, Nicoara, and Arifin, 2007). After interpolation, the procedure extracts approximately equal length traversals. Figure 4.2 shows two extracted traversals starting from the interpolated zero distance position.

The relative positions of the PIE explain the amount of misalignment between the two signals. The negative peak and the positive peak of the PIE reflects the maxima and minima of the vertical acceleration profile from traversing the bump.
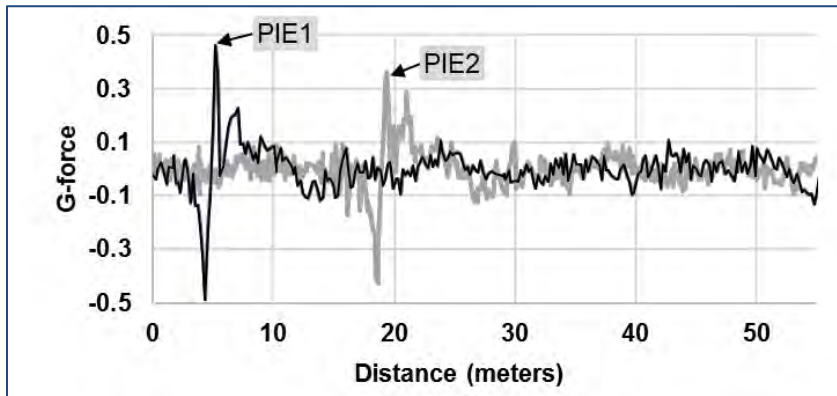
Figure 4.2 Position misalignment of PIEs in the interpolated signal

4.3   Signal Alignment

Signal alignment means that the signal peaks occur at the same distance position. This section briefly explains the development of two signal alignment techniques, heuristic alignment and correlation alignment. Each method uses distance interpolation with velocity and sample period updates to increase the position resolution substantially, and consequently improve signal alignment. Distance interpolation also accommodates the non-uniform sampling of the inertial signals.

4.3.1   Heuristic Alignment

This research introduced and evaluated four heuristic alignment techniques: dual geofence, midway midpoint, centroid asymmetric, and centroid symmetric.

Before position alignment, the geospatial pre-processing algorithm specifies the road or rail segment of interest by contracting the starting and ending geofences. The algorithm then extracted the data for all traversals contained within the geofence boundaries. The data used in this study are collected from a sedan shown in Table 4.1. The dataset did not show any missing GPS updates. That means the variation in the relative distance of the PIEs is due to the misalignment observed in that environment. In the dataset, there is a given signal feature, called known ground truth, that produces isolated PIE for each traversal.

After the traversal's extraction, distance interpolation is done using equation (3). However, the distance interpolation needed at least one GPS block from which to begin. Therefore, all traversals must have at least one GPS block to the left and right of a reference distance. The algorithm then creates two inner geofences, $R0$ and $R1$, perpendicular to the traversal path, which serve as a common distance reference.

Next it located the optimum position for each geofence, $R0$ and $R1$, at the mid-distance point of the segment, and then moved them outward in opposite directions. The reference distance $d0 = 0$ is the first sample of the GPS block that is closest to the left of $R0$.

Reference method. The known ground truth produced the isolated PIE for each traversal. Therefore, the reference method can remove the position errors from GPS estimates to evaluate residual errors. The algorithm interpolates distances to the left and right of the negative peak of each PIE, and then applies the align procedure to set the first and last samples, respectively, of the aligned traversals. The residual errors are from estimates of the speed and sample intervals.

Dual geofence. For all traversals, this method begins by finding the GPS block GL that is closest to the left of $R0$ and setting the index of the first sample within the block as the interpolation initialization sample where $n = 0$. The algorithm then increments index n from left to right until it identifies the first sample of the aligned signal, which is the index that satisfies the optimization shown in Figure 4.3.

$$
\begin{aligned}
&\text{Minimize} \\
&\text{subject} \qquad x_n - \left\| G_{\mathrm{L}} - \overleftrightarrow{R0} \right\|^{\perp} \\
&\qquad\qquad \text{to} \\
&x_n = x_{n-1} + v_n \times \Delta\tau_n,\ n = \{1, 2, \cdots\} \\
&\qquad\qquad \text{and} \\
&\qquad\qquad x_0 = 0.
\end{aligned}
$$

Figure 4.3 Optimization of aligned signals

The algorithm then finds the GPS block GR that is closest to the right of $R1$ and sets the index of the last sample in that block as the interpolation initialization sample where $k = 0$. Next, the algorithm increments index k from right to left until it identifies the last sample of the aligned signal. Finally, the algorithm interpolates the distance of the aligned traversal by setting the distance of the first sample to zero and ending the interpolation with the last sample.

Midway midpoint. This method starts by spatially joining a centerline $C0$ that is at a distance *midway* between $R0$ and $R1$ and is perpendicular to the traversal path. The algorithm then locates the GPS block GMj that is *closest* to the centerline on either side. The algorithm then sets the distance of the first sample in that GPS block to zero and interpolates to find the *midpoint* distance xMj K of that block. The *align* procedure then sets the block's midpoint distance as the interpolation initialization sample where $k = 0$ and interpolates distances to the left and right. The indices of the samples to the left and right of the interpolation initialization sample are negative and positive integers, respectively. Next, the procedure determines the minimum of the distances Lj to the left of the center point as $Lmin$ and the minimum of the distances Rj to the right of the center point as $Rmin$. Subsequently, the samples of each traversal with distances closest to $Lmin$ and $Rmin$ become the first and last samples, respectively, of the aligned traversal. The notation for the first and last samples of traversal j is xLj and xRj, respectively. Figure 4.4 shows the mid-way midpoint algorithm.

Figure 4.4 Midway midpoint algorithm

Centroid Asymmetric. This method identifies the single GPS block of each traversal that is closest to the $C0$ line and then calculates a centroid from the geospatial positions of that set of coordinates. Subsequently, the method spatially joins a line $C1$ that bisects the centroid and is parallel to C0. Next, the algorithm finds the left GPS block GLj that is closest to $C1$ and then interpolates the distance from the first sample $x0$ of that GPS block to find sample $X^{\frac{cj}{K}}$ that satisfies the optimization shown in Figure 4.5.

The algorithm then calls the align procedure with $X^{\frac{cj}{K}}$ as the input. The asymmetric aspect of this algorithm is that it selects the GPS block immediately to the *left* of the centroid rather than the GPS block that is closest on either side.

Figure 4.5 Optimization algorithm

Centroid Symmetric. This method follows the procedure of the *Centroid Asymmetric* method with the exception that it starts with the GPS block that is closest to the centroid geofence, on either side.

4.3.1.1 Normality Tests

The performance of all four heuristic methods are evaluated through normality tests of the PIE distributions resulting from each method of alignment. The normality tests included the Kolmogorov–Smirnov, Anderson–Darling, and Cramer–von Mises tests. The Kolmogorov–Smirnov, Anderson–Darling and Cramer–von Mises tests for normality are based on the empirical distribution function (EDF) and are often referred to as EDF tests (Yap and Sim, 2011). The tests reject the null hypothesis if the p-value of the test-statistic is less than 0.05.

The empirical distribution function is defined for a set of n independent observations $X_1 \ldots X_n$ with a common distribution function F(x). Under the null hypothesis, F(x) is the normal distribution. The empirical distribution function, $F_n(x)$, is defined as

$$F_n(x) = \begin{cases} 0, & x < X_{(1)} \\ i/n, X_i \le x \le X_{(i+1)}, & i = 1, \ldots, n-1 \\ 1, & X_{(n)} \le x \end{cases} \tag{4}$$

Note that $F_n(x)$ is a step function that takes a step of height $1/n$ at each observation. This function estimates the distribution function $F(x)$. At any value x, $F_n(x)$ is the proportion of observations less than or equal to x, while $F(x)$ is the probability of an observation less than or equal to x. EDF statistics measure the discrepancy between $F_n(x)$ and $F(x)$. The EDF tests make use of the probability integral transformation $U = F(X)$. If $F(X)$ is the distribution function of X, the random variable U is uniformly distributed between 0 and 1. Given n observations X(1), ... X(n), the values U(i) = F(X(i)) are computed. These values are used to compute the EDF test statistics, as described in the three tests.

Kolmogorov-Smirnov Test: The Kolmogorov-Smirnov Statistics (D) is defined in Equation (5)

$$D = sup_x |F_n(x) - F(x)| \tag{5}$$

The Kolmogorov–Smirnov statistics belong to the supremum class of EDF statistics. This class of statistics is based on the largest vertical difference between $F(x)$ and $F_n(x)$. The Kolmogorov–Smirnov statistic is computed as the maximum of D+ and D-, where D+ is the largest vertical distance between the EDF and the distribution function when the EDF is greater than the distribution function, and D- is the largest vertical distance when the EDF is less than the distribution function.

$$
\begin{aligned}
D^+ &= max_i \left(\frac{i}{n} - U_{(i)} \right) \\
D^- &= max_i \left(U_{(i)} - \frac{i-1}{n} \right) \\
D &= max(D^+, D^-)
\end{aligned}
\tag{6}
$$

Anderson-Darling Test: The Anderson–Darling statistic and the Cramer–von Mises statistic belong to the quadratic class of the EDF statistics. This class of statistics is based on the squared difference $(F_n(x) - F(x))^2$. Quadratic statistics have the following general form:

$$Q = n \int_{-\infty}^{+\infty} \left(F_n(x) - F(x) \right)^2 \varphi(x) dF(x)$$

The function $\varphi(x)$ weights the squared difference $(F_n(x)-F(x))^2$. The Anderson–Darling statistics (A^2) is defined as

$$A^2 = n \int_{-\infty}^{+\infty} \left(F_n(x)-F(x) \right)^2 [F(x)(1-F(x))]^{-1} dF(x) \tag{8}$$

