

**TRANSPORTATION SYSTEM PERFORMANCE MEASURES
USING INTERNET OF THINGS DATA**

by

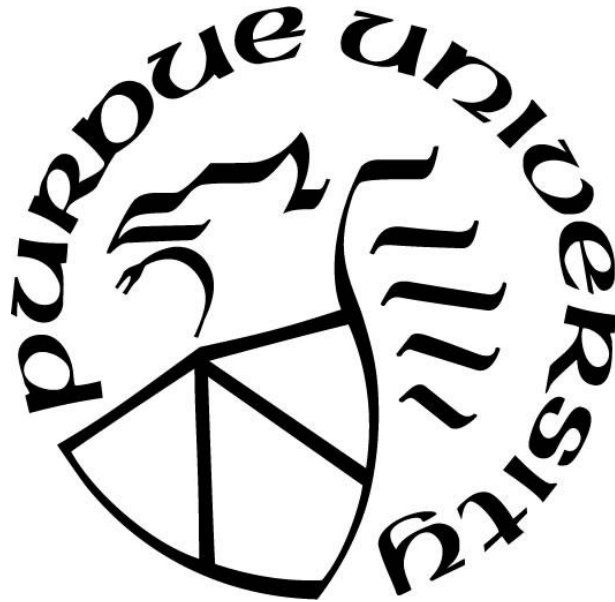
Mathew Jijo Kulathintekizhakethil

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Department of Civil Engineering

West Lafayette, Indiana

May 2018

**THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL**

Dr. Darcy Bullock, Co - Chair

Department of Civil Engineering

Dr. Sarah Hubbard, Co - Chair

Department of Aviation and Transportation Technology

Dr. Jon Fricker

Department of Civil Engineering

Dr. Andrew Tarko

Department of Civil Engineering

Approved by:

Dr. Dulcy Abraham

Head of the Departmental Graduate Program

ACKNOWLEDGMENTS

I am extremely grateful to my advisors Dr. Darcy Bullock and Dr. Sarah Hubbard for their continuous support, motivation and encouragement throughout my Ph.D. research. They have inspired me to give my best every day and I am truly indebted to them in my growth as a researcher. Their guidance and immense knowledge have always helped me during the course of my research and the writing of this dissertation.

Besides my advisors, I would like to thank my doctoral committee members, Dr. Andrew Tarko and Dr. Jon Fricker for their insightful comments and suggestions that helped me complete this research. A special mention to Dr. James Krogmeier for his support and words of wisdom during the rumble strip project.

My sincere thanks to Chris Day, Howell Li and Steve Lavrenz, who shared their expertise on a wide array of topics, which helped me, improve my technical and software skills. I also thank the valuable contributions from Wesley Major, Andrew Balmos, Drake Krohn and Maggie McNamara in some of the studies discussed in this dissertation. Thanks to Aaron Madrid for sharing the bike data and Joerg Wolf from Volkswagen ERL for helping us with the connected vehicle research.

Special thanks to all the past and current members of the Traffic Lab, including Steve Remias, Lucy Richardson, Michelle Mekker, Wayne Bunnell, Lou Rymarscuk and Greg Wisecaver for hanging out as a team both inside and outside research hours.

This research would have been impossible without the enormous support from the JTRP and INDOT. Big thanks to Debbie Horton, Bridget Brunton, Kym Pelfree, Katie Hendyrx, Dana Plattner and Tim Wells for all the help.

My Purdue family has been my pillar of strength over these four years. A huge shout-out to my friends Anup, Aswathy, Femi, Cibin, Krishna, Vishnu, Lakshmy and Rahul for being with me during my ups and downs.

I also express my profound gratitude to my parents for their unconditional love and support. Last, but not the least, I thank God Almighty for all his blessings throughout my life.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
ABSTRACT	xiii
CHAPTER 1. INTRODUCTION	1
1.1 Motivation	2
1.2 Scope and Organization	3
CHAPTER 2. PERFORMANCE MEASURES FOR EVALUATION OF RUMBLE STRIPS	4
2.1 Evaluation of Sinusoidal Rumble Strips in the Roadway Environment	4
2.1.1 Introduction.....	4
2.1.2 Motivation and scope.....	6
2.1.3 Methodology and data collection.....	7
2.1.4 Results and analysis.....	13
2.1.5 Conclusions.....	22
2.2 Evaluation of Aviation Rumble Strips	23
2.2.1 Introduction.....	23
2.2.2 Motivation and scope.....	24
2.2.3 Methodology and data collection.....	25
2.2.4 Results.....	30
2.2.5 Conclusions.....	36
2.3 Summary	36
CHAPTER 3. ARTERIAL PERFORMANCE MEASURES USING COMMERCIAL PROBE VEHICLE DATA	38
3.1 Introduction	38
3.2 Motivation and Scope.....	39
3.3 Methodology and Data Collection	39
3.3.1 Data.....	39
3.3.2 Methodology.....	40
3.4 Results	52

3.4.1	Before-and-after assessments	52
3.4.2	Incidents.....	63
3.5	Summary	64
CHAPTER 4. PERFORMANCE MEASURES FOR AIRPORT SECURITY		
SCREENING		66
4.1	Introduction	66
4.2	Motivation and Scope.....	68
4.3	Methodology and Data Collection	68
4.3.1	Airport.....	68
4.3.2	Data.....	68
4.3.3	Methodology.....	69
4.4	Results	69
4.4.1	Wait time	69
4.4.2	Reliability	71
4.4.3	Hours with high median wait time.....	72
4.4.4	Performance dashboards.....	74
4.5	Summary	76
CHAPTER 5. PERFORMANCE MEASURES FOR THE ASSESSMENT OF A		
UNIVERSITY BIKE SHARE SYSTEM		
5.1	Introduction	78
5.2	Motivation and Scope.....	79
5.3	Methodology and Data Collection	79
5.4	Performance Measures	81
5.4.1	User rentals	81
5.4.2	Daily user rentals	82
5.4.3	Daily rental duration	83
5.4.4	Cumulative plot of rental duration.....	84
5.4.5	Bike usage.....	85
5.4.6	Bikes in use.....	86
5.4.7	Bikes on ground.....	86
5.5	Summary	87

CHAPTER 6. DEVELOPMENT OF A CURATED DATASET TO EVALUATE THE PERFORMANCE OF OUTLIER FILTERING ALGORITHMS	89
6.1 Introduction	89
6.2 Motivation and Scope.....	90
6.3 Methodology and Data Collection	91
6.3.1 Data.....	91
6.3.2 Methodology for developing the curated dataset.....	92
6.3.3 Outlier filtering methods.....	102
6.4 Results	104
6.4.1 Comparison of outlier filtering algorithms	104
6.4.2 Performance during real-time estimation/predictions	105
6.5 Summary	107
CHAPTER 7. TRAFFIC SIGNAL PERFORMANCE MEASURES FOR CONNECTED VEHICLES	109
7.1 Introduction	109
7.2 Motivation and Scope.....	112
7.3 Methodology and Data Collection	112
7.4 Results	116
7.5 Performance Measures to Improve Prediction	120
7.5.1 Split diagrams and probability of expected clearance time	120
7.5.2 Distribution of green for a phase	125
7.6 Summary	128
CHAPTER 8. CONCLUSIONS	130
8.1 Rumble Strips	130
8.2 Arterial Probe Data Performance Measures.....	131
8.3 Airport Security Wait Times	131
8.4 Bike share Programs.....	132
8.5 Outlier Filtering Algorithms and Curated Dataset	132
8.6 Traffic Signal State Prediction	133
8.7 Summary of Transportation Modes Analyzed and Contribution	134

APPENDIX A. STATISTICAL MODELLING OF RUNWAY INCURSION OCCURRENCES IN THE UNITED STATES	136
APPENDIX B. METHODOLOGY FOR BENEFIT COST ESTIMATION	158
APPENDIX C. SAMPLE CURATED DATASET	161
GLOSSARY OF TERMS	162
REFERENCES	168

LIST OF TABLES

Table 1. Sound level comparisons with NCHRP recommendations	20
Table 2. Summary of Durability Testing at FAA William J. Hughes Technical Center ..	34
Table 3. Legend for before-and-after date ranges.....	49
Table 4. Characteristics of selected corridors	53
Table 5. Hypothesis testing of the before and after travel times for corridor A3	57
Table 6. Summary of annual CO ₂ emission reductions for the five corridors	62
Table 7. Descriptive statistics of the expedited and standard wait times.....	71
Table 8. Frequency table for median wait time on hourly basis for 16 Dec 2015.....	73
Table 9. Comparison of outlier filtering algorithms	105
Table 10. Number of false spikes reported for travel time difference above 5 minutes.	107
Table 11.Split times for all phases from high resolution data	122
Table 12. Summary of transportation areas and modes analyzed.....	134

LIST OF FIGURES

Figure 1. Traditional and sinusoidal rumble strip.....	7
Figure 2. Profile of alternative rumble strip configurations.	8
Figure 3. Sinusoidal rumble strip testbed	9
Figure 4. Test vehicles	10
Figure 5. Accelerometer installation.....	11
Figure 6. Deployment of sound meters during data collection.....	12
Figure 7. Test scenarios	13
Figure 8. Acceleration traces for test vehicles on 12” sinusoidal rumble strips.....	15
Figure 9. Average maximum RMS of accelerometer dynamic magnitude for all test vehicles	16
Figure 10. Sound power traces on centerline incursion for 12 in sinusoidal rumble.....	17
Figure 11. Sound level comparison for all vehicles on centerline rumble.....	19
Figure 12. Retroreflectivity data collection on edge and center line	20
Figure 13. Close-up view of thermoplastic marking on sinusoidal rumble strips	21
Figure 14. CFDs of retroreflectivity readings on edge line rumble strips	21
Figure 15. CFDs of retroreflectivity readings on centerline rumble strips.....	22
Figure 16. Different types of rumble strips.....	26
Figure 17. Rumble strip test bed from 2015 at KLAF.....	27
Figure 18. Aircraft tested	28
Figure 19. Rail mount configuration.....	29
Figure 20. Seat and rail mount configuration on a Cessna 172	29
Figure 21. FAA airfield heavy vehicle simulator	30
Figure 22. Acceleration plots for Cessna 172 at 5 knots	31
Figure 23. Peak-to-peak acceleration ranges on rail.....	32
Figure 24. IQR plots for Cessna 172 at 5 knots.....	33
Figure 25. Rail-to-seat ratio comparison for Cessna 172 at 5 knots.....	33
Figure 26. Edge raveling and crumbling on saw cut	34
Figure 27. Sweeper and blower test on temporary rumble strips	35
Figure 28. Cracked thermoplastic rumble strip after winter operations	35

Figure 29. Overview of the real-time data ingestion process	40
Figure 30. Travel time comparison tool.....	41
Figure 31. Travel times on a corridor during a 24-hour period	42
Figure 32. Plotting CFDs from timing plan 4.....	42
Figure 33. CFDs comparing before and after retiming.....	43
Figure 34. Median travel time improvement of 1 minute after signal retiming	43
Figure 35. Multi-criteria arterial ranking tool.....	44
Figure 36. Median travel time and speed limit travel time on Newtown Bypass and US-1 during a period of one week.....	45
Figure 37. Normalized travel time on Newtown Bypass and US-1.....	46
Figure 38. Ranking the corridors using the median normalized speed limit travel time ..	47
Figure 39. 25 th and 75 th percentiles.....	47
Figure 40. Ranking the corridors based on the IQR	48
Figure 41. Speed limit and IQR normalization for one date range.....	49
Figure 42. Speed limit and IQR normalization for before and after date ranges.....	50
Figure 43. Arterial congestion ticker	51
Figure 44. A 15-minute time slice of the graph generated by the arterial congestion ticker and its respective speed breakdown over the length of the corridor.....	51
Figure 45. Travel delay monitor graph over a 1-week period	52
Figure 46. Selected corridors in the Greater Philadelphia area	53
Figure 47. Corridor maps of intersections running adaptive control	54
Figure 48. Travel times before and after installation of adaptive signal control on corridor A3 on weekdays between 17:00 and 18:00.....	56
Figure 49. Arterial ranking tool showing the before-and-after travel time and reliability trends for the selected corridors on weekdays 17:00 and 18:00.....	58
Figure 50. Weekday median travel times and IQR ranges by hour	60
Figure 51. Weekend median travel times and IQR ranges by hour	61
Figure 52. Summary of annual user cost benefits for the five corridors.....	62
Figure 53. I-76 West accident impact and expected Gulph Rd detour	63
Figure 54. Travel time impacts on westbound Gulph Rd	64
Figure 55. Congestion heat map of westbound Gulph Rd	64

Figure 56. Median wait time and checkpoint volume in 2015	70
Figure 57. CFD comparing the standard and expedited wait times	72
Figure 58. Median wait time for each hour on 16 Dec 2015	73
Figure 59. Proportion of checkpoint wait time for the calendar year 2015	74
Figure 60. Average hours of median wait time at standard checkpoint by month	75
Figure 61. Weekly summary of average hours of median wait time	75
Figure 62. Hours of median wait time for week 51 in December 2015.....	76
Figure 63. Campus map with location of bike share docking stations	80
Figure 64. Variation in pricing scheme over the study period.....	81
Figure 65. Rentals by user	82
Figure 66. Total daily rentals	82
Figure 67. Median temperatures during spring 2016 and 2017	83
Figure 68. Daily rental duration.....	84
Figure 69. Cumulative plot of rental duration before policy changes	84
Figure 70. Average bikes in use by 15-min periods	85
Figure 71. Pareto chart of rental duration by bike	86
Figure 72. Cumulative days unused by bike.....	87
Figure 73. Over 150,000 waypoint data from probes for a day on I-94, IN	92
Figure 74. Northern Indiana I80/I94 corridor	94
Figure 75. Estimation of travel time by matching trips in each zone	95
Figure 76. Raw travel time after trip matching for EB and WB directions	96
Figure 77. Missed checkpoint and waypoint outside roadway (WOR) trips.....	100
Figure 78. Non-outlier trips with high travel time	101
Figure 79. Curated dataset after removing missing checkpoint and indeterminate trips	102
Figure 80. Graphical representation of box and whisker plot.....	104
Figure 81. Performance of outlier filtering algorithms compared to ideally-filtered (IF) and unfiltered (UF) data.....	106
Figure 82. DSRC mode of communication for connected vehicles.....	110
Figure 83. Cellular mode of communication for connected vehicles	110
Figure 84. Dashboard with countdown timer for the next green	111
Figure 85. US 231 & River Road intersection diagram.....	113

Figure 86. Video camera setup for data collection	114
Figure 87. Extracting timestamps for comparison.....	115
Figure 88. Residual plot comparing prediction and actual	117
Figure 89. Late prediction.....	118
Figure 90. Early prediction	119
Figure 91. State change event diagram for one cycle	121
Figure 92. Split diagram for ring 1	122
Figure 93. Probability of split for each phase	123
Figure 94. Fixed time with clearance.....	124
Figure 95. Actuated coordination with clearance	125
Figure 96. Green distribution for phase 2 (northbound through) movement.....	126
Figure 97. Green distribution on all phases	127

ABSTRACT

Author: Kulathintekizhakethil, Mathew Jijo Ph.D.

Institution: Purdue University

Degree Received: May 2018

Title: Transportation System Performance Measures Using Internet of Things Data

Co-Advisor: Darcy Bullock

Co-Advisor: Sarah Hubbard

The transportation system is undergoing a rapid change with innovative and promising technologies that provide real-time data for a variety of applications. As we transition into a technology-driven era and Internet of Things (IoT) applications, where everything is connected via a network of smart sensors and cloud computing, there will be an increasing amount of real-time data that will allow a better understanding of the transportation system. Devices emerging as a part of this connected environment can provide new and valuable data sources in a variety of transportation areas including safety, mobility, operations and intelligent transportation systems. Agencies and transportation professionals require effective performance measures and visualization tools to mine this big data to make design, operation, maintenance and investment decisions to improve the overall system performance. This dissertation discusses the development and demonstration of performance measures that leverage data from these emerging IoT devices to support analysis and guide investment decisions. Selected case studies are presented that demonstrate the impact of these new data sources on design, operation, and maintenance decisions.

Performance measures such as vibration, noise levels and retroreflectivity were used to conduct a comprehensive assessment of different rumble strip configurations in the roadway and aviation environment. The results indicated that the 12 in sinusoidal wavelength satisfied the National Cooperative Highway Research Program (NCHRP) recommendations and reduced the noise exposure to adjacent homeowners.

The application of low-cost rumble strips to mitigate runway incursions at general aviation airports was evaluated using the accelerations on the airframe. Although aircraft are designed for significant g-forces on landing, the results of analyzing accelerometers

installed on airframes showed that long-term deployment of rumble strips is a concern for aircraft manufacturers as repeated traversal on the rumble strips may lead to excessive airframe fatigue.

A suite of web dashboards and performance measures were developed to evaluate the impact of signal upgrades, signal retiming and maintenance activities on 138 arterials in the Commonwealth of Pennsylvania. For five corridors analyzed before and after an upgrade, the study found a reduction of 1.2 million veh-hours of delay, 10,000 tons of CO₂ and an economic benefit of \$32 million.

Several billion dollars per year is expended upon security checkpoint screening at airports. Using wait time data from consumer electronic devices over a one-year period, performance dashboards identified periods of the day with high median wait times. The performance measures outlined in this study provided scalable techniques to analyze operating irregularities and identify opportunities for improving service. Reliability and median wait times were also used as performance measures to compare the standard and expedited security screening. The results found that the expedited screening was highly reliable than the standard screening and had a median wait time savings of 5.5 minutes.

Bike sharing programs are an eco-friendly mode of transportation gaining immense popularity all over the world. Several performance measures are discussed which analyze the usage patterns, user behaviors and effect of weather on a bike sharing program initiated at Purdue University. Of the 1626 registered users, nearly 20% of them had at least one rental and around 6% had more than 100 rentals, with four of them being greater than 500 rentals. Bikes were rented at all hours of the day, but usage peaked between 11:00 and 19:00 on average. On a yearly basis, the rentals peaked in the fall semester, especially during September, but fell off in October and November with colder weather. Preliminary results from the study also identified some operating anomalies, which allowed the stakeholders to implement appropriate policy revisions.

There are a number of outlier filtering algorithms proposed in the literature, however, their performance has never been evaluated. A curated travel time dataset was developed from real-world data, and consisted of 31,621 data points with 243 confirmed outliers. This dataset was used to evaluate the efficiency of three common outlier filtering algorithms, median absolute deviation, modified z-score and, box and whisker plots. The

modified Z-score had the best performance with successful removal of 70% of the confirmed outliers and incorrect removal of only 5% of the true samples.

The accuracy of vehicle to infrastructure (V2I) communication is an important metric for connected vehicle applications. Traffic signal state indication is an early development in the V2I communication that allows connected vehicles to display the current traffic signal status on the driver dashboard as the vehicle approaches an intersection. The study evaluated the accuracy of this prediction with on-field data and results showed a degraded performance during phase omits and force-offs. Performance measures such as, the probability of expected phase splits and the probability of expected green for a phase, are discussed to enhance the accuracy of the prediction algorithm. These measures account for the stochastic variations due to detectors actuations and will allow manufacturers and vendors to improve their algorithm.

The application of these performance measures across three transportation modes and the transportation focus areas of safety, mobility and operations will provide a framework for agencies and transportation professionals to assess the performance of system components and support investment decisions.

CHAPTER 1. INTRODUCTION

A safe, efficient and sustainable transportation system is essential for a city's economic competitiveness and quality of life. The primary role of a transportation system is to move people and goods in a safe, integrated, reliable and efficient system that effectively leverages and connects all modes of transport. A number of urban traffic solutions including Intelligent Transportation Systems (ITS) play an eminent role in mitigating the various transportation problems faced by cities. In the absence of widely accepted performance measures, it is challenging to evaluate the effects of specific technologies and their applications in the transportation system.

Performance measures provide quantitative metrics to determine the progress towards defined organizational tasks and objectives. State Department of Transportation (DOTs) are increasingly using performance measures to help guide resource allocation decisions at the program level [1] [2]. By monitoring the key performance measures of a system, state and local officials can develop strategies and actions for improvement. Information about system performance at a national level can ensure decisions to improve the system.

Early performance measures to evaluate the congestion suggested by the Highway Capacity Manual (HCM) include volume to capacity (V/C) ratio, average intersection delay and level of service (LOS) [3]. Travel time, speed and delay were also suggested as a common measure of effectiveness [4], [5]. The Highway Safety Manual also reports a number of performance measures such as crash frequencies, crash modification factors (CMF) and safety performance functions (SPF) to evaluate quantitative safety analyses [6]. According to a performance report by the US Department of Transportation (USDOT), the most common performance measure concerning safety are the fatalities per 100 million vehicle-mile of travel and the number of accidents per 100 million vehicle-miles of travel [7]. Studies on performance measures have also addressed signal systems [8], intermodal transportation [9] and multimodal transportation systems [10]. In recent years, highway monitoring and performance measure requirements have also been increasingly emphasized for federal transportation funding mandates such as MAP-21 [11].

In the past decade, technology has evolved at a tremendous pace, paving the way for new and innovative programs in the field of transportation. With the advent of technology and ease of access, bike sharing programs and ride-hailing services have had a huge impact on the transportation system. Vehicles, pedestrians, bikes, public transit, and infrastructure are connected through a series of smart sensors, which provide data to improve performance and assessment of the transportation system. In 2011, the USDOT announced the “Connected Vehicle Program” that allows vehicles to communicate with other vehicles (V2V), infrastructure (V2I) and other devices (V2X) to improve safety, mobility and reduce environmental impacts like fuel consumption and emissions [12], [13]. The USDOT also launched the Smart City Challenge in 2015 that integrates big data applications and Internet of Things (IoT), where everything is connected via a network of smart sensors, to develop an intelligent and connected transportation system which moves people and goods quickly, cheaply and efficiently [14].

This research focuses on developing performance measures using data from both proven and new transportation technologies. Evaluating the performance using data from these technologies will help transportation officials make better decisions and improve the nation’s transportation system.

1.1 Motivation

As we transition into a technology-driven era and IoT applications, the devices connected to this network can provide new and valuable data sources that can help us better understand the transportation system. Agencies and transportation professionals require effective performance measures and visualization tools to mine this big data to make design, operation, maintenance and investment decisions to improve the overall system performance. It is also necessary to have a robust system that provides a good science on the development of these IoT performance measures. This dissertation discusses the development and demonstration of performance measures that leverage data from these emerging IoT devices to support analysis and guide investment decisions in the areas of transportation mobility, safety and operations. Performance measures are proposed for various studies within these areas, which will assist federal and state

agencies as well as manufacturers, vendors and other researchers to assess the performance of their system and identify opportunities to enhance them.

1.2 Scope and Organization

The scope of this research is the development of new and existing performance measures from IoT data that can be used to guide policy-making and investment decisions as well as provide new research opportunities in the following case studies:

- To evaluate the application of sinusoidal rumble strips as an alternative to standard rumble strips in reducing the noise impact at residential areas (chapter 2).
- To examine the feasibility of aviation rumble strips as a potential mitigation measure to enhance pilot situational awareness and reduce runway incursions (chapter 2).
- To develop a suite of web dashboards to assess the mobility impacts of signal modernization, traffic diversions and maintenance activities on arterials as well as to prioritize resources based on the arterial performance (chapter 3).
- To analyze operating irregularities at airport security checkpoints and identify opportunities to examine resource allocations for improving service (chapter 4).
- To monitor the progress of a bike share program and evaluate possibilities to improve operations, policies and asset management (chapter 5).
- To create a curated travel time dataset and performance measures to evaluate the efficiency of outlier filtering algorithms (chapter 6).
- To improve the accuracy of traffic signal state predictions for connected vehicle applications (chapter 7).

Finally, chapter 8 presents the summary and findings of this dissertation.

CHAPTER 2. PERFORMANCE MEASURES FOR EVALUATION OF RUMBLE STRIPS

This chapter discusses performance measures for the evaluation and application of rumble strips in the roadway and aviation environments. Some of the information presented in Section 2.1 regarding sinusoidal rumble strips in the roadway environment was published in 2017 Transportation Research Board annual meeting compendium of papers [15]; some of the information in Section 2.2 regarding aviation rumble strips was submitted to the Federal Aviation Administration [16].

2.1 Evaluation of Sinusoidal Rumble Strips in the Roadway Environment

2.1.1 Introduction

Roadway rumble strips are safety countermeasures that use tactile vibration and audible rumbling to alert inattentive drivers of potential danger. Rumble strips are typically laid in two different formats – longitudinal and transverse. Longitudinal rumble strips are useful in providing lane-departure warnings when a vehicle drifts off a lane, whereas transverse rumble strips are more useful in providing advance warnings such as in the case of a slowdown or of an approaching construction zone [17]. Typical placement of longitudinal rumble strips is along the centerline and edge line. Centerline rumble strips are effective in reducing head-on collisions and opposite-direction sideswipes, especially in the case of drivers crossing center lines of two-lane roads [18]. Shoulder or edge line rumble strips are commonly used in narrow roads to warn drivers when they drift off from their lanes. They are primarily efficient in reducing run-off-the-road crashes [19].

Rumble strips can be rolled, formed, milled and raised [20]. Rolled rumble strips are rounded, or V-shaped grooves pressed into the asphalt pavements during construction. Formed rumble strips are similar to rolled, except they are made by pressing forms into concrete shoulders. Milled rumble strips are grooves (typically 5 to 7 inches wide with a 12 in spacing and 0.5in depth) cut into the pavement by a machine with a rotary cutting head. Raised rumble strips are round or rectangular markers or thermoplastic strips

(typically 2 to 12 in wide and 0.25 to 0.5in high) which adhere to new or existing pavements. The application of raised rumble strips is limited in areas where snowplow operations are predominant during winter [21]. Milled rumble strips are found to be more common due to their ease of constructability, durability, and cost-effectiveness. Studies have also found that milled rumble strips produce more noise than rolled and formed [22]. In recent years, longitudinal rumble strips based on a new sinusoidal design have been reported to provide effective lane departure warnings to a driver with lower exterior noise.

A pilot study conducted by the Danish Road Institute found that the sinusoidal pattern led to an increase of only 0.5 to 1 decibels (dB) over normal road noise [23]. The Minnesota Department of Transportation [24] evaluated the noise on three different rumble strip types:

- *California design*: sinusoidal shaped with a flat crest (14 inch center to center, 1/32 – 5/8 inch depth and 8-inch width)
- *Pennsylvania design*: sinusoidal (24 inch center to center, 1/8 – 1/2 inch depth and 8-inch width)
- *Minnesota design*: traditional milled rumble strips (12 inch center to center, 3/8 – 1/2 inch depth and 16-inch width)

Tests were conducted using three vehicle types at speeds of 30, 45, and 60 mph. The in-cabin noise levels for the Pennsylvania design (sinusoidal) were found to be 3 to 5 dBA higher than the Minnesota design (standard milled) in the test car and 14 to 19 dBA higher in the test pick-up truck.

Rumble strips are painted with a retroreflective coating to increase the visibility of the pavement edges and centerline, at night and during adverse weather conditions. These rumble strips are known as rumble stripes [25]. Various studies have been conducted to evaluate the performance of rumble stripes. Researchers from Indiana Department of Transportation (INDOT) and Purdue University conducted a study to compare the retroreflective characteristic of rumble strips and standard painted lines [26]. The study also evaluated the durability of both after winter snowplowing operations. The results showed that rumble stripes were effective in providing increased night time visibility in dry and wet conditions, as well as increased durability after snowplow operations.

However, recent deployments have exhibited low retro-reflectivity characteristics, perhaps due to new fog seal treatment procedures.

A study conducted by Virginia DOT compared the durability of six different pavement marking technologies over a period of 23 months and found that the markings installed in grooves or rumble strips retained more reflectivity and received less damage than those on the surface of the roadway [27]. A study by the Minnesota Department of Transportation evaluated the retroreflectivity of the rumble stripes on 14 different roadways, 12 months after their installation. The results showed that more than half of the sites had 90% their retroreflective readings in excess of the arbitrary benchmark set for performance [28].

A comprehensive study of the various designs of sinusoidal patterns that affect the noise and vibrations on vehicles are yet to be found in the literature. Although studies have established rumble stripes to be effective in providing increased visibility during night time, as well as improved durability after snow plow operations, their performance on sinusoidal rumble strips has not been evaluated.

2.1.2 Motivation and scope

Studies have suggested rumble strips as an alternative to raised pavement markers (RPM's), particularly during periods of decreased visibility and/or adverse weather conditions [29]. Studies have also established that rumble strips reduce vehicle crashes by 35% to 45% [22], [29]. However, when a vehicle engages the strips, a loud exterior noise is generated in addition to the alerting in-cabin noise. The extraneous exterior noise may travel several hundred feet and be considered a nuisance by nearby residents. Studies have found that traditional milled and rolled rumble strips increase the exterior noise levels from 100 to 150 feet away from the centerline of the roadway [30]–[32]. The National Cooperative Highway Research Program (NCHRP) Report 641, Guidance for Design and Application of Rumble Strips, recommends that rumble strips should produce an in-cabin sound level increase of 10 to 15 dBA. To limit exterior noise near residential land uses, an exterior increase of 6 to 12 dBA is considered acceptable [33]. Sound propagation may vary depending on the installation method [22], width and

spacing [31], [34] speed, type of vehicle [22], [30] and environmental conditions such as air temperature, humidity and wind speed [35].

Although the sinusoidal rumble strip has been reported to produce low exterior noise while still providing adequate warnings for drivers [23], [24], there are still questions regarding the impact of the waveform parameters: wavelength, depth and amplitude on noise volume and alerting of drivers departing from their lane. The visibility of these rumble strips during night time, and inclement weather conditions are also uncertain. This study uses three performance measures to evaluate the sinusoidal rumble strips (Figure 1(b)): vibration, noise levels, and retroreflectivity. Six different vehicles were used for the vibration and noise level testing. For comparison purposes, the standard INDOT rumble strips (Figure 1(a)) are also studied.



(a) Traditional rumble strip

(b) Sinusoidal rumble strip

Figure 1. Traditional and sinusoidal rumble strip

2.1.3 Methodology and data collection

2.1.3.1 Rumble strips

The study evaluated the performance of three different sinusoidal wavelengths (12, 18, and 24 inch). Figure 2 shows the geometric construction details. The rumble strips have fixed amplitude ($3/16$ inch) and depth ($5/16$ inch).

