MOUNTAIN-PLAINS CONSORTIUM

MPC 19-384 | P. Lu, R. Bridgelall, D. Tolliver, L. Chia, and B. Bhardwaj

Intelligent Transportation Systems Approach to Railroad Infrastructure Performance Evaluation: Track Surface Abnormality Identification with Smartphone-Based App

MOUNTAIN-PLAINS
CONSORTIUM REGION 8

Intelligent Transportation Systems Approach to Railroad Infrastructure Performance Evaluation: Track Surface Abnormality Identification with Smartphone-Based App

Prepared By

Dr. Pan Lu
Associate Professor/ Associate Research Fellow
Pan.lu@ndsu.edu
Tel: 701-212-3795

Dr. Raj Bridgelall
Assistant Professor/ Program Director
Raj.bridgelall@ndsu.edu
Tel: 408-607-3214

Dr. Denver Tolliver
Director
Denver.Tolliver@ndsu.edu
Tel: 701-231-7190

Leonard Chia
Research Assistant/ Ph.D. student
Leonard.Chia@ndsu.edu

Bhavana Bhardwaj
Research Assistant/ Ph.D. student
Bhavana.Bhardwaj@ndsu.edu

Department of Transportation, Logistics and Finance/College of Business
Mountain-Plains Consortium/Upper Great Plains Transportation Institute
North Dakota State University

July 2019

Acknowledgements

Disclaimer

The authors alone are responsible for the preparation and the facts and accuracy of the data presented herein. The document is disseminated under the sponsorship of the Mountain-Plains Consortium in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof. This report does not constitute a standard, specification, or regulation.

ABSTRACT

Federal track safety regulations require railroads to inspect all tracks in operation as often as twice weekly. Railroad companies deploy expensive and relatively slow methods using human inspectors and expensive automated inspection vehicles to inspect and monitor their rail tracks. The current practices are not only expensive and decrease rail productivity by taking away track time to perform inspection, but also increase the safety risk for railway inspection workers.

Sensors, such as inertial sensors, accelerometers, gyroscopic sensors, and global positioning system (GPS), are carried on a railway vehicle to continually monitor and inspect rail assets to meet the growing safety improvement needs for reliable and low-cost rail operations. Smartphones use such sensor networks, including wireless communication microchips. In this research, smartphone-based signaling data collection applications, data fusion algorithms, and data processing algorithms to detect a wide variety of possible track surface abnormalities are developed and validated.

The research methods will not rely on adapting sensor configurations, and will require only a data upload capability. The new sensors will compress and upload their geo-tagged inertial data periodically to a centralized processor. Remote algorithms will combine and process the data from multiple train traversals to identify abnormal track surface symptoms, and localize their positions. Track surface abnormality identification will enable asset managers to allocate the appropriate specialists to scrutinize the abnormality location.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1   Background

Railroads spend billions of dollars on infrastructure maintenance and condition monitoring each year [1]. Federal laws currently specify the type and regularity of full track inspections. Railroad companies deploy relatively slow and expensive methods using human inspectors and automated inspection vehicles to search for possible abnormalities. The expense and labor requirements of these existing non-destructive evaluation (NDE) methods limit their ability to scale for continuous and network-wide monitoring for risk mitigation and safety improvements. Hence, the industry could save billions of dollars if sensors aboard regular rolling stocks could screen the infrastructure for track surface abnormalities automatically and continually. Such a solution would enable asset managers to focus inspections on areas with a high likelihood of surface abnormality without closing lines to search for developing issues.

Federal track safety standards require railroads to inspect all tracks in operations as often as twice weekly. However, with resources thinly stretched and the rate of abnormality formation escalating with traffic load-density, railroads seek to enhance the efficiency of inspections and maintenance of way. The current inspection practices not only decrease rail productivity by taking away track time to perform inspection and maintenance, but also increase the safety risk for railway inspection workers. Based on FRA reports, there were 358 railroad employee-on-duty fatalities from 1999 to 2017; 309 cases occurred on or near a track, and 130 of those were track inspection and maintenance employees. Track surface abnormality screening sensors on rolling stock will automatically and continually provide track condition profiles and characterize potential abnormalities without risking maintenance employees' lives. Such a solution would free up more track time and capacity previously reserved for manual inspections while improving safety for railroad workers.

The inertial responses of a railcar are symptoms of possible track surface abnormality. Although not the case in all track surface abnormalities, a significant majority of them produce accelerated car movements in all directions [2]. Inertial sensors that monitor vehicle-track interactions (VTI) exist, and are widely used due to their small size, low cost, low power consumption, and robustness [3]. However, they do not classify their level of severity [4]. Existing sensor-based approaches generally produce alerts directly when acceleration magnitudes exceed fixed thresholds. Inspectors must then travel to the estimated locations of every alert event to search for associated abnormalities. Technicians must also pre-configure such sensors with thresholds based on their experiences or intuition. Without an objective and consistent approach to setting thresholds, some significant track abnormalities could go unnoticed, whereas minor ones could result in unnecessary and expensive inspections. The inertial responses of vehicles will vary with train speed, gross weight, suspension system design, and weather conditions. Therefore, thresholds must adapt to the specific circumstances to improve accuracy and reduce false positives. However, threshold adaptation can be complex and expensive. Adaption that uses complex algorithms on remote servers will require that the sensors support two-way communications, thus increasing their cost.

The method developed in this research will not rely on adapting sensor configurations, and will require only a data upload capability. The new sensors will compress and upload their geo-tagged inertial data periodically to a centralized processor. Remote algorithms will combine and process the data from multiple train traversals to identify abnormal symptoms and localize their position. Track surface abnormality or unevenness classification will enable asset managers to allocate the appropriate specialists to scrutinize the uneven surface location.

1.2   Research Objectives

The primary goal of this study is to research, develop, and evaluate an automated symptom screening system for railroad tracks and equipment. The system will locate and characterize the possible track surface abnormality by analyzing the inertial dynamics of hi-rail vehicle. The research team will include at least two Ph.D. students who will learn and practice the required signal processing, data processing, modeling, and signal classification techniques. The research output will feed into the development of new curricula in multimodal intelligent transportation systems for a transportation and logistics (TL) program. Students will be able to use the framework established in this research as part of laboratory exercises and to identify additional research projects that will refine and build upon its initial capabilities. An ability to demonstrate the full solution via smartphone embedded sensors will lead to outreach activities that would encourage manufacturers of ruggedized industrial sensors to engage in technology transfer for the commercialization of a dedicated or customized solution. The following summarizes the primary research objectives:

1) Develop and modify a smartphone data logging application (app) to collect data onboard a hi-rail vehicle or/and in-service vehicle
2) Develop, demonstrate, and validate the signal compression, track surface abnormality, or unevenness detection algorithms
3) Train two Ph.D. students on the various theoretical and simulation methods employed
4) Develop publications and associated outreach material suitable for engaging sensor manufacturers and service providers in possible technology transfer activities for full product commercialization
5) Utilize the research findings and techniques developed to draft classroom instructional material, laboratory exercises, and student project plans for new curricula in multimodal intelligent transportation systems for TL programs.

These research objectives will further the overall goals of promoting economic development, safety, interdisciplinary education, workforce development, and technology transfer that serves the critical needs of the Mountain-Plains Region.

1.3   Report Organization

This section introduces the organization of the report:

Section 2 describes track surface abnormalities and conducts a complete literature review on the current practice of track condition monitoring. Techniques reviewed include the current practice in NDE and approaches that use inertial sensors and digital signal processing.

Section 3 summarizes the specifications and developments of a smartphone data logging app for rail track condition monitoring.

Section 4 introduces the data collection plan and the data cleaning and data fusion techniques developed in this research.

Section 5 introduces and demonstrates feature extraction methods to identify and locate track surface abnormalities. Methods are validated with ground truth.

Section 6 summarizes the conclusions and recommendations from the study.

2. LITERATURE REVIEW

A complete literature search to understand state-of-practice and state-of-the-art on sensor-based rail track monitoring is critical to comprehend the body of knowledge. The review will cover research background, types of surface abnormalities, their measurements, current railroad practice, and inertial sensor-based systems.

2.1 Significance

Studies show that a combination of track and equipment failures causes over half of all rail accidents around the world (Astin, 2011). U.S. train accidents declined 71% from 1980 to 1990. However, the track and equipment caused accident rate has leveled off since then (GAO, 2010). This implies that breaking the past plateau will require additional resources and/or efficiency improvements of existing approaches for locating and remediating track abnormalities as soon as they form. Railroads will benefit economically from finding and fixing abnormalities quickly, before they cause derailments or costly delays.

Railroads have been downsizing since 1980 while their operational efficiency improved. They now have fewer than half as many employees per mile of road operated. (Tolliver and Bitzan, 2005). At the same time, the rate of traffic growth has steadily increased (FRA, 2011). Previous studies show that track abnormality forms at a rate directly proportional to accumulated trainloads (Besuner et al., 1978). According to data compiled from the American Association of Railroads (AAR), ton-miles per track-mile have tripled since 1980, as shown in Figure 2.1. This indicates a widening gap between the rate of abnormality formation and the labor resources available to discover, diagnose, and fix them. Data from the Surface Transportation Board (STB) show that Class I railroads currently assign nearly one in four employees to the maintenance of way and structures (AAR, 2011).

In the last decade, there were nearly 26,000 accidents, which resulted in 446 deaths and 5,300 injuries, costing the railroad industry more than $3.8 billion. In 2013, there were 1,854 train accidents, and track-related issues accounted for more than 30% of train accidents. Derailments accounted for almost 60% of the train accidents from 2008 to 2017. Hence, railways invest 15% of their total revenue on maintenance and condition monitoring every year.