Here the weight function is $\varphi(x) = [F(x)(1-F(x))]^{-1}$. The Anderson–Darling statistic is computed as

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} \left[(2i-1)\log U_{(i)} + (2n+1-2i)\log(\{1-U_{(i)}\}) \right]$$

Cramer-Von Mises Test: The Cramer-Von Mises statistic (W^2) is defined as:

$$W^2 = n \int_{-\infty}^{+\infty} \left(F_n(x) - F(x) \right)^2 dF(x)$$

Here the weight function is $\varphi(x) = 1$. The Cramer–von Mises statistic is computed as

$$W^2 = \sum_{i=1}^{n} \left(U_{(i)} - \frac{2i-1}{2n}\right)^2 + \frac{1}{12n}$$

Table 4.2 Summary Statics of All Methods

	PIE		P-value			Length
Method	STD	Skew	KS	CVM	AD	STD
Reference	0.09	0.159	>0.150	>0.250	>0.250	0.09
Centroid asymmetric	4.04	0.335	>0.150	0.128	0.15	0.13
Dual geofence	4.84	1.546	<0.010	<0.005	<0.005	5.59
Centroid symmetric	6.32	0.522	0.101	0.099	0.08	0.24
Midway midpoint	7.24	0.69	0.125	0.072	0.05	0.20

From the summary statics of all methods in Table 4.2, it is clear the reference method outperforms the proposed method with all statistics including skewness. Among the four methods, the centroid asymmetric provided the lowest standard deviation of the PIE and the least skewness. Among the four methods, the centroid asymmetric method performed approximately 17%, 36%, and 44% better than the dual geofence, centroid symmetric, and midway midpoint methods, respectively.

The null hypothesis of the normality test is that there is no significant departure from normality. The commonly accepted approach is to reject the null hypothesis when the p-value is less than 0.05, or 5%. Both the reference method and the centroid asymmetric method fails to reject the hypothesis for all three tests. This result indicates that those methods maximally agree with a Gaussian distribution. Since both the skewness and normality tests agree, this research selects the centroid asymmetric method for signal alignment.

4.3.2   Correlation Alignment

The cross-correlation alignment technique requires that all signals have the same number of samples and represent the same distance positions. Despite approximately equal length distance interpolated traversals, the number of samples may change due to non-uniform sampling. Uniform resampling all the signals at the same uniformly spaced distances fulfill the requirement of the cross-correlation. The uniform resampling algorithm is shown in Figure 4.6. The number of samples S_M and distance D_M of a reference signal are the inputs. The reference signal is selected as the signal with the most samples because it represents the highest distance precision available in the dataset. For each traversal, the uniform signal resample algorithm first defines the function $f(x)$ based on the original distance vector x and the signal vector y. The algorithm next creates a new vector $\hat{x}$ that has the same number of samples and distances as the reference signal. Using the defined function f, the algorithm then estimates a new signal by evaluating the function at the new uniformly spaced distances in vector $\hat{x}$.

```
Uniform Signal Resample
Inputs:
    𝕋                          # Extracted traversal tables [signal, distance]
    SM                         # Number of samples (rows) of reference signal.
    DM                         # Distance of reference signal.
Begin:
    For ∀j ∈ 𝕋
        y_ij = f(x_ij)                              # Define y(x) at the original distance
        x̂_ij = i (D_M / S_M), i = {0,1, … (S_M − 1)}   # Create new uniformly sampled dis
        ŷ_ij = f(x̂_ij)                             # Estimate y(x) at new distances
End:
Outputs:
    𝕋                          # Uniformly resampled traversals (Signal, Distance)
    R                          # Index of the reference signal in the dataset
```

Figure 4.6 Resampling algorithm

After resampling at the same and uniformly spaced distance, all extracted traversals are now ready for correlation alignment.

A cross-correlation performs between the reference signal and the signals of all the other traversals. Figure 4.7 shows the correlation alignment algorithm. For the correlation alignment, uniformly resampled signal, distance, index of reference signal, and number of samples S_M in the reference signal are input values.

```
Correlation Alignment
Inputs:
    𝕋                          # Uniformly resampled signal tables [signal, distance]
    R                          # Index of reference signal
    S_M                        # Number of samples in the reference signal.
Begin:
    For ∀j ∈ 𝕋
        r_kj = Σ_{n=0}^{S_M−1} x̂_{n,j} × x̂_{n−k+S_M−1,R} ,  k = {0,1, … (2S_M − 2)}   # Cross-correlate
        s_j ← ⟨max_{∀k}(r_kj)⟩ − S_M + 1          # Shift position of maximum cross-correlation
    If s_j > 0              # Positive shift
        x_ij = x̂_ij,  i = {s_j, s_j + 1, … (S_M − 1)}       # Advance distance vector
        y_ij = ŷ_ij,  i = {0,1, … (S_M − 1 − s_j)}          # Truncate signal vector
    Else
        x_ij = x̂_ij,  i = {0,1, … (S_M − 1 − |s_j|)}        # Truncate distance vector
        y_ij = ŷ_ij,  i = {|s_j|, |s_j| + 1, … (S_M − 1)}   # Advance signal vector
End:
Outputs: 𝕋                  # Aligned signal tables [signal, distance]
```

Figure 4.7 Correlation alignment algorithm

In Figure 4.7, $\max_{\forall k}(r_k)$ is the function that shifts the position of maximum cross-correlation signal. It is possible that there will be many values in the cross-correlation vector rk that are equal to the maximum value. As a result, the operator ⟨·⟩ is used to return the index of the first maximum value. The shift position is relative to the distance of the first sample of the reference signal, which is zero. Hence, the position of zero shift is when the distance labels of both vectors overlap. This is the default shift position when there are no signals to correlate, which means that all the sensor samples contain only noise. The positive and negative shifts are the shifts to the right and left of the overlapping position. The shifted signal inherits the distance labels of the reference signal. When the shift is positive, the first sample of the shifted signal will inherit the distance labels starting at the shifted number of positions to the right of the distance vector of the reference signal. This is equivalent to advancing the distance vector of the signal by the number of shift positions for maximum cross-correlation. Since the signal is advanced, the tail will be

21

```
Uniform Signal Resample
Inputs:
    𝕋                          # Extracted traversal tables [signal, distance]
    SM                         # Number of samples (rows) of reference signal.
    DM                         # Distance of reference signal.
Begin:
    For ∀j ∈ 𝕋
        y_ij = f(x_ij)                                  # Define y(x) at the original distance
        x̂_ij = i (D_M / S_M), i = {0,1, … (S_M − 1)}     # Create new uniformly sampled dis
        ŷ_ij = f(x̂_ij)                                 # Estimate y(x) at new distances
End:
Outputs:
    𝕋                          # Uniformly resampled traversals (Signal, Distance)
    R                          # Index of the reference signal in the dataset
```

Figure 4.6 Resampling algorithm

After resampling at the same and uniformly spaced distance, all extracted traversals are now ready for correlation alignment.

A cross-correlation performs between the reference signal and the signals of all the other traversals. Figure 4.7 shows the correlation alignment algorithm. For the correlation alignment, uniformly resampled signal, distance, index of reference signal, and number of samples S_M in the reference signal are input values.

```
Correlation Alignment
Inputs:
    𝕋                          # Uniformly resampled signal tables [signal, distance]
    R                          # Index of reference signal
    S_M                        # Number of samples in the reference signal.
Begin:
    For ∀j ∈ 𝕋
        r_kj = Σ_{n=0}^{S_M−1} x̂_{n,j} × x̂_{n−k+S_M−1,R} ,  k = {0,1, … (2S_M − 2)}   # Cross-correlate
        s_j ← ⟨max_{∀k}(r_kj)⟩ − S_M + 1          # Shift position of maximum cross-correlation
    If s_j > 0              # Positive shift
        x_ij = x̂_ij,  i = {s_j, s_j + 1, … (S_M − 1)}       # Advance distance vector
        y_ij = ŷ_ij,  i = {0,1, … (S_M − 1 − s_j)}          # Truncate signal vector
    Else
        x_ij = x̂_ij,  i = {0,1, … (S_M − 1 − |s_j|)}        # Truncate distance vector
        y_ij = ŷ_ij,  i = {|s_j|, |s_j| + 1, … (S_M − 1)}   # Advance signal vector
End:
Outputs: 𝕋                  # Aligned signal tables [signal, distance]
```

Figure 4.7 Correlation alignment algorithm

In Figure 4.7, $\max_{\forall k}(r_k)$ is the function that shifts the position of maximum cross-correlation signal. It is possible that there will be many values in the cross-correlation vector rk that are equal to the maximum value. As a result, the operator ⟨·⟩ is used to return the index of the first maximum value. The shift position is relative to the distance of the first sample of the reference signal, which is zero. Hence, the position of zero shift is when the distance labels of both vectors overlap. This is the default shift position when there are no signals to correlate, which means that all the sensor samples contain only noise. The positive and negative shifts are the shifts to the right and left of the overlapping position. The shifted signal inherits the distance labels of the reference signal. When the shift is positive, the first sample of the shifted signal will inherit the distance labels starting at the shifted number of positions to the right of the distance vector of the reference signal. This is equivalent to advancing the distance vector of the signal by the number of shift positions for maximum cross-correlation. Since the signal is advanced, the tail will be

21

ahead of the last sample of the reference signal, and thus, be truncated. When the shift position for maximum cross-correlation is negative, the reverse procedure is applied.

After the traversals position alignment, the signals are now ready for the ensemble average. The ensemble average algorithm is shown in Figure 4.8. Before the algorithm starts calculating ensemble average, it trims the aligned signals such that the beginning and ending positions are equal to the maximum and minimum values. Lastly, Y_k is the mean value of the signal across all traversals at position k.