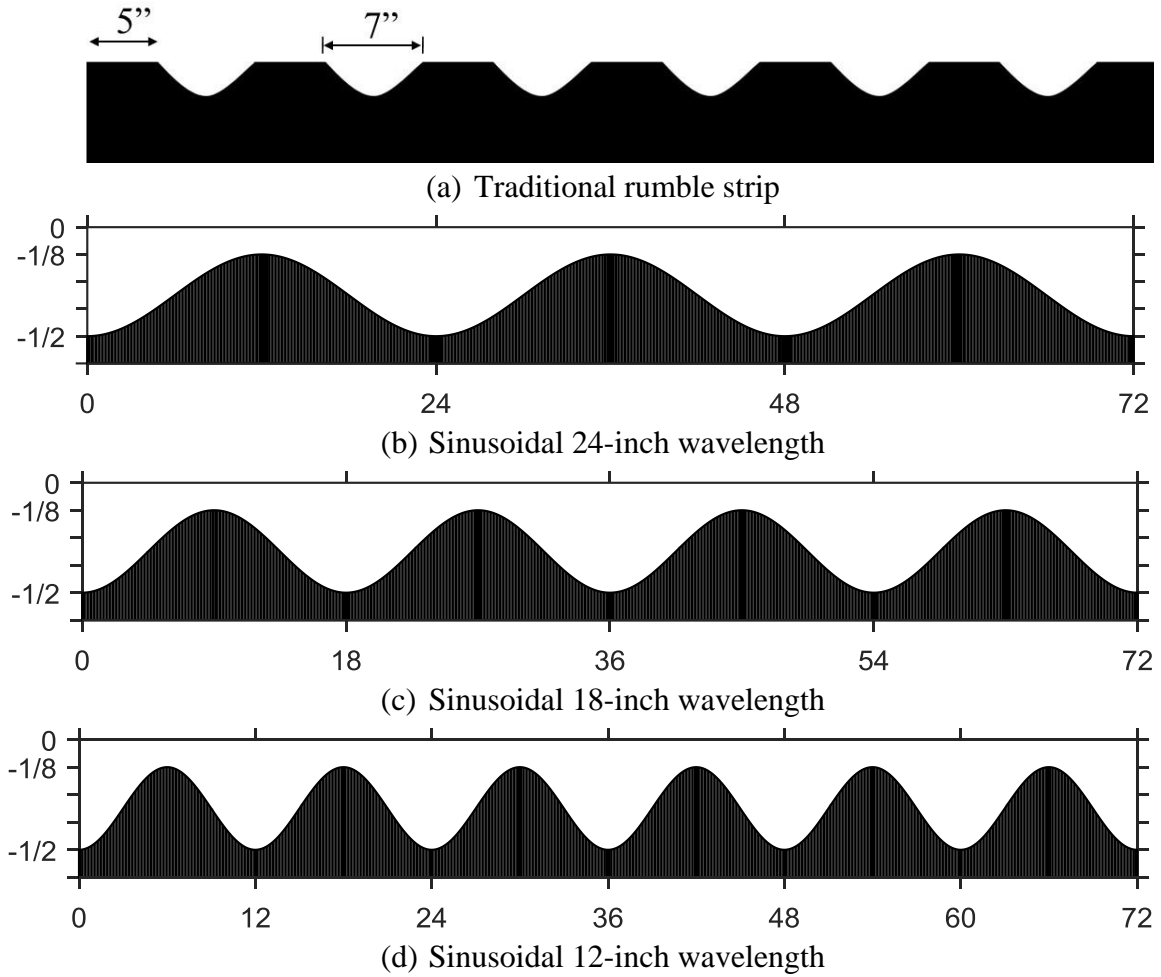


Figure 2. Profile of alternative rumble strip configurations (not to scale). The y-axis represents the depth of the rumble strips with zero being the top of the pavement.

2.1.3.2 Test location

Data collection was at three sites along IN 1, Fort Wayne, IN one for each sinusoidal configuration (Figure 3), and one site along SR 25, Shadeland, IN for the standard milled configuration. The sinusoidal test beds were approximately 2 to 4 miles long.

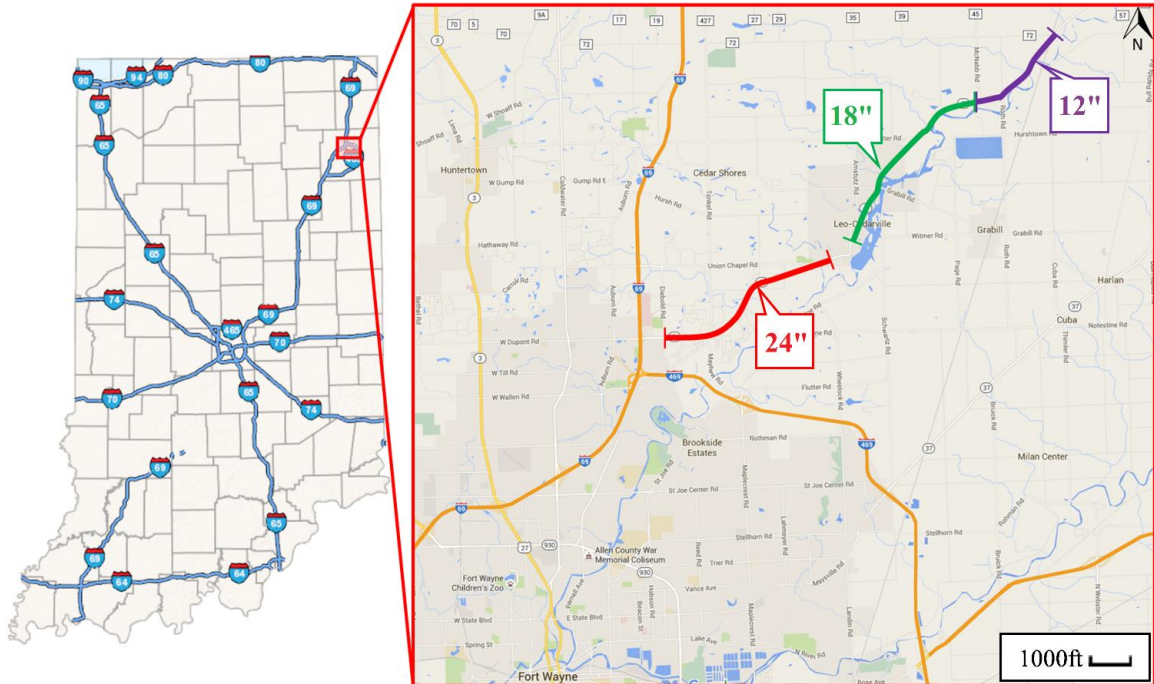


Figure 3. Sinusoidal rumble strip testbed

2.1.3.3 Test vehicles and test speed

Tests were performed with six different vehicles: semi trailer-truck, single axle truck, tandem axle truck, Ford E-150 (minivan), Chevrolet Suburban (SUV) and Chevrolet Impala (sedan) (Figure 4). For the standard milled rumble strips, only the small vehicles (minivan, Suburban, and Impala) were used for data collection. The vehicles were tested at a speed of 50 mph, the speed limit on the road. A total of three runs were performed for each test vehicle.



Figure 4. Test vehicles (a) Tandem axle (b) Single axle (c) Semi-trailer (d) Chevrolet Suburban (e) Chevrolet Impala and (f) Ford E-150 Mini Van

2.1.3.4 Test sensors

The sensors consisted of a 3-axis accelerometer and class 1 sound level meters. In particular, a GCDC X2-2 tri-axial USB accelerometer [36], with a sampling frequency of 512 Hz, was mounted on the driver side seat frame (Figure 5 (a)). The sound level meters were Larson–Davis Model 831 Type 1 units (Figure 6 (b)), with audio recording functionality. Additionally, a camcorder was used to record each test event. All the sound meters and accelerometers were calibrated and time synchronized.



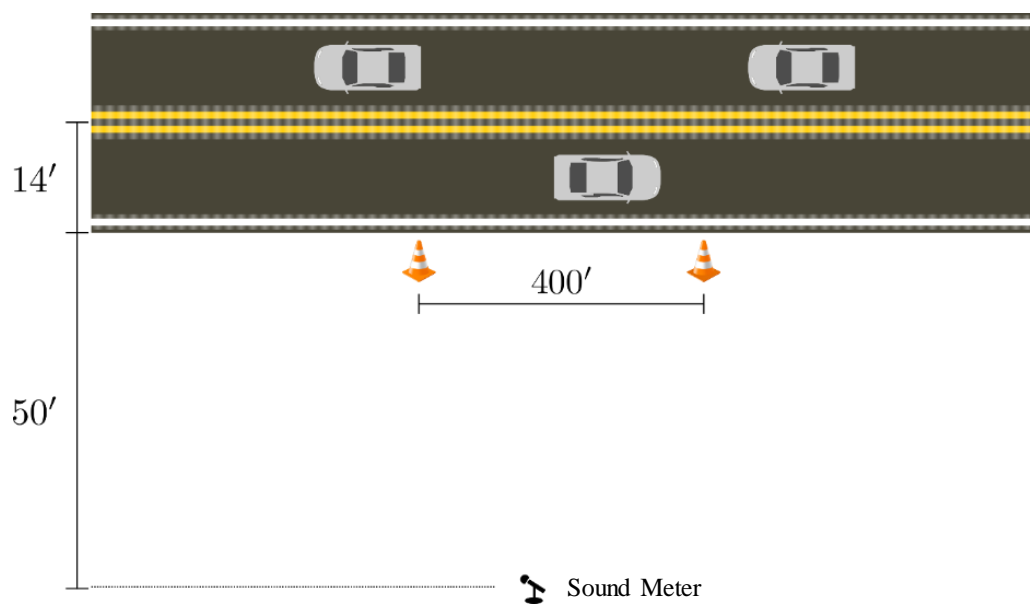
(a) Installing accelerometer on driver side seat frame



(b) Final setup

Figure 5. Accelerometer installation

Figure 6 shows the placement of the exterior and in-cabin sound level meters. The exterior sound meters were placed at a distance of 50ft from the closest edge line rumble strip and at a height of 4ft from ground level (Figure 6 (a)). The in-cabin sound meter was mounted inside the vehicle near the driver's ear as seen in Figure 6 (d) and (e). Traffic cones were placed at a distance of 200 feet on either side of outside sound meter to provide reference locations for the driver and video logs.



(a) Layout of the exterior sound meter at test site



(b) Larson-Davis Model 831 Type 1 sound meter



(c) Sound meter at distance of 50ft off the edge line and at a height of 4ft



(d) In-cabin sound meter in heavy vehicles



(e) In-cabin sound meter in smaller vehicles

Figure 6. Deployment of sound meters during data collection

2.1.3.5 Test scenarios

Road noise is generated by passing vehicles under a variety of conditions that include rumble strip incursions as well as pass-by traffic with no rumble strip incursions. To exclusively capture the noise generated from the rumble strips, and for safety reasons, the tests were conducted using short-term flagging operations to restrict traffic temporarily. Two major test scenarios were evaluated:

- a) *Centerline*: Incursion on the far side centerline rumble strip (Figure 7 (a))
- b) *No incursion or Baseline pass-by run*: Normal pass-by of the vehicle without any incursion on the rumble strips (Figure 7 (b))

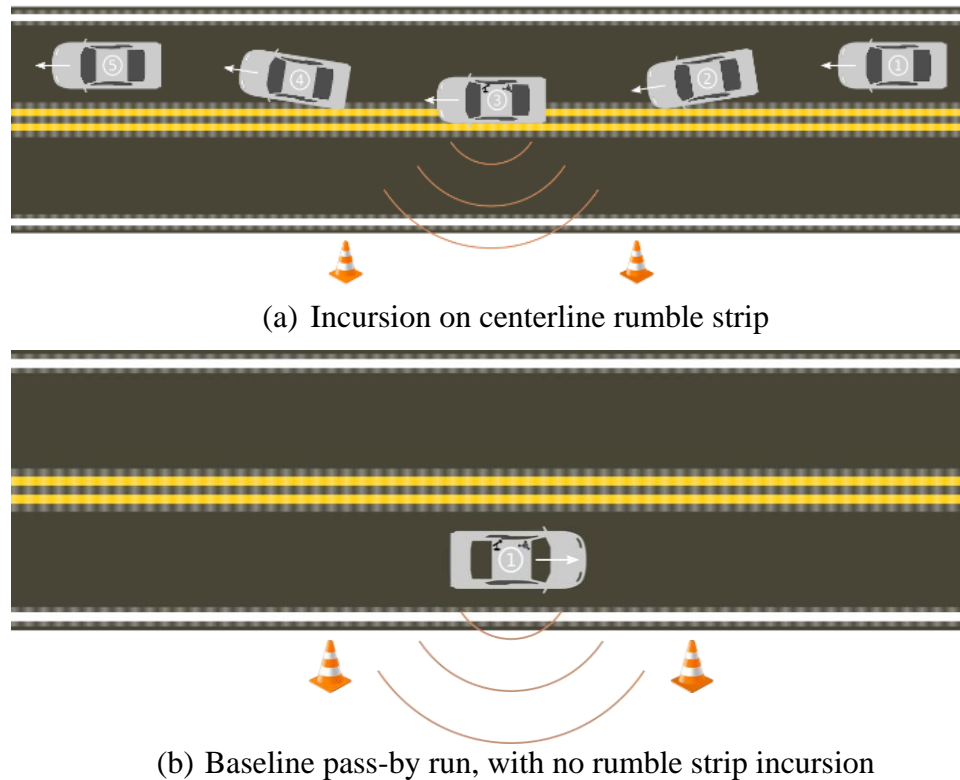


Figure 7. Test scenarios

2.1.4 Results and analysis

2.1.4.1 Accelerometer analysis

The accelerometer measures acceleration in the sensor's three-dimensional frame using units of gravity (g). This study is predominantly interested in acceleration caused by vibration; therefore, the constant acceleration of Earth's gravity must be accounted for

in the study. However, that is not easily accomplished because the coordinate frame of the sensor is not precisely known. Instead, for each incursion, the constant acceleration is subtracted from each of the sensor's 3-axis. A time-series "dynamic magnitude" trace, $a_d[n]$ is computed from the result:

$$a_d^N[n] = \sqrt{(a_x[n] - \mu_x)^2 + (a_y[n] - \mu_y)^2 + (a_z[n] - \mu_z)^2} \quad (1)$$

where $a_\alpha[n]$ is the acceleration, μ_α is the mean acceleration in the α direction during the N -th trial, and n is the discrete-time index with a sample rate of 512 Hz (data was collected at a frequency of 512 Hz).

The design of the experiment requires the test vehicle to travel at a constant speed during a rumble strip incursion. Therefore, the only constant acceleration present during the data collection is gravity. In this fact, the gravity components are estimated by averaging all data collected during a particular trail. By subtracting this average from the original signal, the dynamic portion of the total vibration is estimated.

Figure 8 compares some example dynamic magnitude traces collected from the 12 in wavelength rumble strips (blue signal) and the baseline pass-by run (orange signal) for all the test vehicles. As seen, the engine vibrations are dominant across the heavy vehicles, and it is difficult to separate the rumble strips from the baseline traces (Figure 8 (a) – (c)). As for the smaller vehicles (Figure 8 (d) – (f)), there is a clear distinction between the acceleration traces from the rumble strips and the baseline (no rumble). Acceleration traces on the 18 and 24-inch rumble strips also yielded similar results.

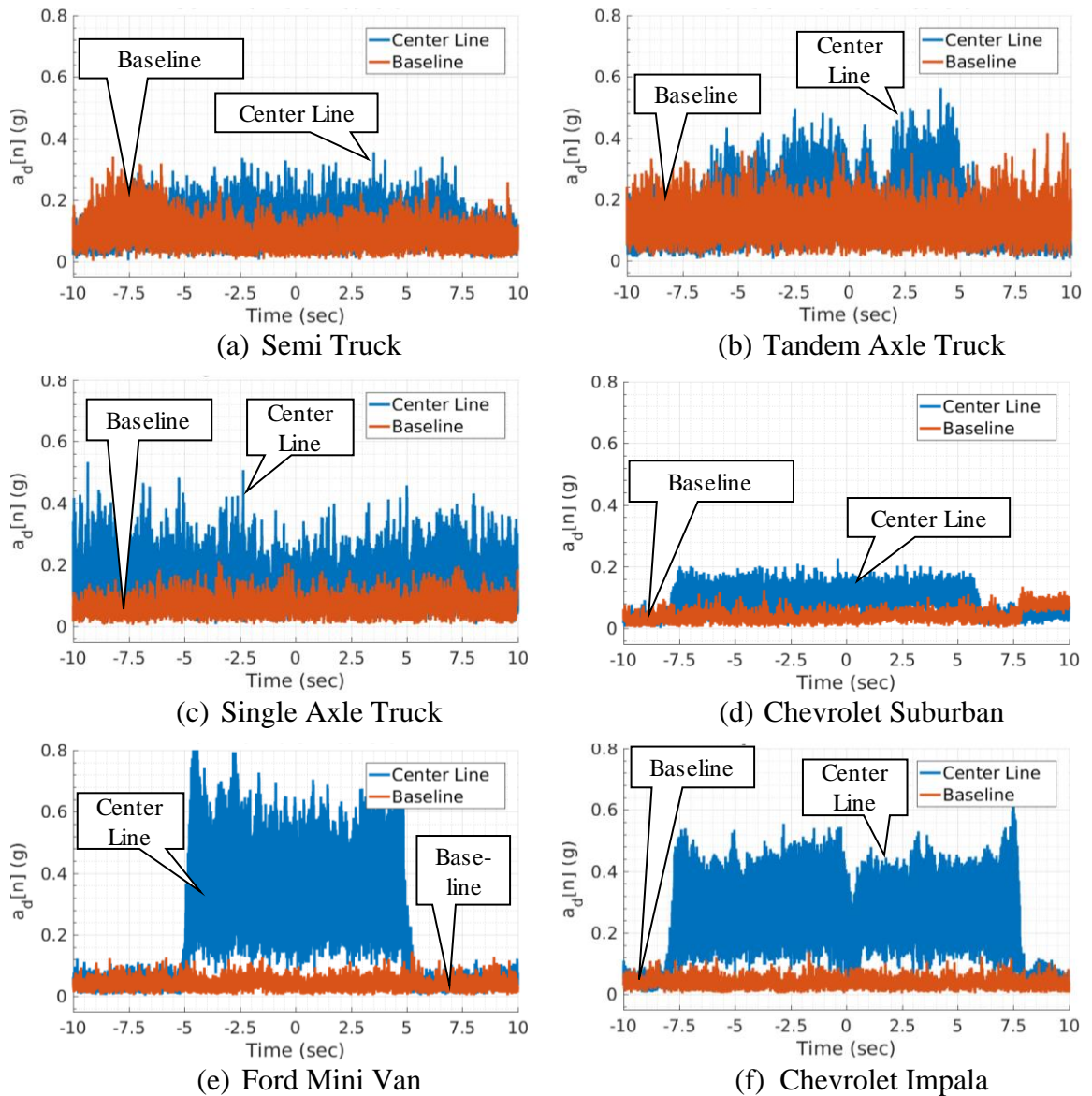


Figure 8. Acceleration traces for test vehicles on 12'' sinusoidal rumble strips

Figure 9 compares the average maximum of observed root-mean-squared (RMS) dynamic acceleration across the experiment runs. The RMS is computed using a 125 millisecond (ms) long moving window. The baseline marks the average maximum RMS of the dynamic acceleration during the baseline pass-by run. The induced vibration is a function of the vehicle's suspension, and therefore no clear pattern emerges. However, the vibration does tend to decrease in the seat frame with increasing wavelength.

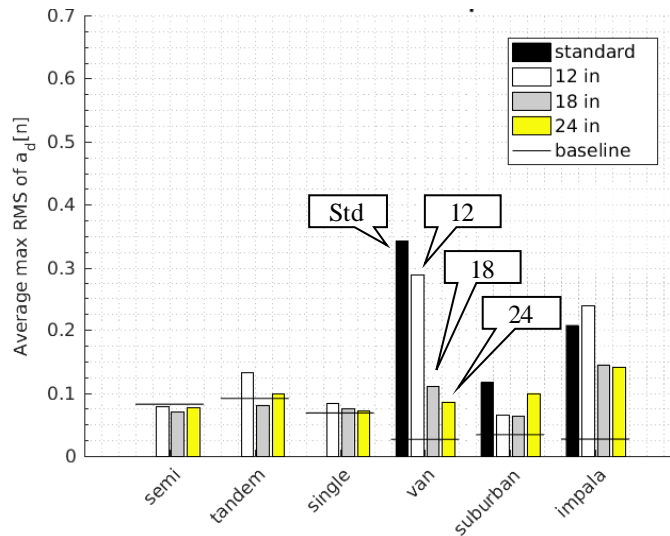


Figure 9. Average maximum RMS of accelerometer dynamic magnitude for all test vehicles

2.1.4.2 Sound level analysis

Developing a metric for the perception of a sound is difficult [37], [38] and an appropriate metric has not been widely accepted. The human ear and the auditory processing center in the brain is a very complex organ with many individual parts that each sense its own portion of the sound spectrum. As a result, sound perception is not only dependent on frequency but also other characteristics of the pressure wave, such as the length of time a particular component is present and the overall complexity of the pressure waveform, e.g., a single tone vs. a composition of tones.

In practice, one solution to this problem is to compute the sound power level (SPL) after filtering the waveform by a weighting function that approximates the human ear's response, as shown in equation (2).

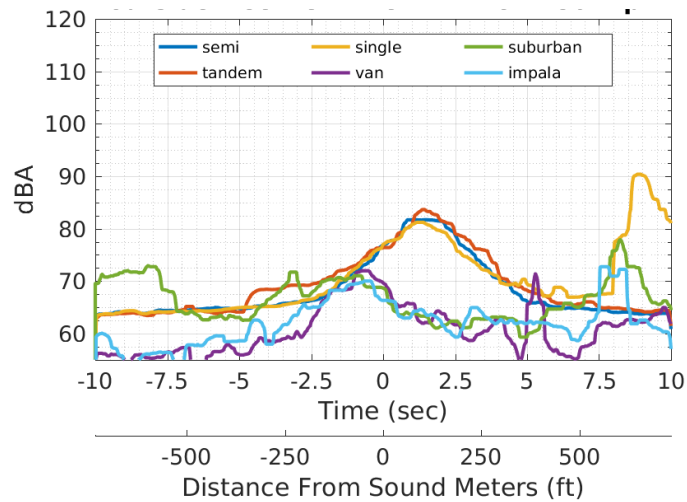
$$L_w[n] = 20 \log_{10} \left(\frac{s[n] \times h_w[n]}{20 \mu\text{Pa}} \right) \text{dB} \quad (2)$$

where $L_w[n]$ is the signal power, weighted by the w weighting function, $s[n]$ is the sound waveform in Pascals, $h_w[n]$ is the w weighting filter's impulse response, and $20 \mu\text{Pa}$ is the standard reference for SPL (often considered the threshold of human hearing).

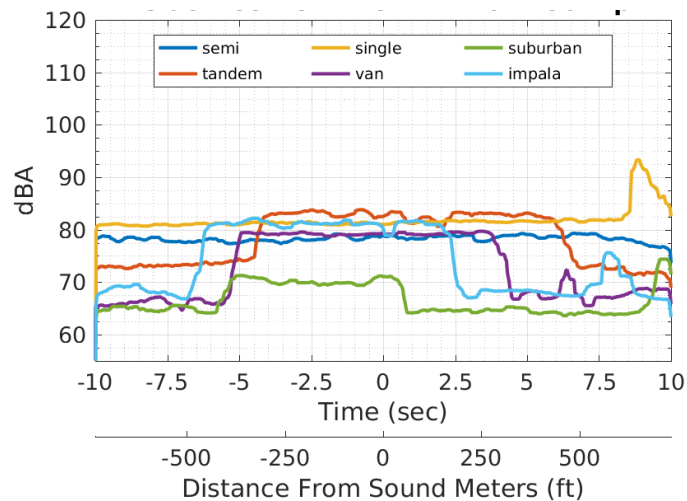
However, because no linear, time-invariant filter can completely capture the processing done by the ear, many different weightings functions have been proposed,

each useful within its own criteria. This study was performed using the A-weighting filter defined by International Electrotechnical Commission (IEC) 61672:2003 [39], which is often used as the required weighting function for many safety and environmental noise standards.

Figure 10 compares sound power traces measured by each sound meter on the center line incursion for all the test vehicles. The signals are averaged with a 125-millisecond moving window, defined by the IEC as the “fast” average [39].



(a) Sound meter outside



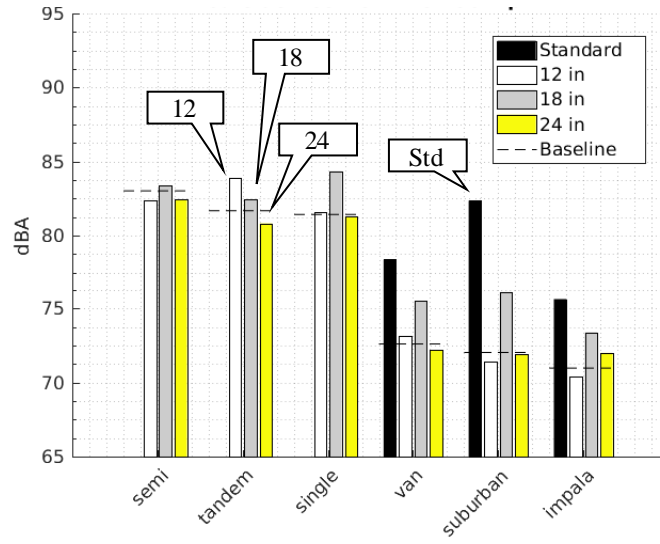
(b) Sound meter inside

Figure 10. Sound power traces on centerline incursion for 12 in sinusoidal rumble strip

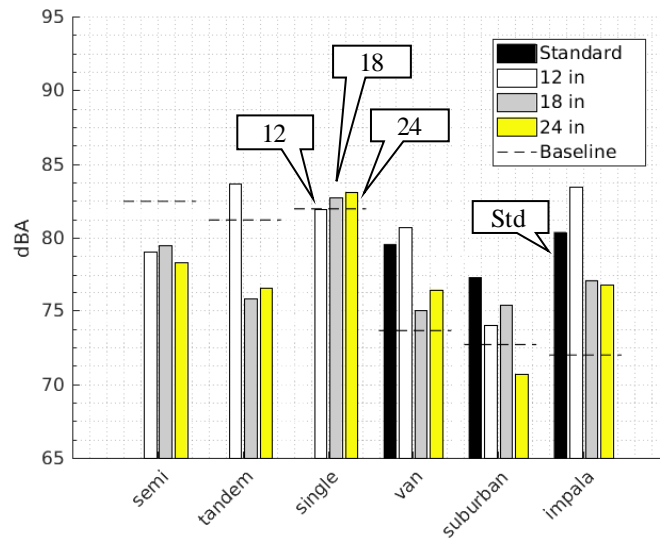
To reduce the time-series traces (Figure 10) to a single metric, the maximum observed power values for each vehicle encounter of a rumble strip, given a particular configuration, was averaged. For the baseline pass-by measurements, the maximum observed power level within a ± 4 second window surrounding the time the vehicle passes the sound meter was averaged. All of the configurations have at least three averaged repetitions. Figure 11 compares the measured sound level on centerline incursions across the experiment's runs. The baseline (dotted line) shown on the sound level plots is the noise level during the baseline (no incursion) runs.

Overall, the in-cabin and exterior sound responses varied across the vehicles. In general, from the exterior, the sinusoidal rumble strips produce less noise than traditional rumble strips, with a reduction in sound power by anywhere between 5 dBA and 11 dBA (Figure 11 (a)). From the interior of the vehicle, they are almost as loud as the standard rumbles but still increase the in-cabin sound level by between 2 and 9 dBA (Figure 11 (b)) as compared to baseline (or no incursion). Some of the data suggest that the 24 in wavelength is actually quieter than the baseline, however, this is a result of stochastic variation. During the experiment, the researchers observed some difficulties in detecting the difference between a 24 in wavelength incursion and a baseline pass-by run from outside of the vehicle. There is also a large drop-off of interior noise for the heavy vehicles, which is likely due to their dominant engine noise and superior vehicle suspensions.

Interestingly, the 12 in wavelength seems to strike a balance between a reduced exterior noise and an increased interior noise. From outside, the 12 in sinusoidal rumble strips were found to be 5 to 11 dBA quieter than standard, and from inside, they were found to produce a sound level increase of 4 to 12 dBA compared to baseline road noise. The 12 in was also found to routinely satisfy the recommendations for in-cabin and exterior sound levels proposed by the NCHRP report [33] (Table 1).



(a) Exterior



(b) In-cabin

Figure 11. Sound level comparison for all vehicles on centerline rumble

Table 1. Sound level comparisons with NCHRP recommendations

NCHRP Recommendations	Exterior sound levels	In-cabin sound levels
	To limit exterior noise near residential land uses, sound should not increase by more than 12 dBA and preferably by less than 6 dBA	In-cabin (inside) sound level should increase by 10 dBA and preferably over 15 dBA.
12"	✔ 0 to 1 dBA above baseline	! 4 to 12 dBA above baseline
18"	✔ 3 to 5 dBA above baseline	✘ 1 to 5 dBA above baseline
24"	✔ 0 to 1 dBA above baseline	✘ 0 to 4 dBA above baseline
Standard	! 5 to 11 dBA above baseline	✘ 5 to 8 dBA above baseline

2.1.4.3 Retroreflectivity results

Retroreflectivity tests were performed on the three sinusoidal rumble strip patterns, a year after their installation. The Delta LTL-M mobile road unit [40] was used to collect the retroreflectivity data. The equipment, mounted on an INDOT vehicle, collected readings every 0.1 mi along the edge lines and center lines (Figure 12). The mobile equipment also logged the GPS coordinates of the data points. Figure 13 shows a close-up view of an 18 in sinusoidal rumble stripe with thermoplastic marking.