2.2 Federal Track Safety Inspections

The Federal Track Safety Standards (FTSS), documented in 49 Code of Federal Regulation (CFR) §213.233, defines both the type and frequency of inspections for each track class. These standards require visual inspections at least twice weekly for most track classes in operation. The Federal Railroad Administration (FRA) rigorously enforces these standards and imposes hefty fines for non-compliance. The FRA operates several automated inspection cars containing a variety of non-destructive-evaluation (NDE) technologies to analyze most tracks in operation each year.
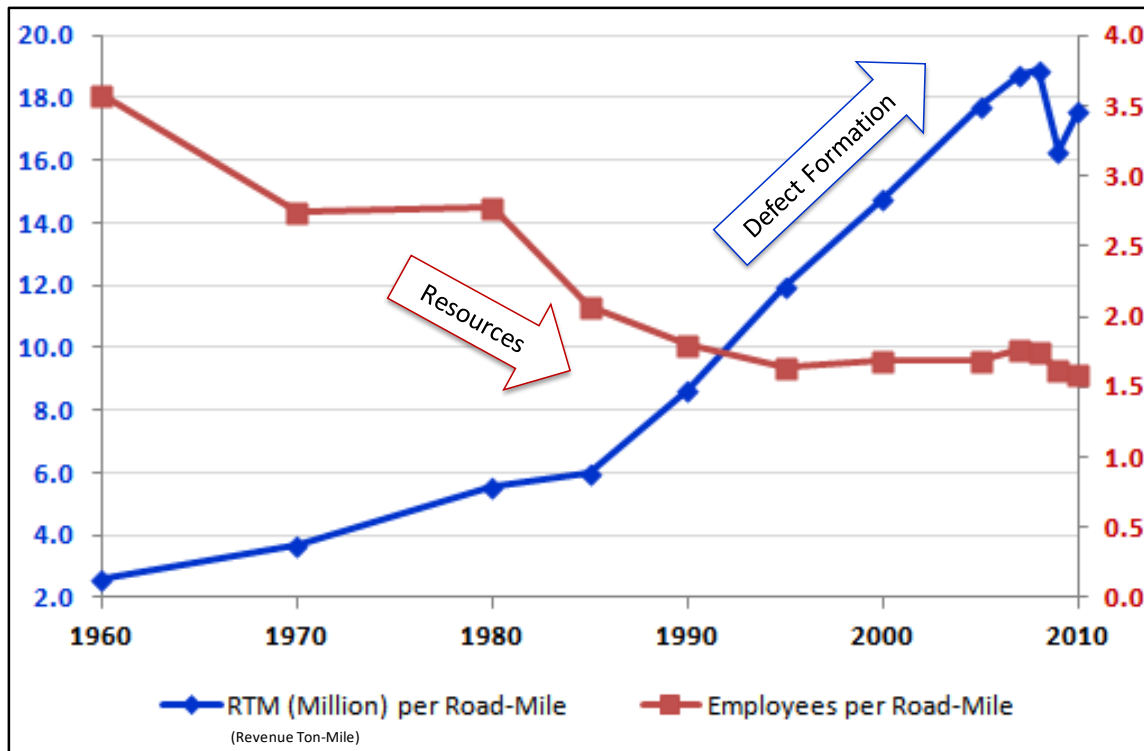
Figure 2.1 Growth in traffic load relative to employment per mile of road

Railroads augment visual inspections with automated NDE equipment to locate developing and mature abnormalities more quickly. However, the overall inspection rate is practically limited because the trailing repair gang must still be able to schedule track time, weather permitting, and keep up with the rate of abnormality discovery. A recent FRA survey found that railroads conduct 94% of visual inspections with the aid of a hi-rail vehicle (Al-Nazer et al., 2011). These contain hydraulic pumps to lower track wheels onto the rail when entering the rail line and raise them when returning to the pavement. Since many defects are not observable from a hi-rail vehicle, inspectors must still patrol the tracks by foot.

2.3   Track Surface Abnormality

Railroad defects or faults are interchangeable terms widely used in the railroad industry, and they form a rather broad topic. A railroad structure consists of the rail track and the substructure. This report will cover only track structure faults, and will categorize them into track fault and track surface abnormality. Rail faults can develop in any type of rail or rail weld as a result of the rail manufacturing process, cyclical loading, and impact from rolling stock, rail wear, and plastic flow. Rail surface anomalies or rail surface abnormalities are interchangeable terms used by researchers in the railroad industry. A rail surface anomaly or abnormality is one that deviates from a standard or normal rail surface. It is an abnormal track surface feature, characteristic, or occurrence in rail track surface. This research only focuses on track surface abnormality detection.

2.3.1   Type of Surface Abnormalities and Their Measurements

The wheel-rail contact geometry relationship is a common problem in the railroad system. A mismatched wheel-rail surface due to abnormal track surface will lead to serious problems for a running train, such as instability and high contact stress between the wheel and rail, which may further result in heavy tread wear, fatigue cracks, high noise, and derailment. Track maintenance agencies require many resources to

4

continually monitor the track surface abnormality condition. Table 2.1 summarizes the common types of surface abnormality, detection tools, and remedial action.

Table 2.1 Summary of Track Surface Abnormalities and Their Measurements

Surface Abnormality	Detection Tools	Measurements	Remedial action
Gauge (56½ inches)	Gauge measurer	56 inches ≤ gauge ≤ 58 inches	Adding spikes or tie
Alignment (62 feet)	String	> 5 inch	
Cross level	Track level	> 3 ½ inch on curves > 2 ½ inch on normal	Tamping fresh ballast
Profile (62 feet)	String	> 3 inch	Tamping fresh ballast
Warp	Track level	> 3 inch	Tamping fresh ballast
Cant	Track geometry recording trolley	< 5 inch	
Twist	Track geometry recording trolley	1 inch in 62 feet	
*Dipped joints	Laser		
Corrugation: Short Pitch Long Pitch	Optical and laser	30mm to 90mm wavelength > 300mm wavelength	Grinding
Squats	Sound / Ultrasonic		Grinding or replace
Turnouts (Chipped)	Optical	> 7/8 inch ¾-inch deep and longer than 4 inches 5/8-inch deep and longer than 7 inches	Repair or replace

2.3.2 Summary

Track surface condition monitoring is highly demanded in the railway industry to meet the ever growing needs of safety, reliability, and lower-cost operations. There are two categories for these modern track condition monitoring systems: track side or on-board systems. On-board monitoring systems have been used most often to monitor for track defects. Sensors typically used for these on-board monitoring systems typically include accelerometers, gyros, noise sensors, and a GPS, which are often used altogether to detect track surface irregularities, vehicle speed and location, and vehicle dynamic behavior.

2.4 Current Railroad Practice

Railroads supplement human visual observations with specially outfitted inspection cars that autonomously analyze the track for abnormalities using NDE technologies. These include electromagnetic, acoustic, optical, and inertial sensing methods, which this research organizes in the taxonomy of Figure 2.2. When deployed on inspection cars, NDE can locate many types of abnormalities, faster and more consistently than most human inspectors can. However, they have significant shortcomings in accuracy, precision, size, and costs. Moreover, inspection vehicles add non-revenue traffic to the tracks, which decreases available capacity and potentially increases downtime.

Railroads deploy NDE methods in two ways: as infrastructure integrated sensors to monitor the local area infrastructure and passing train characteristics, and as rolling stock integrated sensors to continually monitor the traveled infrastructure and vehicle for defects.

The latter method is favored by researchers because of its effectiveness. But the size and cost of these technologies currently limit their deployment to specially constructed automated inspection vehicles that locate internal rail flaws and irregular track geometry, track modulus, and gauge restraint. According to §213.233 of the FTSS, railroads must conduct automated track geometry measurements one to three times annually per section of track.
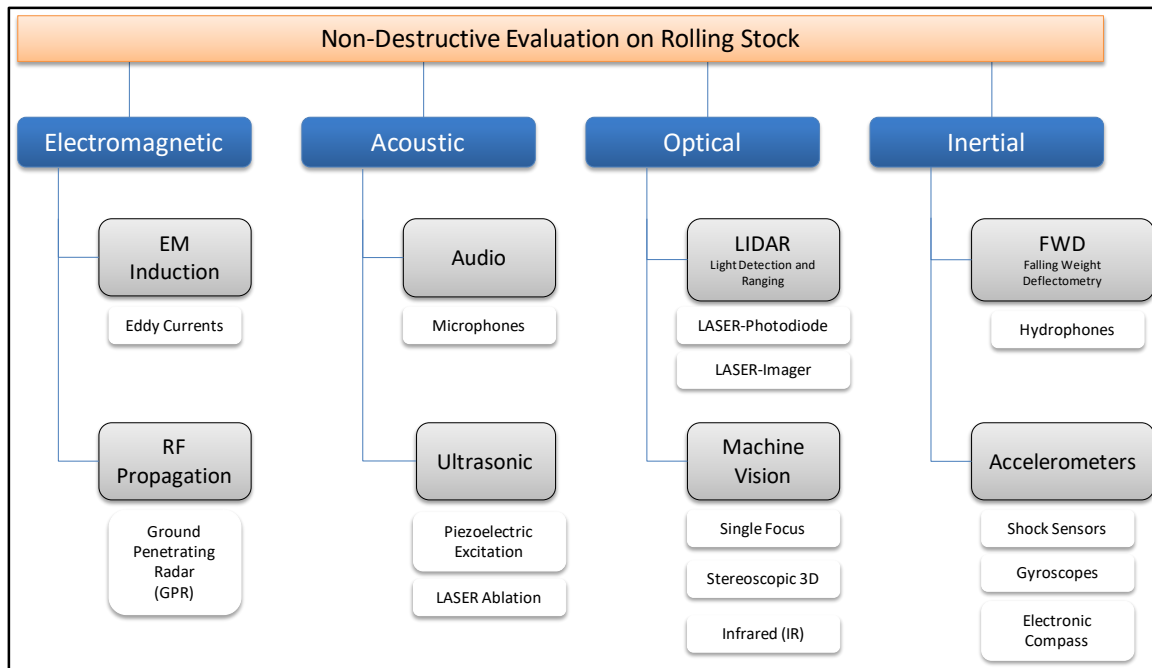
Figure 2.2 A taxonomy of NDE technologies

The FRA believes "the development of measurement technologies fitted on moving equipment has greatly increased the accuracy and speed of inspections, and has been a major factor in the decline of track-caused derailments" (FRA, 2008).

All NDE technologies emit some forms of energy into the track area and sense a response. The transmitted energy can be electromagnetic, optical, acoustic, or kinetic. Although some of these technologies have continually improved over the years, they are still not sufficiently accurate to replace visual inspections. The Amtrak derailment near Flora, Mississippi, on April 6, 2004, is a characteristic case study of this shortcoming (NTSB, 2005). The literature search finds that no single NDE technique offers a complete solution for locating and characterizing all types of defects.

Therefore, all full-scale solutions combine a variety of complementary NDE techniques. These combinations further increase their size, complexity, computational, and maintenance requirements (Papaelia et al., 2008). The most constraining technologies of the combined solution limit the inspection speed well below the average revenue-service train speed (Clark, Gordon, and Forde, 2003). The next few sections examine the functionality and limitations of each type of technology. A complete literature review regarding inertial sensor-based track condition monitoring systems and digital signal processing is summarized by Chia et al. (2019).

2.4.1 Electromagnetic Sensors

Waveforms carrying electromagnetic energy pass through a media, induce eddy currents at the media boundary, or reflect from the media boundary. Probes inject electromagnetic energy into the track area, and sensor arrays pick up the reflected energy or induced currents.

2.4.1.1 Electromagnetic Induction

An alternating current in a coil placed within a few millimeters of the railhead will induce eddy currents in the conductive portions of the rail. A second receiver coil monitors the phase and magnitude of these

eddy currents to detect changes in the electrical conductivity and magnetic permeability of the material. Rail flaws, such as cracks, will show up as parameter changes from normal.

Eddy current approaches are one of the earliest NDE methods used. They have the advantage of being able to detect small cracks near the surface of complex shapes, such as a rail. However, they cannot analyze non-conductive materials. In fact, the surface finish and standoff distance from the rail affects the reading. Consequently, these probes must travel relatively close to the railhead with a consistent standoff distance. This requirement practically limits their speed and potential for deployment on revenue service trains.

2.4.1.2 Radio Wave Propagation

Ground Penetrating Radio Detection and Ranging (RADAR) systems, also known as GPR, provide an interpretation of the load bearing capacity of the underlying track support structure, particularly the ballast, sub-ballast, and sub-grade (Silvast et al., 2010). A GPR produces a radiographic image interpretation of the subsurface layers by sending an electromagnetic energy pulse into the ground and observing reflections of the propagating radio waves. Typical GPR frequencies range from 400 MHz to 3 GHz, depending on the depth and resolution of interest. Decades of research in GPR seem to converge on 2 GHz for characterizing the degree of ballast fouling near the surface, and 500 MHz for identifying sub-ballast and sub-grade anomalies to a depth of about 6 feet.