Ensemble Average

Inputs:
  $\mathbb{T}$ # Tables of aligned signals [signal, distance]
  SM   # Number of samples
Begin:
  For $\forall j \in \mathbb{T}$
    $L_{max} \leftarrow \max_{\forall j \in \mathbb{T}} x_{0,j}$ # Maximum of starting distances
    $R_{min} \leftarrow \min_{\forall j \in \mathbb{T}} x_{S_M-1,j}$ # Minimum of ending distances
    $x_{Lj} \leftarrow \min_{\forall k}|x_{kj} - L_{max}|,\ x_{Rj} \leftarrow \min_{\forall k}|x_{kj} - R_{max}|$ # First and last samples
    $Y_k = \frac{1}{T}\sum_{j=0}^{T-1} y_{kj},\ k = \{0,1,\dots\}$ # Ensemble average signal across position k
End:
Outputs: W   # Vector [ensemble averaged signal, distance]

Figure 4.8 Ensemble average algorithm

The most significant property of the ensemble average is that the signal quality will improve with each additional signal. That means noise in the composite signal will continue to reduce with additional signals. However, the rate of improvement will depend on the alignment accuracy achieved by the algorithm. Moreover, Figure 4.9a shows the position misalignment of three individual signals and Figure 4.9b illustrates the effect of ensemble average on the aligned and non-aligned signals.

Table 4.3 Statistical Evaluation of The Correlation Alignment

Description	ΔPIE (m)		EA Peak
	μ	σ	(g)
Non-aligned signals	0.57	6.07	0.46
Aligned signals	0.11	0.09	0.99
Improvement factor	5.1	67.4	2.14

Table 4.3 represents the statistical summary of the distance differences between the PIE of each signal and that of the reference signal, denoted as ΔPIE. Thus, it is evident that signal alignment and ensemble average enhances the signal and minimizes the noise by reducing the position variation of the signal. Moreover, the alignment decreases the mean and the spread of ΔPIE by a factor of more than 5 and 67, respectively. Also, the ensemble average of the aligned signal is more than double relative to the non-aligned signals.
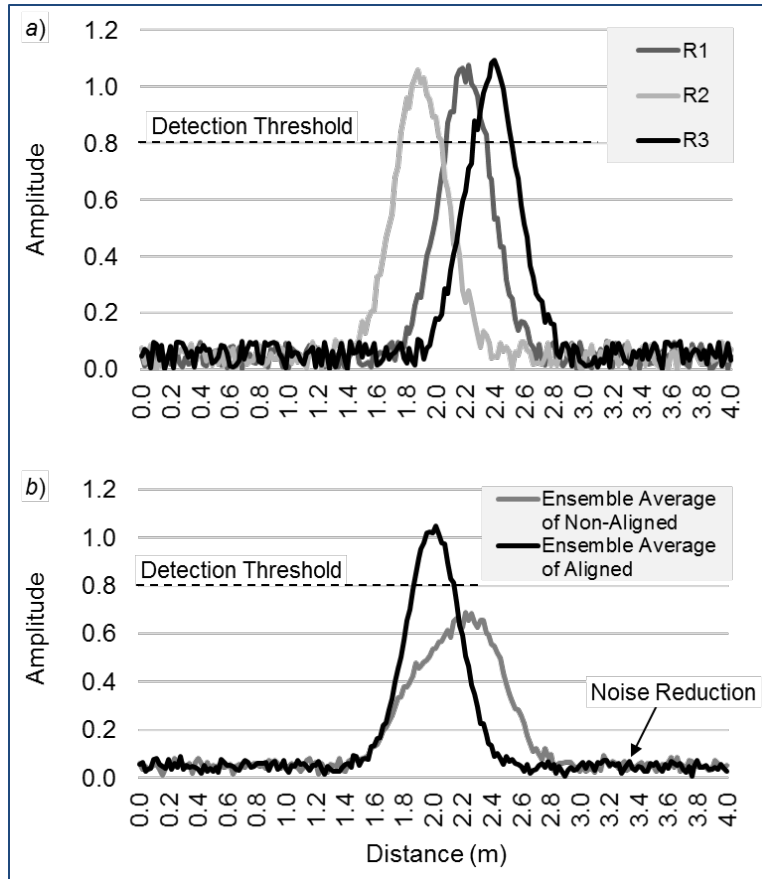
Figure 4.9 a) Three individual misaligned signals b) Effect of
ensemble average on non-aligned and aligned signals.

4.4   Signal Filtering

Some interferences, such as vehicle vibrations and electromagnetic waves, contribute to noise and other unwanted signal components in the inertial sample. Thus, it is necessary to consider the selection of appropriate signal filtering techniques that minimize the noise and maximize the signal (Hasan et al., 2010; Alam and Rohac, 2015). Previously (Hasan et al., 2010; Heirich et al., 2011; Real et al., 2011; Alam and Rohac, 2015; Milne et al., 2016), many attempts had been made to address this problem. Generally, all these methods used and designed standard digital filters, such as IIR filter, FIR filter, and notch filter, that require prior knowledge of frequency response parameters. However, in the absence of such information, it is not possible to design standard filters because of the changing nature of the filter's requirements (Ali, Rawat, and Malvia, 2017). Thus, the inability of such a system to work with limited information makes it less suitable for general application (Oni, Olatunde, and Babatunde, 2018). Subsequently, in such a situation, it is preferred to use an adaptive filter. However, there are some drawbacks of adaptive filtering, such as use of a more complex cost function and bank of filters to evaluate the SNR in a closed-loop manner. Hence, convergence can be slow and require high computational complexity. Also, it operates on a single stream and does not guarantee convergence to a local minimum that would maximize the SNR. Hence, these factors limit their use in practice.

To overcome all these constraints, this research developed a method that aids the cutoff frequency selection for the FIR low-pass filter. The procedure not only maximizes the SNR but lowers the computational complexity and increases the stability for practical use. To achieve this, the method combined the fast Fourier transform (FFT) from signal streams of the same traversal to produce an

23

ensemble average FFT (EA-FFT). Before the EA-FFT algorithm, the data containing distinct inertial signal for all road or railway traversals are aligned so that signal peaks occur at the same distance position. The algorithm then extracted signals from approximately equal length aligned traversals. Figure 4.10 shows the workflow of the cut-off frequency selection algorithm. Each FFT used the mean sampling frequency (Fs) of its individual signal. However, the mean sample frequency varies among traversals. Therefore, the algorithm aligned the frequency bins before averaging the non-zero magnitudes. The algorithm utilized the road data used in the previous subsections.
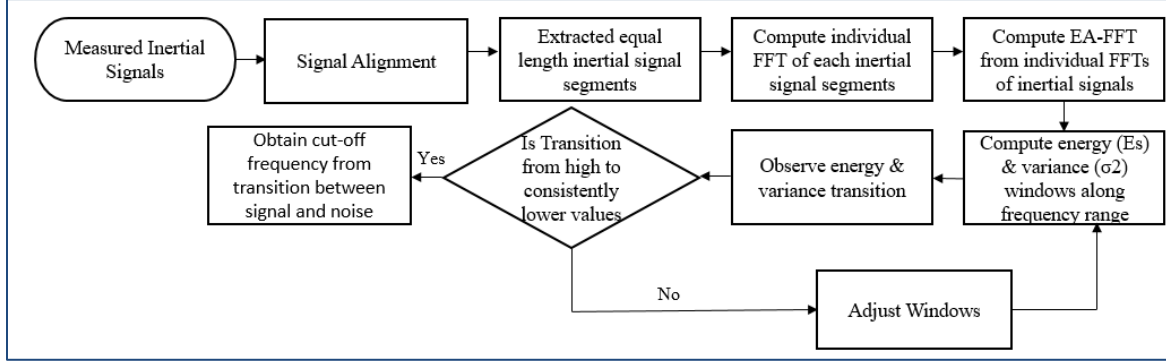
Figure 4.10 Workflow of the signal cut-off frequency

4.4.1 FFT and Ensemble Average FFT

The fast Fourier transform (FFT) (Cooley and Tukey, 2006) is an efficient method of computing the discrete Fourier transform (DFT) X_j of x_l as follows:

$$X_j = \sum_{l=0}^{N-1} x_l W^{-jl} \qquad W = exp\frac{2\pi\sqrt{-1}}{N}, \tag{12}$$

where j= 0, ….,N-1 and each x_l is an equally spaced sample of a function $x(t)$, l represents the time index, and j represents the frequency index. N is an integral power of 2.

The EA-FFT is the mean value of all spectral magnitudes within a frequency bin x such that

$$X_{Nx} = \frac{1}{N}\sum_{i=1}^{N} X_{ix} \tag{13}$$

where X_{Nx} is the ensemble average of N FFTs across bin x, and X_{ix} is the magnitude of the FFT of an individual signal within frequency bin x. Figure 4.11 demonstrates the difference between the FFT of the signals and their EA-FFT.
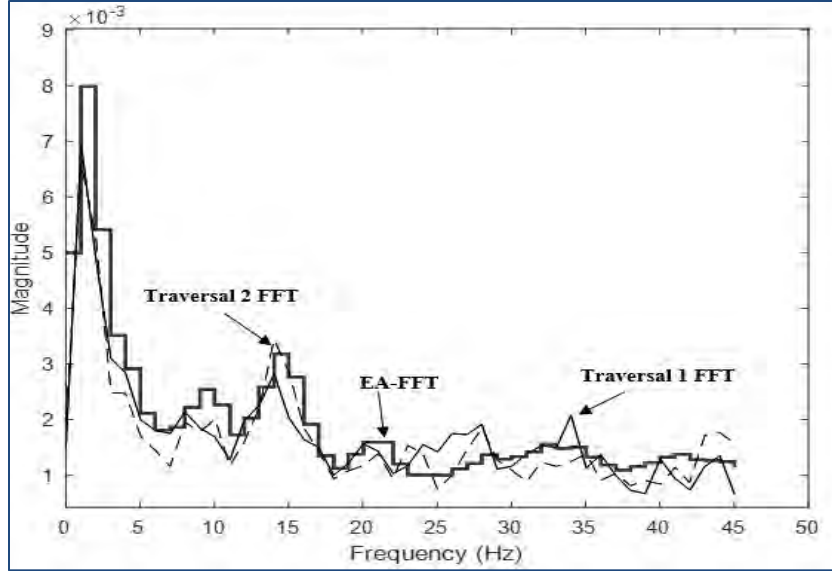
Figure 4.11 FFT of individuals and their ensemble average FFT

4.4.2 Energy

Signal processing (Chaparro, 2011) defines the energy of a finite signal x(n) as

$$E_s = \sum_{n=0}^{N-1}|x(n)|^2 \tag{14}$$

where Es is the energy of a signal, n is the sample number, and *N* is the total number of samples. Figure 4.10a shows the energy of the EA-FFT along 5 Hz windows.