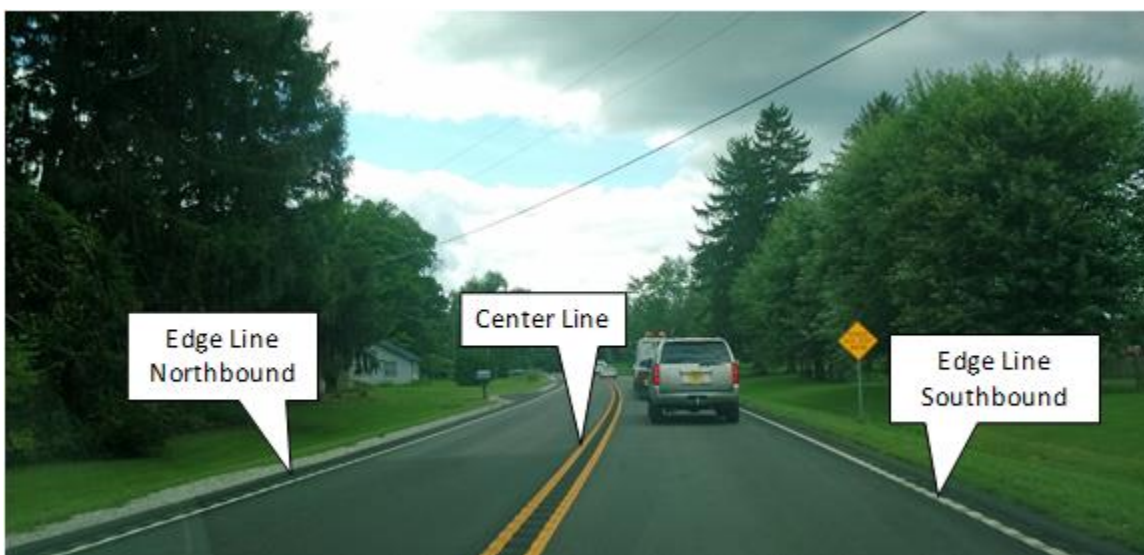


Figure 12. Retroreflectivity data collection on edge and center line



Figure 13. Close-up view of thermoplastic marking on sinusoidal rumble strips

A cumulative frequency diagram (CFD) of the retroreflective data on the three sinusoidal patterns for edge line and centerline is shown in Figure 14 and Figure 15, respectively. Chapter 3 discusses in detail on CFDs. The orange colored ranges in these figures denote the minimum retroreflectivity values specified by INDOT [41]. INDOT specifies a minimum range of 250 to 299 $\text{mcd/m}^2/\text{lx}$ for the white thermoplastic (edge lines) and 150 to 199 $\text{mcd/m}^2/\text{lx}$ for the yellow thermoplastic (center line) material. For the northbound edge line (Figure 14a) and the center line rumble stripes (Figure 15), all the three sinusoidal patterns surpass the minimum retroreflectivity readings. In case of the southbound edge lines, almost all of the readings exceeded the minimum threshold.

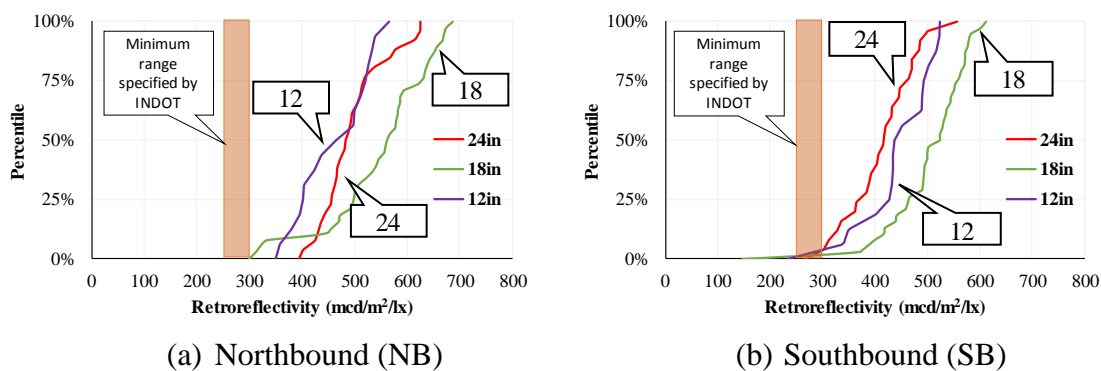


Figure 14. CFDs of retroreflectivity readings on edge line rumble strips

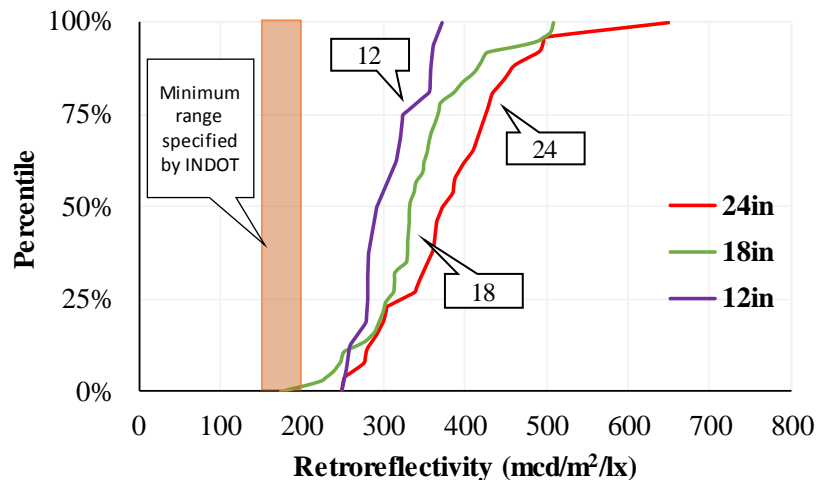


Figure 15. CFDs of retroreflectivity readings on centerline rumble strips

2.1.5 Conclusions

This study utilizes three performance measures, vibration, noise levels and retroreflectivity, to evaluate sinusoidal rumble strips of three different wavelengths (12, 18 and 24 inch). Data was collected from six test vehicles, ranging from a passenger car to a semi-truck. The acceleration data was used to measure the vibrations, whereas the acoustic data from the sound meters was used to compute the noise levels. Comparisons were also made with standard milled rumble strips and pass-by runs (normal movement on road). Retroreflectivity tests were performed to examine the visibility of the rumble strips during the night and adverse weather conditions. The major results from the study are:

- 1) For heavy vehicles, engine noise and vibrations were found to dominate from inside the vehicle.
- 2) Sound responses varied across the vehicles. From the exterior, all three sinusoidal rumble strips were less loud than the traditional rumble strips, with a reduction in sound power by anywhere between 5 and 11 dBA. From the interior of the vehicle, they are almost as loud as the standard rumbles, but with some selected cases increasing between 2 and 9 dBA.

- 3) The retroreflectivity tests on all the three sinusoidal patterns, on both the edge and center lines, were found to exceed the minimum threshold set by INDOT specifications.

Sinusoidal rumble strips are effective given the correct choice of wavelength. Among the three sinusoidal patterns examined, only the 12 in was found to routinely satisfy the recommendations for in-cabin and exterior sound levels proposed by the NCHRP Report 641 on Guidance for Design and Application of Rumble Strips. The results from this study suggest that 12 in wavelength has a desirable decrease in exterior noise while still maintaining reasonable or even, at times, superior (than the standard milled rumbles) lane departure warning to the driver.

The results discussed in this section are based on a speed of 50 mph. It would be desirable to conduct further tests at different vehicular speeds in order to provide a comprehensive assessment of the sinusoidal rumble strips noise levels and vibrations.

2.2 Evaluation of Aviation Rumble Strips

2.2.1 Introduction

Safety is a top priority for all aviation stakeholders, and a key component of aviation safety is runway safety. One significant threat to runway safety is runway incursions. Runway incursions must be reported at all airports having air traffic control (ATC) towers. The Federal Aviation Administration (FAA) defines a runway incursion as, “any occurrence at an aerodrome involving the incorrect presence of an aircraft, vehicle or person on the protected area of a surface designated for the landing and takeoff of aircraft.” [42]. The FAA identifies a reduction in runway incursions as an important component in improving safety [43] and a key metric in the 2015 – 2017 National Safety Plan [44].

The FAA defines three types of runway incursions: operational incidents (OI), pilot deviations (PD), and vehicle/pedestrian deviations (VPD). An OI is a result of the action(s) of an air traffic controller that results in less than required minimum separation between two or more aircraft, or between an aircraft and obstacles, or providing clearance to incorrect/unsafe operations. A PD occurs when the action(s) of a pilot violates any

Federal Aviation Regulation (FAR). A VPD results when pedestrians or vehicles enter any portion of the airport movement areas (runways/taxiways) without authorization from ATC [45].

In 2014, 1264 runway incursions were reported to FAA [46], a 10% increase since 2012 and more than 30% increase since 2009 [47], [48]. The number of incursions reported to FAA probably understates the potential risk. Since the majority of general aviation (GA) airports do not have an ATC tower, runway incursions at GA airports may go unreported, and as a result, the incidence and severity of the problem may not be fully recognized. Previous research reveals that 50 to 80% of all aviation accidents are caused by human error [49], [50]. Moreover, approximately 80% of total incursions that took place in 2014 are the result of a combination of PD (60%) and VPD (18%), so strategies to raise situational awareness regarding upcoming entrances onto live runways may be appropriate and effective [46].

Situational awareness can be enhanced by the runway status light program (RWSL), the electronic flight bag (EFB) and NextGen technologies such as Airport Surface Detection Equipment Model X (ASDE-X) and the Airport Movement Area Safety System (AMASS) [43], [51]. However, GA airports may not have the technology infrastructure and financial resources to support these NextGen solutions which enhance the situational awareness.

Another strategy under evaluation to raise situational awareness and warn pilots of upcoming dangers is the use of rumble strips. It may be appropriate to place rumble strips at the beginning of the enhanced taxiway centerline marking on an entrance taxiway to provide a tactile warning of the approaching runway and runway threshold. Highway rumble strips have been successfully deployed in roadway applications and have significantly reduced targeted crashes by up to 45% [22], [29], [34], [52]. The documented success of rumble strips in the roadway environment warrants exploration of the application to aviation.

2.2.2 Motivation and scope

In June 2015, the FAA announced the Runway Incursion Mitigation (RIM) program to identify airport risk factors that might contribute to a runway incursion and

develop strategies to help airport stakeholders mitigate those risks [53]. Airport characteristics are not congruent across categories, and the causes of runway incursions may vary depending on the size of the airport, aircraft fleet, and resources available for mitigation. As a result, mitigation measures for large hub airports are very different than for GA airports.

The motivation for this study comes from research that was conducted using econometrics based modeling techniques to identify statistically significant factors that affect runway incursions at different airport categories ([54], Appendix A). The study suggested that there is no single solution to prevent runway incursions, and in fact, the most appropriate countermeasures may vary depending on the category of airport. At smaller airports, including GA and non-hub airports, where incursions due to PD and VPD were found to be more common, low-cost solutions may be an effective way to enhance pilot situational awareness and reduce incursions. Given the documented success of highway rumble strips, the low cost and the potentially significant benefits, the purpose of this research was to examine the application and performance of rumble strips for enhancing aviation safety. Two performance measures are used for evaluation purposes, acceleration on the airframe and the durability of rumble strips.

2.2.3 Methodology and data collection

2.2.3.1 Rumble strips

This research examined three different types of rumble strips: temporary, saw cut and thermoplastic. The temporary/portable rumble strips have been widely used in the highway setting to alert the drivers of a temporary construction zone or potential lane closures. Saw cut rumble strips are similar to milled roadway rumble strips with permanent cuts made in the pavement using abrasive blades. The thermoplastic rumble strip was tested due to its proven application in the roadway environment. All the rumble strips were at least 18 feet wide to accommodate the aircraft landing gear. Figure 16 shows the different rumble strips.



(a) Temporary/Portable



(b) Saw Cut



(c) Thermoplastic

Figure 16. Different types of rumble strips

2.2.3.2 Test bed

The test bed was confined to a non-movement area of the Purdue University Airport (KLAF). Figure 17 shows an overview of the test bed configuration at KLAF.



Figure 17. Rumble strip test bed from 2015 at KLAf

2.2.3.3 Aircraft and test speeds

Seven aircraft were subjected to data collection (Figure 18): Cessna 152, Cessna 172, Cirrus SR20, Piper Warrior PA-28, Piper Seneca PA-34, Sky Arrow L600 and CRJ 200. Aircraft speeds of 5, 10, 15, 20 and 25 knots were tested over each rumble strip configuration. Two trials were conducted for each aircraft at each speed.



(a) Cessna 152



(b) Cessna 172



(c) Cirrus SR20



(d) Piper PA28



(e) Piper PA34



(f) Sky Arrow L600



(g) Canadair Regional Jet (CRJ) 200

Figure 18. Aircraft tested

2.2.3.4 Test sensors and setup

A three-axis accelerometer mounted on the seat and rail of the aircraft was used for the vibration assessment. The rail mounting used FAA certified cargo rail tie downs combined with a bracket to hold the accelerometer (Figure 19). Figure 20 shows the seat and rail mounting configuration on a Cessna 172 aircraft.



(a) Cargo rail tie down



(b) Rail mount setup

Figure 19. Rail mount configuration



Figure 20. Seat and rail mount configuration on a Cessna 172

2.2.3.5 Durability tests

Durability testing of the temporary and saw cut rumble strips was carried out at the FAA William J. Hughes Technical Center in Atlantic City, New Jersey. The rumble strips were subjected to loads for a number of passes with the airfield heavy vehicle simulator to mimic the impact of aircraft loadings over time (figure 21). Durability assessment after winter operations was performed at KLAFL.



Figure 21. FAA airfield heavy vehicle simulator

2.2.4 Results

2.2.4.1 Accelerometer analysis

Figure 22 shows the acceleration plots from the accelerometer for a Cessna 172 at 5 knots. This plot has four traces: the x (blue), y (red) and z (green) components of the three-axis accelerometer, and the resultant magnitude (violet) of all the forces. Although all three axes showed some response to the rumble strips, the vertical (z) component was found to have the most pronounced impact.

Figure 22 (a) displays the acceleration response from the rail whereas figure 22 (b) shows the response from the seat accelerometer. Compared to the rail, there is a considerable decrease in the amplitude of the acceleration that reached the pilot seat. At higher aircraft taxi speeds, the acceleration signatures become more compressed and the amplitude increases. The highest amplitude was recorded for the saw cut rumble strips.

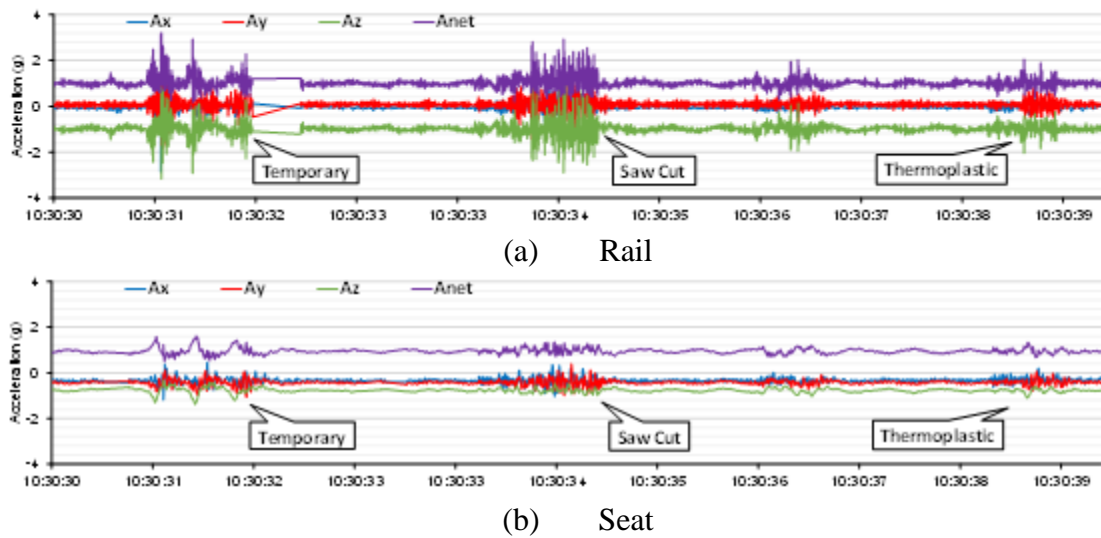


Figure 22. Acceleration plots for Cessna 172 at 5 knots

A peak-to-peak analysis compared the difference between the maximum and minimum net acceleration values and was performed at a range of speeds to study the impact of rumble strips on different aircrafts (Figure 23). As seen from the figure, the acceleration ranges increase at higher speeds. The Cessna 152 at a taxi speed of 25 knots was found to have the maximum peak-to-peak of 2.5g. The impact of the rumble strips varied across aircraft types, and it was difficult to establish a particular trend.

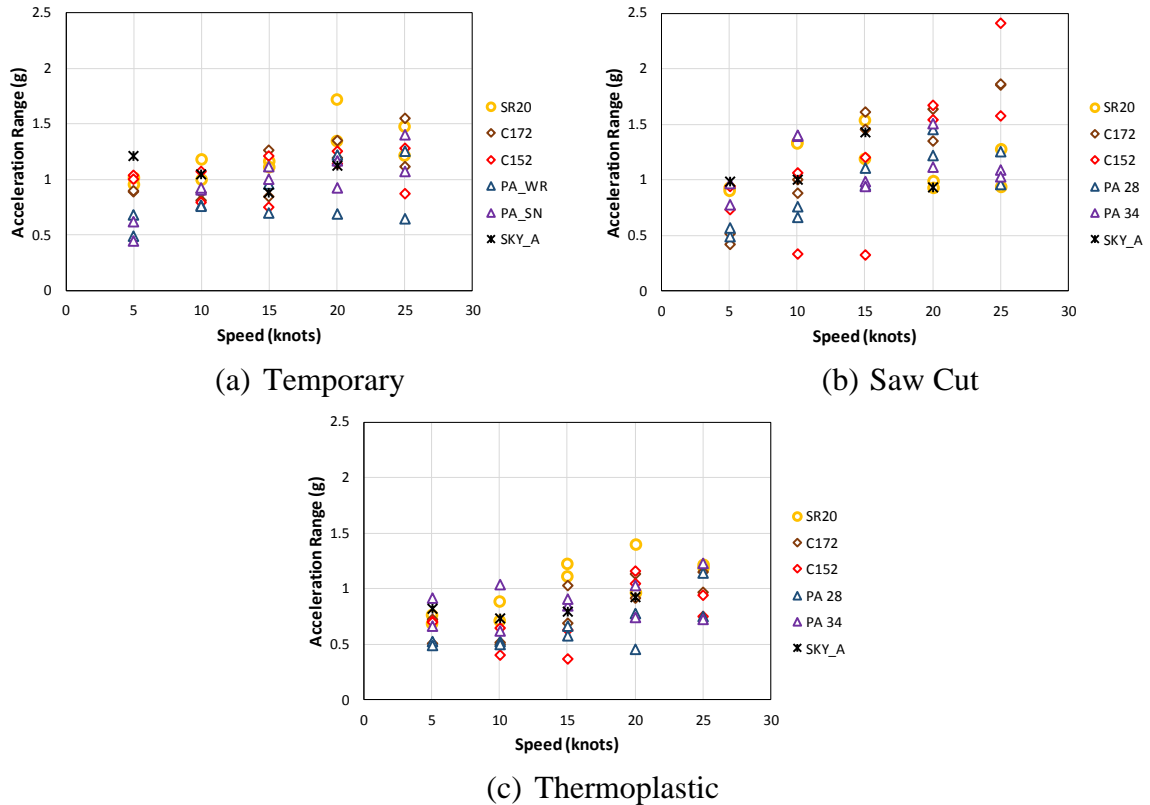


Figure 23. Peak-to-peak acceleration ranges on rail

Figure 24 compares the interquartile range (IQR) values of the seat and rail for the z-axis acceleration resulting from the trial of the Cessna 172 at 5 knots. As expected, the IQR values for the rail are higher than the seat, illustrating that a higher acceleration range acts on the rails. Inspecting the ratio of the rail-to-seat IQR values (figure 25), the saw cut rumble strips had the maximum ratio, with the IQR of the rail values more than 2.5 times higher than the IQR values for the seat. Based on feedback from subject matter experts, high acceleration ratios could trigger concerns with respect to airframe fatigue over time.

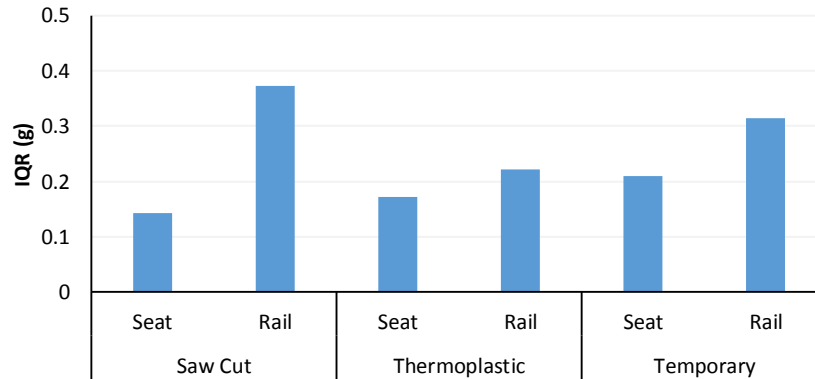


Figure 24. IQR plots for Cessna 172 at 5 knots

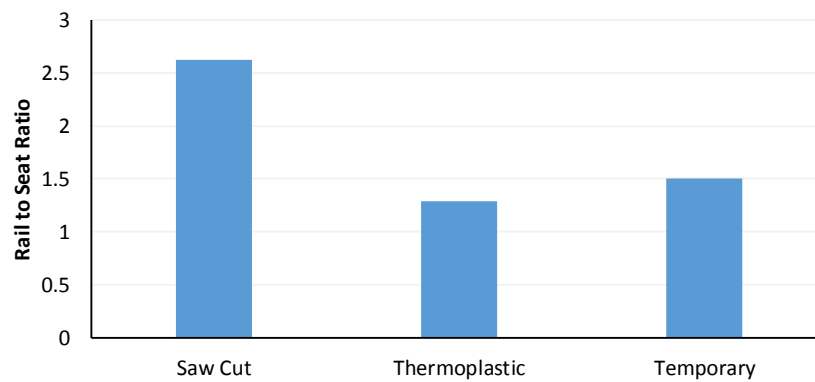


Figure 25. Rail-to-seat ratio comparison for Cessna 172 at 5 knots

2.2.4.2 Durability

Table 2 shows the results of the loading tests at the FAA William J. Hughes Technical Center. The straight edge saw cut rumble strip did not withstand the durability tests and showed signs of degradation, edge crumbling and raveling, resulting in foreign object debris (FOD) (Figure 26). The presence of FOD can create safety hazards and impact operations by requiring removal and risking aircraft damage when not removed [55]. The temporary rumble strips held up well to the test and started to show signs of wear only at heavier loads.

Table 2. Summary of Durability Testing at FAA William J. Hughes Technical Center

Load/Aircraft Simulated	Number of Passes	90 degree	Temporary
10k lbs./E190	1000	Began crumbling at 500 passes	No change
15k lbs./CRJ-700	1000	Continued crumbling	No change
30k lbs./757-200	1000	Heavily worn	Sign of warn, but no damage
45k lbs./738/A321	500	Heavily worn	Connection point becoming loose
60k lbs./ Air Force One	10	Heavily worn	~1/8" depression



(a) Formation of FOD



(b) Edge raveling/crumbling

Figure 26. Edge raveling and crumbling on saw cut

The temporary rumble strips also underwent additional durability tests. To simulate the high winds generated by the turbine engines and to evaluate the performance during winter operations, the temporary rumble strips were subjected to wind speeds more than 300 mph from the International 4800 equipped with a Sweepster 3000 Series Windrow Sweeper (Figure 27). In addition to the blower mechanism, the rotary broom was used directly on the portable rumble strips, traveling both perpendicular and parallel. The rumble strips were found to have minor to no significant lateral movements following the test. Researchers still recommend periodic checks, even though there was

no significant movement of the rumble strips during the winter operations test and routine rumble strip test.



(a) Sweeper



(b) Blower

Figure 27. Sweeper and blower test on temporary rumble strips

The durability of thermoplastic rumble strips was not evaluated at the FAA's William J. Hugh's Technical Center. However, researchers were able to examine the durability at the KLAF location. Following the first storm of the year and the resulting winter operations, the thermoplastic rumble strips cracked and had to be removed (Figure 28).



Figure 28. Cracked thermoplastic rumble strip after winter operations

2.2.5 Conclusions

Two performance measures, acceleration and durability, were used to evaluate the application of rumble strips as a low-cost solution to mitigate runway incursions at GA airports. From an operational perspective, based on the findings of this research, permanent rumble strips are not recommended for implementation at airports. Rumble strips that are large enough to be noticeable to pilots may result in forces that are large enough to cause airframe structural damage due to fatigue. These findings are also consistent with other studies [56].

Temporary rumble strips may be appropriate for further study as an alternative for short-term deployments in areas where aircraft volumes are limited, and speeds are low. The study does not recommend long-term deployment because the repetitive traversal of the rumble strips and the associated airframe stress cycles may result in airframe fatigue for an aircraft. Based on the results of this testing, aviation rumble strips do not appear to be an appropriate mitigation measure to reduce runway incursions.

2.3 Summary

This chapter discussed performance measures for evaluating the application of rumble strips as a potential safety countermeasure in highway and aviation environments.

Measures such as vibration, noise levels, and retroreflectivity were used to examine the performance of sinusoidal rumble strips in the roadway environment. Results showed that sinusoidal rumble strips are a promising technology well suited for lane departure warning in residential areas. Among the three sinusoidal patterns, the 12 in wavelength had decreased exterior noise than the standard rumble strips while still providing adequate lane departure warning to the driver. In some cases, the exterior noise from the 12 in was even on par with normal noise generated by traffic on the travel lane.

The potential application of aviation rumble strips to enhance aviation safety was evaluated using two performance measures (vibration and durability). Results of the testing suggest that aviation rumble strips may not be the best solution to mitigate runway incursions. Rumble strips that were significant enough to warn the pilot may cause

potential problems concerning airframe fatigue. Temporary rumble strips may be an acceptable alternative for further study for short-term deployments in areas where aircraft volumes are limited, and speeds are low to minimize airframe impact and mitigate airframe fatigue concerns.

CHAPTER 3. ARTERIAL PERFORMANCE MEASURES USING COMMERCIAL PROBE VEHICLE DATA

This chapter discusses the development of a data system and performance measures to support web dashboards that use commercial probe vehicle data for arterial performance measures to evaluate and monitor traffic conditions. Some of the information presented in this chapter is published in “Implementation of Probe Data Performance Measures” [57].

3.1 Introduction

In recent years, highway monitoring and performance measure requirements have been increasingly emphasized for federal transportation funding mandates such as MAP-21 [11]. These mandates have led to an increased need for system performance reporting at both state and local levels. It is highly likely that future legislation will require data-driven performance measures, as well. Historically, this has been a challenge due to the data collection infrastructure required for wide-scale deployment. Recently, advances in connected and probe vehicle technologies have resulted in data through third-party commercial vendors; this data must be transformed before it can be used to analyze performance based on agency goals and objectives.

Corridor progression is a common objective of traffic signal system operations on arterial highways and is often measured using travel time, which can be measured using a number of techniques, such as Global Positioning System (GPS) travel time runs [58], or vehicle re-identification [59], [60], or estimated using segment speed data from connected vehicles obtained from private-sector data providers [61]–[64]. Several researchers have recently explored the viability of private sector speed data for analysis of arterial travel times [61]–[63]. While results have varied, the growing consensus is that such data is viable on corridors with higher traffic volumes. A recent study demonstrated the scalability of this approach by applying it to a large inventory of corridors on a state-wide basis; in this case the corridors were ranked by travel time and travel time reliability for multiple times of day [64].

In May 2016, the Pennsylvania Department of Transportation (PennDOT) sponsored a 12-month research project at Purdue University to develop, implement and assess three web dashboards and a data system that use commercial probe data to produce arterial performance measures to evaluate and monitor traffic conditions. Real-time and historic traffic speed data was downloaded from the commercial vendor to populate roadway speeds at a nominal 0.3-mile resolution. The dashboards mapped the speeds to 138 corridors in the five-county region of District 6, including Bucks, Chester, Delaware, Montgomery, and Philadelphia counties, and produced travel time and reliability metrics, cross-corridor rankings, and a congestion monitoring tool on a web-enabled user platform.

3.2 Motivation and Scope

Outcome assessment is an important part of traffic signal modernization and this project addresses the need to rank the performance of 138 corridors in the Commonwealth of Pennsylvania using arterial performance measures developed from commercial probe vehicle data. Three web dashboards that use various performance measures including travel time, reliability, delay and congestion are introduced to provide cross-corridor rankings and impacts of maintenance activities. A before/after assessment is also performed using these tools on five corridors in Philadelphia that underwent signal retiming and adaptive control installations.

3.3 Methodology and Data Collection

3.3.1 Data

The speeds were obtained on a minute-by-minute basis from an aggregation of individual vehicle speeds determined from timestamped positions of GPS-enabled devices, including fleet telematics and cellular phones. For this study, segment definitions from the data provider were used, each approximately 0.3 mile in length, known as “XD segments”.

A web Application Programming Interface (API) was used to retrieve real-time speed data for the corridor segments. A Windows service application was developed to automate the process on the server and retrieve the data once-per-minute. Every three

months, archived data was also available for the previous quarter. A semi-automatic process was developed to extract and process this data into the Microsoft SQL Server database. There were approximately 30 billion data records for one year.

Analysis was performed using 15-minute periods of aggregated data to improve the computational performance of the dashboards. A programming service runs every 15-minutes to generate aggregated statistics for each XD segment over the interval. The client-side chart graphics were generated using a JavaScript library and the data queried from database defined by user inputs. Figure 29 shows the overview of the real-time data ingestion process.

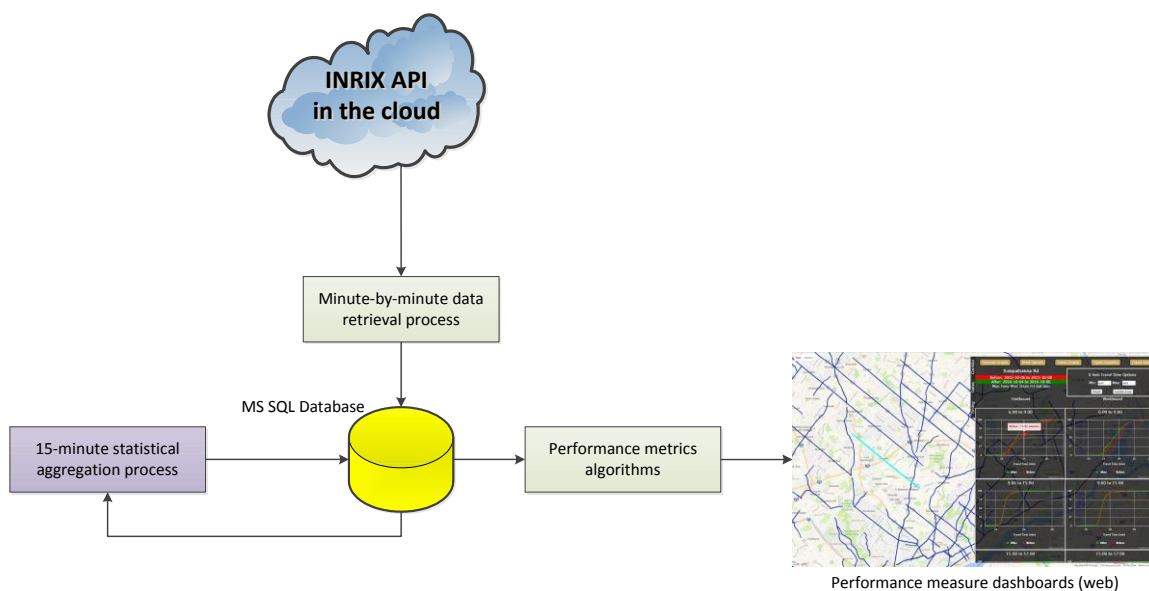


Figure 29. Overview of the real-time data ingestion process

3.3.2 Methodology

Three arterial performance dashboards were developed to evaluate and monitor the traffic conditions on a number of corridors. The methodology behind these dashboards is explained in the below sections.