The FCC limits the radio frequency power output at various frequencies. This constrains the convergence time needed for the background noise filters, which practically limits the inspection vehicle speed to about 25 mph. Surveying smaller areas between ties for more detail will constrain the speed to about 8 mph. Reducing the scan depth to a few inches, and quadrupling the antennas and controllers to decrease the scanning area per antenna, can push the inspection speed to about 125 mph (Roberts et al., 2009). However, the size of these high directivity antennas are inversely proportional to the radio frequency wavelength and can measure several feet in each dimension, as shown in Figure 2.3. The high output power also requires physically large power supplies and power conditioning circuitry to minimize noise contamination in the faint backscatter signals.

When combined with mechanistic models, GPR data can estimate vertical modulus to predict failures (Narrayanan et al., 2004). However, as with other NDE techniques such as eddy currents, considerable expertise is necessary to effectively design, conduct, and interpret GPR surveys. High conductivity materials, such as clay and salt contaminated soils, limit GPR performance. Rocky soils also excessively scatter the signals and reduce the information content. In general, GPR solutions are presently too large, slow, and expensive for integration on revenue generation trains.

2.4.1 Acoustic Sensors

2.4.2.1 Audio

Wheel bearing defects and wheel flats produce acoustic signatures that either wayside or rolling stock microphones can detect. Although they cannot detect all bearing defects, they tend to outperform infrared heat detecting solutions, because defective bearings will generate a characteristic acoustic signature much earlier and heat much later in their deterioration cycle. Most commercially available solutions are wayside monitors that sense the acoustics of passing vehicles (Rasmussen, 1998). Wayside monitors can be effective in identifying equipment defects, but do occasionally miss a few that have actually led to derailments (Southern et al., 2004). Although considered a well-established inspection technology, on-board devices are almost non-existent, most likely because of their limited capabilities and narrow application focus.
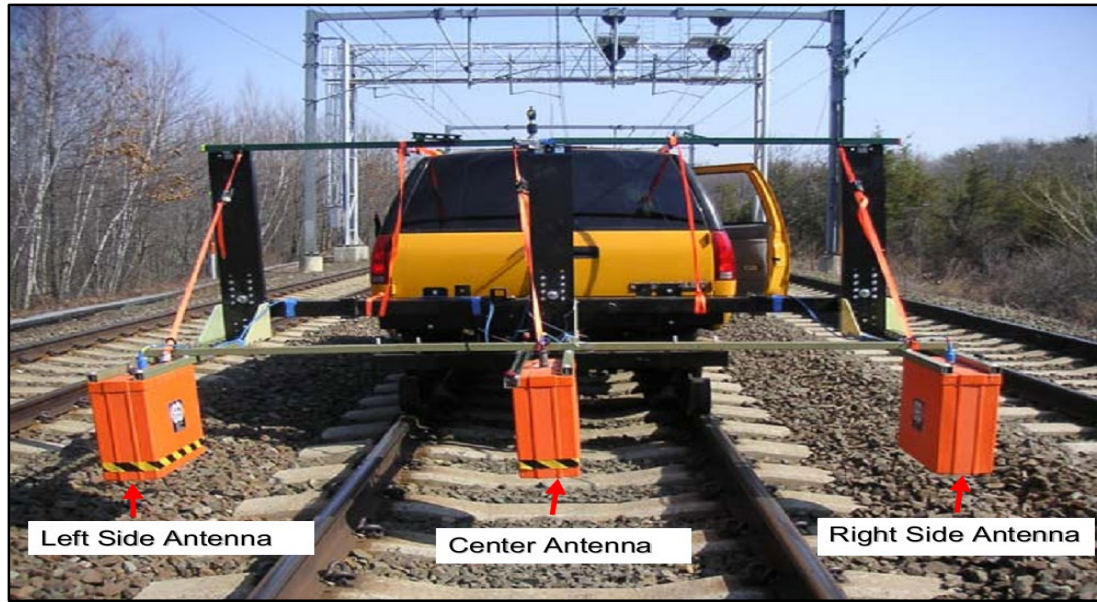
Figure 2.3 GPR attachment for a hi-rail [*image credit: Federal Railroad Administration (FRA), 2009*]

2.4.2.2 Ultrasonic

Thermal expansion creates longitudinal forces in the rail that promote fatigue cracking, welding separation, and fretting corrosion in bolt-jointed parts. Compressive stress from thermal contraction can lead to track buckling under dynamic trainloads and velocities. The speed of ultrasonic wave propagation along the track is proportional to such longitudinal rail stresses (Kube, 2010). However, measuring the waveform propagation speed to the desired accuracy requires a sufficiently wide separation between the exciter and receiver, and hence multiple probes.

Most ultrasonic flaw detection applications utilize frequencies from 500 KHz to 10 MHz. At frequencies in the megahertz range, sound energy does not travel efficiently through air or other gases, but it travels freely through most liquids and solids. Therefore, systems designed for inspection cars use liquid-filled rubber wheels to couple the excitation energy into the rail. However, this approach limits inspection speed to about 40 mph (Phillips, 2006). Minor variations in the wheel probe position, water path length, and internal fluid temperature significantly affect wheel probe results. The technique also misses cracks in the rail-web and rail-foot (Garcia, 2011).

Ultrasonic NDE is a relatively new technology, and there is still much to learn about the behavior of guided waves in complex structures, such as railroad tracks. There are also numerous performance limitations. For example, the presence of residual layers from wheel burns can shadow internal defects. The backscattered waveforms require complex signal processing and experts to interpret them. The technique is better suited for flaw detection in the material core and relatively poor at detecting surface or near-surface defects, where most of the surface abnormalities are located. Therefore, most NDE equipment includes electromagnetic probes to compensate for this deficiency (Zahran, Shihab, and Al-Nuaimy, 2002). In addition to the limited testing speed, multiple probe types increase the equipment size and power consumption, making them ill-suited for integration on revenue service trains.

2.4.3 Optical Sensors

Optical systems use light emitters to illuminate the surface and image sensors to capture the reflected light. Some systems use laser sources to measure the distance from the surface while others use high-

powered exotic gas light sources to illuminate the surface for image registration. Environmentally insulated and explosion proof cabinets must protect these devices from the harsh environments, making the construction bulky and expensive.

2.4.3.1 Light Detection and Ranging (LIDAR)

LIDAR measures the light reflected from a transverse laser beam emitted across the travel path to create a surface depth profile. Rotating mirror assemblies move a laser light spot across the travel surface. A series of lenses focus the reflected light onto position-sensitive light detectors (photodiode arrays) that translate the distance from the surface to a proportional electrical signal.

Variations in track modulus cause dynamic loading, which reduces the life of track components. Modulus is the supporting force per unit length of rail per unit of vertical deflection, or simply track stiffness. Lower quality rails, ties, rail joints, ballast, and sub-grade exhibits lower track modulus (Selig and Li). Traditional methods to measure modulus require track crew traveling the track to apply known loads with a falling weight deflectometer (FWD) and measuring the resulting deflection (Huille and Hunt, 2000). Automated methods use LIDAR to estimate the modulus by measuring the amount of rail displacement from a tangential horizontal plane above the wheel contact point.

The laser scanning action produces sufficiently high transversal surface resolution, but longitudinal resolution decreases with increasing train speed. The high transversal resolution is an advantage of LIDAR, but the limited longitudinal resolution is a major shortcoming. Manufacturers can overcome this limitation, to some degree, by using sophisticated components that increase the scan rate, thereby increasing the inspection speed beyond 40 mph. However, such additional performance requires larger and more expensive construction. Overall, the physical limitations in signal bandwidth, signal-to-noise ratio, sample rate, power consumption, and processing speed ultimately provide diminishing returns in all such optical systems. These systems must also operate relatively close to the tracks, and their relatively large construction and hardening to withstand the harsh environmental conditions make them less attractive for installation on revenue service trains.

2.4.3.2 Machine Vision Systems

With the appropriate level of surface illumination, machine vision systems can capture an image and process it to extract features that would identify and characterize surface abnormalities. Image feature analysis can identify obvious abnormalities, such as missing rail fastener components, rail surface deterioration, cracked ties, broken rails, broken switch points, mud spots, and excessive ballast vegetation. More recent systems add additional cameras to create stereoscopic vision or 3D images for depth information. Adding infrared filters will shift the spectral sensitivity toward longer wavelengths to detect cold wheels, hot wheels, and hot journals. Wheels colder than others can indicate poor brake performance. Relatively hot wheels can indicate skidding or sticking brakes. Hot journals can indicate impending bearing failure or overheating from ceased bearings, which can cause a derailment.

The main advantages of machine vision systems include greater objectivity and consistency than human inspectors. However, they have numerous disadvantages. In general, machine vision solutions require large storage capacity for the images, and an ample light source with sun shielding for consistent image quality. Image processing is computationally intensive and often involves self-learning algorithms to detect specific objects in the image frame. High-power xenon lights or lasers can improve the lighting condition for a subset of image analysis types, but they add significant cost, bulk, and power consumption. Another shortcoming of car-mounted cameras is that the longitudinal resolution depends on the frame capture rate, which in turn limits the car operating speed. Technology advancements can increase frame rate at higher cost, but their difficulty coping with unusual or unforeseen circumstances,

such as occlusion from precipitation, leaves, or debris, ultimately limits their accuracy (Sawadisavi, 2008). Even at high frame rates, these systems cannot detect abnormality under conditions of low illumination or line-of-sight obstructions. Most of the systems reported in the literature provided roughly 80% detection accuracy for the specific abnormality they targeted, nonetheless, those were under conditions of good lighting and reasonably high image resolution.

2.4.4 Inertial Sensors

This general class of NDE relates to systems that measure the impulse response from mechanical energy directed into the track structure.

2.4.4.1 Falling Weight Deflectometer

Methods of structural capacity estimation in the 1980s used an FWD to direct kinetic energy into the track support structure while observing the surface deflections with an impulse response sensor, such as a geophone (Burrow, Chen, and Shein, 2007). An inspection car or Hi-Rail typically hauls a trailer containing the FWD equipment. The test speed is limited by the impulse response duration. With the appropriate signal processing, it is possible to measure the impulse response from the weight of the rolling stock itself to estimate of track modulus (Kobayashi et al., 2008). Early approaches investigated the possibility of measuring track modulus and lateral alignment with gyroscopic sensors mounted on the bogie (Weston et al., 2003). The method based its estimate on the principle that double integration of the acceleration signal produces vertical displacement. However, the offset cancellation and required calibration became impractical. This method still holds promise and could yield better results with adaptive advanced signal processing concepts.

2.4.4.2 Vehicle-track interaction monitors

Inertial sensors that analyze impulse responses from the vehicle-track interaction are the least developed of all NDE technologies currently in use. The FRA sponsored the development and testing of a GPS-accelerometer-based device in 1996 to monitor the vibration of wheel and axle assemblies. ENSCO, Inc. commercialized the technology in 1998 as a vehicle track interaction (VTI) monitor, and has since deployed it on about 250 freight and passenger trains in North America and Australia (Stevens, 2009). The installation and processing requirements are complex.