4.4.3 Variance

Signal variance indicates how much a signal varies about the average value (Smith, 1997). The variance of a signal is

$$\sigma^2 = \frac{1}{N-1}\sum_{i=0}^{N-1}[x_i - \mu]^2 \tag{15}$$

where the signal samples are x_i, μ is the signal mean, *N* is the number of samples, and σ^2 is the signal variance (Smith, 1997). Figure 4.10b shows the variance of the EA-FFT along 5 Hz windows.

4.4.4 Finite Impulse Response (FIR) Low-Pass Filtering

The definition of a low-pass FIR filter of order N is (Punskaya, 2005):

$$y[n] = b_0 x[n] + b_1 x[n-1] + \cdots + b_N x[n-N] \tag{16}$$
or
$$y[n] = \sum_{i=0}^{N} b_i x[n-i] \tag{17}$$

where $x[n]$ is the input signal, $y[n]$ is the output signal, and N is the filter order. Each coefficient b_i is the value of the impulse response at the ith instant for $0 \leq i \leq N$. The impulse response of the filter is:

$$h[n] = \sum_{i=0}^{N} b_i \delta[n-i] = \begin{cases} b_n & 0 \le n \le N \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

The phase response of the FIR filter causes a delay in the filtered signal. Here, it is not an issue as the filter will delay all signals similarly. Therefore, the algorithm estimates and corrects the distance offset accordingly. Figure 4.12c compares the unfiltered (black) and filtered (gray) signals.
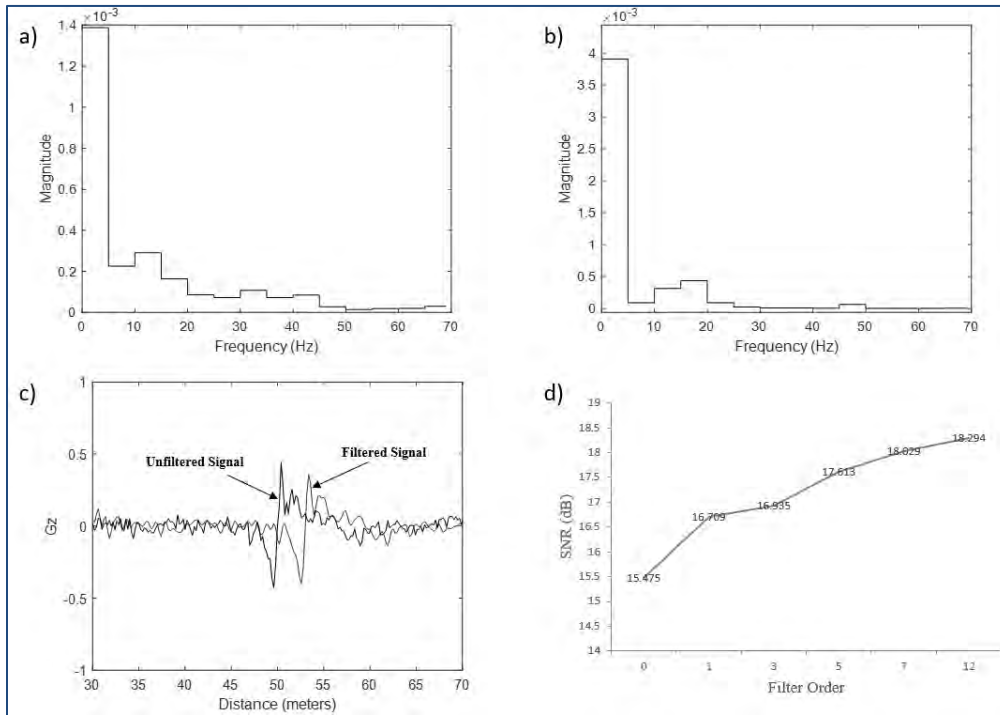
Figure 4.12 a) Energy windows of the EA-FFT, b) variance windows of the EA-FFT, c) the unfiltered and filtered signals, and d) filter order demonstrates the effectiveness of the cutoff frequency selection.

4.4.5   Filtering Results and Discussion

The filter cut-off frequency is necessary in improving the SNR. From the individual FFTs, it is hard to observe a clear pattern that separates signal from noise. The results showed that the method of EA-FFT with statistical decision criteria is effective for informing a cut-off frequency to a digitally low-pass filter. It is clear from Figure 4.11 that there is agreement among the EA-FFTs on the frequency at which both the energy and variance in the spectral windows reach a minimum before rising again slightly. More precisely, in Figure 4.12a and Figure 4.12b, the energy and variance windows of the EA-FFT show such a transition at 25 Hz.

Figure 4.12c represents the filter effectively removing noise while preserving the quality and strength of the inertial signature. However, a poor choice of cut-off frequency could result in the attenuation of both signal and noise, and subsequently no improvement in SNR. Therefore, observing SNR improvement with the increasing number of filter order could validate the cut-off frequency selection. After applying a low-pass FIR filter with the selected cut-off frequency, the SNR steadily increased with the order of the filter as shown in Figure 4.12d.
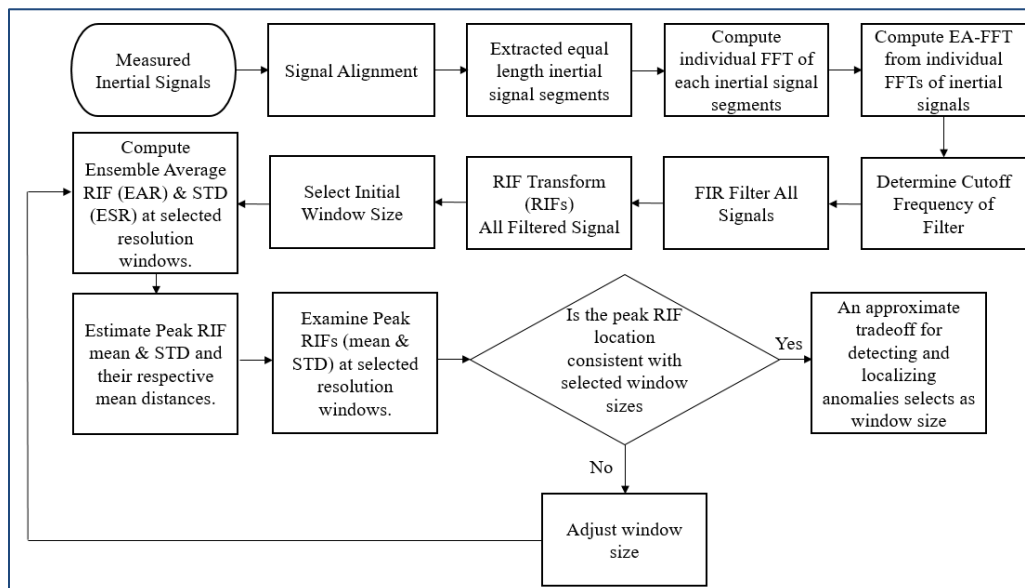
4.5   Ensemble Average Feature

The inertial sensor directly measured the roughness that riders experience either on road or rail tracks. The roughness represents the deviation of a uniform track geometry that affects vehicle dynamics and ride quality. Roughness and irregularities are the interchangeable terms in this research. Such roughness produces at least one maxima in the inertial signal. Figure 4.13 shows the comparison of the inertial signals from two random traversals of an isolated roughness. The roughness produced a peak inertial event (PIE) as illustrated.
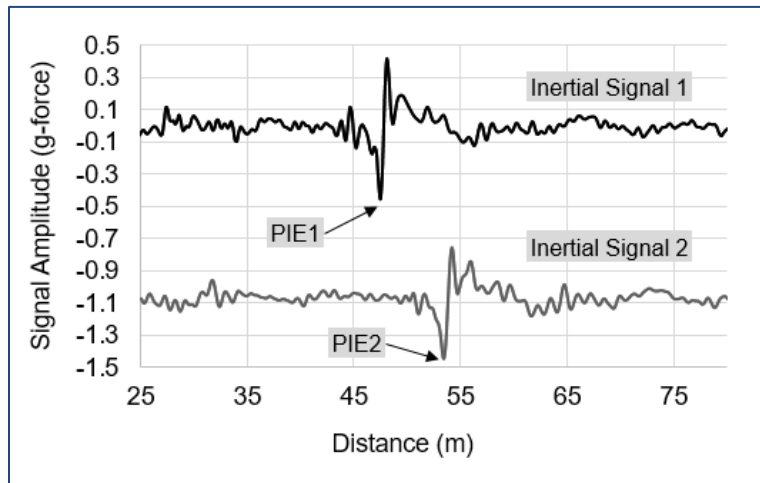
Figure 4.13 Variation of the peak of the two inertial signals

This section introduced a scalable approach that extracts, transforms, and combines roughness features of the inertial signals obtained from multiple traversals of a track segment into a reliable single-index summary of road or railway roughness called EAR-index. Figure 4.14 illustrates the complete workflow of ensemble average feature extraction. The peak EAR values are inherently the best estimators for both the roughness intensity and position.