3.3.2.1 Travel time comparison tool

The travel time comparison tool compares samples of travel times on any of the corridors during a user-specified before-and-after time range. The performance metric used in this dashboard is the cumulative frequency diagram (CFD) of all estimated travel

times within the selected periods on the corridor, typically shown as two before-and-after curves on a single chart. A screenshot of the application screen is shown in Figure 30.

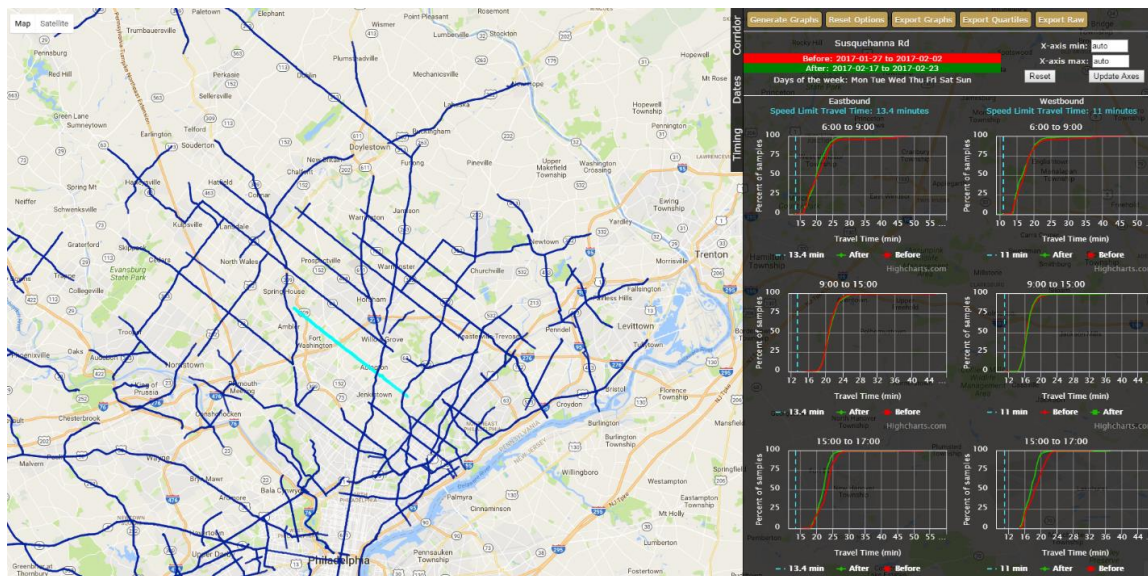


Figure 30. Travel time comparison tool

The CFDs can be used to perform a before-and-after assessment of the travel times on a corridor. For example, this could be used to evaluate the impact of a signal retiming project before and after retiming. The CFDs are plotted with the travel time along the x-axis and the cumulative frequency percentage along the y-axis. A near-vertical curve suggests the range of travel times are narrow, and therefore travel times are more consistent throughout the analysis period, whereas a wider curve suggests a wider range of travel times throughout the analysis period, indicative of a less reliable travel time.

Figure 31 shows the 1-minute travel time samples in a corridor from Monday to Friday, combined in a single 24-hour chart. The color-coded boxes represent the timing plans operating during the day, reflecting different traffic signal control parameters during each period. For example, Timing Plan 5 operates between 15:00 and 19:00, which represents the “PM Peak” for this particular corridor.

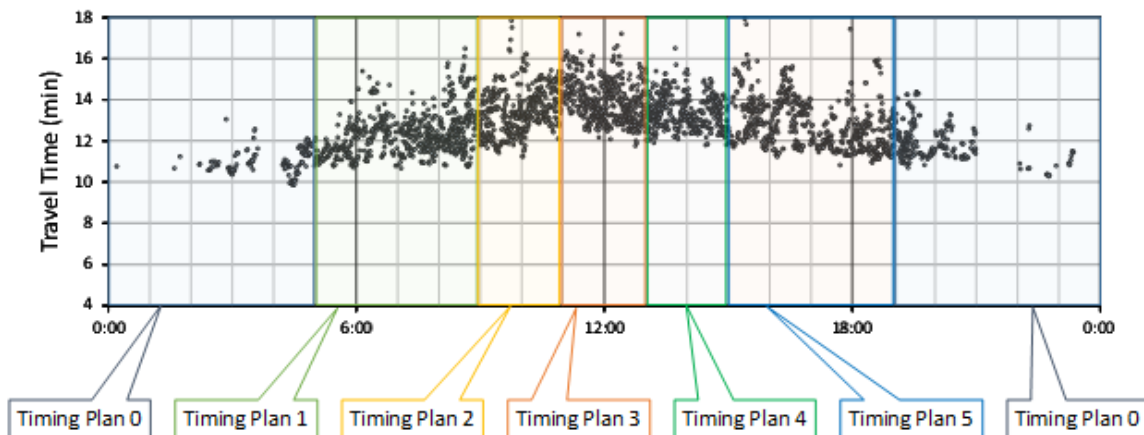


Figure 31. Travel times on a corridor during a 24-hour period

The cumulative frequency diagram for Timing Plan 5 reflects all travel time samples during the defined time-of-day over the five weekdays, as shown in the black line in Figure 32, which also includes a histogram.

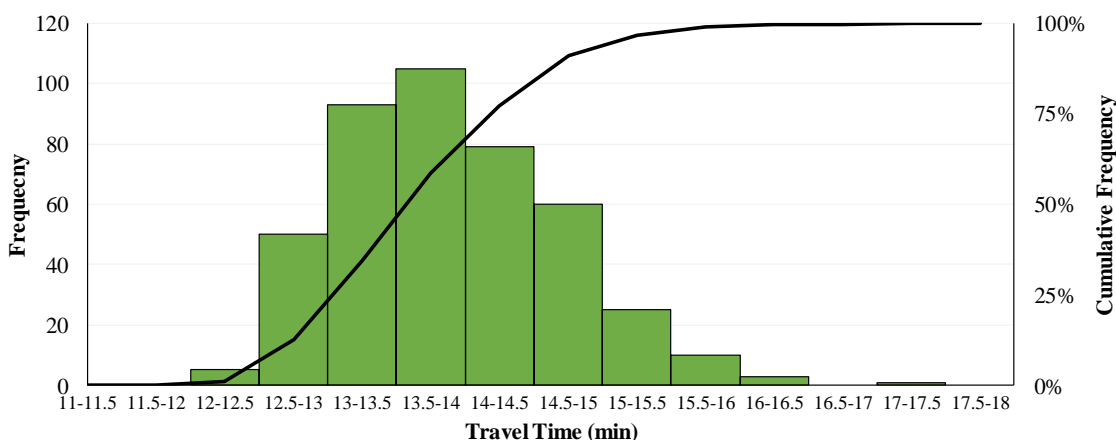


Figure 32. Plotting CFDs from timing plan 4

The CFDs can be used to evaluate the impacts of signal retiming on a specific corridor (Figure 33) and to illustrate how travel times have changed (Figure 34). In this case, there was a median travel time reduction of nearly 1 minute after the signal retiming, which is demonstrated by the green “after” curve being positioned to the left of the red “before” curve with a separation of 1 minute along the x-axis, at the median (50%) line indicated by the green arrow in Figure 34.

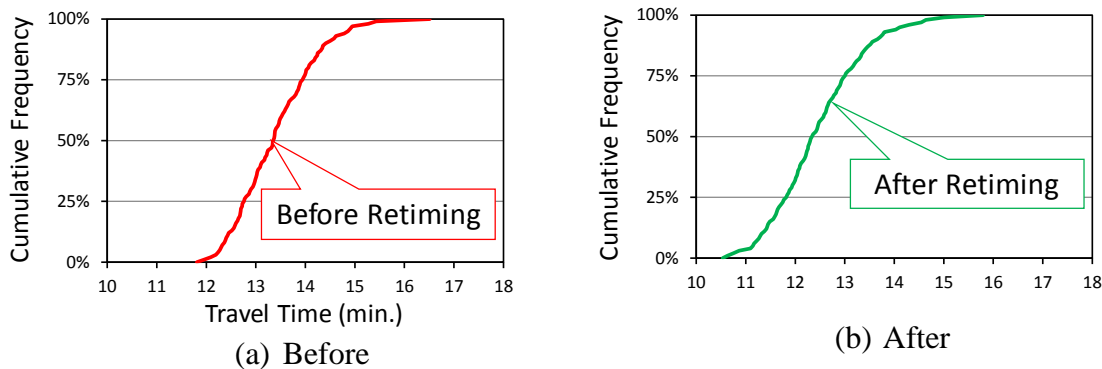


Figure 33. CFDs comparing before and after retiming

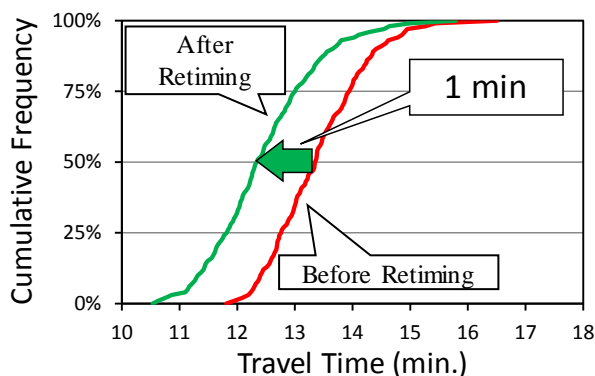


Figure 34. Median travel time improvement of 1 minute after signal retiming

3.3.2.2 Multi-criteria arterial ranking tool

The multi-criteria arterial ranking tool evaluates the performance of multiple corridors in a county or region, allowing objective comparison using posted speed limits in each corridor. For corridors with multiple posted speed limits, a weighted average was used to compute the speed limit. This dashboard ranks corridors by the central tendency of travel time, using the median value, and the reliability of travel time, using the IQR. The median and IQR are robust to outliers [65] and are hence preferred over to mean and standard deviation. The travel times in this tool are normalized using the travel time based on the speed limit to facilitate comparisons among multiple corridors. The tool produces sorted bar charts (Figure 35) based on either performance measure, or a scatter-plot using both criteria.

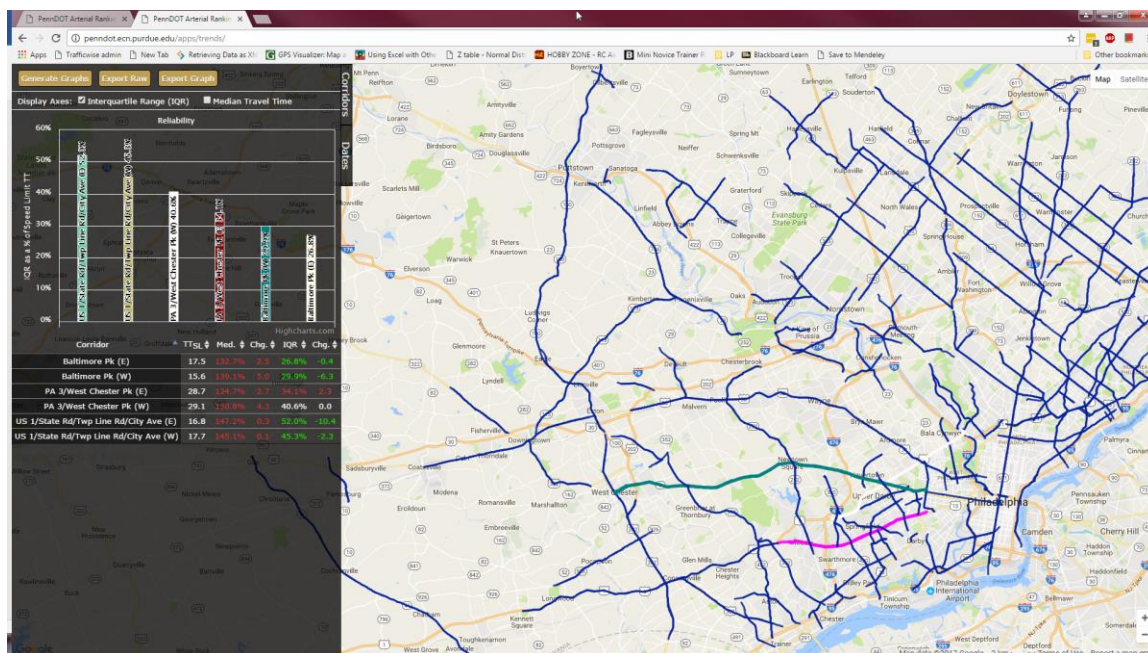


Figure 35. Multi-criteria arterial ranking tool

Corridor travel time performance can be evaluated by developing distributions of measured travel speeds and using metrics from percentiles of the distribution or other statistical properties [66], as used in the before-and-after CFDs of the travel time comparison tool, with speeds normalized based on a weighted average of the posted speed limit in the corridor. The speed limit travel time is not necessarily the “free flow” travel time, since free flow may occur at faster (or slower) speeds than the posted speed limits but it represents the ideal travel time for a vehicle traveling at the legal maximum speed unimpeded by delays due to traffic control, road work, congestion, weather, geometry, or other causes.

3.3.2.2.1 Speed limit travel time normalization

Consider the travel times on two separate corridors over a study period. The days-of-week analyzed are Monday through Friday. The first corridor, Newton Bypass, runs over a length of 4.2 miles with 11 signals and an AADT of 35,015. The second corridor, US-1/State Rd/Township Line Road/City Ave, is 10 miles with 40 signals and an AADT of 35,628. The speed limits on Newtown Bypass and US-1 are 45 mph and 55 mph respectively. Figure 36 shows their median measured travel times and speed limit travel time over a 24-hour analysis period during one week (solid and dotted lines represent

each direction of travel). Since lengths and speeds limits vary across the corridors, comparisons without normalization might be misleading.

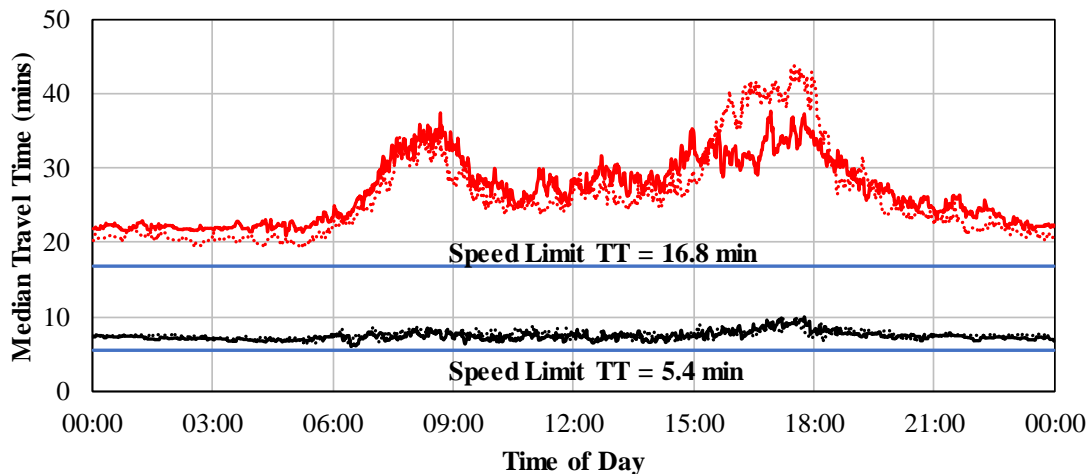


Figure 36. Median travel time and speed limit travel time on Newtown Bypass (shown in black) and US-1 (shown in red) during a period of one week

As mentioned earlier, the normalized travel time (TT) is computed as follows:

$$\text{Normalized TT} = \frac{\text{Median TT}}{\text{Speed limit TT}} \quad (3)$$

Figure 37 shows the normalized travel times of the two corridors from Figure 36. The blue line signifies the travel time at the speed limit (100%), after normalizing for each corridor's distance and speeds limit.

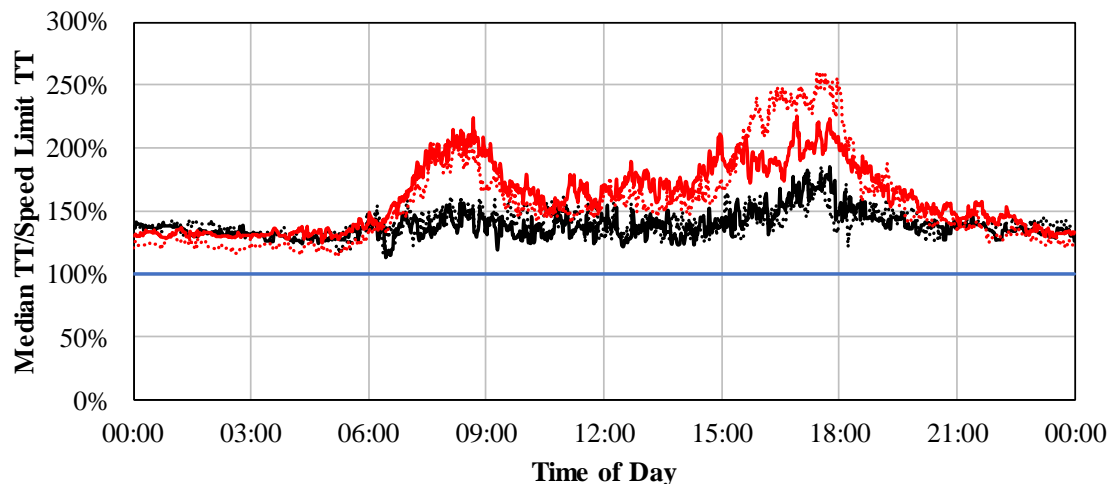


Figure 37. Normalized travel time on Newtown Bypass (shown in red) and US-1 (shown in black)

The normalized travel time provides a measure of the delay that allows comparison between corridors and may provide a way to prioritize corridors for resource allocation. Figure 38 shows an example in which ten corridors in a county are ranked on the basis of the normalized travel time for morning (0600-0900) data during over a one week period. Here, the median travel time for the corridor US52 – US231 to Sagamore Pkwy in the eastbound direction is 68.9% higher than the speed limit travel time, which denotes the lower speeds on the corridor during this period.

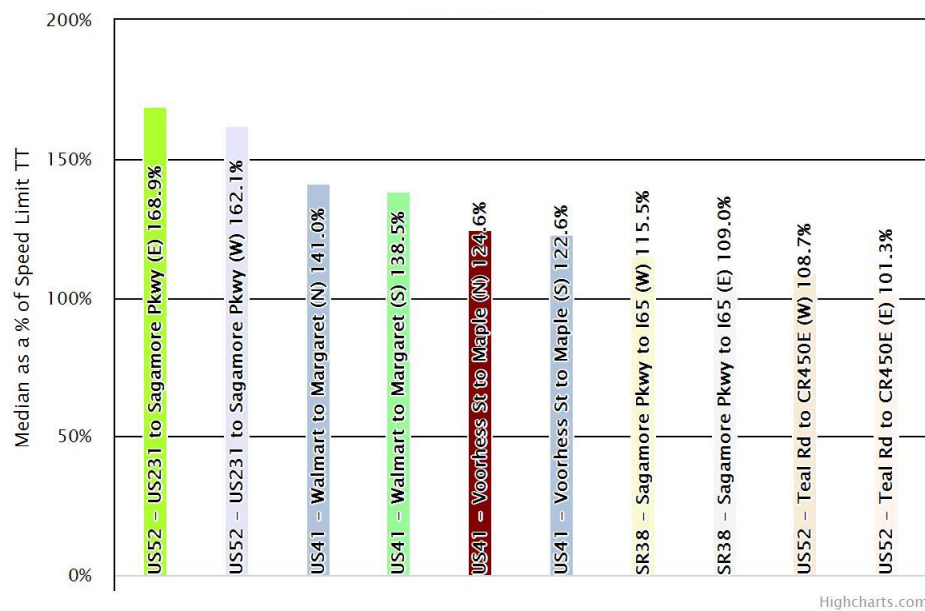


Figure 38. Ranking the corridors based on the median normalized speed limit travel time

3.3.2.2.2 IQR normalization

The interquartile range (IQR) is the difference between the 75th and 25th percentile travel time for a period of analysis, and can be used to quantify the reliability of travel time of each corridor. A higher IQR indicates a less reliable travel time. Figure 39 shows the 75th and 25th percentile values obtained from a hypothetical CFD.

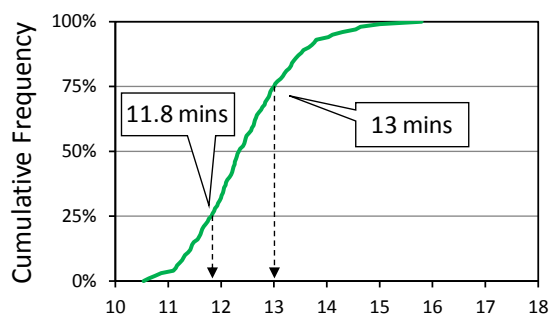


Figure 39. 25th and 75th percentiles

The weighted speed limit in the corridor is then used to normalize the IQR. The normalized IQR is calculated as follows:

$$\text{Normalized IQR} = \frac{(75^{\text{th}} \text{ percentile TT} - 25^{\text{th}} \text{ percentile TT})}{\text{Speed Limit TT}} \quad (4)$$

Rankings based on normalized IQR ranking are shown in Figure 40 for the same ten corridors shown in Figure 38. Higher percentages denote lesser reliability on the corridor. The IQR on the corridor US52 – US231 to Sagamore Pkwy in the eastbound direction is 21.6% of the speed limit travel time, indicating reasonable reliability and indicates the IQR of likely travel times varies by 21.6% of the posted travel.

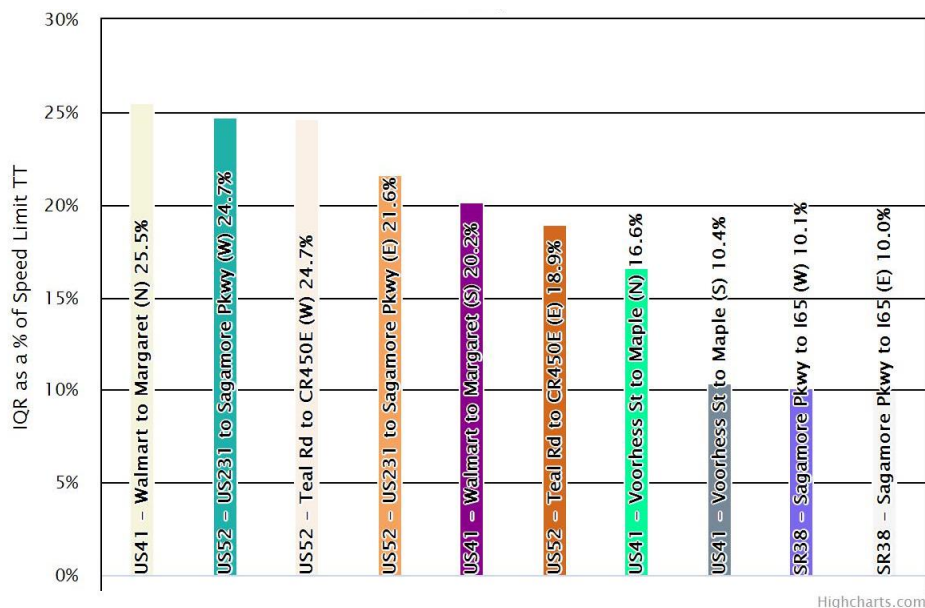


Figure 40. Ranking the corridors based on the IQR

3.3.2.2.3 Speed limit and IQR normalization

The speed limit and IQR normalization metrics can be combined on a two-axis scatter plot to identify corridors with higher travel time tendencies and lower reliability. Figure 41 shows such a plot; the corridors in the upper-right-hand region are relatively slower and less reliable than corridors in the lower-left-hand region. Note the US52 – US231 to Sagamore Pkwy corridor in the eastbound direction from the previous sections, with high median travel time and IQR values, plotted on the upper right hand corner in Figure 41.

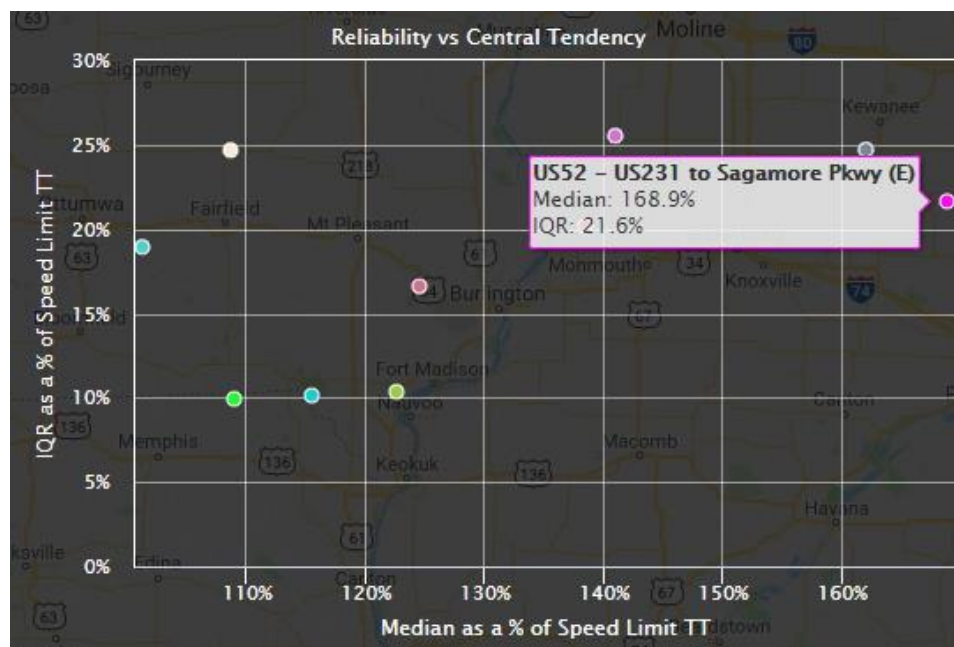





Figure 41. Speed limit and IQR normalization for one date range

The tool can be used to assess the performance of the corridors for before-and-after analysis. Table 3 shows a legend for interpreting this performance measure tool for the two date ranges. The performance of the “before” period is plotted as a dot, while the “after” period is plotted as a triangle. A line is drawn connecting the two periods associated with the same route. Routes with improvements in both the median and IQR during the after period are colored green, while routes with decline in both median and IQR are colored red. Corridors with mixed results are colored orange. Figure 42 shows an example in which the median and IQR values improved to 128% and 8.2%, respectively, during the after period.

Table 3. Legend for before-and-after date ranges

Before = ●	After = ▲
Decrease in both median and IQR	
Increase in both median and IQR	
Mixed results	

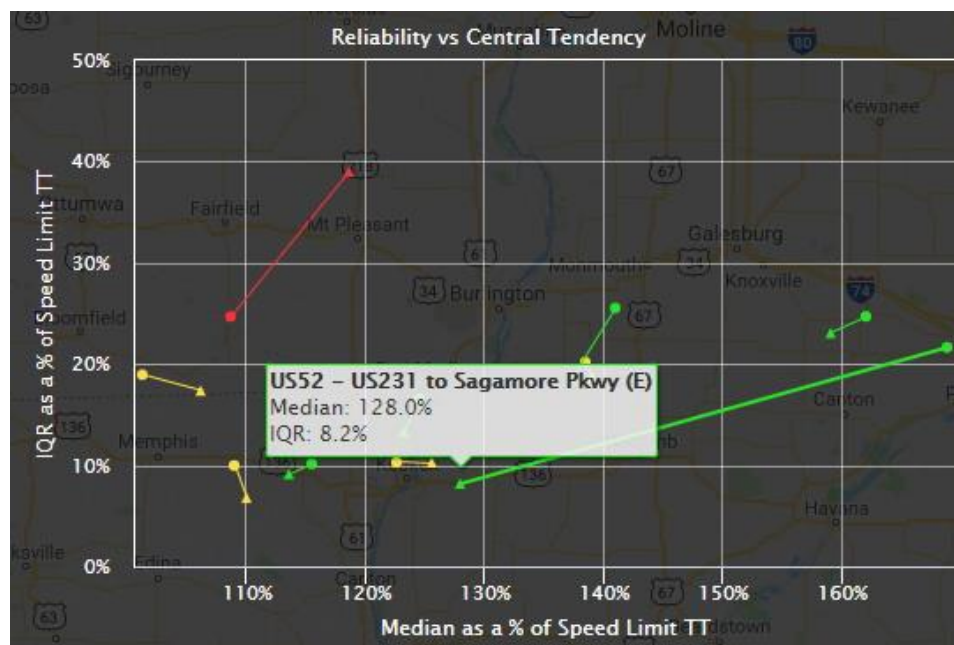


Figure 42. Speed limit and IQR normalization for before and after date ranges

3.3.2.3 Arterial congestion ticker

The third web application tool is the arterial congestion ticker, also known as the “travel delay monitor”, which looks at the cumulative number of miles in a corridor operating under different speeds (Figure 43). This tool generates a time-series plot for one or more selected corridors that shows the cumulative miles of the corridor(s) color-coded to reflect the corridor speed during the period. It shows the number of miles within a corridor length that is operating under a specified speed range. The chart allows the user to identify trends or “spikes” in the profile that can be correlated to specific times. The tool can also display a ticker for an “after” condition to facilitate comparison of two periods.