The VTI varies with the type of track irregularity and quality of the rail, tie, ballast, and sub-grade. Testing demonstrated that the system could detect 84% of FTSS "exception" conditions by using a neural network to establish the optimum shock-level thresholds (FRA, 2000). It is otherwise difficult to establish these thresholds analytically or by trial and error. Possible shock levels vary widely, due to a combination of the specific abnormality type and the VTI characteristics under variations of carload, configuration, speed, and condition.

Early testing showed that VTI sensors produced a high false positive rate (20%) for vehicle suspension faults (Tuzik, 2005). However, with the appropriate sample rate and signal processing, it should be possible to improve their performance.

Based on the unit design, construction details, and required deployment configuration, an experienced engineer's estimate would be roughly several thousand dollars per VTI sensor, not including installation, configuration, cellular connectivity, or maintenance costs. This high unit cost is likely a significant factor in their limited deployment.

Railroads are seeking to enhance the efficiency of inspections and maintenance of way. Instead of continually searching for track abnormality throughout the system during precious track time, this paper introduces a new low-cost wireless smartphone-based sensor technology and method for guiding inspectors to the likely locations of track abnormality for more efficient diagnosis and remediation.

2.5   Inertial Sensor Based Systems

Inertial sensors, such as accelerometers and gyroscopes, have been widely used in asset management since the early 1990s. A gyroscope measures the sensor's angular velocity while an accelerometer measures the specific external force acting on the sensor. This section focuses on an extensive and systematic review of inertial sensor applications, and challenges in transportation health monitoring.

2.5.1 Inertial Sensor Applications

The inertial sensor applications reviewed in this section focused on vehicle-track interaction monitor applications. The application normally contains two steps: data acquisition and data analysis. Inertial sensors normally collect vibration signals through the vehicle-track dynamic interaction due to track irregularities and surface abnormalities. And signal data processing techniques will be selected for further acceleration signals data analysis to identify track irregularities and detect faults or surface abnormalities.

Inertial sensors are widely used as data acquisition tools due to their small size, low power consumption, and robustness. A wide variety of inertial sensors are available on the market, ranging from uniaxial accelerometer/gyroscopes to inertial measurement units (IMUs) with 6 degrees of freedom. The dynamic range of the measurements varied, depending on the sensor type and implementation characteristics. Inertial sensors and their configurations are summarized in Table 2.2.

Table 2.2 Inertial Sensor Application Specifications

| Year | Number of Sensors | Types and Specification | | | | Sampling Frequency (Hz) | Position of the Sensor |
| | | Accelerometer | | Gyroscope | | | |
		Measuring Axes	Measuring Range	Measuring Axes	Measuring Range		
2000	4	2		3			Axle-box
2006	15	3	±100g; ±10; ±4g	3		4000	Bogie frame
2006	3	3				2000	Vehicle body, axle-box
2006	3	3				2000	Vehicle body and axle-box
2006		3		3			Bogie, axle-box
2007	4	2		3			Bogie
2007	4	2	±10g	3	±50 ° s^{-1}		Bogie
2007	9		±250g			3000	Each axle-box of bogie
2010	3	3		3		2000	Car body and bogie
2010	3	3					Bearing box of wheelset
2011		3					Axle-box
2011	3	3		3			Car body
2011		3		3		2048	Axle-box and bogie
2011	1	3				100	Overhead luggage rack
2011	4	3		3		4000	Axle boxes
2011	5	3					Axle-box, bogie
2012	2					2048	Axle-box and bogie
2012	4	3	±500g			100	Lead bogie axle-box
2013	2	3	±3g				Bogie
2013	3	3		3		1000	On board
2014	3	3		3			Bogie
2014	2	3					Driving wheels; free running wheels
2014	3	3		3			On-board
2014	2	3		3		82	On-board
2014	3	3;3;2	±70g; ±18g;±1.7			79,200 4000	Axle-box; bogie; car-body
2014	4	3					Lead bogie axle-boxes
2015	4	3				2000	Car body
2015		3					Wheel axles boxes
2016	2	3	±60g				Motor bogie, trailer bogie
2016	8	3	2g, 10g			2000	Car body floor, bogie frame
2016	4	3	±50g			2000	Axle-box of bogie
2017	5	2,3				1,600	Wheel truck; inside cabin

*note, the missing value in the table indicates the corresponding information is not available in the application report or journal.

From Table 2.2, we can see the dynamic range of accelerometers used in asset condition monitoring was from ±2 g to ±500 g, where g is the g-force unit. Inertial sensor collection sampling frequency ranges from 82 Hz to 4000 Hz, and it worth noting the actual data were collected under a non-uniform sampling frequency rate. Also note there is no clear guidance on sensor configuration selection, number of sensors, and the location of the sensors, but most of the research has multiple sensors (more than one) and inertial sensors are normally mounted on axle boxes and bogies.

The signal data were filtered or processed after they were collected. The signals from inertial sensors contain noise and other unwanted signal components, including low-frequency components, such as undesirable offsets. Therefore, the selection of the appropriate signal filtering technique is crucial to the success of the application.

2.5.2 Filtering Techniques

Linear time-invariant (LTI) filters are those whose behavior does not change over time. LTI filters are implemented using finite impulse response (FIR) or infinite impulse response (IIR) methods. The basic structure of the FIR filter requires a multiplier, an addition, and a unit delay. Characteristics of IIR filters are as follows (41):
- Based on transforming a continuous-time analog filter into discrete-time filter
- Incorporates delays of the output signal
- Allows both zeros and poles
- Can be implemented with recursion
- It is a feedback filter

Advantages: leverages decades of experience in designing analog filters; less expensive than FIR filters; allows for greater shaping potential.

Drawbacks: They can have phase distortion and ringing; they are not guaranteed to be stable.

A Kalman filter (KF) is a model-based method. A KF offers more practical implementations based on the recursive concept (33). A KF consists of a predictive block and correction block. Time update uses the dynamic equation to predict the next time step. Measurement update explains how the information available in the new measurement is incorporated into the estimate. The time update of the KF consists of the current state prediction and system process covariance calculation (33). The KF is widely used as an effective health monitoring method for linear systems, and the extended KF is normally used for nonlinear stochastic systems. The KF is a method widely used for state and parameter estimation.

Convolution is the mathematical way of combining two signals to form a third signal that has much better performance than recursion filters, but executes much more slowly (36). They are also called finite impulse response filters. However, there is a delay such that all incoming samples receive the same treatment so that the signal phase relationship is preserved (35). An algorithm called FFT convolution is used to increase the speed of the convolution, which allows the FIR filter to perform faster.

Deconvolution is a filtering implementation that reverses the process of convolution. It recreates the signal as it existed before the convolution took place. Deconvolution works well in the frequency domain. It usually requires knowing the characteristics of the convolution, which is the impulse response and the output vector (spectra). Researchers (36, 42) have applied deconvolution filtering to measure railcar body vertical accelerations. The authors used Laplace transform theory and pole-zero plots to achieve the deconvolution. In essence, deconvolution filtering eliminates or at least reduces the amplification and attenuation effects of the railcar suspension system. Deconvolution filtering advantages are that they are often linear, deterministic, non-iterative, and fast. The disadvantages are their sensitivity to noise, which can result in noise amplification and the difficulty of incorporating available a priori information.

2.5.3 Limitations and Challenges

The sampling collection rate of the inertial sensors often varies during data collection non-uniformly due to the variation in signal processing speeds. Moreover, various inertial sensors are designed at different sampling rates, thus different inertial sensors will capture vibrations at various frequency bands.

Aliasing is an effect that causes different signals to become indistinguishable when sampled. In other words, high-frequency input signals will appear as low-frequency signals at the output. If aliasing occurs, no signal processing operation downstream of the sampling process can recover the original continuous time signal. It occurs when a signal is sampled too slowly. To prevent aliasing, the signal must be sampled at least twice as fast as the highest frequency component (Nyquist criterion).

Group delay distortion occurs when signals at different frequencies take a different amount of time to pass through a filter. If the phase response is linear, the group delay of the filter is constant, which means that each frequency component experiences the same delay. Otherwise, the frequency components have different delays, which cause a smearing phenomenon in the time-domain signal. It is important to describe a filter's passband characteristics or evaluate the filter's phase nonlinearity.

In practice, due to tradeoffs in roll-off steepness, filter delay, and computational complexity, the maximum cutoff frequency should be set much lower than half the minimum sample rate of the sensor. Therefore, some practices use as high a sample rate as possible, within some reasonable considerations for minimizing the data accumulation rate and the sensor power consumption.

The inertial sensors suffer from bias instability, noisy output, and insufficient resolution. Digital filtering is an important technique used to address such issues. Much research has been conducted or proposed to obtain "good" digital filters with high performance. However, in practical applications, few research results are used efficiently. Since the research field of digital filters have become diverse and complicated, there is a lack of understanding of what constitutes a good digital filter and how can we obtain it.

3.   APP TO USE SMARTPHONE AS SENSORS

In lieu of having no low-cost and low-power commercial sensors, the research team developed a smartphone application capable of autonomously collecting and uploading data from hi-rail vehicles, geometry cars, locomotives, and end-of-train cars where power is available. The technology transfer phase will inform commercialization partners about the best approaches to develop a lower-cost and self-sufficient version of the sensor system deployed during the research.

3.1   RIVET App

The RIVET app 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. Online algorithms are required to process the data from the inertial, gyroscopic, and GPS sensors. Applications include road and rail performance evaluations and condition monitoring. The app is available for research use only, and only by written request from the principal investigator.

3.2   App Installation

Send a Gmail containing the .apk file to yourself. Open Gmail on your Android device and tap on the attachment. This will begin the process of installing the app.

3.3   App Setup

3.3.1   Descriptions

In the "Settings" screen, enter the two description information about your data collection.

3.3.2   URL

The URL for your local server must begin with http:// and end with a forward slash /. The following is an example:
 http://dotsc.ugpti.ndsu.nodak.edu/PAVVET_SMARTSe/PL/

This assumes that you have already set up your local server and it has a valid URL. The server must have a verifiable security certificate. The app makes an HTTP post to the URL provided and responds with an "Invalid URL" for HTTP status codes of 404, 405, or greater than 500.

See the appendix for instructions in setting up a local server in Windows. You can map the URL as a network drive to view and download the .CSV files. See instructions for mapping a network drive here:

https://www.ndsu.edu/its/help_desk/collaboration_and_storage/file_services/map_drives/window s7/

3.3.3   Screen Lock

Activate this to keep the app active during data logging. If the screen falls asleep, some Android phones will throttle down its performance. This will slow down data logging and uploading, and perhaps produce other unexpected behaviors.

 When using the app in unattended mode, turn down the intensity level to the minimum to conserve battery life if power is lost. The device will stop logging and uploading files if power is lost and will

continue to log and upload files when power returns. This feature is essential when automatically logging data in vehicles (trucks, trains, etc.) with the phone plugged in, and without having to ever attend to the device.

3.3.4 Calibrate Inertial Sensors

Use of this feature is not recommended:
- When tapped, the app gets an average value of the accelerometer data for one second, and then subtracts that value from the logged variables. Note carefully that this includes the Az value, which typically reflects the 1-g pull of earth. This value persists throughout app launches.
- Use of the calibration function is not recommended because sensor offsets can be removed with offline processing and filtering. The app posts a message if calibration was done, and the timestamp of the last calibration is located in the header (fifth column of the first row). There is a known issue where the first 60 samples are static before the calibrated values are applied. Hence, use of this feature is not recommended.