Figure 4.14 Ensemble average feature extraction workflow

Prior to calculating EAR, the algorithm first interpolates distance from a spatial reference position to align the signals and to extract approximately equal length segments and computing an FFT and ensemble average of the FFT (EA-FFT) across multiple signals to inform an appropriate filter cutoff frequency and applying a FIR low-pass filter described in the previous subsections. Subsequently, the algorithm computes the ensemble average of RIF indices (EAR) across the distance windows of multiple traversals. The significant benefit of the EAR is that it reduces noise while enhancing feature detection by leveraging the correlation properties of uncorrelated noise and correlated signals, respectively. As a result, EAR enhances the SNR that reduces both false positive and false negative feature detection errors.

4.5.1　Feature Extraction

The RIF transform reduces the data into features that are proportional to the roughness within a distance window ΔL such that

$$R^{\Delta L} = \sqrt{\frac{1}{\Delta L} \sum_{n=0}^{N-1} \left| G_{z[n]} v_n \right|^2 \delta t_n} \tag{19}$$

where the RIF index $R^{\Delta L}$ is the average g-force magnitude per unit of distance ΔL traveled. $G_{z[n]}$ is the vertical acceleration for signal sample n. The sample period instant is δt_n and the instantaneous traversal speed is v_n.

4.5.2　Ensemble Average of RIF Indices (EAR)

An ensemble average of the RIF-indices (EAR) within a selected distance window along the traversal path and across N traversals produced an estimate of the average roughness E^w experienced in window w, at any speed where

$$E^w = \frac{1}{N} \sum_{x=1}^{n} R_x^w. \tag{20}$$

R_x^w is the RIF index from traversal x within distance window w. The intensity of the peak EAR is proportional to the amount of roughness produced from traversing roadway or railroad track anomalies. Figure 4.15a and Figure 4.15b represent the EAR of the traversal data for spatial resolution windows of 1, 5, 15, and 20 meters.

4.5.3　Ensemble RIF STD

The ensemble RIF (ESR) STD is

$$\sigma_{RIF} = \sqrt{\frac{\sum_{x=1}^{N}(R_x^w - E^w)^2}{N}} \tag{21}$$

It is a measure of the variability of the roughness intensity across traversals within a distance window. The ESR of the traversal data with distance windows of 1, 5, 15, and 20 meters are shown in Figure 4.15c and Figure 4.15d.
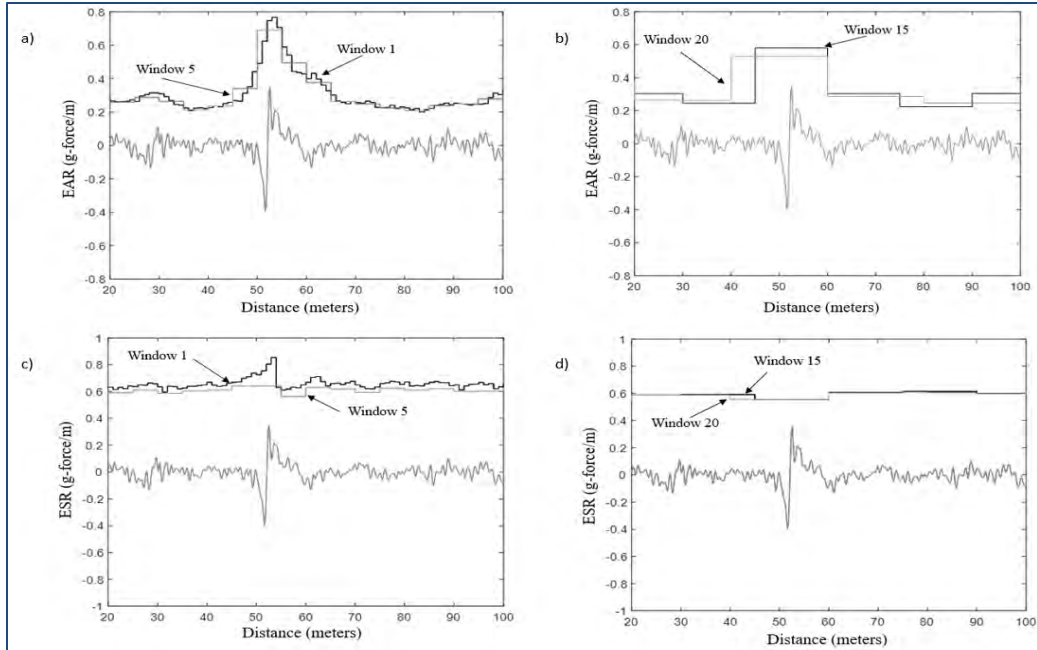
Figure 4.15 a) EAR at resolution window of 1 and 5 meters b) EAR at resolution window of 15 and 20 meters c) ESR at resolution window of 1 and 5 meters d) ESR at resolution window of 15 and 20 meters

4.5.3.1 Margin-of-Error

The margin-of-error (MOE) is calculated with a 95% confidence interval (CI) as a function of the number of traversals within the selected window sizes such that

$$\text{MOE} = z * \frac{\sigma}{\sqrt{n}} \tag{22}$$

where z is the critical value of 1.96 and σ is the standard deviation of the intensity of the peak RIF within the selected distance window sizes. The total number of traversals within the window size is n. The MOE of the peak RIF for distance windows of 1, 2, 5, 10, 15, 20, and 40 meters shown in Figure 4.16.
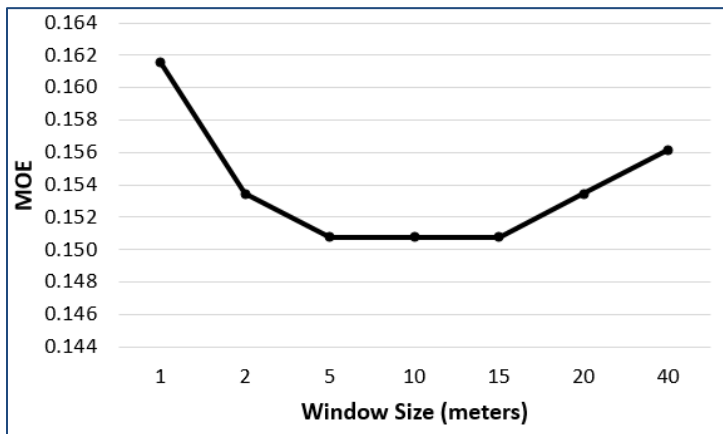

Figure 4.16 MOE of the peak RIF for various distance windows

4.5.4 Ensemble Average Peak RIF

The multi-resolution feature of the RIF-transform localizes roughness within an adjustable spatial resolution window. Increases in traversal volume improve the precision of anomaly localization by reducing the variance of the RIF-index. Table 4.4 summarizes the statistical parameters of peak RIF within a selected resolution window relative to the beginning of the traversals. The table reveals some important outcomes; first, an optimum window size minimizes the STD of estimating the RIF intensity. That means the smaller the window size the higher the precision of locating irregularity. Second, a lower STD informs the uniformity of detecting irregularity, thus reducing the false positive and false negative detection errors. Subsequently, Figure 4.15c and 4d show the fall-off the ESR with the increase in window size. Nonetheless, the position of the maximum EAR in Figure 4.15a and b is estimate of the position of the anomaly.

Table 4.4 Statistical Parameters of Peak Rif within Selected Resolution Window

Window Size	Peak RIF Mean	Peak RIF STD	Peak RIF Dist Mean	Peak RIF Dist STD
1	1.48	0.60	53.38	4.02
2	1.21	0.57	53.59	4.02
5	0.87	0.56	54.47	8.50
10	0.70	0.56	56.79	15.91
15	0.60	0.56	56.15	20.15
20	0.53	0.57	51.98	5.91
40	0.43	0.58	101.22	56.57
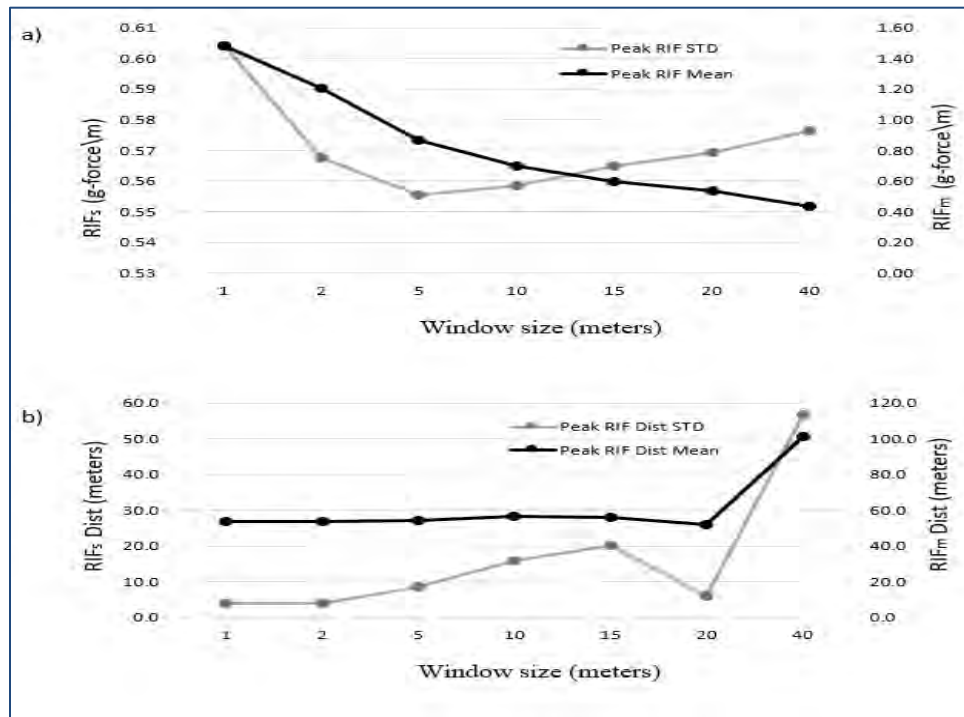
Figure 4.17 Shows the stable estimate of peak RIF position at a minimum STD

Additionally, from Figure 4.17 it is clear that keeping the distance window size below 5 meters maintains a stable estimate of the position of the peak RIF (approximately 53 meters) and minimizes the STD of that estimate.