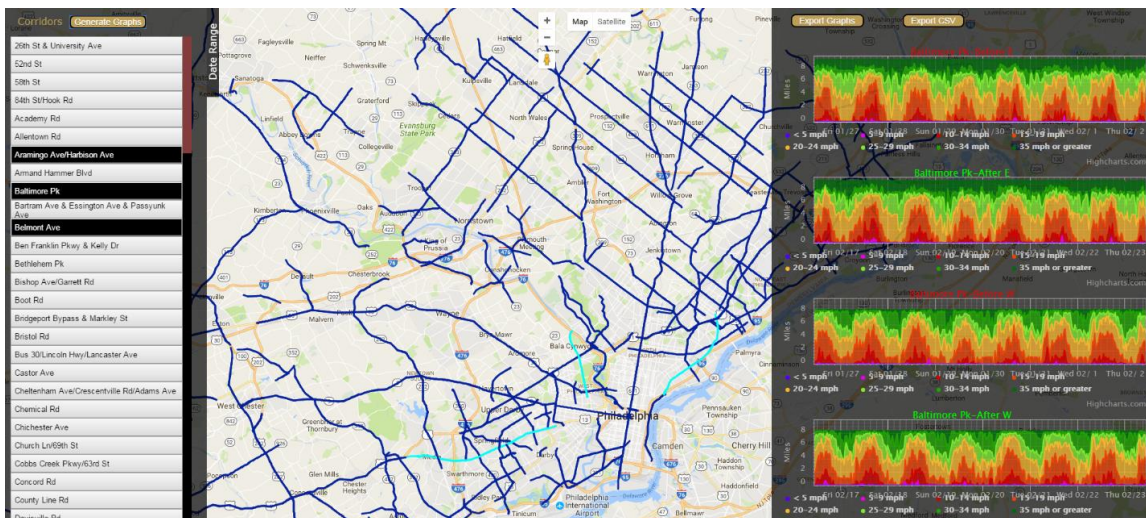


Figure 43. Arterial congestion ticker

Figure 44 shows a further breakdown of the color-coded time slice, where the vertical slice from the graph is represented as a pie chart. Each slice of the pie represents the percentage of the corridor’s total length that is operating at the specified speed range. The entire pie graph represents the total length of the corridor.

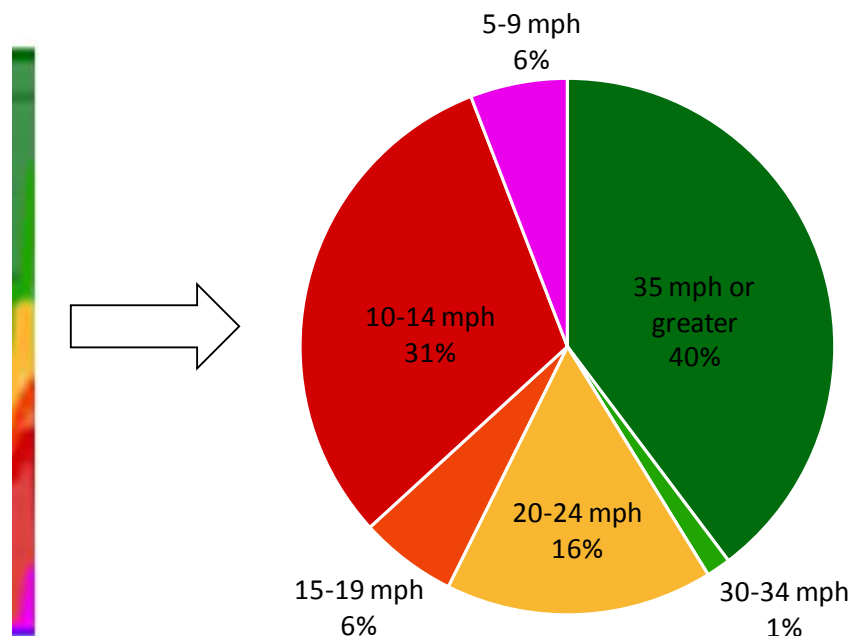


Figure 44. A 15-minute time slice of the graph generated by the arterial congestion ticker and its respective speed breakdown over the length of the corridor

Figure 45 shows the travel delay monitor tool for the northbound direction on Ben Franklin Parkway and Kelly Drive corridor between Friday, January 27 and Thursday, February 2 2016. For the majority of the period, the corridor operates above 24 mph for about 5 out of the 6 miles. Callouts in the figure point to possible instances of congestion where more than two miles of the corridor operated under 25 mph.

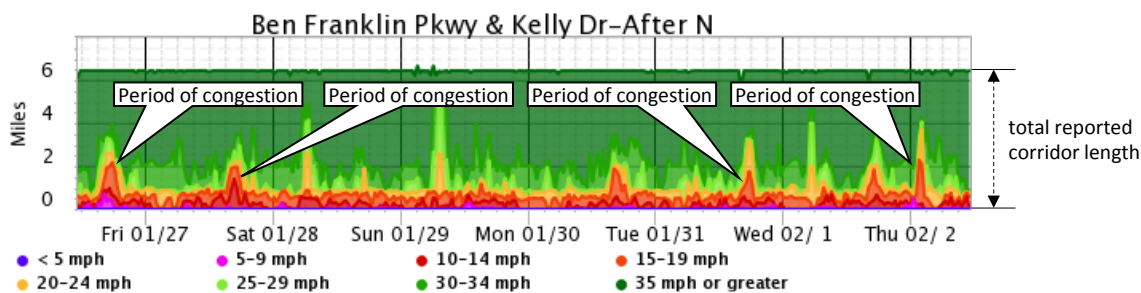


Figure 45. Travel delay monitor graph over a 1-week period

3.4 Results

This section provides case studies to illustrate the application of the performance measure dashboards.

3.4.1 Before-and-after assessments

A case study including a before-and-after analysis was carried out on five corridors in Philadelphia, to assess the performance of adaptive signals that were installed [67]. Figure 46 highlights the five selected corridors in the region where 61 out of 186 signals had adaptive control installed (Figure 47). Table 4 provides an overview of the characteristics, including the annual average daily traffic (AADT) of the selected corridors. The study analyzes the impact of these adaptive deployments on arterial travel times using the travel time comparison and arterial ranking tool. A benefit/cost analysis is also performed to evaluate the cost impacts of the adaptive installation.

Table 4. Characteristics of selected corridors

Corridor ID	Corridor Name	AADT	Length (mi)	Average Speed Limit (mph)	Signal Count (Adaptive Signals)	Before Date Range	After Date Range
A1	PA 132 / Street Rd	33,965	15.2	45	50 (21)	10/12/2015–11/23/2015	1/4/2016–2/15/2016
A2	PA 332 (Newtown Bypass)	35,015	4.8	53	12 (12)	2/22/2016–4/4/2016	4/25/2016–6/6/2016
A3	US 1/State Rd/Township Line Rd/City Ave	35,268	10.0	36	40 (4)	10/12/2015–11/23/2015	3/7/2016–4/18/2016
A4	US 202/Wilmington Pkwy	46,553	8.6	45	16 (9)	9/4/2015–10/26/2015	1/4/2016–2/15/2016
A5	PA 611/Old York Rd/ Easton Rd	30,919	16.3	42	4/27/2015–6/8/2015	1/4/2016–2/15/2016	

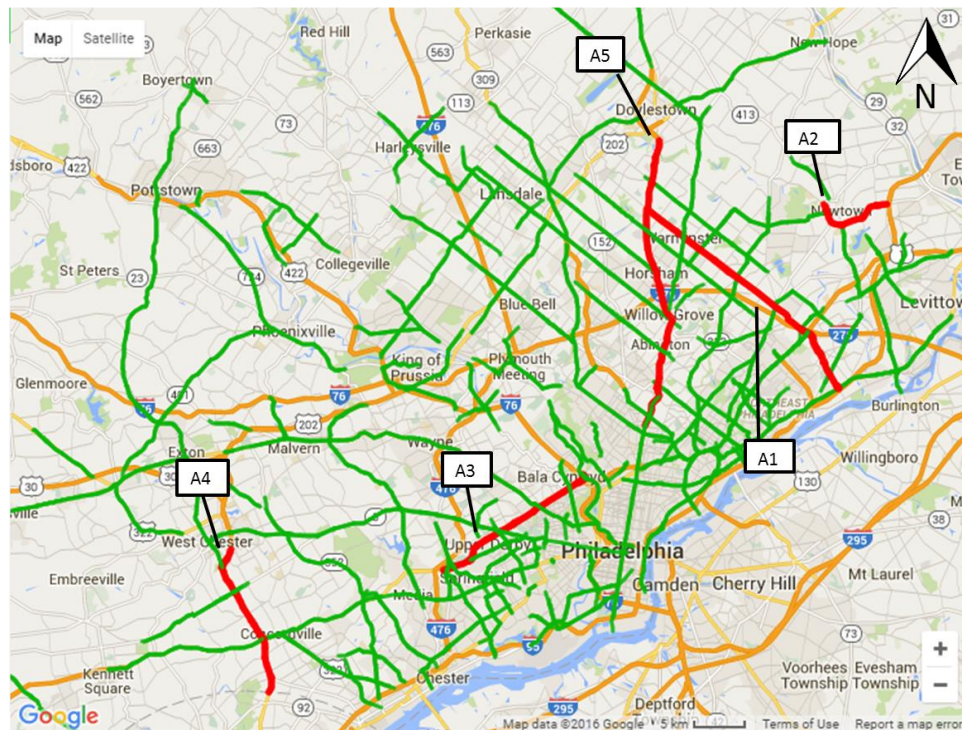
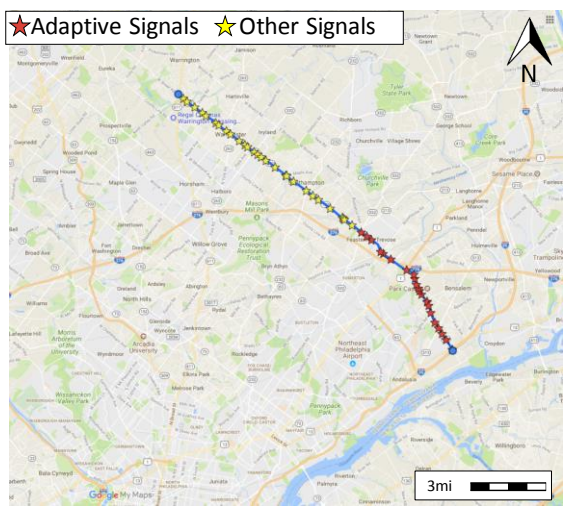
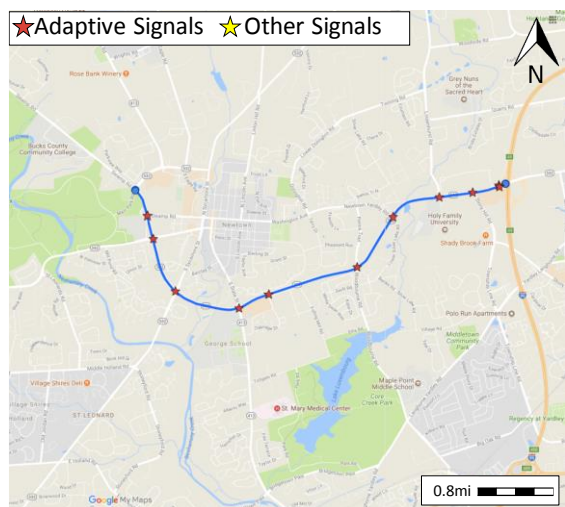


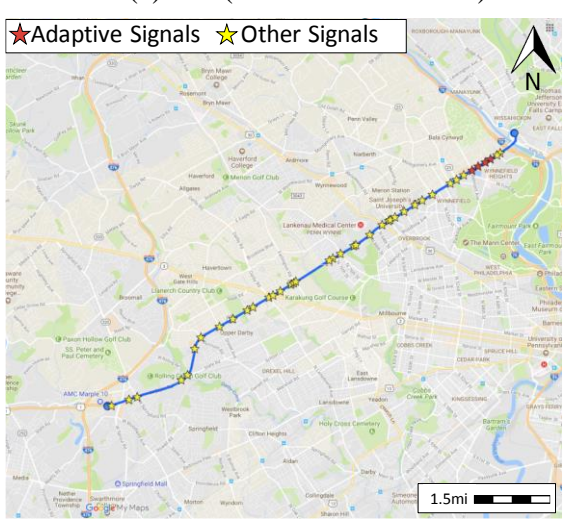
Figure 46. Selected corridors in the Greater Philadelphia area



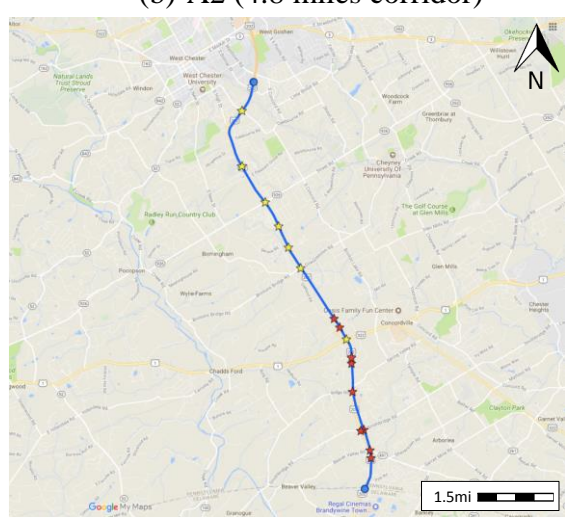
(a) A1 (15.2 miles corridor)



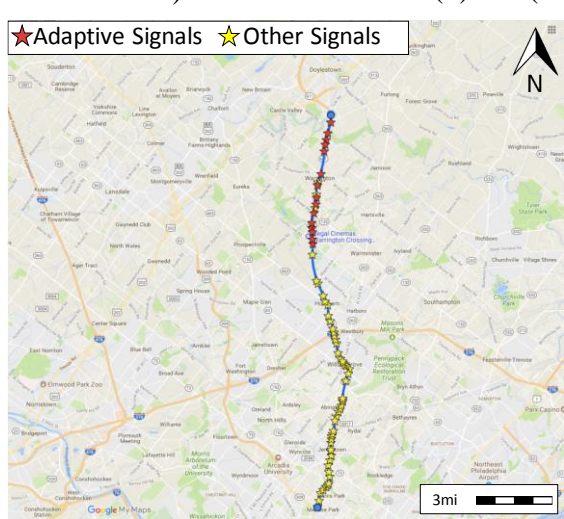
(b) A2 (4.8 miles corridor)



(c) A3 (10 miles corridor)



(d) A4 (8.6 miles corridor)



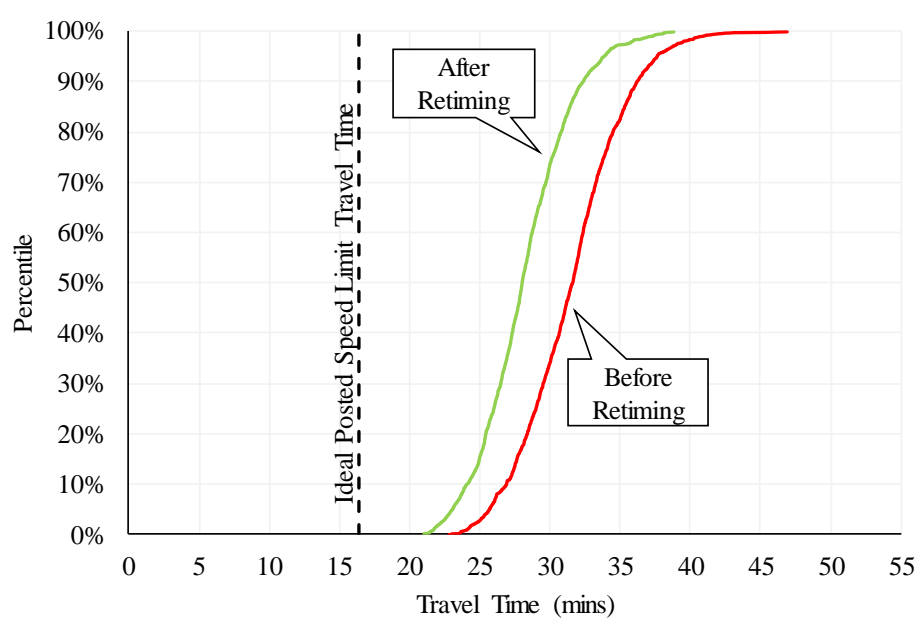
(e) A5 (16.3 miles corridor)

Figure 47. Corridor maps of intersections running adaptive control

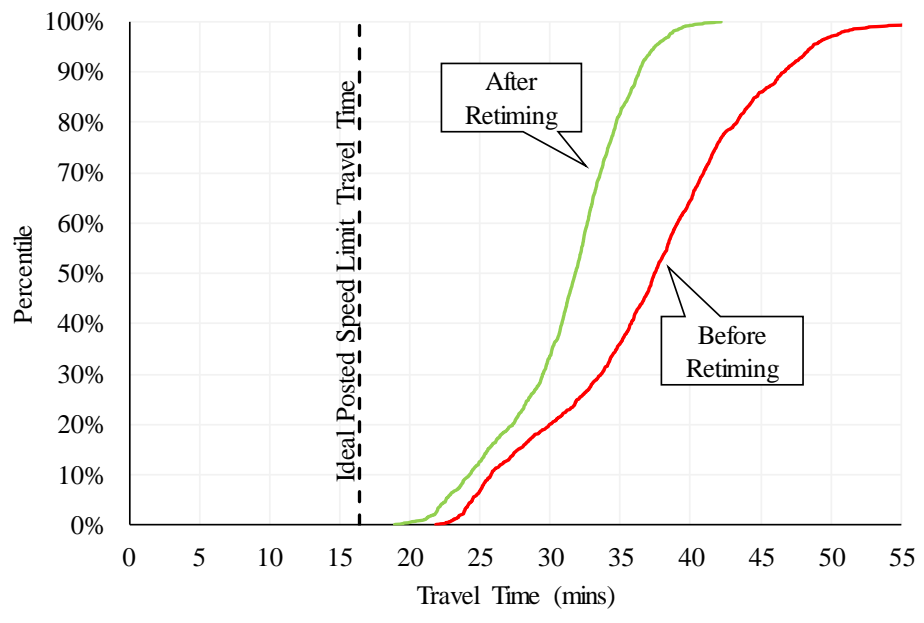
3.4.1.1 Travel time comparison

Figure 48 illustrates two CFDs for corridor A3 comparing a six-week period in October/November 2015 (before adaptive control deployment) and another six-week period in March/April 2016 (after deployment), for the two directions of travel. The analysis period includes weekdays from 17:00 and 18:00 (PM peak period), exclusive of holidays and periods when winter weather was likely to affect operation. The vertical dotted line in each figure represents the travel time in each direction at the speed limit; the red and green curves illustrate the CFDs for operation during the before and after periods.

The CFDs in Figure 48 (a) show an overall reduction in estimated travel times in the eastbound direction after deploying adaptive signal control. The median travel time improved by 3.6 minutes. The travel times range from 22.8 to 46.9 minutes in the before period, compared to 21.0 to 38.9 in the after period, indicating that travel times were more reliable, with less variation and a slightly steeper curve. Figure 48 (b) shows an even more significant improvement in the westbound direction, with a median travel time improvement of 5.6 minutes. The travel times range from 21.9 to 61.2 minutes in the before period, compared to 18.9 and 42.2 minutes in the after period, showing a substantial improvement in reliability.



a) Eastbound



b) Westbound

Figure 48. Travel times before and after installation of adaptive signal control on corridor A3 on weekdays between 17:00 and 18:00

A one-sided hypothesis test was also conducted on the above data to determine whether the travel times improved after installation of the adaptive signals. The null and alternate hypothesis are as follows:

$$H_0: \mu_a - \mu_b \geq 0$$

$$H_a: \mu_a - \mu_b < 0$$

where μ_a and μ_b are the means for the “after” and “before” period travel times, respectively. Table 5 shows the results of the tests for both eastbound and westbound direction.

Table 5. Hypothesis testing of the before and after travel times for corridor A3

	Eastbound		Westbound	
	After	Before	After	Before
Mean	28.19 mins	31.54 mins	31.10 mins	37.01 mins
Variance	10.37	13.15	19.55	54.04
Observations	1800		1800	
t-stat	-29.32		-29.24	
t-critical ($\alpha = 0.01$)	2.327		2.327	
p-value	<0.0001		<0.0001	

The absolute value of test statistic for both eastbound and westbound directions is much larger than 2.327, the critical value for a one-tailed test at the 1% significance level. As a result, the null hypothesis is rejected and there is evidence that the travel times have significantly improved during the after period.

3.4.1.2 Arterial ranking

Figure 49 shows the before and after conditions in the five study corridors based on normalized median travel times and IQR performance measures. For each directional route, distinct 6-week “before” and 6-week “after” periods were used to assess the changes in performance associated with the deployment of the adaptive systems. The corridor with the largest improvement from an adaptive implementation was westbound A3, which had 34.2% and 24.9% decreases in the normalized median and IQR travel times, corresponding to the CFD shown in Figure 48 (b).

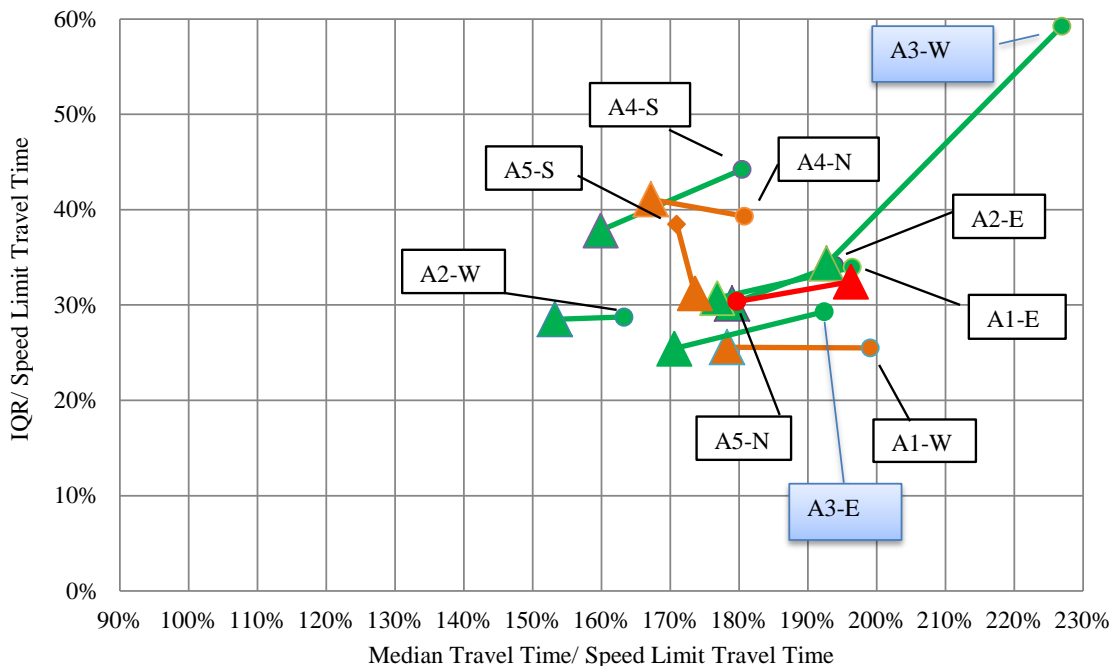


Figure 49. Arterial ranking tool showing the before-and-after travel time and reliability trends for the selected corridors on weekdays 17:00 and 18:00

3.4.1.3 Hour-by-hour median and IQR evaluation

Analysis carried out in one-hour intervals from 6:00 AM to 8:00 PM characterized the performance of the corridor throughout a day as shown in Figure 50 and Figure 51 respectively.

On weekdays (Figure 50), majority of the improvements were seen during peak hours:

- While corridor A1 also witnessed improvements during the non-peak hours, travel times on corridor A2 mostly remained unchanged
- Corridor A3 had the maximum improvement during the peak hours (callout i and ii) and 25% travel time decreases for most weekday hours.
- Corridor A4 had improved reliability through a reduction in the 75th percentile travel time.
- Corridor A5 saw increased travel times during both morning and evening peak hours. This was later reported due to some unanticipated maintenance and construction activities during the “after” evaluation period.

Figure 51 shows that a majority of weekend improvements came in the afternoon and evening hours, which is expected. The greatest improvements for weekend traffic can be seen in corridor A1 and southbound A4. As with the weekday travel times, weekends for corridor A2 were very consistent throughout the day. Corridor A3 saw maximum improvements during evening peak hours, with travel time savings of more than 10 minutes. Corridor A5 continued to have low reliability throughout the weekend, especially in the evenings for northbound travel.

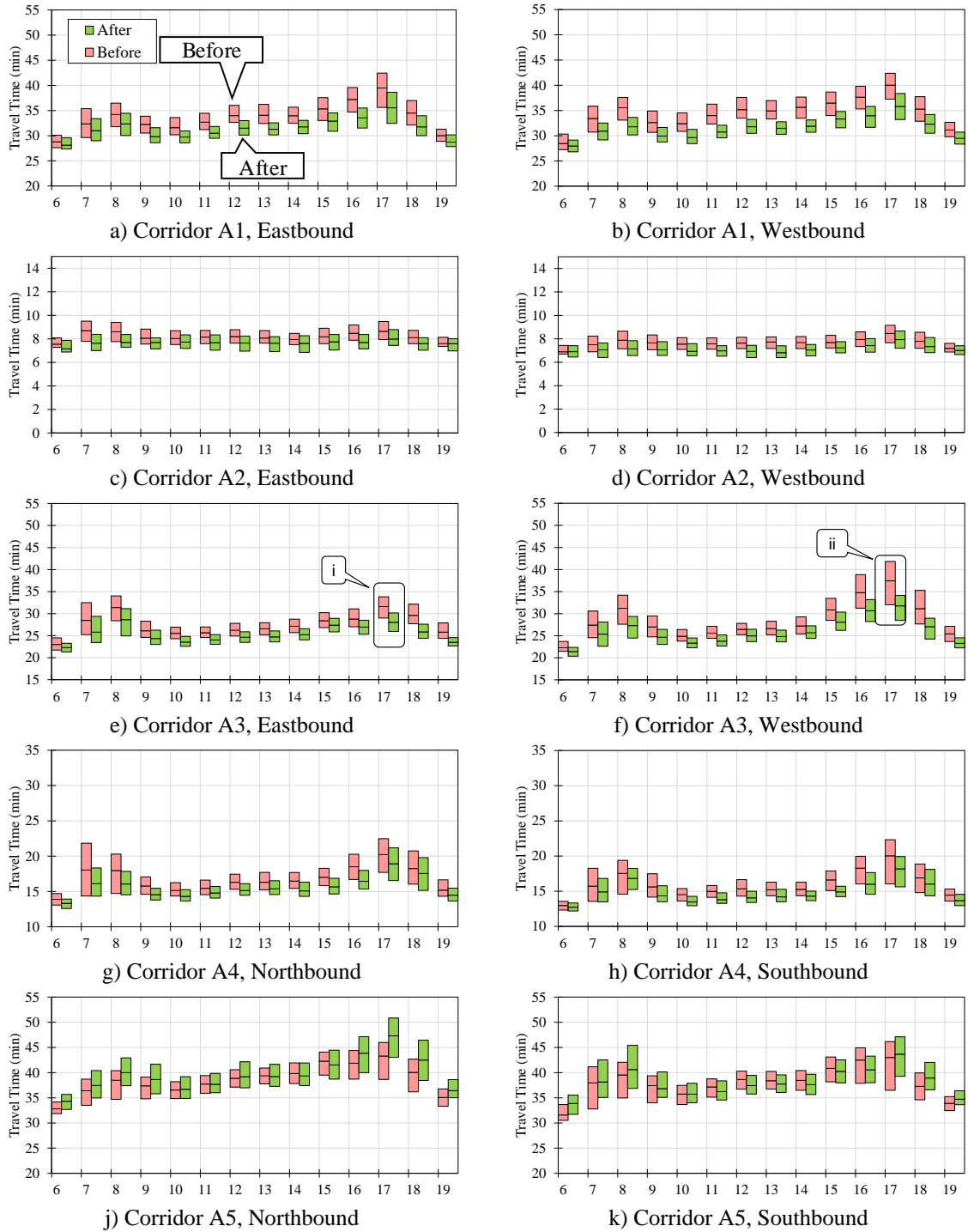


Figure 50. Weekday median travel times and IQR ranges by hour

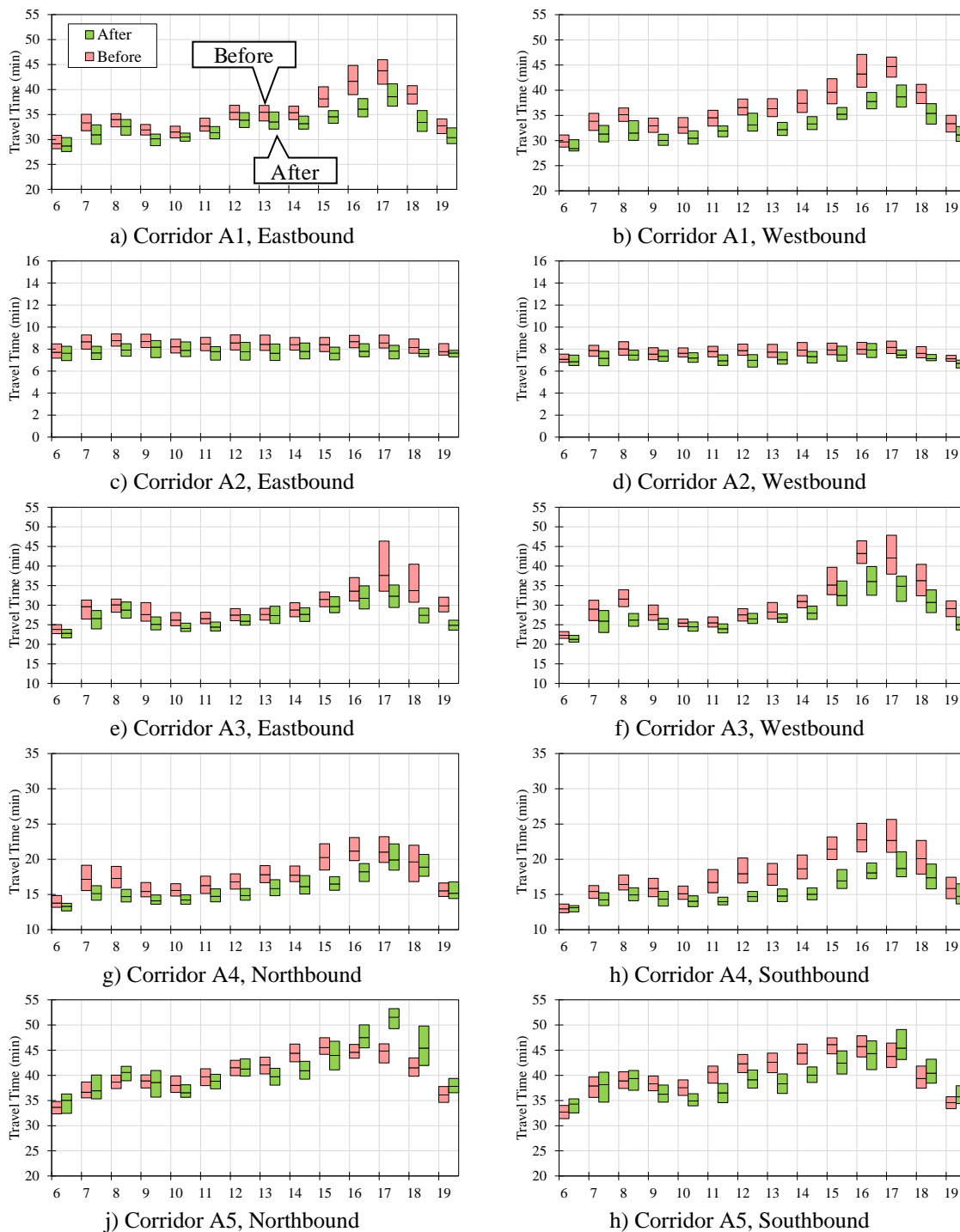


Figure 51. Weekend median travel times and IQR ranges by hour

An economic evaluation of the travel time improvements on these five corridors revealed a \$32 million annual user benefit (Figure 52), a savings of \$369,000 CO₂ emissions (Table 6) and a reduction in 1.2 million veh-hours of delay. Negative user benefit and CO₂ savings from corridor A5 are a result of the increased travel times during

the after period (corridor A5 in Figure 50 and Figure 51). This was later found to be because of unanticipated maintenance and construction activities during the “after” evaluation period. Overall, the investments showed a positive return for user benefits on four of the five selected corridors.