3.4   Main Screen

The values shown are:
- Accelerometer update rate in Hz averaged over a one-second interval.
- GPS update rate in Hz averaged over a one-second interval.
- Speed in meters/second. This is the instantaneous sample (not averaged).
- Accelerometer in the three dimensions: Ax, Ay, and Az. They are the instantaneous samples (not averaged) in units of meters/second. They must be converted to g-force when used to produce the RIF.
- Gyroscope angles in three dimensions: Azimuth (or Yaw), Pitch, and Roll. They are the instantaneous samples (not averaged) in units of degrees. They must be converted to radians when producing the resultant accelerations from a rotated device.
- Magnetic field strength in three dimensions: Mx, My, and Mz. They are the instantaneous samples (not averaged) in units of micro-Tesla.

Note that the device logs the gyroscope angular velocities in three dimensions: Rx, Ry, and Rz, but does not display them on the main screen. They are in units of radians per second.

3.5   File Upload

At about 400 Hz, each file size is about 4,600 KB each minute.

3.5.1 Automatic Uploads

- Tap "Start Logging" or simply plug in the device to initiate data logging and uploading. The files automatically upload approximately every minute.
- Data logging continues and files are queued if the network connection is lost while logging. When the network connection returns, the device uploads whatever files were pending during the outage, and continues to upload files every minute.
- If the network connection is lost during a file upload, the device may attempt to re-transmit the same file when connection returns. This can lead to a single file having two headers. The data cleaning process must detect this and remove the redundant data.

3.5.2 Manual Uploads

- Tapping "Stop Logging" causes any pending files logged to automatically upload.
- If logging was stopped during a network outage, toggling "Start Logging" and "Stop Logging" uploads any queued files, plus the one created due to "Start Logging"

3.6 Power Management

Plugging into power at any time initiates a new data logging session. Unplugging the device stops data collection and uploading.

3.6.1 Plugged In

- When a "Start Logging" session is initiated while plugged in, the device logs and uploads data continually until "Stop Logging" is tapped.
- With the app launched, plugging the device into power automatically initiates data logging and uploading.
- If started while plugged into power, the app logs and uploads data normally. It stops logging data if power is lost. It restarts logging data when power returns.
- Android Pixel devices can stay alert and monitor for power returning for more than 24 hours. This also assumes the screen intensity is turned down to minimum, with screen-lock activated.

3.6.2 Unplugged

While unplugged and logging data, a plug-in event will restart the data logging

3.7 Data Files

The organization of the data file is two rows of information and labels, followed by the data. Each data row is a sample of the accelerometer data. GPS data update more slowly, so it is repeated until updated.

3.7.1 Header Row 1

- Column 1: Device Unique ID
- Column 2: Data entered into the settings "Location" field
- Column 3: Data entered into the settings "Comment" field
- Column 4: Timestamp of file creation in epoch time format for milliseconds
- Column 5: Timestamp of last calibration in epoch time format for milliseconds

3.7.2 Header Row 2

Labels for the 16 data fields

3.7.3 Data Rows

- Column 1: Timestamp of data in epoch time format for milliseconds (DateTime)
- Column 2: Latitude in decimal degrees (Lat)
- Column 3: Longitude in decimal degrees (Lon)
- Column 4: Speed in meters-per-second (Speed)

- Column 5: Accelerometer x-value in meters-per-second-squared (Ax)
- Column 6: Accelerometer y-value in meters-per-second-squared (Ay)
- Column 7: Accelerometer z-value in meters-per-second-squared (Az)
- Column 8: Gyroscope yaw angle in degrees (Azimuth)
- Column 9: Gyroscope pitch angle in degrees (Pitch)
- Column 10: Gyroscope roll angle in degrees (Roll)
- Column 11: Gyroscope rotation rate around the x-axis in radians-per-second (Rx)
- Column 12: Gyroscope rotation rate around the y-axis in radians-per-second (Ry)
- Column 13: Gyroscope rotation rate around the z-axis in radians-per-second (Rz)
- Column 14: Geomagnetic field strength along the x axis in micro-Tesla (Mx)
- Column 15: Geomagnetic field strength along the y axis in micro-Tesla (Mx)
- Column 16: Geomagnetic field strength along the z axis in micro-Tesla (Mx)

3.7.4 Application Notes

- Mount the devices flat to the vibrating surface to maximize the contact area. Secure it with tape to prevent bouncing around.
- Disable all other apps on the device to prevent any interruptions.
- Disable all automated app updating.
- Disable automatic OS updating.
- Disable all notifications of any other apps so that vibrations or audio alerts do not register falsely as inertial events.
- For application in calculating the RIF, divide the accelerometer values by -9.06 to get g-force.
- Convert the degrees to radians when performing the matrix rotation.
- The recommended practice for using PAVVET algorithms with RIVET data files is to first convert the RIVET data into the PAVVET format.

3.8   Setting Up the Local Server

3.8.1 Setting Method 1

Step 1: Install Internet Information Services (IIS). See the following or a similar link for instructions: http://www.howtogeek.com/112455/how-to-install-iis-8-on-windows-8/

In addition to IIS, CGI also needs to be installed. It is found in the IIS selections in the Windows Features dialog box. Select to enable it, as it is not enabled when IIS is first selected.

Also see the following link for setting up a basic website in IIS: https://support.microsoft.com/en-us/help/323972/how-to-set-up-your-first-iis-web-site

PHP also needs to be installed for the upload PHP script to work properly.

- Download PHP here: http://windows.php.net/download/
- Extract the zip file to a file you create: C:\PHP
- Open IIS Manager, select the hostname of your computer in the Connections panel, and then double-click Handler Mappings.
- In the Action panel, click Add Module Mapping.
- In Request path, type *.php.
- From the Module menu, select FastCgiModule.

- In the Executable box, type the full path to Php-cgi.exe; for example, C:\PHP\Php-cgi.exe.
- In Name, type a name for the module mapping; for example, FastCGI.
- Click OK.

Step 2: In the C:\inetpub\wwwroot, create the directory where you want the data to go. This is where you will place the upload.php file.

Step 3: In the IIS Manager, you should see the folder you created under an existing website. If you do not have a website created, you should be able to right-click on the server in the Connections window and Add Web Site.

Step 4: The physical path for your website would be your wwwroot folder. Validate this.

Step 5: When you click on the folder you created in the Connections pane, the center pane should have a number of icons for you to click on for setting adjustment.

Double-click "Default Document" and click on "Add in the Actions Pane" to the right.

Add "upload.php." This will create the web.config file in that folder you created.

Step 6: Uncheck the encryption property for upload.php file.

3.8.2 Setting Method 2

Step 1. Install Java environment on Windows. The latest version is located at:
http://www.oracle.com/technetwork/java/javase/downloads/jdk9-downloads-3848520.html

After installing Java, go to installation directory: C:\Program Files\Java. Rename the existing folders to jdk jre

Step 2. Go to Control Panel --- System --- Advanced system setting:

Click "Environment Variables" under "System variables," find Path Edit/Add the address of the above two folders at the end of the existing one. It should look something like this: C:\Program Files\Java\jdk\bin;C:\Program Files\Java\jre\bin

Step 3. Download tomcat 9.4 from this link: https://tomcat.apache.org/download-90.cgi   Under Core select either 32-bit or 64-bit Windows zip file, unzip the folder and move it to C drive and rename it tomcat9.4 There should be a folder under C:\tomcat9.4 containing several files/folders such as (bin, conf ….)

Step 4. Now go to Control Panel → System →Advanced system setting Click "Environment Variables" Under "System variables" find Path Create two new variables as follows, Variable name: JAVA_HOME → Variable value: C:\Program Files\Java\jdk Variable name: CATALINA_HOME→ Variable value: C:\tomcat9.4

Step 5. Test if every step was done okay so far. Open terminal (Command Prompt) and type "javac." You should not get any error –□, which means that Java is installed successfully.   Open another terminal and go to C:\tomcat9.4\bin Type startup.bat. It will start the server. If it asks for permission click yes. Open your browser and type localhost: 8080. It should show Apache homepage. Then shut it down by typing showdown.bat.

Step 6. Download the Quercus zip file from here:

https://drive.google.com/open?id=1CIsAnCvUsTwBm2FxpUwEtCc3SEoWERZP.

Unzip the Quercus zip file and then place the quercus-4.0.39 folder and quercus-4.0.39.war into the folder: C:\tomcat9.4\webapps.

When first running the localhost server, the system would automatically remove the folder (pavvet-web) from the quercus-4.0.39 folder, hence creating a problem. To resolve this issue, recopy the folder (pavvet-web) into the folder (quercus-4.0.39 folder) in your system.

Step 7. In the Phone setting, set the URL or IP as: http://xxx.xxx.x.xxx:8080/quercus-4.0.39/pavvet-web

When done, files will be saved in this location in .csv format C:\tomcat9.4\webapps\quercus-4.0.39\pavvet-web.

Step 8. Whenever you want to make connection between your phone and computer, you should run startup.bat; when you are done, shut down the server.

3.9   Orientation Conventions

Orientation is positive in the clockwise direction. Angles are in degrees (unlike rotation rate). 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. They are shown in Figure 3.1.
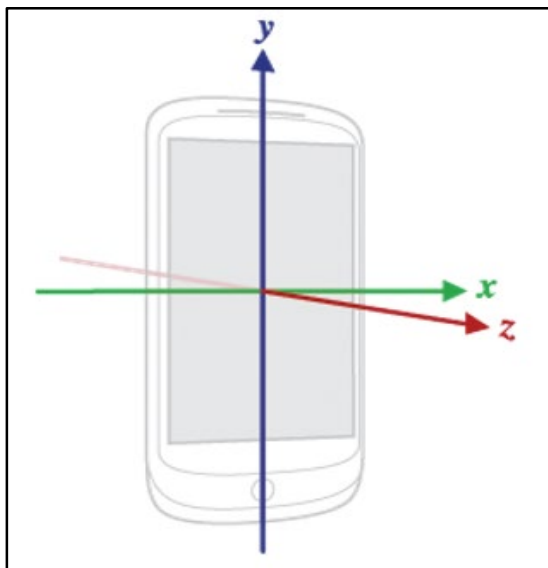

Figure 3.1 Orientation of the mobile device with three angular quantities

4. DATA COLLECTION AND PROCESSING

4.1 Data Collection Plan

Data collection is a crucial step in the process of the project, because valid and reliable data are the backbone of this research. Before collecting data, we need have a solid understanding of the purpose of the data collection to meet our desired research requirements and to design a detailed list of the specific measures we will consider in future data analysis. This chapter introduces the data collection plan, the data cleaning, and the data fusion techniques developed in this research

4.1.1 Identify Collaborating Railroad

Northern Plains Railroad & Rail Services owns Northern Plains Railroads (NPR). NPR is a customer-focused regional railroad operating a 350-mile network in North Dakota and western Minnesota. Since 1997, NPR's operations have primarily evolved into grain and miscellaneous industrial product handling and transportation systems. NPR&RS provide various in-house and third-party rail services, including track maintenance, repair, and construction. Due to its professional expertise in track maintenance and repair, NPR has been recognized through a comprehensive rail satisfaction survey conducted by the North Dakota Public Service Commission, scoring the highest in customer satisfaction of railways operating safely in the state.