4.5.4.1 Margin-of-Error of Peak RIF

Table 4.5 MOE of Peak RIF within Selected Window Size

Window Size	No. of Traversals	MOE Peak RIF
1	53	0.162
2	53	0.153
5	53	0.151
10	53	0.151
15	53	0.151
20	53	0.153
40	53	0.156

Table 4.5 Summarizes the MOE of peak RIF within the selected window size. From Figure 4.16 & Table above, it appears that the MOE declines most substantially between window sizes of 1 and 5 meters. Therefore, MOE validates confidence in the measurements and purposes to an optimal window size selection. It also indicates the window size of 5 meters is within the visual range for inspector to locate the anomaly.

4.6   Section Conclusion

A low-cost and reliable model for detecting and identifying the position of irregularity can enhance the efficiency of conventional track monitoring by focusing and deploying inspection resources on high-risk sites. The use of sensors aboard regular service vehicles enhances coverage and monitoring frequency and reports roughness intensity. But the GPS and accelerometer sensor errors diminish the attainable accuracy and precision of detecting the presence and location of an irregularity. It results in signal position misalignment and additive signal noise. Consequently, it reduced the SNR needed to detect and localize track irregularity reliably. In particular, false positives and false negatives can increase when SNR decreases. Appropriate signal alignment, filtering, and ensemble averaging the signals from multiple traversals can degrade the noise level and, thereafter, reduce the false positive rate of the anomaly detection.

This research introduced a framework that includes the methods of reducing detection error and enhances the quality of the signals and extracted features. At the same time, it increases the SNR and reduces the false positives and false negatives. The developed technique uses distance interpolation, heuristic, and correlation alignment to align the signals across the traversals. The algorithm then extracts the feature from the aligned and filtered signals.

Based on the performance statistics, among the four heuristic alignment methods, the centroid asymmetric provides the lowest standard deviation of the PIE and least skewness. The centroid asymmetric method performed approximately 17% and 44% better than other methods, respectively. The correlation alignment method shows that the mean deviation and the spread of the reference signal are improved by factors of more than 5 and 67, respectively. The improvement in the aligned signal from the ensemble average of all the aligned signals is more than two compared with the ensemble average of the non-aligned signals.

This research developed an approach that shows appropriately aligned signal filtering for ensemble average could be effective in S/R improvement for subsequent feature extraction. In particular, determination of the filter cut-off frequency is necessary. From the individual FFTs, it is hard to differentiate signal patterns from noise. The ensemble average of the individual FFTs (EA-FFTs) from the

approximately equal length and positional aligned inertial signals can enhance the clarity of the underlying signal pattern. Subsequently, two statistical methods to frequency windows of the EA-FFT offer a quantitative rationale for selecting a frequency threshold where the signal transitions to noise.

Next, the approach employs the road impact factor (RIF) space-time transform, which provides the advantage of data reduction and feature extraction within adjustable and consecutive distance windows along the traversal path. The ensemble average of RIF-indices within each window offers an estimate for the intensity and position of an irregularity. The experiment demonstrated the reliability and accuracy of the employed method. Tradeoff analysis found that a window size of 5 meters provided a good balance among data reduction, accuracy, and precision of anomaly detection. The MOE of peak RIF values also validates confidence in the measurements and points to an optimal window size selection.

The applied approach suggests that the methods have low computational complexity and are easy to use in practical application. Hence, the framework methods with ensemble averaging can be generalized to any sensor data from multiple traversals.

5. RAIL TRACK ANOMALY POSITION LOCALIZATION ASSESSMENT: A CASE STUDY

This section describes and investigates the potential use of low-cost sensors aboard a regular service vehicle to monitor automatically and continuously for the inertial events produced by irregular track geometry. The section also introduces a signal processing and statistical method to estimate the position of peak inertial events from multiple traversals. The methods are validated by comparing the estimated positions of detected irregularity with the actual positions of irregularities reported by railroad inspectors.

5.1 Anomaly Position Localization Assessment

With the increase of the train's load and speed, wheel-train interaction unavoidably accelerates the rail track geometry irregularities, such as profile, alignment, and warp. When a train traverses such an irregularity, it produces a peak inertial event (PIE). So the magnitude of PIEs is an effective means of extracting track condition patterns.

The method processes the GPS tagged inertial data (shown in Table 3.1) collected from smartphones aboard a hi-rail vehicle to estimate the positions of track irregularity. The authors selected an RIF transform that extracts track condition patterns as features from inertial sensor signals to reveal the signature of a track irregularity called warp. That is, the RIF features are proportional to the amount of "cross level" reported by a railroad inspector. The cross level is the elevation difference between rails at a given position along the track. A geographical information system (GIS) platform is used for the visualization of PIEs, which are RIF values in the high 5 percentiles of its distribution.

5.1.1 Road Impact Factor

The RIF, denoted as RIF_{Rt}, is an intensity that is proportional to the resultant rotation rate about x and y axis calculated by replacing Gt_n in equation 1 with Rt_n the resultant rotation signal sample n in units of the radians-per second such that RIF transform becomes
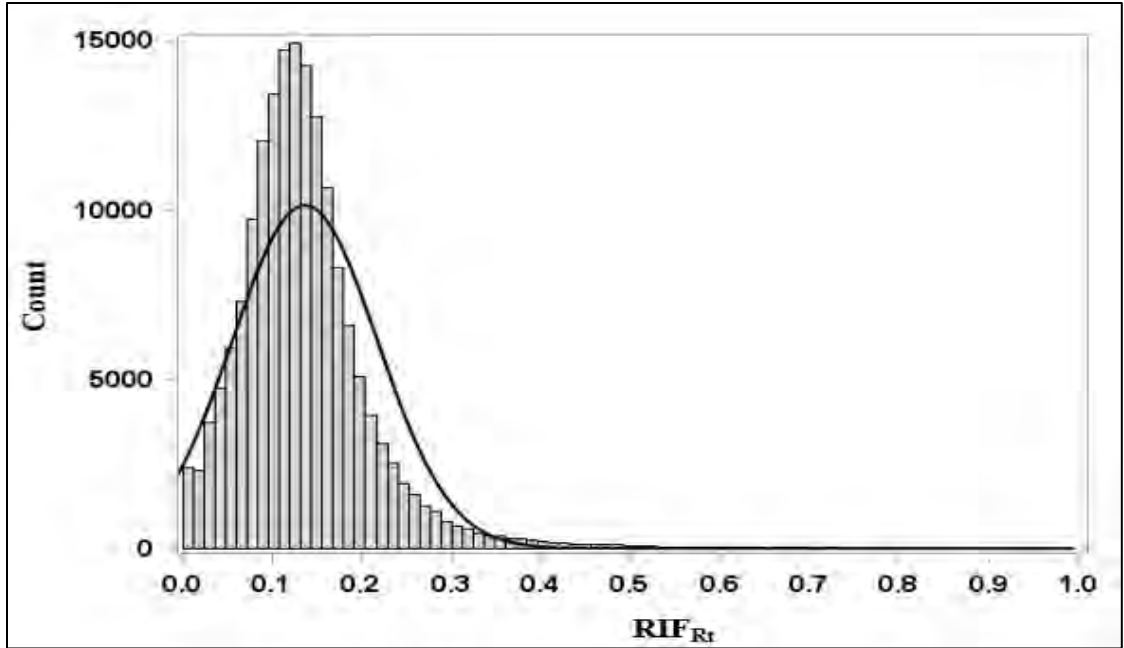
$$RIF_{Rt} = \sqrt{\frac{1}{L}\sum_{n=0}^{N-1}|Rt_n \, v_n|^2 \, \delta t} \tag{23}$$

Where, Rt is the vector sum of Rx and Ry, angular rotations around X and Y axis of the phone

$$Rt = \sqrt{(R_x)^2 + \left(R_y\right)^2} \tag{24}$$

5.1.2 RIF_{Rt} Statistical Distribution

As shown in Figure 5.1, the RIF_{Rt} feature follows normal distribution with a truncated left tail.

Figure 5.1 RIF_{Rt} feature distribution

In the previous section, Table 3.2 defined a color-coded scheme based on percentile thresholds of the RIF distribution to visualize PIEs using a GIS platform. Figure 5.2 shows ArcGIS plots the RIF features at their corresponding longitude and latitude position of each GPS blocks as small circles with intensity levels in the range of red and orange described in a color-coded scheme.

Figure 5.2 A map of PIEs that correspond to severe track irregularity

5.1.3 Position Localization Assessment

Localizing the position of an irregularity involved computing the centroid from the PIEs detected at any speed across multiple traversals. For this, the method accumulates the PIEs that are within two GPS updates of the reported anomaly from all available traversals from either direction, respectively. Fundamentally, the offset bias in one direction will tend to cancel the offset bias in the other direction.