Accepted assumptions from the Texas Transportation Institute and Argonne National Laboratory was used to estimate the economic benefits. The methodology adopted for the economic evaluation is shown in Appendix B.

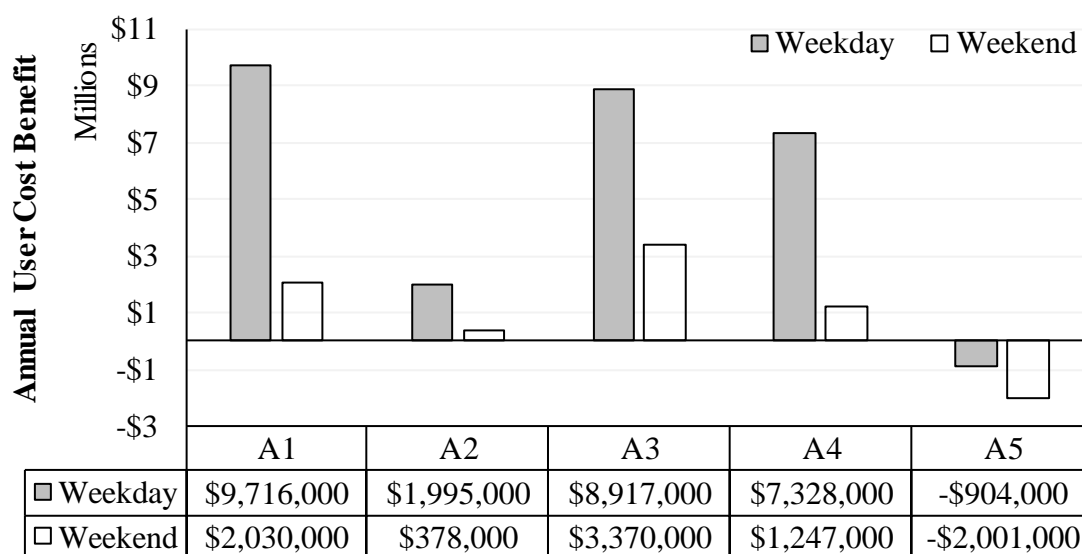


Figure 52. Summary of annual user cost benefits for the five corridors

Table 6. Summary of annual CO₂ emission reductions for the five corridors

Corridor	Weekday CO ₂ Savings		Weekend CO ₂ Savings	
	Tons	Dollars	Tons	Dollars
A1	3120	\$112,000	650	\$23,000
A2	640	\$23,000	120	\$4,000
A3	2890	\$104,000	1080	\$39,000
A4	2320	\$84,000	400	\$14,000
A5	-310	-\$11,000	-650	-\$23,000
Total	8660	\$213,000	1610	\$58,000

3.4.2 Incidents

The travel time comparison tool and arterial congestion ticker can be used to compare atypical traffic patterns that arise from an incident, special event, construction and maintenance activities.

The red line in Figure 53 shows a road closure on I-76 West due to a multi-vehicle accident. As a result, traffic was diverted along one of the parallel arterials, Gulph Road (denoted by the blue line in Fig xx).

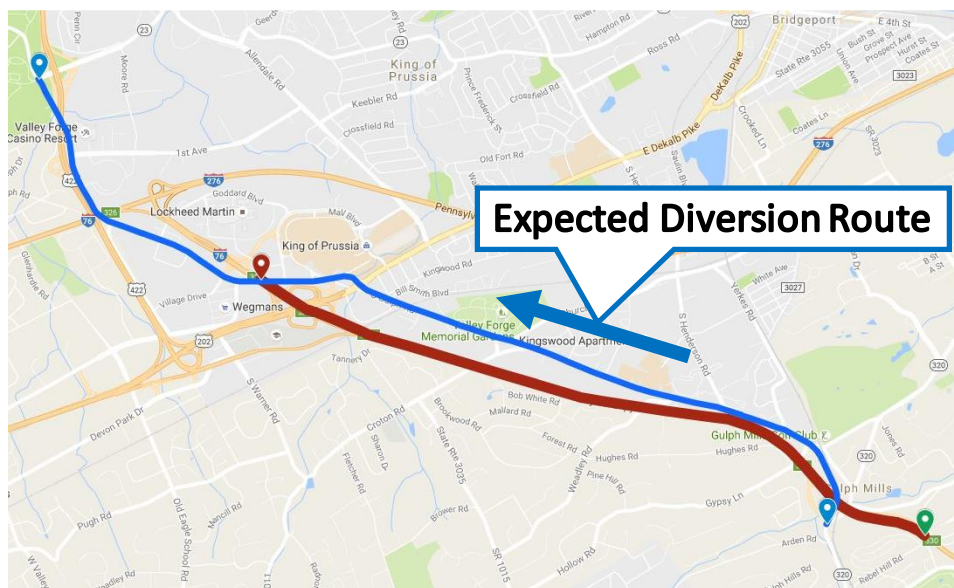


Figure 53. I-76 West accident impact (red) and expected Gulph Rd detour (blue)

The impact of this incident on the travel time along this arterial can be analyzed by comparing it to another typical day, as shown by the travel time comparison tool in Figure 54. As illustrated by the graph, there is a 2.5 minute increase in the median travel time during the detour period in the westbound direction. Figure 55 compares the impact of the detour in the westbound direction using the travel delay monitor. During this period, a major portion of the corridor is operating under 25 mph, with nearly 1 mile having speeds less than 10 mph.



Figure 54. Travel time impacts on westbound Gulph Rd

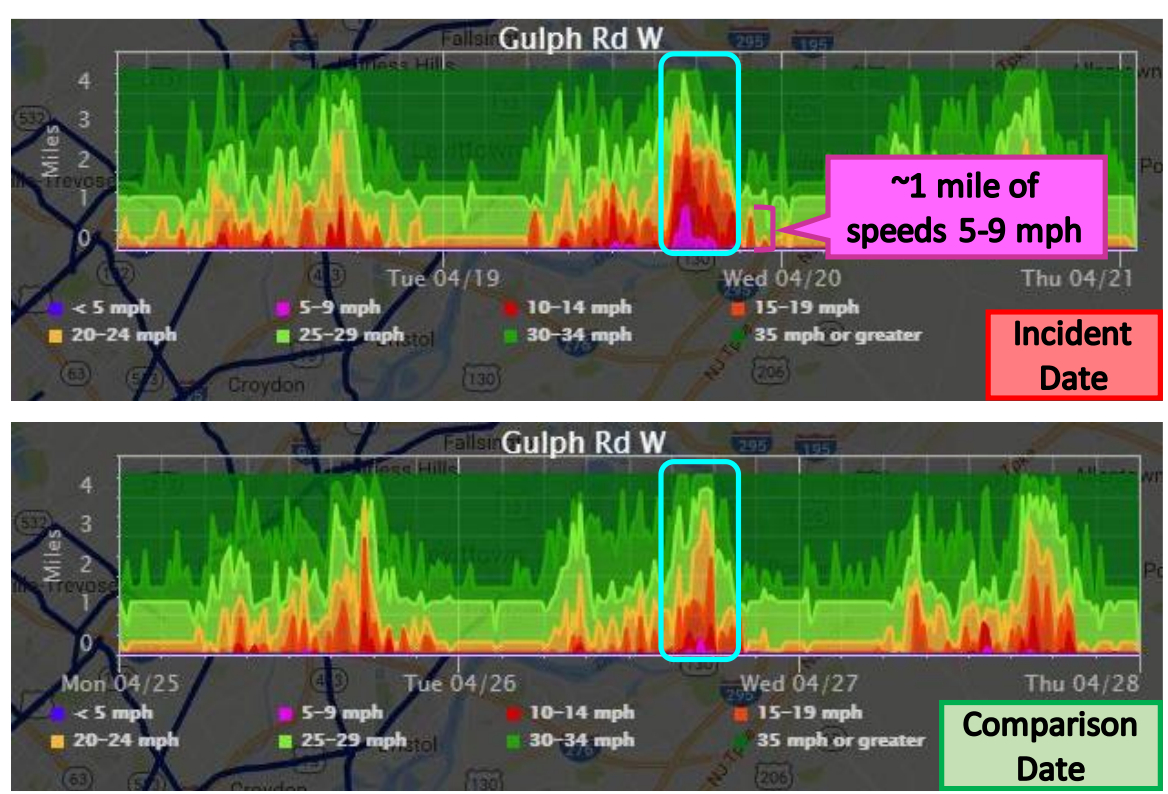


Figure 55. Congestion heat map of westbound Gulph Rd

3.5 Summary

This chapter presents arterial performance measures developed from commercial probe vehicle data; the application of these performance measures is illustrated for the evaluation of traffic conditions on signalized corridors. Three real-time web dashboards were developed:

- An arterial travel time comparison tool, which allows a comparison of travel times during “before” and “after” conditions, using filtering by date, day of week and time of day. The tool uses travel time and CFDs of travel times to illustrate differences between the before and after conditions. This tool is instrumental for assessing the effects of maintenance, operational changes, capital programs and adaptive deployments.
- An arterial ranking tool, which enables the user to view performance of several corridors and rank the corridors according to their normalized median and IQR travel times; the normalized IQR provides a measure of the travel time reliability. The tool produces sorted bar charts based on either performance measure, or a scatter-plot using both criteria.
- An interactive arterial congestion ticker produces a chart of speed distributions on selected arterial routes over time. The cumulative miles of the corridor operating under different speeds is used as performance measure to identify trends. Users can use filtering criteria to focus on specific times of day or segments with specific speed characteristics.

Results from a before/after assessment using these tools on five corridors in Philadelphia showed improvements in travel time and reliability due to the signal improvements. Economic evaluations revealed a user benefit in four of the five corridors, with total annual user savings of \$32 million in these four corridors where improvements were realized. These tools also were useful to allow the identification of the impact of incidents, including the diversion of traffic from an interstate highway to a parallel arterial. The resulting increases in travel time on the arterial was evident through the use of the travel time comparison tool and the congestion ticker.

With the increasing emphasis on data-driven, outcome-oriented performance analysis, tools such as the ones described in this chapter are useful to for transportation agencies to evaluate the performance of the system at both network level and individual corridor level. The performance measures also provide better information about facility operations, which allows informed decisions about system investments and independent analysis of operations that are improved by investments.

CHAPTER 4. PERFORMANCE MEASURES FOR AIRPORT SECURITY SCREENING

This chapter discusses performance measures to identify trends and support resource allocation at airport security checkpoints using wait time data collected from the Cincinnati/Northern Kentucky Airport (CVG). Some of the information presented in this chapter was presented at the 12th Intelligent Transportation Systems European Congress meeting held in Strasbourg, France in 2017 [68].

4.1 Introduction

The Transportation Security Administration (TSA) and airports face challenges deciding how to allocate scarce resources to assure both safety and positive travel experience. In 2009, there were approximately 46,000 full-time equivalent security screeners employed by the Transportation Security Administration (TSA); this number was limited by legislation in 2003 [69]. The fully loaded estimated cost of these screeners, in 2009 dollars was \$2.6 billion [70]. Based upon this 2009 data, worldwide annual spending on airport security is on the order of \$10B (2017 dollars).

Security protocols have undergone rapid changes over the past decade. Programs to expedite travelers through security screening such as TSA's PreCheck (introduced in 2011) and Customs and Board Protection's Global Entry (pilot program announced in 2008) [71] reduced security wait time for passengers and allowed TSA to operate within staffing constraints. The Managed Inclusion program, which allowed passengers that were not on known traveler lists or given PreCheck based on automated risks assessments to utilize expedited screening under certain circumstances [72], also facilitated passenger screening activities.

Airport security has faced a number of challenges. In 2015, internal tests by TSA indicated that 95% of explosives, weapons and banned items were able to make it through screening undetected [73]. In the fall of 2015, TSA underwent revisions in the security screening protocol that reduced the use the Managed Inclusion program. Under the revised protocol, fewer passengers were included in the Managed Inclusion program,

and additional TSA resources were required to allow pre-screening by TSA canines or selection by a TSA behavior detection officer who administers explosives trace detection tests prior to using the TSA expedited screening. These changes came while passenger enplanements continued to increase and while TSA staffing requirements remained constrained due to legislative caps.

Increasing passenger enplanements, PreCheck enrollments that did not meet projections, and rollbacks for participants in the Managed Inclusion program resulted in increased security wait times. This became dramatic by Spring 2016, when security wait times not only affected individual travelers but also affected airline operations. American Airlines testified that more than 70,000 passengers had missed flights, and more than 40,000 were delayed, as a result of the TSA security challenges [74]. TSA increased staffing levels and partnered with American Airlines, United Airlines, and Delta Airlines at major airports to improve operations. Airlines provided additional staffing for non-security related tasks such as stacking bins, and partnered to implement more efficient security configurations at major hub airports [75], [76].

In May 2016, Congress approved a request from TSA to reprogram \$34 million to hire 768 new security officers by June, and to pay overtime to existing security officers, to enable the screening of 220,000 additional passengers per day. In June 2016, Congress approved \$28M in additional funds to hire additional full-time screeners and add security lines [77]. These additional funds, along with the changes that resulted from the involvement of the airlines and the addition of modern security lines at major hub airports have substantially reduced wait times, which averaged 30 minutes or less for 99% of travelers, and 15 minutes or less for 93% of travelers [78].

Using present technology, many major airports in the US are capable of measuring security wait times in real-time [79], [80]. Most systems are oriented towards providing real-time displays, and very little technical literature has been developed on the best practices for management oriented performance dashboards. This study uses 12 months of data from the CVG airport to develop performance measures that present an outlier based reporting model for both standard and expedited (TSA Pre-Check) security lanes.

4.2 Motivation and Scope

Many airports now provide real-time information on parking and security wait time, and in some cases real-time travel time within the terminal from the passenger's current location to the gate. Although TSA uses traditional data collection techniques for many reporting requirements, several airports have deployed systems to collect security wait time data for their own use. Security checkpoint wait time data can exceed 5,000,000 records/yr for large multi-terminal airports; however little has been done regarding scalable techniques that utilize this "big data" for management decisions and to identify opportunities to improve the predictability (reliability) of airport security wait times. This study examines over 700,000 wait time records collected over 12 months at CVG to compare the efficiency and reliability of expedited screening relative to standard screening. The study also develops specific performance measures to identify trends and support resource allocation at the airport security checkpoints.

4.3 Methodology and Data Collection

4.3.1 Airport

CVG is a public international airport located in Hebron, Kentucky, and serving the Greater Cincinnati metropolitan area in the Tri-State region of Indiana, Ohio, and Kentucky. CVG has more than 170 daily departures to over 50 non-stop destinations around the world, and is the only airport in the region to provide non-stop flights to Europe. It is the largest airport in the region considering both airline service and geographic size (it covers 8,000 acres) CVG was the first airport in the Tri-State region to introduce the TSA Pre-Check service in 2012 [81]. During 2015, there were 20 carriers at CVG; Delta was the primary carrier (27% market share) followed by Endeavour (17%) and Frontier (10%). CVG underwent a 15% increase in passenger arrivals as well as a 31% increase in freight during 2017, compared to 2016 [82].

4.3.2 Data

This airport provides an excellent case study for security wait time due to its early adoption of automated wait time data collection and the availability of historical wait

time data. The wait times for both the standard and expedited service were obtained from a collection of sensors located throughout the security area that anonymously monitored the movement of consumer electronic devices. The passenger volumes discussed in this paper are based on data collected and reported hourly by TSA.

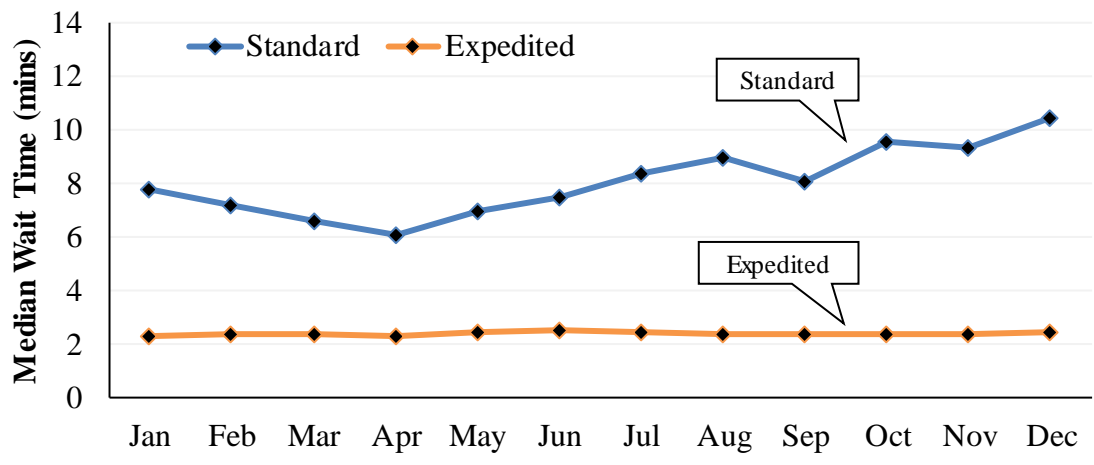
4.3.3 Methodology

Data obtained from the airport was aggregated over one-minute intervals. The statistical distribution of this data was used to compare the efficiency and reliability of expedited screening compared to standard security screening using descriptive statistics such as bar charts, line graphs and CFDs. The study also compared the fluctuation of wait times with the checkpoint volumes. Hypothesis testing was conducted to check whether there was a significant difference in the expedited and standard wait times. Finally, the median wait time is used as a performance measure to identify periods with operating irregularities.

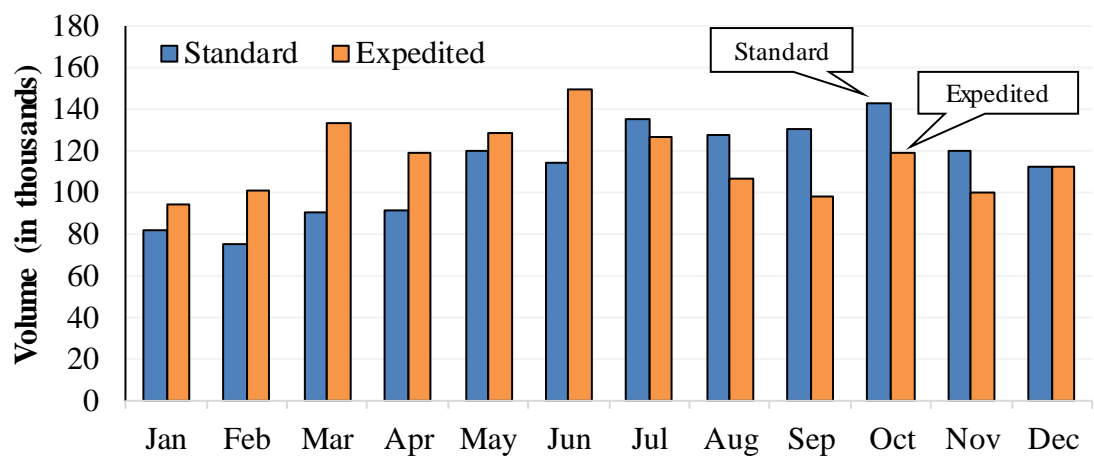
4.4 Results

4.4.1 Wait time

Figure 56 shows the median wait time and the corresponding checkpoint volumes observed during each month for both the standard and expedited screening. The expedited service handled the majority of the passenger volume and had lower wait times compared to standard service. Although the standard wait times fluctuate throughout the year, the expedited wait time remains relatively constant, even during holiday season periods when there are high passenger volumes.



(a) Average Wait Time



(b) Checkpoint Volume

Figure 56. Median wait time and checkpoint volume in 2015

A one-sided hypothesis test was conducted to determine whether the mean wait time for the expedited screening is lower than the mean wait time for standard screening. The test was conducted using the wait time data aggregated for each hour (from 5:00 to 20:00) during the 12 month period. Based on the central limit theorem, the distribution of the means is approximated by the normal distribution when the sample size is over 30 [83]. The null and alternate hypothesis are as follows:

$$H_0: \mu_e - \mu_s \geq 0$$

$$H_a: \mu_e - \mu_s < 0$$

where μ_e and μ_s are the mean for the expedited and standard security wait times, respectively. Descriptive statistics for wait times for the expedited and standard service are shown in Table 7.

Table 7. Descriptive statistics of the expedited and standard wait times

	Expedited	Standard
Mean (\bar{X})	2.43 mins	8.84 mins
Variance (s)	0.13 min	15.86 min
Standard Deviation	0.36 mins	3.98 mins
Observations (n)	5932	5917
Median	2.37 mins	7.82 mins
25 th percentile	2.19 mins	6.21 mins
75 th percentile	2.59 mins	10.29 mins

The test statistic is given by:

$$Z^* = \frac{(\bar{X}_e - \bar{X}_s) - (\mu_e - \mu_s)}{\sqrt{\frac{s_e^2}{n_e} + \frac{s_s^2}{n_s}}} = \frac{(2.43 - 8.84) - 0}{\sqrt{\frac{0.13}{5932} + \frac{15.86}{5917}}} = -123.30 \quad (5)$$

The absolute value of test statistic is much larger than 2.326, the critical value for a one-tailed test at the 1% significance level, and so the null hypothesis is rejected and there is evidence that the expedited wait times are lower than the standard wait times.

4.4.2 Reliability

Security screening should be quick, reliable and consistent. The reliability of the wait times can be characterized by examining the standard deviation and the IQR. As mentioned in Chapter 3, the IQR is the difference between the 25th percentile and 75th percentile of all the wait time observations. The reliability can also be graphically assessed by analyzing the CFDs. A more vertical CFD suggests that the wait times are reliable and consistent throughout the data, whereas an S-shaped or more horizontal curve indicates a higher variance in in the data.

Examining the descriptive statistics from Table 7, the mean wait time for the standard service is 8.8 minutes with a standard deviation of 3.9 minutes, whereas the

mean wait time for expedited is 2.4 minutes with a standard deviation of 0.3 minutes. The low standard deviation of the expedited service illustrates a high reliability, which translates to a very predictable travel time through security. The IQR value for the expedited service is less than 0.5 minutes compared to the 4.0 minutes for standard service (Table 7), indicating that expedited service is substantially more reliable. This is also graphically illustrated by the near vertical expedited CFD compared to the S-shaped standard CFD in Figure 2.

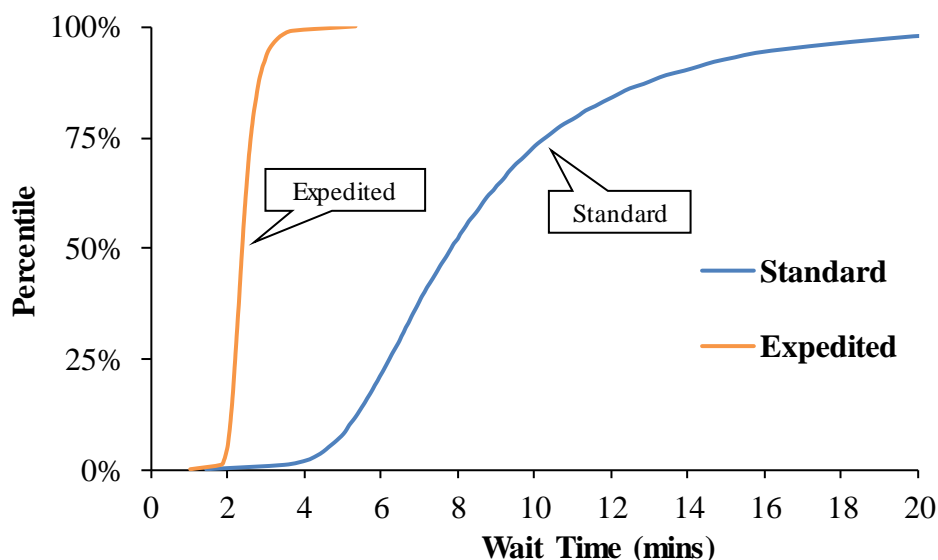


Figure 57. CFD comparing the standard and expedited wait times

4.4.3 Hours with high median wait time

Although the summary statistics shown in Figure 56 provide a high-level overview of trends, the large number of off-peak hours in a day can mask peak hour performance characteristics. Many airports have a stated goal of providing wait times less than 20 minutes [84]. This performance threshold offers a reasonable basis for identifying operating irregularities. Figure 58 shows the hourly median wait times for both standard and expedited security lines for one day (in this example, 16 Dec 2015). The boundaries of the wait time ranges are shaded dark green, light green, yellow, orange, and red corresponding to wait times of 0-5, 5-10, 10-15, 15-20 and >20 min. Table 8 shows summary statistics for these same ranges and summarizes the number of hours the standard and expedited security lanes operate in each of these wait time conditions. For

example, the wait time for the standard security line plotted in Figure 58 is less than 5 minutes for only one hour of the day, the 19:00 hr. The wait time for the expedited security line is less than 5 minutes for all 15 hours of operation. There are five different hours with wait times over 20 minutes (5:00, 6:00, 8:00, 10:00 and 17:00) for the standard security line, and there is no hour in which the median wait time for expedited screening exceed five minutes.

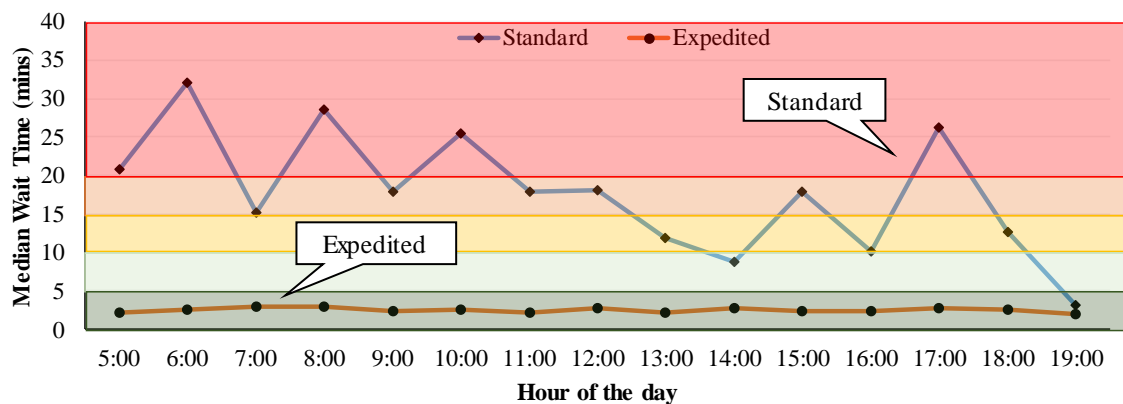


Figure 58. Median wait time for each hour on 16 Dec 2015

Table 8. Frequency table for median wait time on hourly basis for 16 Dec 2015

Wait time (mins)	Standard (hrs)	Expedited (hrs)
< 5	1	15
5 – 10	1	0
10 – 15	3	0
15 – 20	5	0
> 20	5	0

Figure 59 depicts the proportion of hours that the standard and expedited security lines operate with different wait time ranges during the entire calendar year. The expedited screening provided wait times less than 5 minutes for 99.9% of the year, again underlining its reliability and consistency over the standard screening.

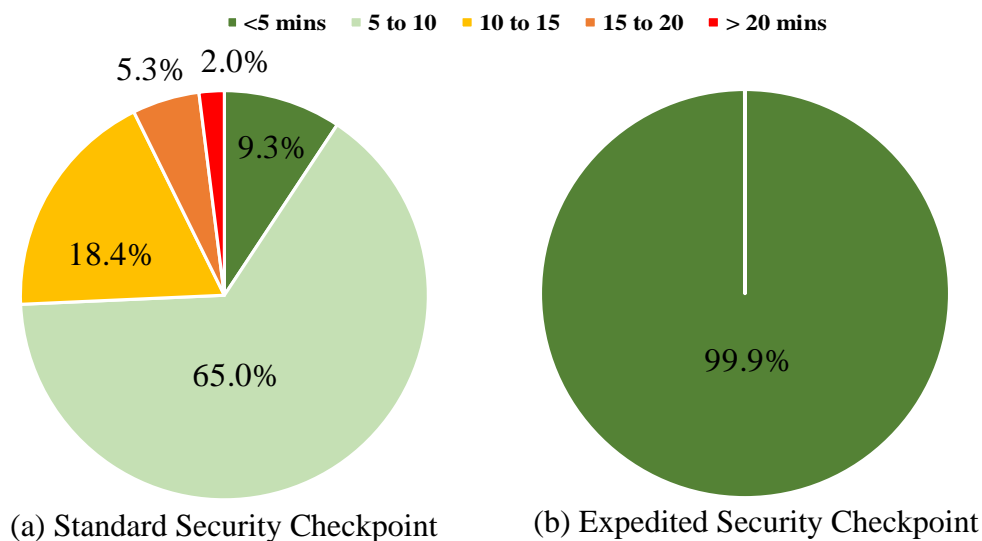


Figure 59. Proportion of checkpoint wait time for the calendar year 2015

4.4.4 Performance dashboards

Figure 59 (a) illustrates that approximately 2% of the hours operate with a wait time greater than 20 minutes for standard service; additional information regarding wait time for each month is shown in Figure 60. October, November, and December all have more hours with median wait times greater than 20 minutes. This could result from policy changes that limited the use of managed inclusion and a higher volume of travellers during the holiday season in November and December. The month of December averages 1.5 hours per day with wait times greater than 20 minutes.