4.1.2 Data Collection Plan

The Federal Railroad Administration (FRA) serves as safety agency to issue safety rules covering issues specifically connected to the railroad industry and the federal Occupational Safety and Health Administration (OSHA). The FRA also issues general safety and health regulations that apply to the railroad industry. Moreover, each railroad company has its own corporate safety standards and advisory for their railroad workers and visitors to ensure workplace safety. Due to liability concerns, most railroads do not run a hi-rail vehicle over the same line every day. Perhaps they may do so once per week. So it is a challenge for researchers to collect a statistically significant sample size over a short time period where the rate of track deterioration is relatively slow.

The alternative plan is to collect data across a specific track section with known surface abnormalities over an extended data collection period, which accumulates to the number of needed traversals. For instance, if the hi-rail collects data once per week over the selected track segment, then the period to accumulate data from 15 traversals will be around four months. In that case, the researchers will need to attach Android phones to the hi-rail vehicle, secure it in a semi-permanent position with line-of-sight to GPS satellites, plug it into the vehicle power supply, and launch the data collection app to automatically collect and upload data. The app will automatically start and stop data collection, without manual intervention, when power returns and leaves, respectively. This method assumes that a) the rate of deterioration over the four-month period is negligible, and b) the vehicle will be started within 24 hours to recharge the battery on the phone.

This alternative automatic data collection plan requires the following two action if the railroad agrees: 1) at least two researchers (Pan and Leonard) visit NPR to install the phones on the identified hi-rail, connect it to the vehicle power supply, and activate the app on that day; and 2) if the app ever stops working (that is, we observe no uploaded data when expected), then one identified researcher (Leonard) will contact Michael to restart the phone, verify that it is charging, re-launch the app, and verify that it is uploading data.

In this data collection plan, the researchers will need a critical supportive document from NPR that lists the known abnormality profile, the detailed abnormality association information, and their location in geospatial coordinates of latitude and longitude.

4.1.3 Proposed Data Collection Locations and Phone Installations

A hi-rail vehicle is typically a modified road truck that can operate both on rail tracks and conventional roads. They typically keep their normal rubber tires, but are equipped with additional flanged steel wheels to run on rail tracks, as shown in Figure 4.1. The interior of a hi-rail vehicle is a typical truck interior. Due to the limited truck interior space and power supply resources, a three-phone data collection system is proposed. The rationale is to accommodate the limited space and to provide enough information to verify location effects on the algorithms.

Figure 4.1 A typical hi-rail vehicle

The following proposed potential locations are recommended considering the different kinetic energy responses: 1) on top of the driver side dashboard, 2) on top of the passenger side dashboard, 3) on the floor of the rear passenger side, 4) on the floor of the front passenger side, 5) on the floor under driver seat, 6) on truck floor under passenger seat. Then the final three locations can be selected among those proposed potential locations based on power supply resources, installations, and GPS signal receiving strength.

The hi-rail ride on the track will be not as smooth as on pavement, so it is important to securely fix the phones on hi-rail. The objective is to secure the phones so they do not move around during the data collection period. One option is to use duct tape. Another option might involve drill installation.

After the installation, the researchers need to generate a phone location map, including their level relative to the floor of the vehicle, and the level of the vehicle's floor to the track and on pavement.

4.2 Data Cleaning and Data Fusion

4.2.1 Data Collection Setups

Harsco Rail's LD515 HY-RAIL® is the vehicle used in this research to collect data. It features hydraulic locks, push button actuation, electrical track insulation, and mechanical safety pins. The vehicle is shown in Figure 4.2 a). The HY-RAIL® is a remodel of a 2015 Ford F-350 light duty truck equipped with Harsco model 1515 rail gear with steel wheels. The detailed specifications of the Harsco LD1515 are shown in Table 4.1.

Figure 4.2 a) HY-RAIL® vehicle used; b) Phone 1 installed on dashboard; c) Phone 2 installed under driver seat; d) Phone 3 installed under passenger seat

A hi-rail vehicle is typically a modified road truck that can operate on both rail tracks and roadways. Short-line railroads typically use such hi-rail vehicles for daily manual inspections. The smartphone location setup considered the different kinetic energy responses, power supply resources, and the GPS signal strength. Figure 4.2 shows the hi-rail vehicle used in the research and the three locations of smartphone installations.

Table 4.1 HY-RAIL LD1515 Specifications

OID	Item	Descriptions
1	Truck Weight Classification	Light duty up to 11,500 lbs GVWR
2	Maximum Rail Travel Speed	45 mph
3	Unit Weight	320 lbs
4	Load Capacities	Steel wheels: front and rear units – 3,000 lbs each Rubber tread wheels: 1,400 lbs each
5	Guide Wheel Load	Steel wheels: front and rear units – 1,500 lbs each Rubber tread wheels: 700 lbs each
6	Suspension	Coil spring with over load protection
7	Guide Wheels	10 in. tread diameter, rubber tread or steel tread
8	Wheel Bearings	Sealed maintenance free bearings
9	Safety Pin Locks	Mechanical pins plus hydraulic PO check locks
10	Raise/Lower Operation	Electrical/hydraulic actuation with push button controllers
11	Track Gauge Availability	56.5 in., 63 in., and 66 in.
12	Track Signal Insulation	All guide wheels electrically insulated for track signaling
13	Included Equipment	Mounting brackets, steering wheel lock, de-rail guards, wheel modifications, front rail sweeps, hydraulic power pack, electrical controls
14	Optional Equipment	Steel or rubber tread guide wheels, emergency hand pump, rear rail sweeps, inside electrical controls, wireless remote controller, aluminum wheel modifications, guide wheel grease guards
15	Total Installed Weight	1,000 lbs

4.2.2 Data Collection

The automated track monitoring system consists of Android phones placed in a hi-rail vehicle, as shown in Figure 4.2 b), c), and d). The accelerometers of the phones will generate directional accelerations at a rate of 400 measurements per second. In addition, the phones will generate GPS location information every second. These raw measurements must be wirelessly transmitted from the phones. They are then filtered, archived, and processed and used to generate useful indications, such as yaw, pitch, and roll.

Three Android phones are installed in a hi-rail vehicle for a testing period of three months. One of the phones is secured to the dash, and the other two phones are placed in the rear interior of the vehicle under a seat. There is power access to all three phones. The goal of the calibration project is to collect data from more than 100 traversals over the same track segments. The primary areas of interest are locations where known geometry issues exist, i.e., places that have been observed by the track inspector. Ideally, measurements of warp, profile, and alignment taken by the inspector at these locations can be compared to measurements predicted by the automated in-vehicle system. Ideally, the hi-rail vehicle's route will include some areas of acceptable geometry so that the calibration process will minimize the possibility of false positive findings.

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 publicly shared. Specific location information will also be kept confidential.

4.2.3 Data Fusion

The GPS to record the locations of detected track anomalies is the main location information source for the research; however, several factors limit GPS accuracy: 1) the speed of radio frequency signals is not constant, 2) multipath fading causes path distance errors, and 3) atomic clock discrepancies and receiver noise will also result in GPS errors. Moreover, the relatively slow update rate of GPS receivers relative to the inertial sensors leads to a loss of position resolution. Those factors will cause sensor signal alignment errors. Thus, the signaling data from various sensors, accelerometers, gyroscopes, and GPS should be fused or aligned to increase the position resolution.

On-board sensors allow for the combining of features extracted from inertial sensor signals, across multiple train traversals, to significantly enhance their signal-to-noise (SNR) ratio by ensemble averaging. False positives and false negatives decline as SNR increases. On the other hand, variable train speeds, GPS position registration errors, slow GPS update rates, and the uneven sampling rate of inertial sensors result in the misalignment of extracted features. Hence, subsequent ensemble averaging can actually degrade the SNR.

Many researchers assume there is a given known signal feature (also known as the first major valley or FMV), and estimating the sample position of missing GPS updates by interpolating their relative position to the FMV using time and speed data (Bridgelall et al., 2016). Such assumption of known ground anomaly location is not true for position identification purpose. In this research, a statistical based method named left-of-centroid-out-equal-length (LCOEL) is proposed to align the data and method performance is compared with well-established data fusion method assuming known ground anomaly location (Bridgelall et al., 2016). The same data used in the literature are used to test on the proposed method.

A centroid at the center of the maximum overlapping segments is identified and subsequently creates a centerline that bisects the centroid such that it is perpendicular to the analysis line. The centroid position is given as:

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} , \bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{1}$$

where xi and yi are the coordinates for GPS block i and n is the total number of GPS blocks. The GPS block for all traversals that falls on the left of the center line are recorded. Subsequently, those GPS blocks for each traversal are spatially joined to the center line to obtain a distance. Distances are interpolated using time and speed from the left of the GPS block to match the previously obtained distance on the center line. A new zero position is then introduced at that matched distance. Distances are interpolated to the left and right of the new zero position. All traversals are truncated based on the recorded shortest distance to the left and right. The first sample of all truncated traversals is marked as the new starting zero position. Starting from the new zero position, the distance is interpolated to the new ending position. Distances to the FMV are recorded for all traversals.

Table 4.2 summarizes the statistics of the reference method and LCOEL method. The statistics are the standard deviation of the FMV (STD FMV), the standard deviation of the truncated segment lengths (STD Length), the margin-of-error (MOE) in the 95% confidence interval of the FMV (MOE FMV), the MOE in the 95% confidence interval of the segment length (MOE FMV), and the skewness of the distribution for the FMV (Skewness). Table 4.2 also includes the p-values of three different tests for normality of the statistical distributions of the FMV. MOE is defined in equation (2).

$$MOE = \frac{Z\sigma}{\mu\sqrt{n}} \tag{2}$$

where Z is the z-score associated with a 95% confidence interval, and having the value of 1.96. The mean and standard deviation of the interpolated distance is μ and σ, respectively. The number of samples in the inertial signal data is n.

Table 4.2 Summary Statistics of Each Method

| Method | STD FMV | STD Length | MOE FMV | MOE Length | Skewness | P-Value | | |
						Kolmogorov-Smirnov	Cramer-von Mises	Anderson-Darling
Reference	0.09	0.09	0.024	0.024	-0.15859	>0.150	>0.250	>0.250
LCOEL	4.03	0.13	1.08	0.03	-0.3347	>0.150	0.128	0.149

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

$$F_n(x) = \begin{cases} 0, & x < X_{(1)} \\ i/n, X_i \le x \le X_{(i+1)}, & i = 1,\dots,n-1 \\ 1, & X_{(n)} \le x \end{cases} \tag{3}$$

Note that Fn(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, Fn(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 Fn(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.

Based on the summary statistics in Table 4.2, one can see that the reference method outperforms the proposed method with all statistics, including skewness. However, our research result is very promising with all the proposed algorithms able to produce comparable alignment measurements, considering our method is not based on any known ground anomaly location.

The null hypothesis of the normality test was 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 LCOEL method fail to reject the hypothesis for all three tests. This result indicates those methods maximally agree with a Gaussian distribution. Since both the skewness and normality tests agree, this research selects the LCOEL method for signal alignment.