The method then forms a cluster from the PIEs with the intensity levels in the range of "orange" and "red." PIEs are in the orange range when the vehicle moves more slowly across an anomaly. A PIE cluster accumulated from the limited number of traversals provided an estimate of the position of a measured irregularity.
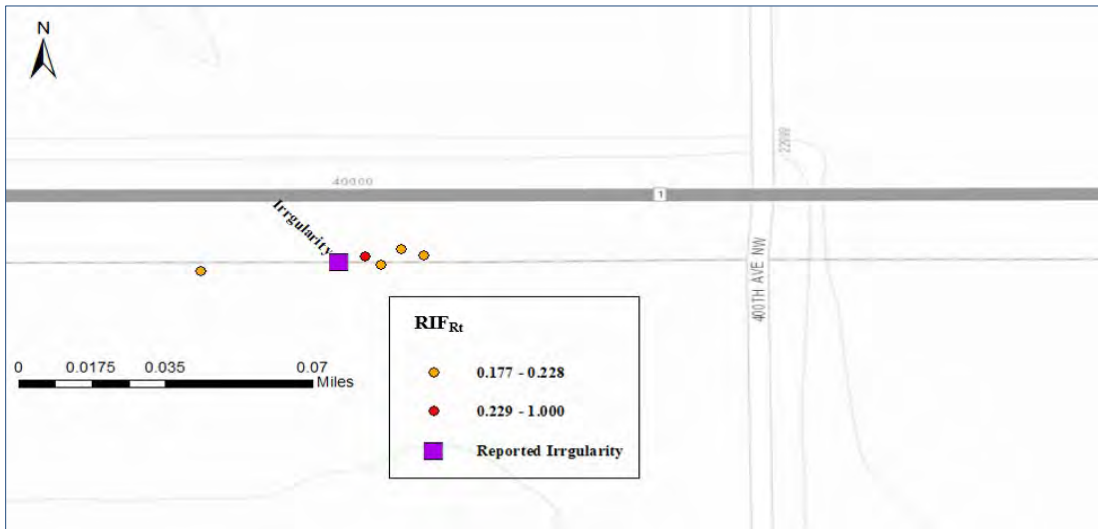

Figure 5.3 PIE cluster near the reported irregularity
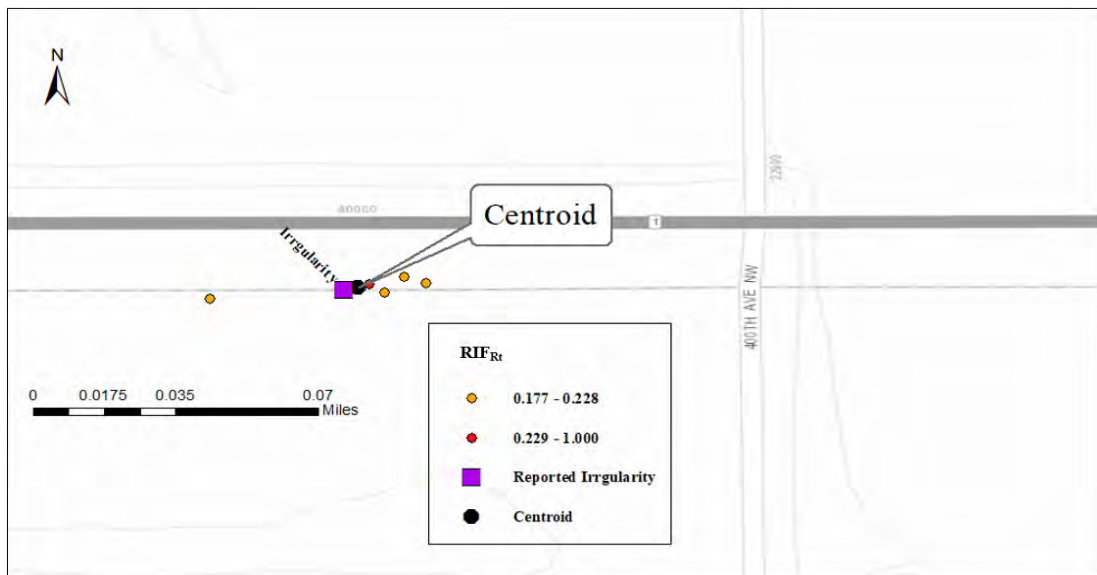

Figure 5.4 PIE cluster centroid relative to the actual irregularity

5.2   Result

From Figure 5.2, it is clear that the red circle is the peak inertial event (Peak RIFs), indicating high probability of irregularity while all other RIF values are shown in different colors depending on their intensity level. To understand the reliability of the anomaly location estimation assessment, the authors considered the actual reported location by a railroad maintenance inspector as a ground truth location for our method validation. Figure 5.3 shows the PIE cluster within two GPS updates of the reported irregularity. Figure 5.4 shows the centroid position relative to the actual irregularity. It is found that the centroid was offset 13.1 feet (3.9 m) from the position of the reported irregularity. These results suggest that the estimated position of the irregularity is within visual distance of the true position of the irregularity.

5.3   Chapter Summary and Conclusion

Signal processing and statistical methods are introduced to estimate the position of peak inertial events from multiple traversals. The study goal is to characterize their accuracy by comparing the estimated position of detected irregularities with the actual positions of irregularities observed by railroad inspectors. The developed mathematical model produced RIFs from the gyroscopic signals and defined a color-coding scheme to visualize PIEs using a GIS platform. A centroid of the PIE clusters accumulated from a limited number of traversals provided an estimate of the position of the irregularity. The result is that the centroid is 13.1 feet (3.9 m) offset from the position of actual irregularity. Therefore, the estimated position of the irregularity is within a reasonable visual distance of the actual position so that it can be seen during a follow-up manual inspection.

6. CONCLUSION AND FUTURE WORK

The frequent and network-wide condition monitoring of the railway infrastructure has become elusive. This research's developed framework and methods will empower railroad agencies to develop and install a more reliable, scalable, and affordable approach using low-cost sensors aboard regular service vehicles. The framework ensemble averaging the signals from multiple traversals can degrade the noise level and, thereafter, reduce the false positive rate of the anomaly detection. This suggests that the signal quality improves continuously with the additionally available data stream. Hence, the developed methods of signal alignments, signal filtering, and ensemble averaging can be generalized for the practical application that produced sensor data from multiple traversals of railroad segments.

This research investigates the potential use of the proposed approach that estimates the peak inertial events (PIEs) position and possible rail track irregularity such as warp. The study characterizes and validates its accuracy by comparing the estimated positions of detected irregularities with the actual positions of irregularities that the railroad inspector observes. The observed irregularity is called the ground truth area (GTA). In particular, the study applies a road impact factor (RIF) transform to extract features from the sensor signals proportional to the amount of track geometry irregularity. The method reduces the data by generating RIF features and visualizing the extracted peak RIF features.

Currently, this research has some extracted features and labeled them as joint track roughness and warp. Subsequently, we are looking for more feature labeling required as an input for future machine learning models. The feature labeling has two values, type and intensity.

Future work will use the extracted features from the composite signals to develop the machine learning models that can be trained to classify the types of detected railway anomalies. The machine learning technique involves supervised and semi-supervised learning that requires labeled features. Another future work is to make the research models and algorithms autonomous and real-time through annotated maps with algorithms. This capability prepares the follow-up inspection to select inspection resources and equipment to validate the type of issue and make informed decisions.

7. REFERENCES

Infrastructure Report Card- Rail (2021) *American Society of Civil Engineers*. Available at: https://infrastructurereportcard.org/wp-content/uploads/2020/12/Rail-2021.pdf. (Accessed: 18 June 2021).

Aboelela, E. et al. (2006) "Wireless sensor network based model for secure railway operations," *2006 IEEE International Performance Computing and Communications Conference*, (January 2005), pp. 623–628. doi: 10.1109/.2006.1629461.

Alam, M., and Rohac, J. (2015) "Adaptive data filtering of inertial sensors with variable bandwidth," *Sensors (Switzerland)*, 15(2), pp. 3282–3298. doi: 10.3390/s150203282.

Ali, F., Rawat, P., and Malvia, S. (2017) "Comparative Analysis and Survey of LMS and RLS Adaptive Algorithms," *International Journal of Computer Applications*, 161(3), pp. 26–29. doi: 10.5120/ijca2017913136.

Alippi, C. et al. (2000) "Composite real-time image processing for railways track profile measurement," *IEEE Transactions on Instrumentation and Measurement*, 49(3), pp. 559–564. doi: 10.1109/19.850395.

Barbosa, R. S. (2016) "New method for railway track quality identification through the safety dynamic performance of instrumented railway vehicle," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*. Springer Berlin Heidelberg, 38(8), pp. 2265–2275. doi: 10.1007/s40430-015-0471-9.

Bistrov, V. (2012) "Performance analysis of alignment process of MEMS IMU," *International Journal of Navigation and Observation*, 2012. doi: 10.1155/2012/731530.

Bridgelall, R. et al. (2016) "Precision enhancement of pavement roughness localization with connected vehicles," *Measurement Science and Technology*. IOP Publishing, 27(2). doi: 10.1088/0957-0233/27/2/025012.

Bridgelall, R., and Tolliver, D. (2020) "Accuracy enhancement of anomaly localization with participatory sensing vehicles," *Sensors*, 20(2), pp. 1–15. doi: 10.3390/s20020409.

Campos-Castellanos, C. et al. (2011) "The application of long range ultrasonic testing (LRUT) for examination of hard to access areas on railway tracks," *5th IET Conference on Railway Condition Monitoring and Non-Destructive Testing (RCM 2011)*, pp. 8B3-8B3. doi: 10.1049/cp.2011.0618.

Carboni, M., and Cantini, S. (2012) "A 'Model Assisted Probability of Detection' approach for ultrasonic inspection of railway axles," *18th World Conference on Nondestructive Testing*, (April), pp. 16–20.

Chaparro, L. F. (2011) "Discrete-Time Signals and Systems," in *Signals and Systems using MATLAB*. 3rd edn. Elsevier, pp. 451–509. doi: 10.1016/B978-0-12-374716-7.00012-0.

Cooley, J. W., and Tukey, J. W. (2006) "An Algorithm for the Machine Calculation of Complex Fourier Series," *Mathematics of Computation*, 19(90), p. 297. doi: 10.2307/2003354.