Figure 61 and Figure 62 provide further details for wait times for individual weeks, and days. As can be seen in Figure 61, October, November, and December combined had nearly 90 hours with more than 10 minutes of wait time. In Figure 62, it can be seen that wait times on Wednesday during week 51 (Dec 13 to 19, 2015) included 5 hours with a median wait times over 20 minutes and another 5 hours with median wait times between 15 and 20 minutes. The Wednesday data in Figure 62 corresponds to the example data introduced in Figure 58 and Table 8.

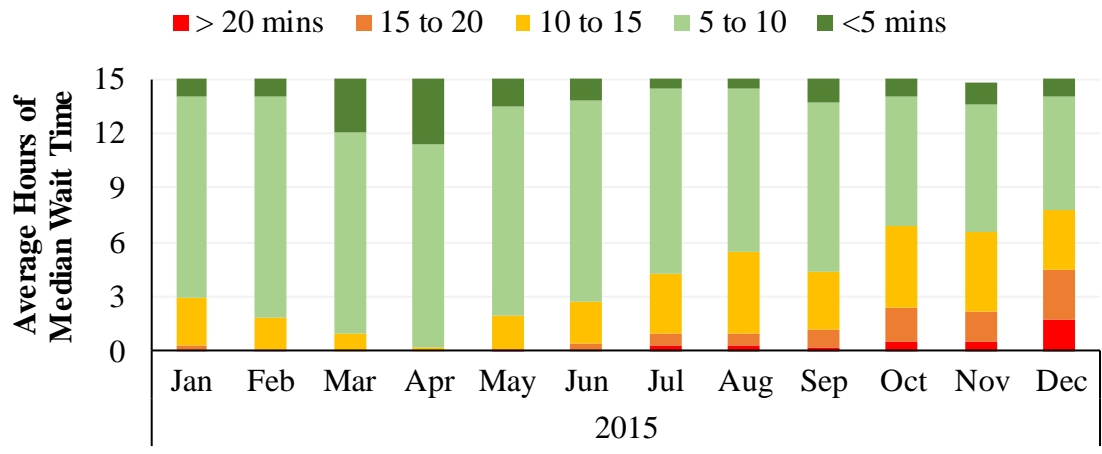
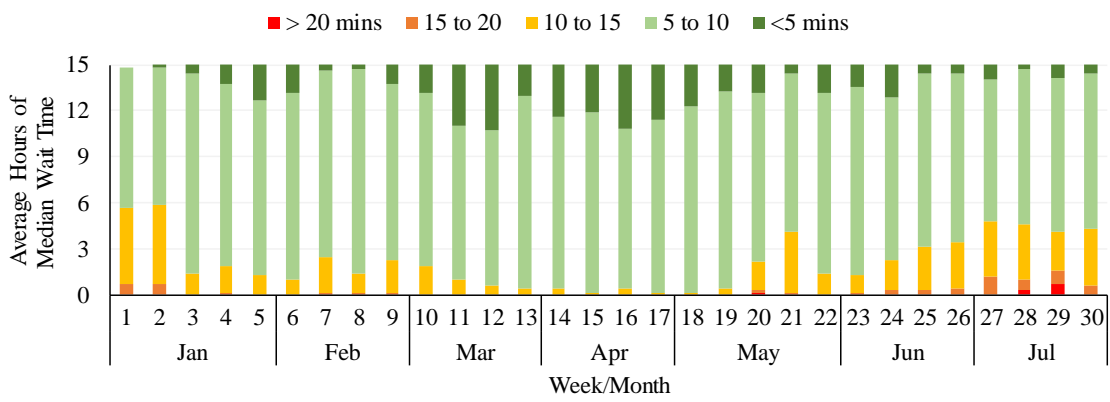
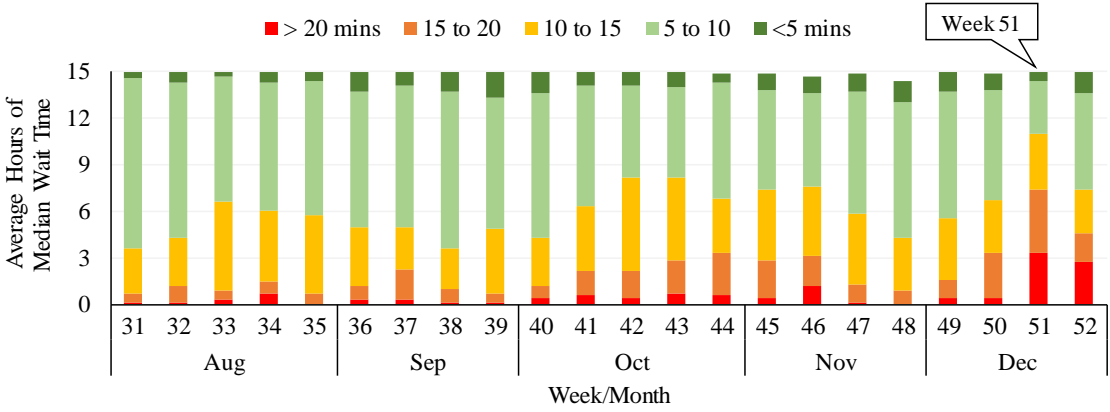


Figure 60. Average hours of median wait time at standard security checkpoint by month



(a) January – June 2015



(b) July – December 2015

Figure 61. Weekly summary of average hours of median wait time

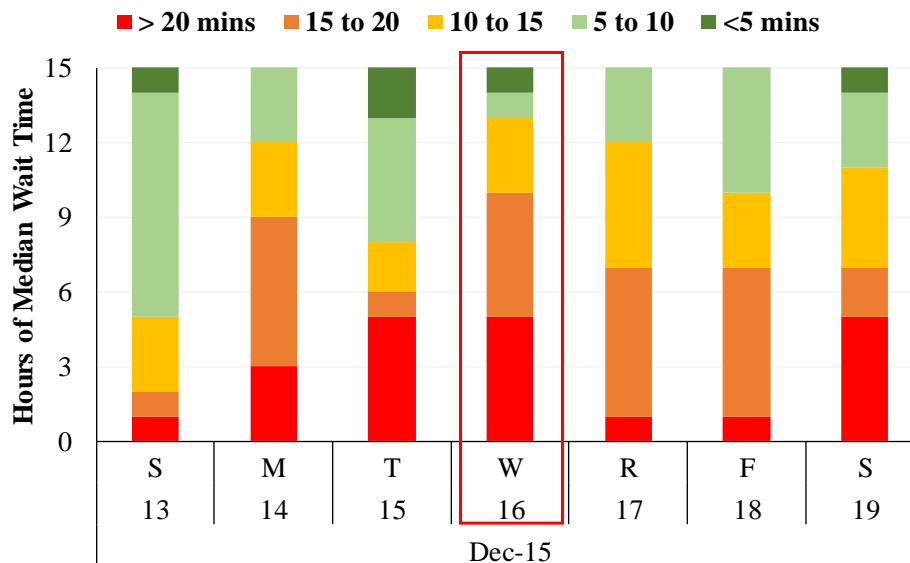


Figure 62. Hours of median wait time for week 51 in December 2015

These performance dashboards can be used to identify times of day with high median wait times. Systematically identifying the periods when wait times are high may be an impetus to examine operational strategies and other opportunities to improve service. With the ultimate goal of reducing screening wait times, one operational strategy will be to identify target wait times for both the standard and expedited screenings. From the 12-month data analysed in this research, nearly 75% of the standard wait time hours were found to provide a wait times less than 10 minutes, whereas 99.9% of the expedited wait time hours were less than 5 minutes. One operational strategy would be to target a wait time fewer than 10 minutes for 90% of the standard screening by improving management practices and resource allocation.

4.5 Summary

This study uses median wait time and reliability as performance measures to compare the expedited screening (including TSA PreCheck) and the standard screening, using 12 months of wait time data collected at the CVG. The wait times for both the standard and expedited service were obtained from Bluetooth sensors in the airport. These sensors anonymously monitored the movement of consumer electronic devices from the non-sterile area (before security) to the sterile area (after security) to provide an estimate of the security wait time. The wait times for the expedited screening were lower

than for standard screening. The standard screening had a median wait time of 7.8 minutes, and the expedited screening had a median wait time of 2.3 minutes, resulting in a median wait time savings of nearly 5.5 minutes for travellers using expedited screening. The expedited screening was also more reliable, with an IQR of 0.4 minutes, compared to 4 minutes for the standard screening.

Agencies have begun to use “big data” to develop dashboards for improving long-term management and capital investment prioritization. Airports and security agencies have followed similar models with wait time data, first displaying real-time wait time on signs and more recently, web pages [85] and mobile applications [86]. This real-time information may be reassuring to passengers when wait times are short, and may also be used by airports and TSA to identify trends and support resource allocation to optimize service. This research also presents a performance dashboard that can be used to identify periods of day with high median wait times. Systematically identifying outliers could provide the framework for identifying opportunities to examine operational procedures that could improve service. The availability of data, combined with clear operational goals, may support management practices and resource allocations that reduce wait times.

One of the stated goals of the TSA is to improve service by reducing wait times at the airport. The analysis and the performance measures discussed in this chapter will help the TSA to address this specific goal of reducing wait times. In order to improve the overall performance of security at airports, proper attention should be provided to other performance measures as well.

CHAPTER 5. PERFORMANCE MEASURES FOR THE ASSESSMENT OF A UNIVERSITY BIKE SHARE SYSTEM

This chapter proposes a series of performance measures that were developed to monitor and assess a bike sharing program initiated in August 2015 at Purdue University.

5.1 Introduction

Bike sharing is an environmental friendly form of public transportation that provides the users with the flexibility of accessing public bikes at unattended stations. Commonly found in large cities, the bike share programs are spread over various neighborhoods with multiple stations that allow the users to check-out and check-in the bikes at different stations.

Bike sharing programs are gaining immense popularity all over the world. It is estimated there are around 1 million bike share bicycles in 712 cities worldwide. In the United States, there were around 3,378 bike share stations operating across 104 cities by 2016 [87]. A more recent article reveals that the bike share program has evolved to 119 cities with 4,789 stations nationwide [88]. In 2008, Washington D.C. introduced SmartBike DC (later launched as Capital Bikeshare) with 10 stations and 120 bikes, the first ever modern bike share program using automated self-serve kiosks and card swipe access. New York's Citibike is currently the largest bike share program in the nation with over 12,000 bikes and 750 stations that spreads across 60 neighborhoods.

More recently, the proliferation of smart bikes, which incorporate all the technology and necessary electric components in the bike itself as opposed to the dock, has improved the pickup and return of bikes to a larger area [89]. Similar to the car-sharing systems, the bike share programs also typically cover the capital costs including the bikes and docking stations as well as providing timely maintenance.

Bike share programs have rapidly emerged as low-cost transportation alternative that can bolster public transit and increase access to opportunities. According to the National Association of City Transportation Officials (NACTO), it estimated that around 88 million bike share trips were made in the United States since 2010 [89]. The study

also found that most of the trips remained short, particularly in a larger system that caters the transfer to other public transit modes. A study conducted on public bike sharing perceptions and travel-behavior changes in four cities found that bike sharing was an enhancement to public transportation and improved transit connectivity [90]. A survey conducted by Zagster, one of the leading bike share programs in the nation, found that a majority of the 3500 respondents felt more connected to their communities as a result of bike sharing [91].

Studies have shown that bike share programs are continuing to be a huge success in both large and small cities [92], as well as universities. A study conducted in the city of Fargo and North Dakota State University (NDSU) found that the bike share program was highly successful in its first two years. The program was the primary or secondary mode of transportation for students and improved the livability in Fargo by providing more travel options for the NDSU students [93].

There are also numerous advantages associated with bike sharing programs including reduced congestion, emissions and fuel use [94]–[96], improved health outcomes [97], [98], and flexible mobility and convenience [99]. As more cities and users adopt bike sharing programs as an eco-friendly alternate mode of transport, it necessary to have performance measures to evaluate and manage the overall progress of the system.

5.2 Motivation and Scope

With the growth of bike share programs, there is little in the way of objective performance measures to assess the programs. This chapter discusses a number of performance measures that can be used to monitor operations and improve asset management and resource allocation of bike sharing programs. This study uses data over a period of 2 years from a bike share program initiated at Purdue University in August 2015.

5.3 Methodology and Data Collection

Figure 63 shows a campus map with the location of docking stations. The program, which initiated with 77 bikes across 13 locations in 2015 expanded to over 170

bikes and 20 locations in 2017 with more than 1600 registered users. During the study period, there were nearly 41,300 rentals over a total duration of approximately 50,000 hours.

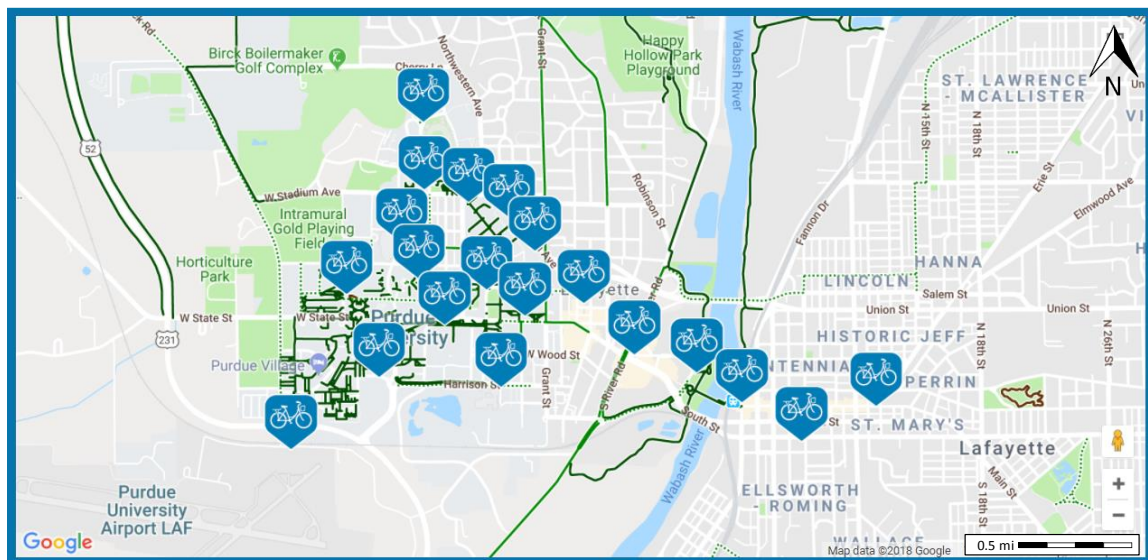


Figure 63. Campus map with location of bike share docking stations

The program kicked off with an annual membership fee of \$25 which allowed free trips for a duration of 3 hours (6 hours on weekends), after which trips were charged with \$2/hr up to a maximum of \$10/ride. The research team performed a preliminary analysis in early 2016 and found that some users were constantly renting the bikes for longer durations (>24 hours). As a result, a revised pricing policy was implemented in August 2016, which resulted in an annual membership fee of \$35 with free trips under a duration of 2 hours. An additional clause was inserted to charge an overtime fee of \$30 for keeping the bikes over a 24-hour period. Fig 2 shows the pricing policies for 2015 and 2016.

Anonymized rental data including user code, bike code and start/stop times of a ride was obtained from the bike share provider. This data was used to analyze the usage patterns as well as identify performance measures for asset management. Performance measures such as user rentals, rental duration over time and bike usage was developed for nearly 2 years of data from August 2015 to May 2017.



Figure 64. Variation in pricing scheme over the study period

5.4 Performance Measures

The following performance measures were developed and detailed in the following section:

5.4.1 User rentals

Of the 1626 registered users, nearly 20% of them utilized the service at least once and around 6% utilized it more than 100 times, with four of them extending to more than 500 times (Figure 65). The number of rentals also varied among users, ranging from a single rental to a maximum of 18 rentals per day.

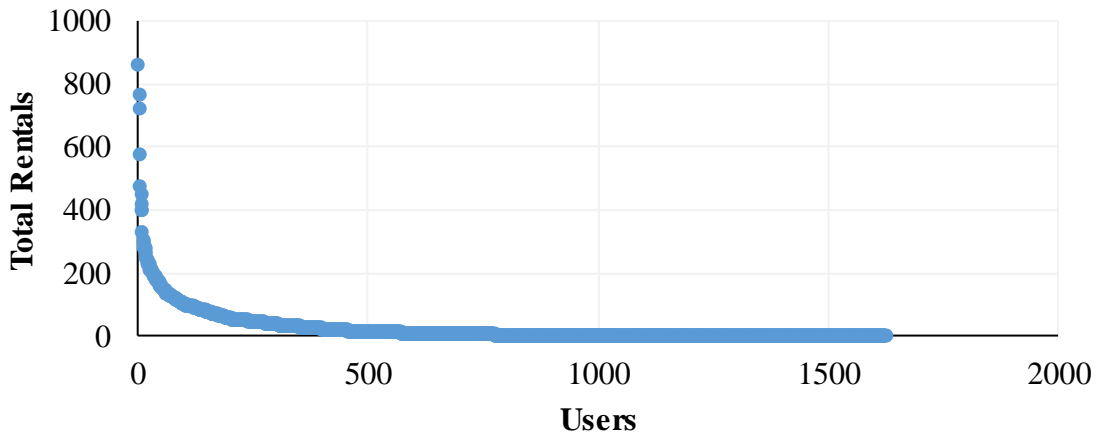


Figure 65. Rentals by user

5.4.2 Daily user rentals

Figure 66 shows the variation of total daily rentals over the study period. On an average there are 71 rentals per day. For the 2015-2016 academic year, rentals peaked in the September (callout i) but fell off in October and November with colder weather (callout ii). Rentals also remained low in the early months of spring semester, but slightly picked up in warmer months of April and May (callout iii). Summer also saw a low number of rentals (callout iv), probably due to fewer students in campus. January and February are winter hibernation periods where the bikes are mostly under maintenance.

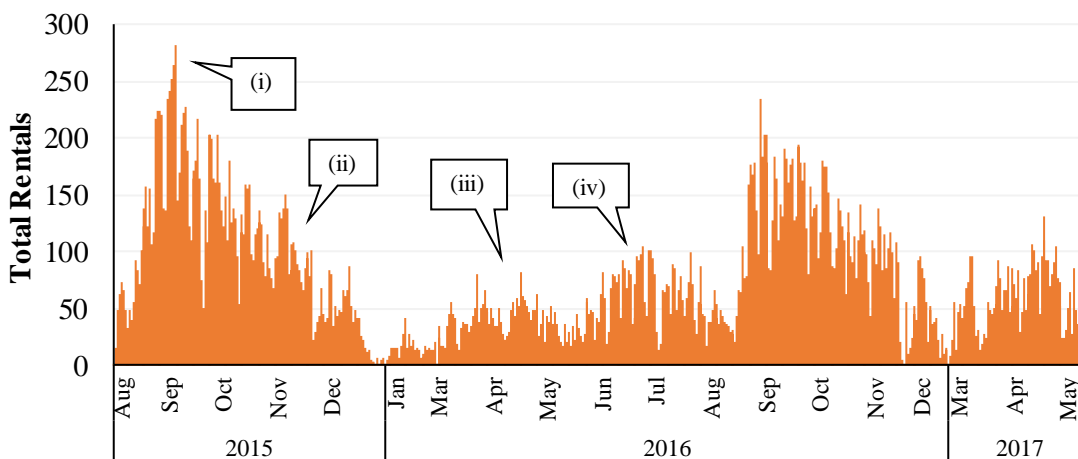


Figure 66. Total daily rentals

Similar trends can also be seen for the 2016-2017 academic year, with slightly higher rentals in spring 2017 compared to spring 2016. The average daily rentals in

spring 2016 was 32.89 in contrast to 57.59 rentals in spring 2017. This could be a result of the better weather conditions in spring 2017. Figure 67 compares the median temperatures [100] during the months of spring 2016 and 2017. This is consistent with earlier studies which found that adverse weather conditions such as cold temperatures, rain and increased wind speeds decrease bike share activity [101], [102].

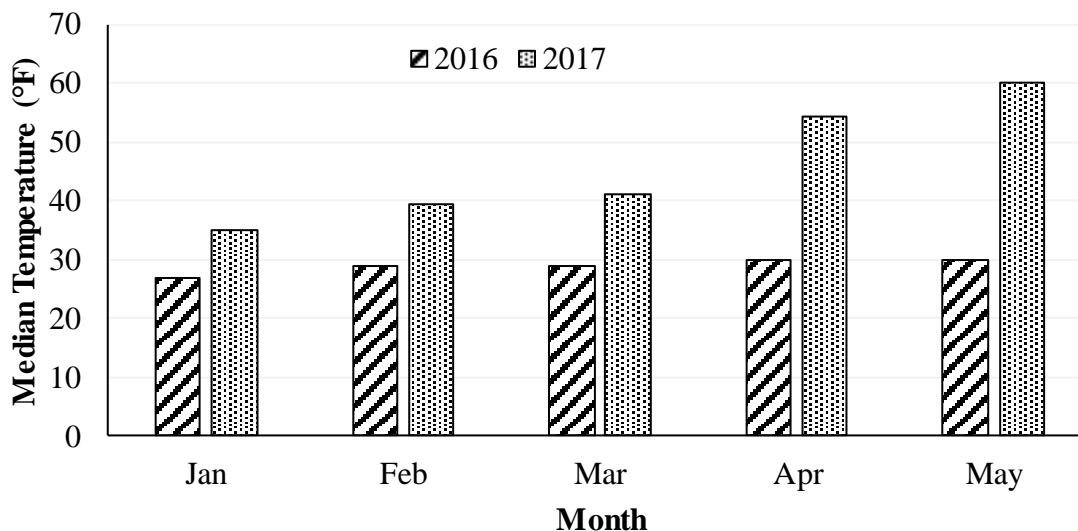


Figure 67. Median temperatures during spring 2016 and 2017

5.4.3 Daily rental duration

Figure 68 provides an overview of the daily rental duration during the study period. This performance measure is an indication of how the usage varied across the users. The trends are similar to the daily rentals (Figure 66) with peak usage during initial months of fall period, which gradually decreases during the winter months. On comparing the two fall periods (callouts i and ii, respectively), it is interesting to note that the usage drops significantly during fall 2016. Even though average number of rentals slightly dropped by 6.86% (102 in fall 2015 compared to 95 in fall 2016), the average rental duration dropped by 41.14% (158 hours in fall 2015 compared to 93 in fall 2016). This is likely a direct impact of the new strategies and pricing policies implemented in 2016 to curtail long rentals that were identified by the project team in early 2016.

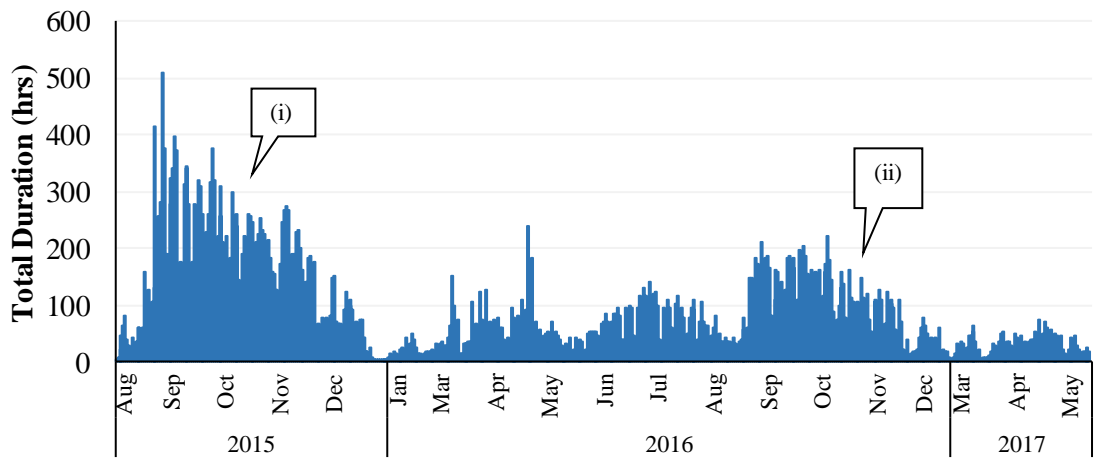


Figure 68. Daily rental duration

5.4.4 Cumulative plot of rental duration

The cumulative plot of the rental duration is another performance measure that can be used to assess the impact of the policy changes on the user behavior. The dashed line on Figure 69 shows the cumulative rental duration for the 2015-16 academic year before the implementation of the policy change. As seen, 95.9% of the users return the bikes before the 3-hour free period (callout ii). Only 27% of the users rented the bikes for small trips with duration less than 15 minutes (callout i). However, there were 25 rentals with a duration more than 24 hours (callout iii), the maximum being 92 hours.

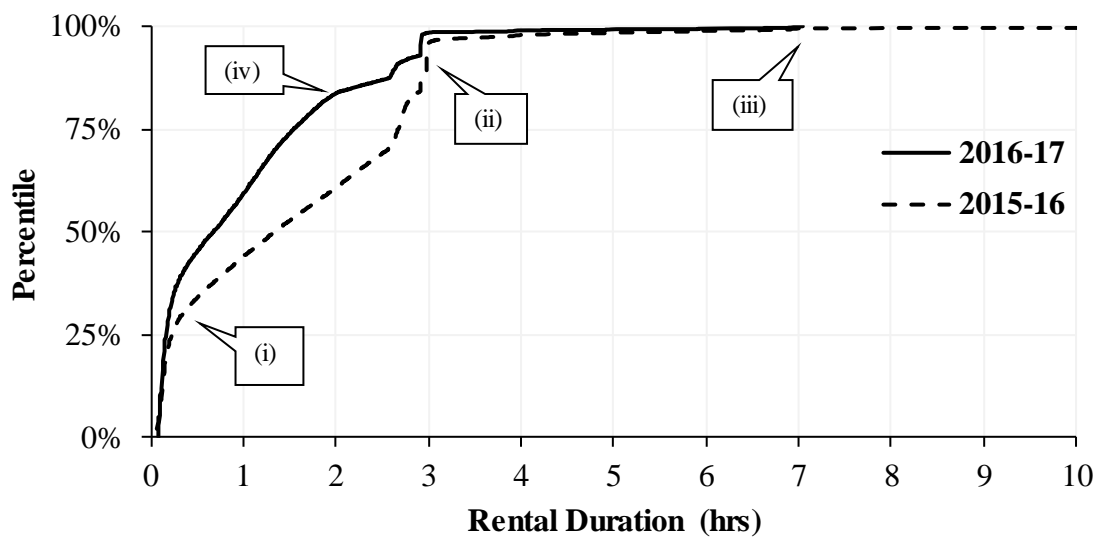


Figure 69. Cumulative plot of rental duration before policy changes

As mentioned earlier, policy changes were initiated in August 2016 to curtail long rentals. As a result, the duration of free rides was limited to 2 hours and a \$30 overtime charge was added for keeping the bikes for over 24 hours. The solid line in Figure 69 shows the cumulative rental duration for the academic year 2016 – 17. The shorter trips increased with around 35% of the users renting the bikes for 15-minute trips (callout i). Only 83.7% of the users returned the bikes before the 2-hour free limit (callout iv), while another 13% returned before the 3-hour period (callout ii), which shows that some users were still unaware of the new policy changes. The biggest impact of the policy change can be witnessed over the longer rentals, with zero rentals over a 24-hour period. The maximum rental duration during this period was 7 hours (callout iii).

5.4.5 Bike usage

Figure 70 illustrates the average number of bikes in use by 15-min periods of the day. This is a strong performance measure that monitors the variation of bike usage over a day. Bikes were rented at all hours of the day, but usage peaked between 11:00 and 19:00 on average, with certain spikes (callout i and ii) that correspond to class schedules. Weekday patterns (Monday-Friday) are very similar, with decreased usage on weekends. Late-night rentals are more likely on Friday and Saturday nights.

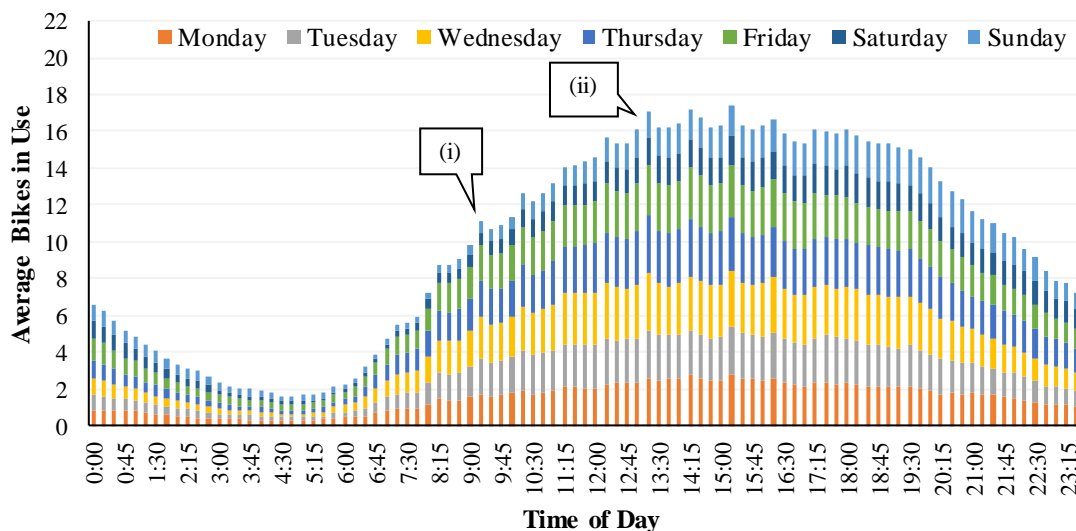


Figure 70. Average bikes in use by 15-min periods

5.4.6 Bikes in use

Figure 71 shows a Pareto chart of the bike duration. This chart is a good indicator of the overall performance of the system. The chart shows around 75 of the 176 bikes being used for more than 80% of the time. Among the other 100 bikes, perhaps some might have maintenance issues or might be located in areas with low demand. This is a perfect opportunity for the agency to perform asset management and resource allocations to improve the overall performance of the system.