5. TRACK SURFACE UNEVENNESS IDENTIFICATION METHODOLOGIES

5.1 Methodologies

The developed method processes the GPS tagged inertial data collected from smartphones aboard a hi-rail vehicle to estimate the positions of track irregularity, and to summarize the accuracy of their position estimates. The method incorporates a signal processing technique that reduces the inertial signals errors associated with the warp to features called road impact factors (RIFs), which the author has previously used to monitor pavement roughness (Bridgelall, 2014). The method uses a geographical information system (GIS) platform for the visualizing peak inertial events (PIEs), which are RIF values in the high 5 percentiles of its distribution.

Rail track surface unevenness and irregularities, such as damaged rail, flattened rail, track alignment, profile, and warp, result in roughness when the train load traverses them at some speed. The smartphone embedded three-dimensional accelerometers sense the induced roughness. A first step in data preparation is the conversion of each accelerometer value in meters/second to g-force. A second step is to remove any signal offset between GPS updates by subtracting the mean signal. A third step is to produce the resultant g-force, Gt in units of the radians per second with the following equation:

$$G_\text{t} = \sqrt{A_{xg}^2 + A_{yg}^2 + A_{zg}^2} \qquad (4)$$

The RIF-transform produces an intensity RIFRt that is proportional to the resultant rotation rate about x and y axis. The RIF-transform is defined as:

$$RIF_\text{Rt} = \sqrt{\frac{1}{L}\sum_{n=0}^{N-1}|Gt_n\,v_n|^2\,\delta tn} \qquad (5)$$

where RIFRt is the average magnitude of rotation rate per unit of distance L travelled (Bridgelall et al., 2017). The sample period δtn is time changes between two inertial samples. A speed sensor produces the instantaneous traversal speed v_n at sample n for N total samples.

As shown in Figure 5.1, the RIFRt feature values follow a normal distribution, with a truncated left tail. After data cleaning, the RIFRt distribution provided a means to visualize the PIEs; if the RIF values are within upper 2.5 percentile, it indicates very high likelihood of severe track surface irregularity, and if the values are within the upper 1 to 2 standard deviation from the mean, it indicates high likelihood of track surface irregularity.
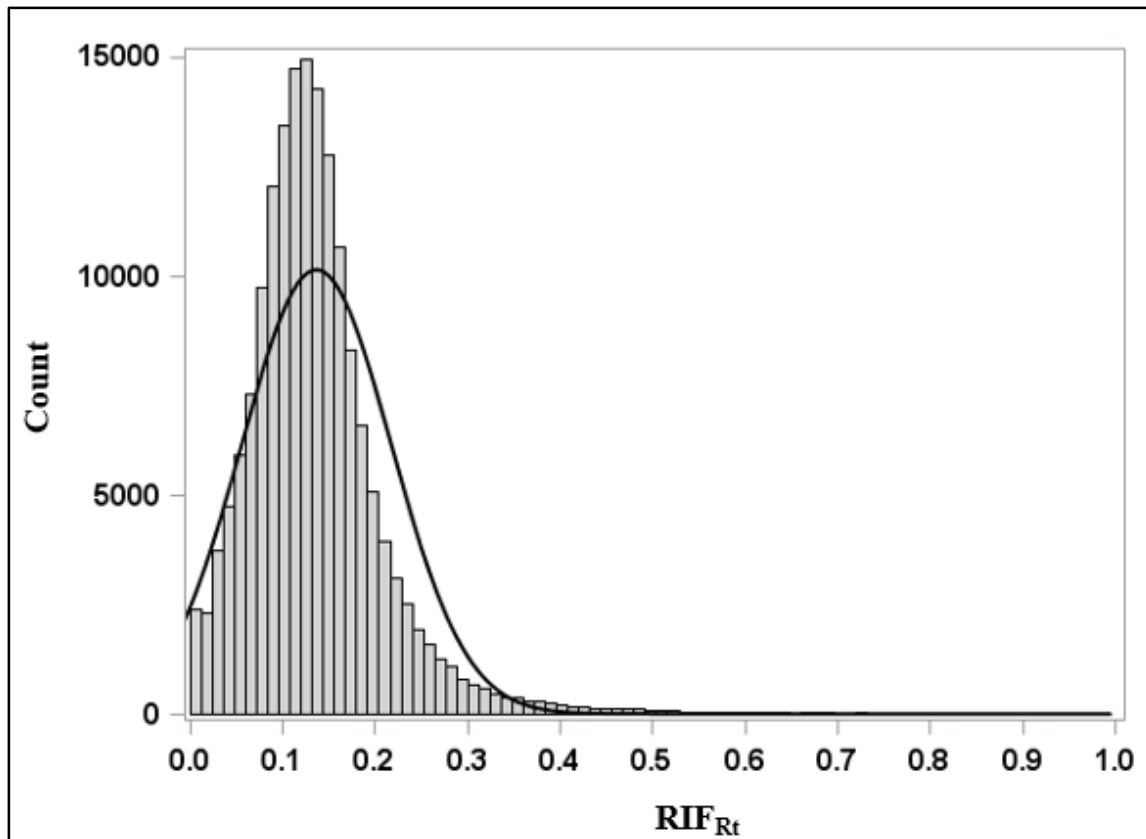
Figure 5.1 Distribution of RIF_{Rt}

The data used for this study are from two traversals collected from the three smartphones installed at three different locations on the hi-rail vehicle. ArcGIS plots the intensity of the RIF features at their corresponding longitude and latitude position of each GPS position update (GPS blocks) as small circles with color-coded levels, and with red and orange described as above—green being within 1 standard deviation from mean, light blue being within the lower 1 to 2 standard deviation from mean, and black being in the lower 2.5 percentile.

Figure 5.2 shows the RIF features from the data collected by one smartphone. We can see there is one location based on our peak RIF value that corresponds to the inertial signals, indicating a very high probability of surface abnormality at that location, shown as a red triangle, while the other reported peak RIF values are coded as other colors.

To better understand our abnormality location estimation result, the authors asked the railroad maintenance engineer if he found any surface anomaly around the time and region the method detected an anomaly. Note that the authors did not release any detailed location and time to the engineer. The engineer did find one abnormality report from his work log on the same day the authors identified the anomaly. Based on the engineer's report, we plotted the location on Figure 5.2, showing it as a big bright green circle. One can tell it is very close to our estimated location, shown as a red triangle.
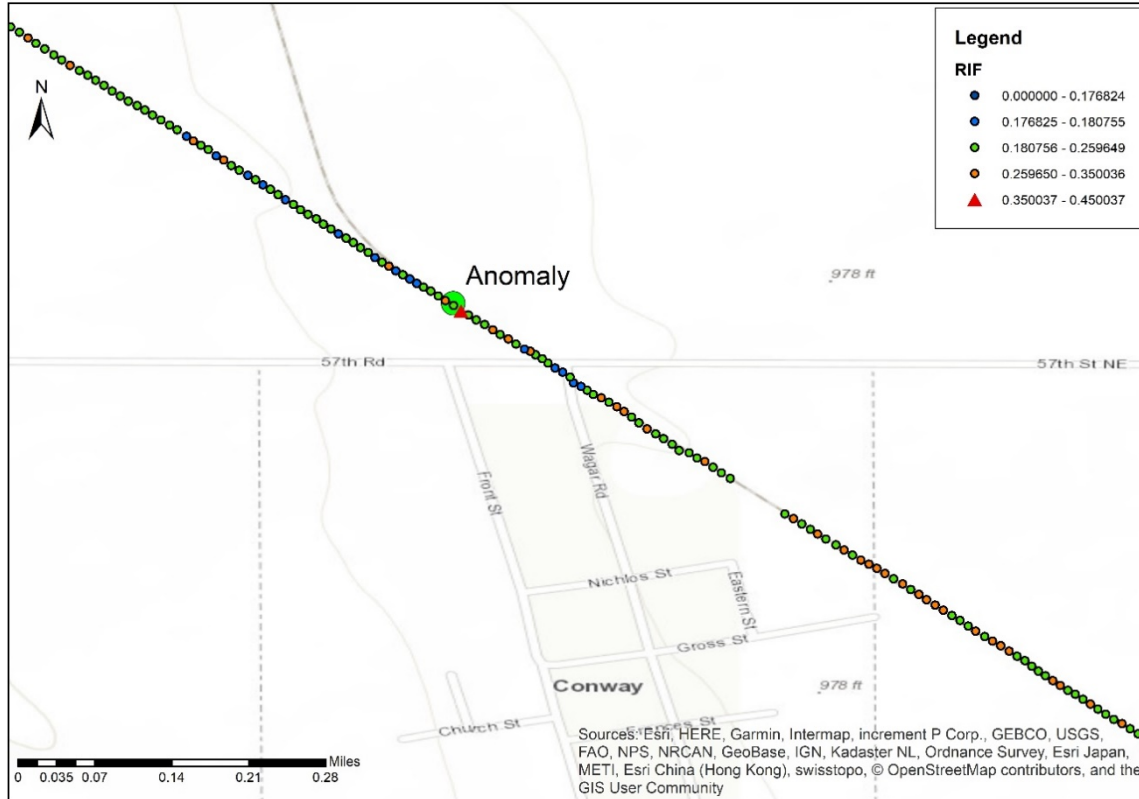
Figure 5.2 RIF_{Rt} features for each GPS block reported by one smartphone with two traversals

The actual reported location will serve as ground truth location for our method validation. The identified anomaly RIF peak values collected from all three phones and both two traversals are summarized in Table 5.1.

Table 5.1 Identified Surface Irregularity from Three Phones and Two Traversals Compared to Ground Truth

Data Resource ID	1	2	3	4	5	6
RIF Peak Values	0.268	0.290	0.292	0.397	0.451	0.488
Distance to Ground Truth Location (M)	9.59	12.38	9.87	10.07	13.59	4.89

Table 5.1 indicates that peak-RIF features are within 14 meters of the ground truth with average of 10 meters. Geometric center (centroid) for all six points is calculated to represent anomaly location estimation for various traversals and/or smartphone reports.

Table 5.2 summarizes the location estimation errors in meters for each traversal and all traversals. Figure 5.3 indicates an example for one traversal from three phones.