Curtin, K., Nicoara, G., and Arifin, R. R. (2007) "A Comprehensive Process for Linear Referencing," *URISA Journal*, 19(2), p. 41.

Eriksen, A., and Gasgoyne, J. (2004) "Train-Mounted GPR for High-Speed Rail Trackbed Inspection," *Electrical engineering and electronics*, pp. 21–24.

Fateh, M. (2000) "Fractal Analysis of Geometry Data for Railroad Track Condition Assessment," *Federal Railroad Administration*, (December), pp. 1–4.

Feng, H. et al. (2014) "Automatic fastener classification and defect detection in vision-based railway inspection systems," *IEEE Transactions on Instrumentation and Measurement*, 63(4), pp. 877–888. doi: 10.1109/TIM.2013.2283741.

Flammini, F. et al. (2010) "Towards wireless sensor networks for railway infrastructure monitoring," in *International Conference on Electrical Systems for Aircraft, Railway and Ship Propulsion, ESARS 2010*. IEEE, pp. 5–10. doi: 10.1109/ESARS.2010.5665249.

Fontul, S. et al. (2016) "Railways Track Characterization Using Ground Penetrating Radar," *Procedia Engineering*. Elsevier B.V., 143(July), pp. 1193–1200. doi: 10.1016/j.proeng.2016.06.120.

FRA, *Track Inspection Time Study* (2008).

Hasan, A. M. et al. (2010) "Wavelet-based pre-filtering for low cost inertial sensors," *Journal of Applied Sciences*, pp. 2217–2230. doi: 10.3923/jas.2010.2217.2230.

Heirich, O. et al. (2011) "Measurement and analysis of train motion and railway track characteristics with inertial sensors," in *14th Internaltional IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, Washington, DC, USA*. Washington, DC, USA, pp. 1995–2000. doi: 10.1109/ITSC.2011.6082908.

Hodge, V. J. et al. (2015) "Wireless sensor networks for condition monitoring in the railway industry: A survey," *IEEE Transactions on Intelligent Transportation Systems*, 16(3), pp. 1088–1106. doi: 10.1109/TITS.2014.2366512.

Hughes, W. J. (2004) "Global Positioning System (GPS) Standard Positioning Service (SPS) Performance Analysis Report 1284 Maryland Avenue SW," *Analysis*, (December 2003).

Hunter, T. et al. (2009) "Path and travel time inference from GPS probe vehicle data," *Traffic*, 32 (section 2), pp. 1–8. Available at: http://www.ce.berkeley.edu/~bayen/conferences/nips09.pdf.

Jarmulak, J., Kerckhoffs, E. J. H., and van't Veen, P. P. (1997) "Case-Based reasoning in an ultrasonic rail-inspection system," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1266, pp. 43–52. doi: 10.1007/3-540-63233-6_477.

Kaleli, F., and Akgul, Y. S. (2009) "Vision-based railroad track extraction using dynamic programming," *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, pp. 42–47. doi: 10.1109/ITSC.2009.5309526.

Kim, S. S. et al. (2009) "Parameter characteristics of rail inspection measurement system of HSR-350x," *Journal of Mechanical Science and Technology*, 23(4), pp. 1019–1022. doi: 10.1007/s12206-009-0332-5.

Komalan, N., and Chauhan, A. (2017) "A survey on cluster based routing, data aggregation and fault detection techniques for locating real time faults in modern metro train tracks using wireless sensor network," *IEEE International Conference on Innovative Mechanisms for Industry Applications, ICIMIA 2017 - Proceedings*, (Icimia), pp. 116–123. doi: 10.1109/ICIMIA.2017.7975584.

Lane, N. D. et al. (2010) "A survey of mobile phone sensing," *IEEE Communications Magazine*, pp. 140–150. doi: 10.1109/MCOM.2010.5560598.

Lederman, G. et al. (2017) "A data fusion approach for track monitoring from multiple in-service trains," *Mechanical Systems and Signal Processing*, 95, pp. 363–379. doi: 10.1016/j.ymssp.2017.03.023.

Milne, D. et al. (2016) "Proving MEMS Technologies for Smarter Railway Infrastructure," in *The 3rd Internal Conference on Transportation Geotechnics (ICTG 2016)*, pp. 1077–1084. doi: 10.1016/j.proeng.2016.06.222.

Moreu, F., Kim, R. E., and Spencer, B. F. (2017) "Railroad bridge monitoring using wireless smart sensors," *Structural Control and Health Monitoring*, 24(2), pp. 1–17. doi: 10.1002/stc.1863.

Oni, E. A., Olatunde, I. D., and Babatunde, K. O. (2018) "Improvement of Audio Signal Quality Using Adaptive Filtering and Its Performance Advantage over Non - Adaptive (Linear) Filtering," *Internal Journal of Electrical and Electronic Science*, 5(2), pp. 39–45.

Otegui, J. et al. (2017) *A Survey of Train Positioning Solutions*, *IEEE Sensors Journal*. doi: 10.1109/JSEN.2017.2747137.

Podofillini, L., Zio, E., and Vatn, J. (2006) "Risk-informed optimisation of railway tracks inspection and maintenance procedures," *Reliability Engineering and System Safety*, 91(1), pp. 20–35. doi: 10.1016/j.ress.2004.11.009.

"Positive Train Control (PTC): Overview and Policy Issues" (2018).

Punskaya, E. (2005) *Design of FIR Filters*. [Online]. Available: www. sigproc. eng. cam. ac. uk/~ op205. doi: 10.1007/978-3-662-04170-3_7.

Real, J. et al. (2011) "Determination of rail vertical profile through inertial methods," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 225(1), pp. 14–23. doi: 10.1243/09544097JRRT353.

Resendiz, E., Hart, J. M., and Ahuja, N. (2013) "Automated visual inspection of railroad tracks," *IEEE Transactions on Intelligent Transportation Systems*, 14(2), pp. 751–760. doi: 10.1109/TITS.2012.2236555.

Sadeghi, J. (2010) "Development of Railway Track Geometry Indexes Based on Statistical Distribution of Geometry Data," *Journal of Transportation Engineering*, 136(8), pp. 693–700. doi: 10.1061/(asce)0733-947x(2010)136:8(693).

Santur, Yunus, Karaköse, Mehmet, and Akın, E. (2016) "Learning Based Experimental Approach For Condition Monitoring Using laser cameras in railway tracks," *International Journal of Applied Mathematics, Electronics and Computers*, pp. 1–5. doi: 10.18100/ijamec.xxxxx.

Sharma, K. et al. (2014) "Railway track breakage detection method using vibration estimating sensor network: A novel approach," *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014*, pp. 2355–2362. doi: 10.1109/ICACCI.2014.6968518.

Singh, M. et al. (2006) "Autonomous rail track inspection using vision based system," *Proceedings of the 2006 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety, CIHSPS 2006*, 2006(October), pp. 56–59. doi: 10.1109/CIHSPS.2006.313313.

Smith, S. W. (1997) "Statistics, Probability and Noise," *in The Scientist and Engineer's Guide to Digital Signal Processing*. 1st ed., *The Scientist and Engineer's Guide to Digital Signal Processing*. 1st ed. San Diego, CA, USA: California Technical Publishing. doi: 10.1049/ep.1971.0259.

Sukkarieh, S., Nebot, E. M., and Durrant-Whyte, H. F. (1999) "A high integrity IMU/GPS navigation loop for autonomous land vehicle applications," *IEEE Transactions on Robotics and Automation*, 15(3), pp. 572–578. doi: 10.1109/70.768189.

Tsunashima, H. et al. (2015) "Development of Track Condition Monitoring System Using On-board Sensing Device," in *Railway Research: Selected Topics on Development, Safety and Technology*, p. 145. doi: http://dx.doi.org/10.5772/57353.

Wang, L., Groves, P. D., and Ziebart, M. K. (2013) "GNSS shadow matching: Improving urban positioning accuracy using a 3D city model with optimized visibility scoring scheme," *Navigation, Journal of the Institute of Navigation*, 60(3), pp. 195–207. doi: 10.1002/navi.38.

Weston, P. et al. (2015) "Perspectives on railway track geometry condition monitoring from in-service railway vehicles," *Vehicle System Dynamics*, 53(7), pp. 1063–1091. doi: 10.1080/00423114.2015.1034730.

Yap, B. W., and Sim, C. H. (2011) "Comparisons of various types of normality tests," *Journal of Statistical Computation and Simulation*, 81(12), pp. 2141–2155. doi: 10.1080/00949655.2010.520163.

Alemi, A., Corman, F., and Lodewijks, G., 2017. "Condition monitoring approaches for the detection of railway wheel defects," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*. 231(8), pp. 961-981.

American Society of Civil Engineers, 2021. 2021 Infrastructure Report Card- Rail. [Online] Available at: https://infrastructurereportcard.org/wp-content/uploads/2020/12/Rail-2021.pdf [Accessed 20 June 2021].

Barbosa, R. S., 2016. "New method for railway track quality identification through the safety dynamic performance of instrumented railway vehicle," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*. 38(8), pp. 2265-2275.

Curtin, K. M., Nicoara, G., and Arifin, R. R., 2007. "A comprehensive process for linear referencing." *URISA Journal*, 19(2), pp. 41-50.

Ngigi, R. W., Pislaru, C., Ball, A., and Gu, F., 2012. "Modern techniques for condition monitoring of railway vehicle dynamics," In *Journal of Physics: Conference Series*, 364(1), p. 012016. IOP Publishing.

Ollier, B. D., 2006. "Intelligent Infrastructure the business challenge," *2006 IET International Conference On Railway Condition Monitoring*. pp. 1-6. IET.

Peters, J. C., and Frittelli, J., 2018. "Positive Train Control (PTC): Overview and Policy Issues," s.l.: Congressional Research Service - R42637.