Table 5.2 Identified Surface Irregularity from Three Phones and Two Traversals Compared to Ground Truth

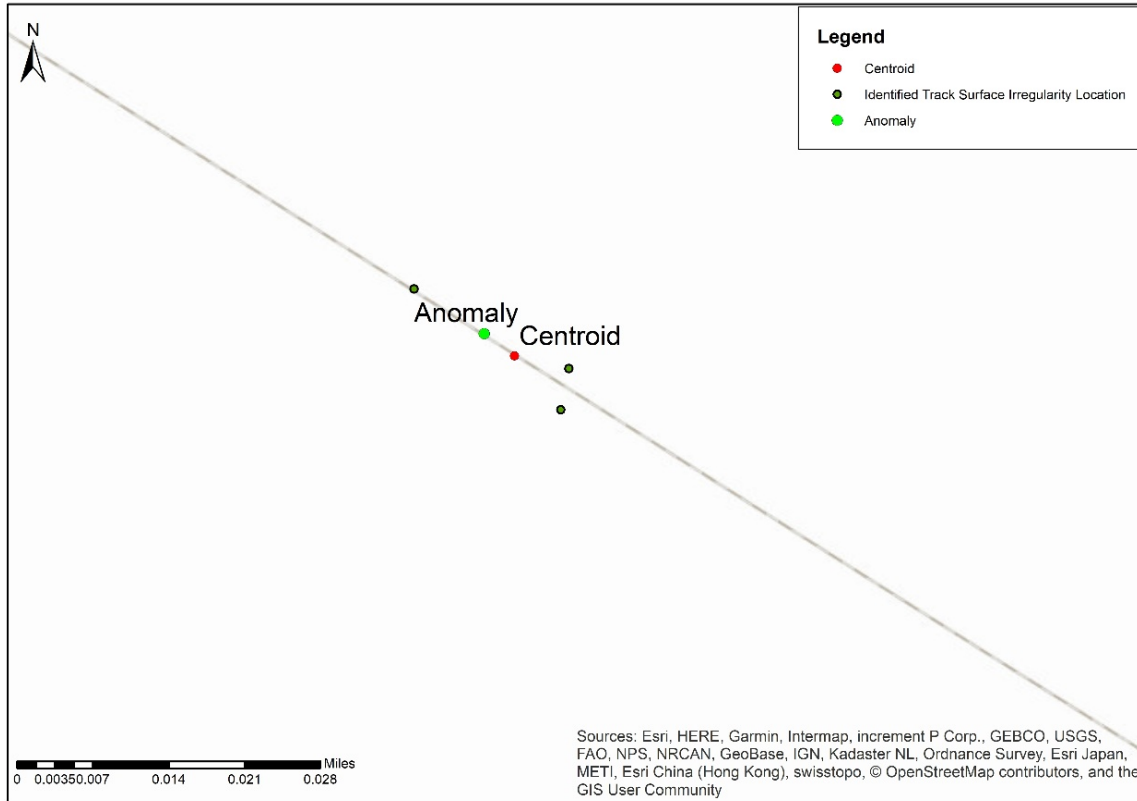| Centroids	1st traversal three-phone data	2nd traversal three-phone data	All data
Distance to Ground Truth Location (M)	4.45	1.54	2.73

29

Figure 5.3 Estimated anomaly location from three smartphones with one traversal

From Table 5.2, one can tell the three-smartphone-based system can improve location estimation error from within 15 meters to within 5 meters with the geometric center method. A centroid of the PIE clusters accumulated from a limited number of traversals (two in the case study) can provide a position of the irregularity estimate within a 3-meter error range. The offset from the true irregularity location is within a reasonable visual sight distance so it can be seen during a follow-up manual inspection.

6. SUMMARY AND FUTURE RESEARCH

6.1   Research Conclusions

It is clearly seen the proposed signal-processing algorithms can identify the location of possible rail surface irregularities. The traversal data from one smartphone will provide reasonable location estimation. With our test data, the errors are within 15 meters, which is within sight distance. This finding is critical because each single run with one phone will narrow down the potential surface anomaly within sight distance. Such a solution would free up more track time and capacity previously reserved for manual inspections while improving safety for railroad workers on duty.

The three-smartphone-based data collection system will significantly improve the location estimation errors; with one traversal, the offset can be improved to within 5 meters, and with two traversals, it can be improved to within 3 meters.

6.2   Future Research

The research introduced a signal-processing algorithm to estimate the position of peak inertial events in order to identify the positions of possible rail surface irregularities. The goal was to characterize the effectiveness of using low-cost sensors by fusing their reported signaling data from accelerometers, gyroscopic sensors, and a global position system (GPS) to enhance extracted peak features and to visualize the inertial events. The research provides very promising surface anomaly identification results to show the feasibility of the proposed system and methods; however, future work will be beneficial.

- Sensitivity analysis on centroid methods to PIE clusters from a larger number of traversals to evaluate the optimal number of traversals needed to provide perfect location estimation is needed when more data become available.
- Statistical and data mining algorithms should be developed to transfer the detected inertial events into the actual measure of warp, profile, and alignment when more data and reported measurements become available.

7. REFERENCES:

United States Department of Agriculture, "Grain Transportation Report," USDA, 2013.

Federal Railroad Administration, "Track Inspector Rail Defect Reference Manual," Federal Railroad Administration: Office of Railroad Safety, Washington, DC, 2011.

C. Li, S. Luo, C. Cole, and M. Spiryagin, "An Overview: Modern Techniques for Railway Vehicle On-Board Health Monitoring Systems," *Vehicle System Dynamics*, Vol. 55, No. 7, 1045 - 1070, 2017.

L. Biess and M. Dick, "Development and Use of V/TI Monitors for Autonomous Track Inspection," in AREMA 2011, Minneapolis, 2011.

Gavin Astin, "Derailment Risk Model—Frequency Analysis and Scenario Development," *Det Norske Veritas* (DNV) Ltd., September 29, 2011

United States Government Accountability Office (GAO), *Rail Safety, Federal Railroad Administration Should Report on Risks to the Successful Implementation of Mandated Safety Technology*, December, 2010

Tolliver, D., J. Bitzan, *Grain Transportation in the Great Plains Region in a Post-Rationalization Environment*, UGPTI Department Publication No. 170, December 2005

Federal Railroad Administration (FRA), Railroad Safety Statistics, Annual Report (Preliminary), June 9, 2011

Besuner, P.M., D.H. Stone, M.A. DeHerrera, and K.W. Schoeneberg, Statistical Analysis of Rail Defect Data (Rail Analysis–Volume 3), AAR Chicago Technical Center, Report Number R-302, 1978

Association of American Railroads (AAP), Rail Time Indicators, February 8, 2011

Al-Nazer, L., T. Raslar, C. Patrick, J. Gertler, J. Choros, J. Gordon, B. Marquis, "Track Inspection Time Study," Federal Railroad Administration, July 2011.

National Transportation Safety Board (NTSB), Derailment of Amtrack Train No. 58, City of New Orleans, Near Flora, Mississippi, Railroad Accident Report, July 26 2005

M.Ph. Papaelias, Roberts, C., Davis, C. I., "A Review on Non-Destructive Evaluation of Rails: Sate-of-the-Art and Future Development," *Journal of Rail and Rapid Transit*, Volume 222, Issue F4, pp 367-384, 2008

Clark, M.R., M.O. Gordon, M.C. Forde, "Issues over High Speed Non-Invasive Monitoring of Railway Trackbed," *TRL*, 15 P, 2003.

M. Silvast, Nurmikolu, A., Wijanen, B., Levomaki, M., "An Inspection of Railway Ballast Quality Using Ground Penetrating Radar in Finland," *Journal of Rail and Rapid Transit*, Vol. 224, Issue F5, PP 345-351, September 2010.

Roberts, R., I. Al-Qadi, E. Tutumluer, J. Boyle, "Subsurface Evaluation of Railway Track Using Ground Penetrating Radar," FRA, April 2009

Ram M. Narayanan, Jakub J.W., Li D., Elias S.E.G., "Railroad track modulus estimation using ground penetrating radar measurements," *NDT and E International*, 37 (2), March 2004, pp. 141-151

G. Rasmussen, "Railway: Wear and Noise," Proceedings of the Institute of Acoustics 1998 Conference, Vol. 20: Part 1, Cranfield University, 1998, p. 175-80.

Carolyn Southern, Rennison, D., Kopke, U. "RailBAM® - An advanced bearing acoustic monitor: Initial operational performance results," CORE – Conference on Railway Engineering, June 20-23, 2004, Darwin, Australia.

Turner, J., C. Kube, "Longitudinal Rail Stress Using Diffuse Ultrasonic Backscatter," Proceedings of the 2010 Joint Rail Conference, April 27 – 19, 2010.

Allen Phillips, "Under Close Inspection: Railroads Scrutinize Flaw Detection and Track Geometry Equipment in Search of More Accurate, Faster Ways to Inspect Rail," *Progressive Railroading*, Volume 49, Issue 2, 2006, pp 20-25.

Greg Garcia, Kalay, S., Elbert, T., "Automatic Ultrasonic Inspection Detects Cracks in Joint Bars," *Railway Track and Structures*, Volume 107, Issue 4, pp 16-18, 2011.

O. Zahran, Shihab, S., Al-Nuaimy, W., "Recent Developments in Ultrasonic Techniques for Rail-track Inspection," NDT 2002, Proceedings of the Annual Conference of the British Institute of Non-Destructive Testing, Southport, September 17-19, 2002, p. 55-60.

Selig, E.T., D. Li, "Track Modulus: Its meaning and Factors Influencing It," *Transportation Research Record* 1470, pp. 47-54.

J. Huille, Hunt, G. A., "Track stiffness: a tool for maintenance?" Proceedings Of The International Conference, Railway Engineering 2000, London, U.K., July 2000.

Steven V. Sawadisavi, Edwards, J. R., Resendiz, E., Hart, J. M., Barkan, C. P. L., Ahuja, N., "Machine-Vision Inspection of Railroad Track," *Transportation Research Board*, November 17, 2008.

Burrow, M. P. N., A. H. C. Chan, A. Shein, "Deflectometer-based analysis of ballasted railway tracks," Institution of Civil Engineers. *Proceedings of Geotechnical Engineering*, Volume 160, Issue 3, 01 July 2007, pp. 169 –177.

M. Kobayashi, Naganuma, Y., Nakagawa, M., Okumura, T., "Digital Inertial Algorithm for Recording Track Geometry on Commercial Shinkansen Trains," Eleventh International Conference on Computer System Design and Operation in the Railway and Other Transit Systems (COMPRAIL08), pp 683-692, 2008.

P. Weston, Roberts, C., Goodman, C. J., Goodall, R. M., Li, P., and Ling, C. S., "Enhanced rail contribution by increased reliability (ERCIR) – instrumenting in-service rail vehicle to monitor vehicle and track," Proceedings of the World Congress on Railway research, WCRR2003, Edinburgh, Scotland, September 28 – October 1, 2003.

Stevens, J., "Autonomous VTI Monitoring Systems," Proceedings of the AREMA Technical Conference, Chicago, IL, September 20-23, 2009.

FRA, "Developed Wheel and Axle Assembly Monitoring System to Improve Passenger Train Safety," *USDOT FRA Research Results*, March 2000.

Bob Tuzik, "Bad Actors aren't always Bad," *Railway Age*, Volume 206, Issue 7, 2005, pp 34-35.

R. Bridgelall, Y. Huang, Z. Zhang, and F. Deng, "Precision enhancement of pavement roughness localization with connected vehicles," *Meas. Sci. Technol.*, vol. 27, no. 2, p. 025012, 2016

Bridgelall, R. (2014). "Connected Vehicle Approach for Pavement Roughness Evaluation," *Journal of Infrastructure Systems*, 20(1), 04013001.

Bridgelall, R., Hough, J., and Tolliver, D. (2017). "Characterising pavement roughness at non-uniform speeds using connected vehicles." *International Journal of Pavement Engineering*, Taylor & Francis, 8436, 1–7.

Leonard Chia, Bhavana Bhardwaj, Pan Lu, and Raj Bridgelall. "Railroad Track Condition Monitoring Using Inertial Sensors and Digital Signal Processing: A Review," *IEEE Sensors Journal*, Volume 19, Issue 1, P 25:33, 2019.