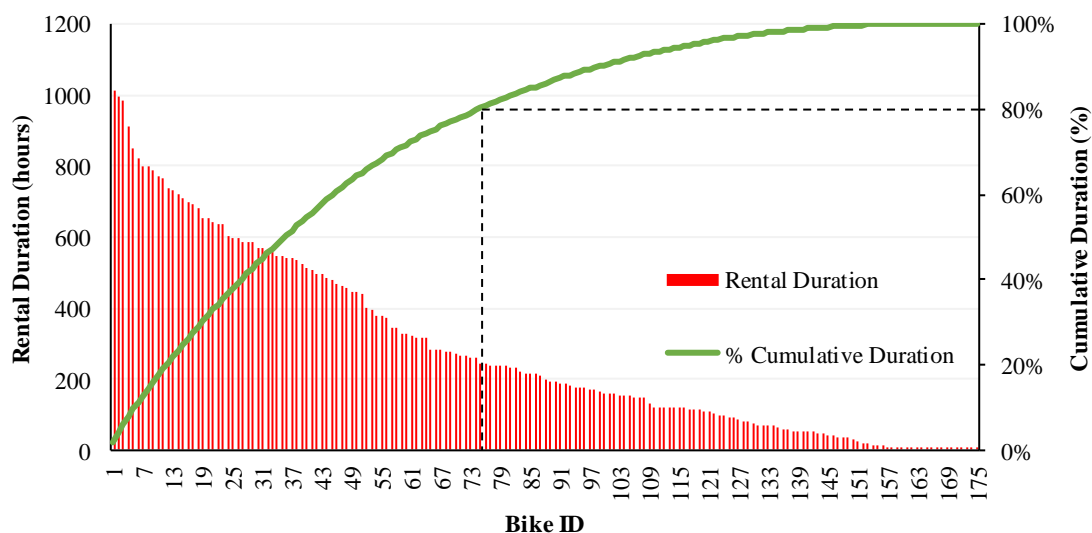


Figure 71. Pareto chart of rental duration by bike

5.4.7 Bikes on ground

The aviation community has a term called aircraft on ground (AOG) to refer to duration that an aircraft is out of service and not in revenue service. We have adopted a similar term called bikes on ground (BOG).

Bikes on ground identifies bikes being unused for a particular cumulative number of days. This is another performance measure that can be used to identify locations with low demand and reallocate these unused bikes to other areas. Figure 72 shows the cumulative number of days for which bikes are unused during the 4-month period between August and December 2015. Each bar within the stacked plot represents a bike. The tool only starts counting when a bike is unused for at least 7 days. Callout i depicts a bike that was unused for up to 100 days and callout ii shows another bike unused for a

period of 69 days. Agencies can set operational strategies to identify bikes with low usage. For example, one operational strategy would be to target bikes unused for a period of one month. Relocating these bikes to areas with higher utilization rates will help in asset management as well as improving the performance of the system. The measurement of elapsed time since a bike was seen will also help to identify potential lost/stolen bikes or those that require maintenance. This metric is expected to be particularly important as the trend increases towards dock less bike sharing and there is risk that bikes are left in remote areas.

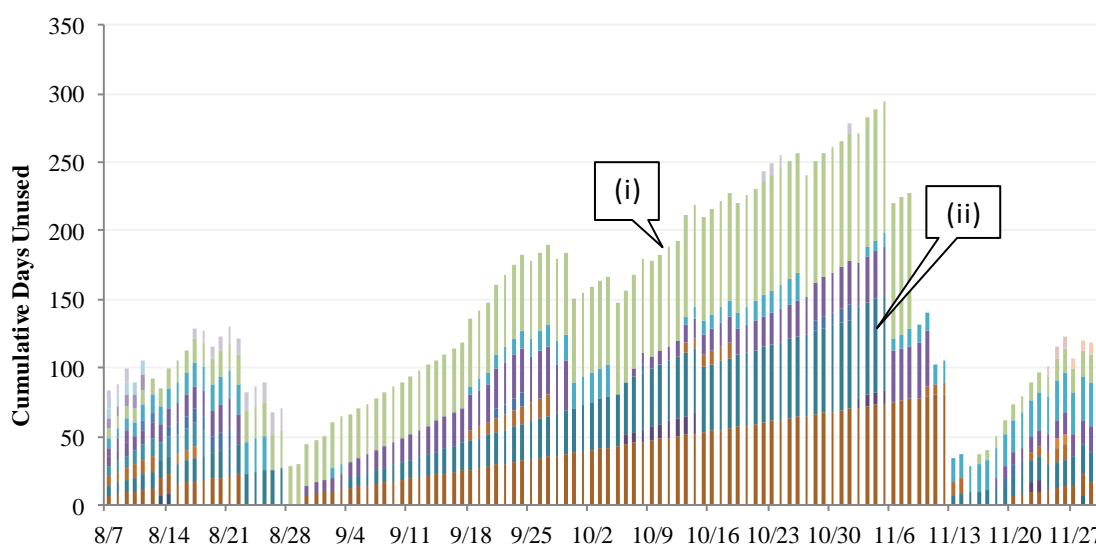


Figure 72. Cumulative days unused by bike

5.5 Summary

This chapter proposes a series of performance measures that were developed to monitor and assess a bike sharing program initiated in August 2015 at Purdue University. Anonymized rental data including user code, bike code and start/stop times of a ride was obtained from the bike share app over a period of 2 years was used to develop the performance measures.

Of the 1626 registered users, nearly 20% of them utilized the service at least once and around 6% utilized it more than 100 times, with four of them extending to more than

500 times. The rentals peaked in September, but fell off in October and November with colder weather.

The performance measures using rental duration provided an insight on the user behavior towards policy changes. There was a sharp decrease in the rental duration when new policies were implemented to curtail long rentals. There were zero rentals with duration more than 24 hours, indicating a total success of the revised policy.

Performance measures on the bike usage was also developed. Bikes were rented at all hours of the day, but usage peaked between 11:00 and 19:00 on average. Weekday patterns (Monday-Friday) are very similar, with decreased usage on weekends. The bikes in use and bikes on ground performance measures can be used to identify inactive and unused bikes that may require maintenance or relocation.

As bike share programs are growing at an astounding pace, city officials and practitioners rely on the performance data from these programs to undertake decisions. With millions of trips made every year, it is necessary to have data visualization tools and performance measures to dissect this big data and evaluate the performance of the system. One direct impact from this study was the cumulative rental duration performance measure, which allowed the agency to understand the user behavior and undertake decisions to improve the system by implementing new pricing policies. The performance measures discussed in this chapter are a valuable tool for decision makers for developing new models that monitor the progress of the system, improve asset management and resource allocations.

CHAPTER 6. DEVELOPMENT OF A CURATED DATASET TO EVALUATE THE PERFORMANCE OF OUTLIER FILTERING ALGORITHMS

This chapter discusses a methodology to develop a curated travel time dataset from high fidelity automatic vehicle location data. This dataset is instrumental in identifying performance measures to evaluate the efficiency of outlier-filtering algorithms. Some of the information presented in this chapter was presented at the 97th Transportation Research Board Annual Meeting in 2018 [103].

6.1 Introduction

Travel time is a key indicator of traffic system performance. Traditional methods of travel time estimation include electronic distance-measuring instruments, automatic license plate readers [104], loop detectors [105], automatic vehicle location [106], floating car techniques [107] and GPS [108], [109].

More recently, other innovative and cost-effective technologies including Bluetooth [110], [111], cellular [112] and dedicated short-range communications (DSRC) [113] have provided reasonably accurate estimates of segment travel time. These methods compute the segment travel time based on matching vehicle IDs at two different locations. This data may include potential outliers with considerable deviation from the average travel time. These outliers can be trip chaining vehicles, devices from non-motorized modes, vehicles using alternate routes, and devices from high occupancy vehicle lanes [114]. Sensor flaws such as faulty communications, time-sync errors and incorrect detections could also lead to abnormal travel times. It is necessary to remove these outliers from a valid dataset to improve the accuracy of the travel time estimates. Various outlier-filtering methods utilizing simple statistical tests have been proposed and adopted in the literature, including median absolute deviation (MAD) [115]–[117] and modified z-score (Mod Z) [118], [119], which deploys percentile and deviation filters to remove outliers based on the variation of travel time from the normal. These filters may not provide satisfactory results when there are few data points, such as during off-peak hours when the sample penetration of the probe vehicles is poor. Adaptive algorithms that vary

across the sampling windows were found to provide better estimates of the travel time [117][116][120]. A report from the Strategic Highway Research Program 2 (SHRP 2) efforts also proposes a robust filtering algorithm to remove the outliers [121].

Most algorithms improve the travel time estimates, but removing outliers is still a challenge since it is often difficult and time-consuming to identify the nature of a trip without a detailed assessment of the waypoints between the origin and destination. Moreover, this level of detail may not be available across all data sources. Simulation has been used to evaluate the performance of specific outlier filtering algorithms [122], but it is challenging to simulate real-world outlier scenarios. This chapter uses high-fidelity crowdsourced GPS data to develop a curated travel time dataset that distinguishes outliers associated with trip chaining. In this study, outliers are defined as trips that leave and later re-enter the roadway of interest, such as those undertaking trip chaining activities or following a route other than the roadway of interest (but within the same endpoints). This curated dataset is then used to examine the performance of three common outlier filtering algorithms. The curated dataset is archived in a digital repository [123] for researchers to analyze the performance of their outlier filtering algorithms.

6.2 Motivation and Scope

Agencies use a variety of technologies and data providers to obtain travel time information. GPS tracking is technically feasible but presents challenges regarding privacy, storage requirements and analysis. More frequently, agencies collect or purchase travel time data based on vehicle identification data, which requires filtering techniques to remove outliers associated with trips that did not follow the route while retaining data associated with incidents or congestion. Although a number of outlier filtering algorithms have been proposed in the literature, their performance has never been evaluated. The objective of this study is to develop a curated dataset (where the outlier status of every data point is known) that can be used to assess the performance of outlier filtering algorithms. Two performance metrics are also including the proportion of true samples rejected, the proportion of outliers correctly identified, are proposed to evaluate the efficiency of three common algorithms, MAD, Mod Z and box/whisker plots. The effectiveness of these algorithms in providing accurate real time estimates of travel time

is also studied. Perhaps more importantly, the publication of this unique curated travel time dataset provides essential open access data for other researchers to use in future studies.

6.3 Methodology and Data Collection

6.3.1 Data

Commercial traffic data providers collect anonymized probe vehicle location data from GPS equipment in vehicles, including personal mobile devices. The data is described by a timestamped latitude and longitude waypoint with a precision of four decimal points (about 36 feet), and a unique arbitrary identifier to link the succession of waypoints generated by a distinct device. A vector of waypoints made by the same device constitutes a trip. This study uses a dataset from a six-week period from 1 May to 12 June 2016, containing over 12 million trips and 980 million waypoints. The data is stored in a relational database with spatial indexing on the waypoint attribute for improved query performance. Figure 73 shows an example of the waypoint data reported during a day on I-94 in northwestern Indiana from the dataset. In this region, 851 trips consisting of 150 thousand waypoints are identified by a virtual geographic bounding box that is 12 by 5 miles.

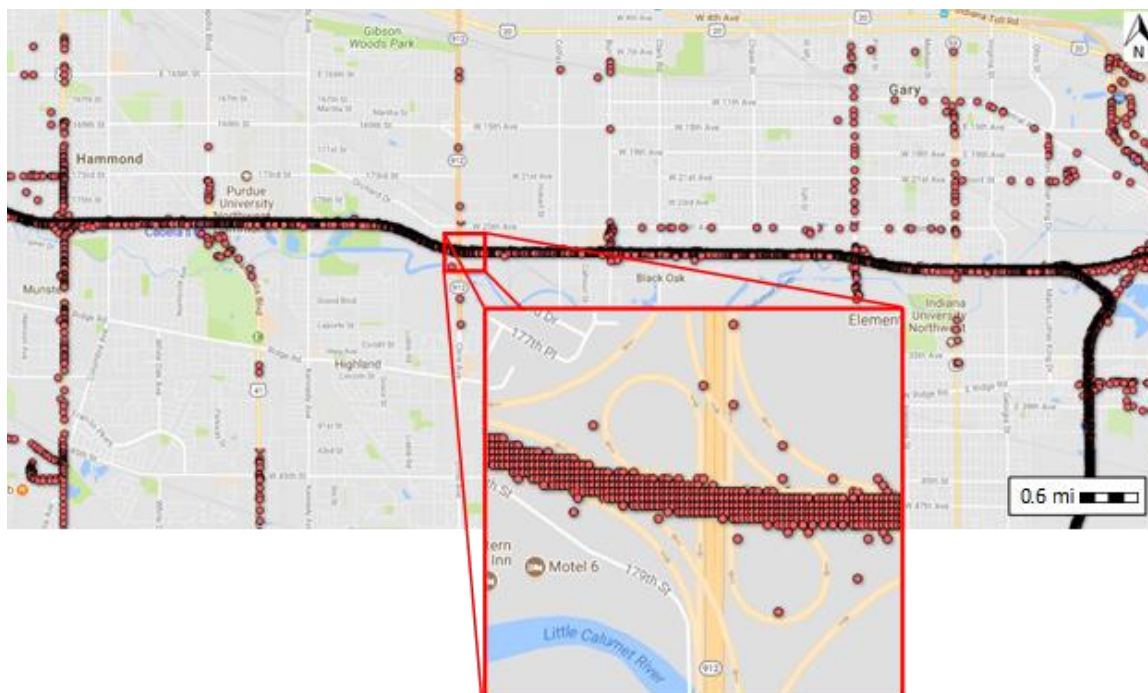


Figure 73. Over 150,000 waypoint data from probes for a day on I-94, IN

6.3.2 Methodology for developing the curated dataset

6.3.2.1 Study Area

The corridor chosen for this study is an 8-mile stretch of roadway between exits 2 and 11 on I-80/I-94 in northwestern Indiana. This is a heavily travelled and congested section with AADT over 190,000; with most of the eastbound traffic heading into Chicago and the westbound traffic towards Detroit and Indianapolis (Figure 74 (a)). The average travel time is 6 to 7 minutes. There are 4 interchanges along each direction of travel, where the motorists can enter and exit the interstate.

6.3.2.2 Defining Zones and Cordons

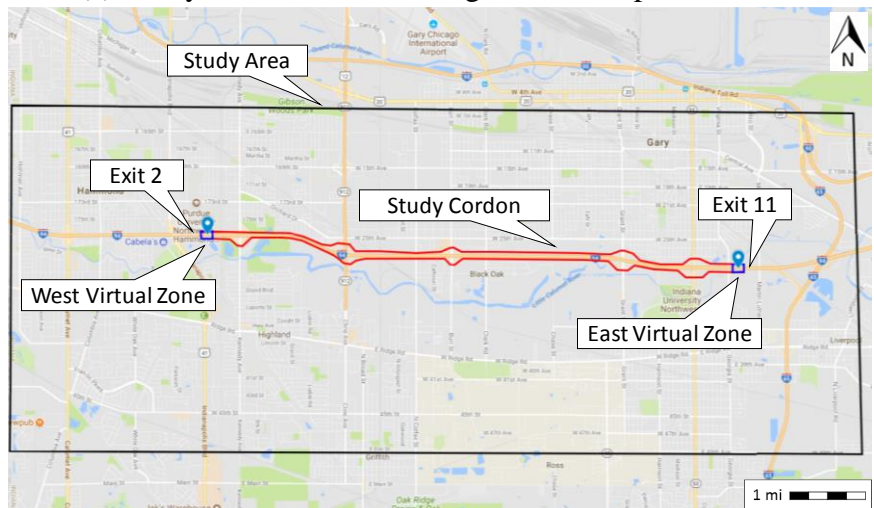
The frequency of the GPS data emitted by the devices depends on a number of factors, including the signal strength, device type, software features, user intervention (powering device on/off), and weather. The frequency varies from one waypoint per second to one per minute. To emulate vehicle re-identification with sensors, two 1000-ft virtual zones were defined on the study route, at each endpoint: the east virtual zone near exit 2, and west virtual zone near exit 11 as shown in Figure 74 (b).

The ultimate origin and destination of a trip are unimportant as long as some part of the trip follows the route through the two zones. The primary focus of this research is to identify vehicles that leave and later re-enter the roadway—that is, trips which passed through the endpoints but did not follow the route of interest. The study route is I-94, as shown in Figure 74 (b). In addition, Figure 74 (b) also shows a “Study Area” as a box encompassing a wider area around the study route. Constraining the GPS analysis to this limiting area greatly reduced the required computation time.

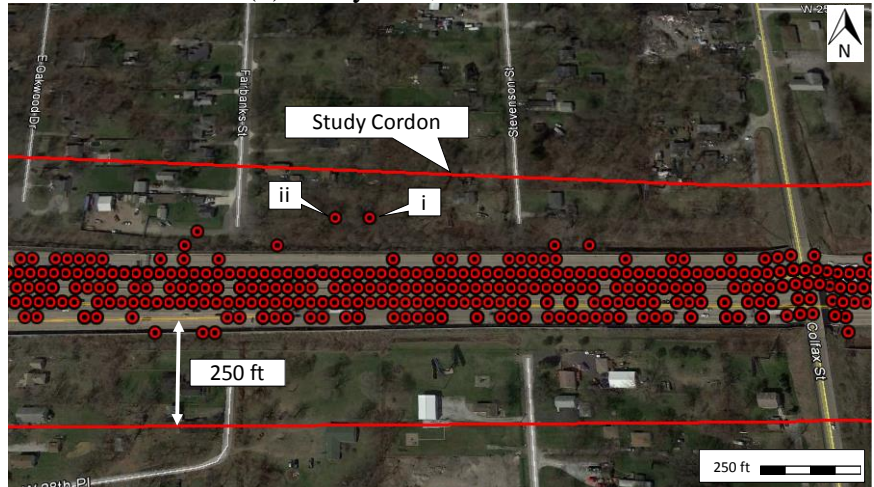
To distinguish between on-roadway and off-roadway waypoints, a cordon line was drawn around the study route. To eliminate false off-roadway points from being generated by GPS errors (from low precision instruments), a rather generous cordon width of 250 ft was used, as shown in Figure 74 (c). With four decimal places in the latitude and longitude, the data has an accuracy of 36 ft. However, there are still occasional errors, as seen in callouts (i) and (ii) in Figure 74 (c). Because the study route was a controlled-access facility, this cordon size was sufficient to filter GPS errors, but any vehicle path that could possibly enter or leave the roadway would have to cross the cordon line.



(a) Study area connects Chicago to Indianapolis and Detroit



(b) Study area, cordons and zones

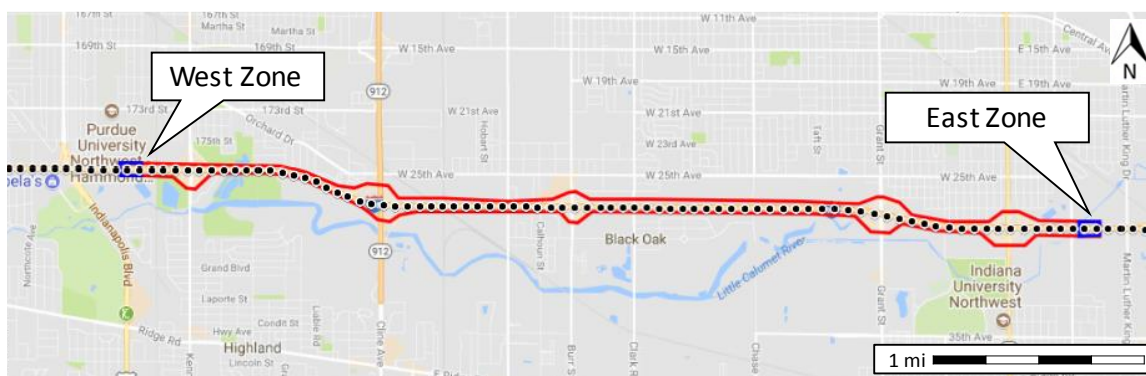


(c) Study cordon drawn 250ft from shoulder to capture waypoints with poor precision

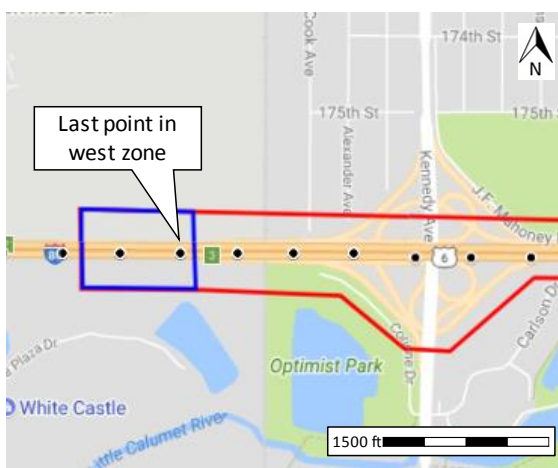
Figure 74. Northern Indiana I80/I94 corridor

6.3.2.3 Estimation of Travel Time

An algorithm was developed to estimate the travel time using the waypoint data. The process first identifies trips with at least one waypoint in each endpoint zone (Figure 75 (a)). The timestamps of the waypoints relative to those zones were then examined to identify the direction of travel. The travel time between the zones was estimated as the difference between the last observation time in the origin zone and the first observation time in the destination zone. Figure 75 (a) shows an eastbound trip with waypoints along the study corridor. The travel time is the difference between the timestamps of the last waypoint in the west zone (Figure 75 (b)) and the first waypoint in the east zone (Figure 75 (c)).



(a) Waypoint data from a single trip passing through both virtual zones



(b) West Zone

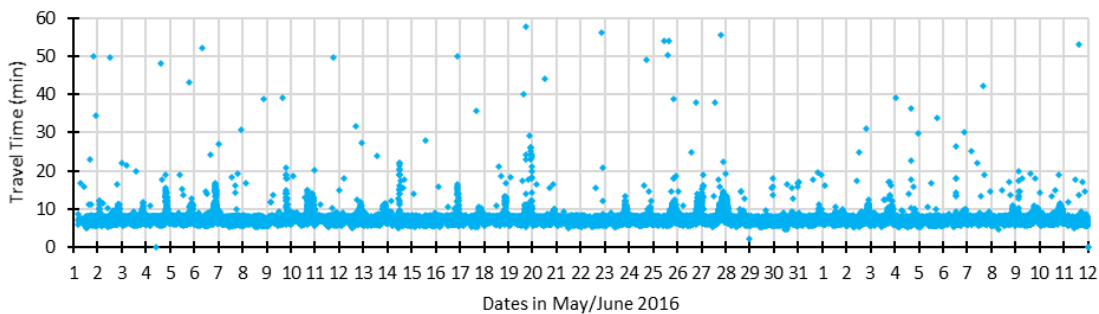


(c) East Zone

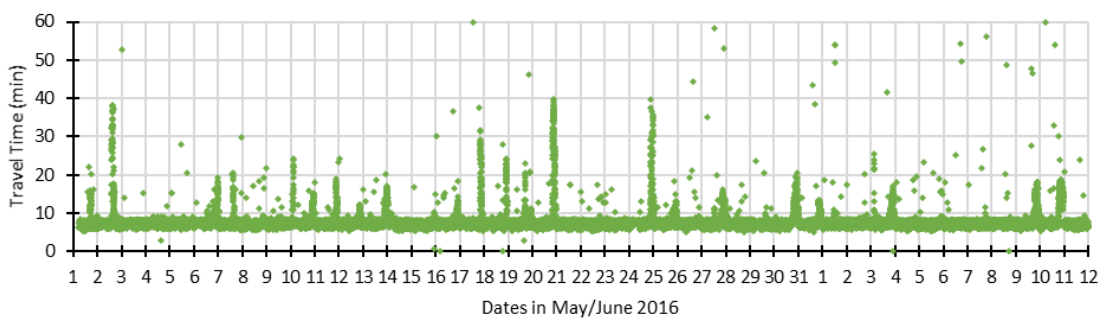
Figure 75. Estimation of travel time by matching trips in each zone

There were 31,878 unique trips that passed through both the zones during the study period. Scatter plots displaying the travel times of these trips in the eastbound

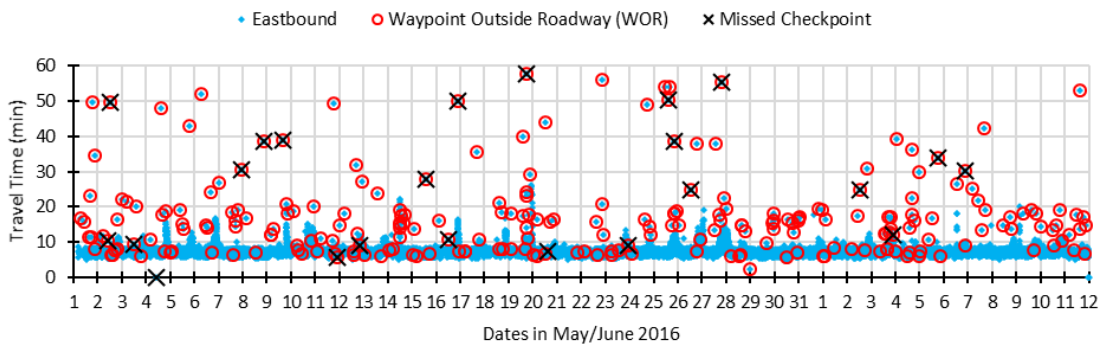
(15,613 trips) and westbound (16,265 trips) direction are shown in Figure 76 (a) and (b) respectively.



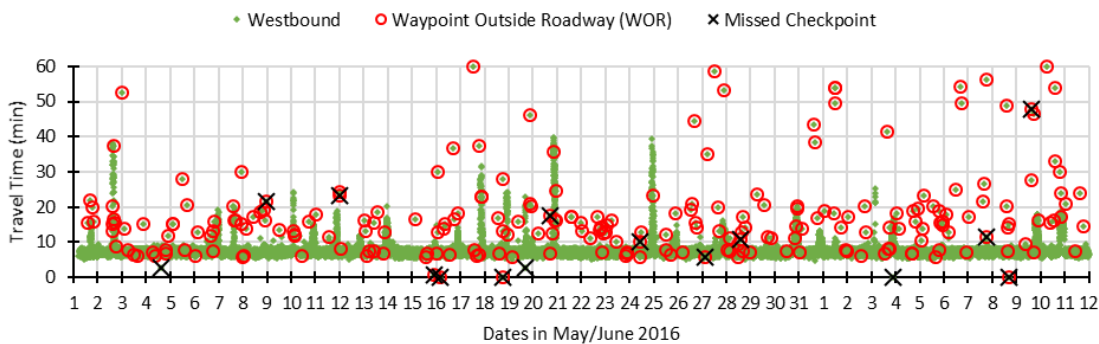
(a) Raw travel time plot for I-94/I-80 eastbound



(b) Raw travel time plot for I-94/I-80 westbound



(c) Possible missed checkpoint trips and outliers in eastbound direction



(d) Possible missed checkpoint trips and outliers in westbound direction

Figure 76. Raw travel time after trip matching for EB and WB directions

6.3.2.4 Missed Checkpoint Trips

Although every trip traveled through both endpoint zones, in some cases they did not do so in a sequence that corresponded to directional travel along the study route. Such a trip, for example, might have been seen in the east zone, then outside of the zone, but then in the east zone again, and then may be seen in the west zone many hours later. The waypoint trajectory shown in Figure 5 (a) is an example of such a trip, with a travel time of 145 minutes, captured after travelling a distance of 90 mi (Figure 5 (b)). A total of 196 such trips (represented by cross marks in Figure 76 (c) and (d)) were removed from the data set.

6.3.2.5 Identification of Outliers

In this study, outliers are defined as trips that left the cordon area, based on the GPS data. Despite the use of a wide cordon, these trips could still potentially include non-outlier trips, due to GPS errors (*i.e.*, “false positives”). The initial data set of trips with at least one waypoint outside roadway (WOR) are highlighted in Figure 76 (c) and (d). These were investigated for further analysis. There are three possible scenarios, as explained below.

6.3.2.5.1 Confirmed outlier

These are trips that actually left the roadway, such as chained trips or use of alternative routes. Figure 77 (c) shows an eastbound trip that left the roadway (red dots). The distance-time plot (Figure 77 (d)) confirms that the vehicle stopped for nearly 5 minutes during the trip, strongly suggesting trip chaining. Vehicles could also take an alternate route, as shown in Figure 77 (e). There were 243 confirmed outliers (124 eastbound and 119 westbound) in the entire data set. These were kept in the data set and flagged as outliers.

6.3.2.5.2 Non-outlier

As mentioned earlier, there is a possibility that GPS error could cause some trip to be identified as outliers, even though they did not leave the roadway. Figure 77 (g) shows an example where only one point (callout i) fell outside the cordon. The travel time is

probable for the study route (10.5 minutes), and the time-distance plot suggests continuous travel on the route (Figure 77 (h)). Such trips were identified as non-outliers (total of 118) and kept in the dataset.

6.3.2.5.3 Indeterminate

These trips did not have enough GPS data to determine whether they left the roadway. Although they contained some WOR data, the GPS record was inadequate, likely due to weak signal strength from the device. Figure 77 (i) shows an example of such a trip with only one point (callout ii) outside the cordon, with gaps in the GPS data, and having a travel time of 19.5 minutes (Figure 77 (j)). These trips (total of 42) were rejected from the data set.

6.3.2.6 Identification of Non-outliers

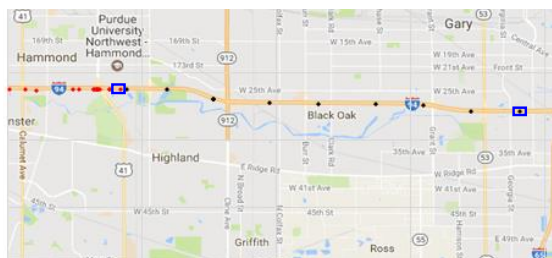
The non-outliers are those trips that did not leave the roadway. The initial dataset comprised those trips with no WOR. Some trips had high travel times, possibly due to congestion or incidents. However, it is also possible that the trip might have left the roadway but did not report any WOR due to infrequent GPS reporting. Trips with travel time above a threshold of 10 minutes were further examined to screen for “false negatives”.

6.3.2.6.1 Congestion/delay

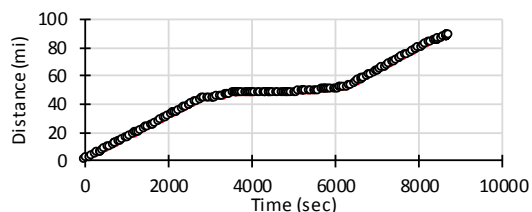
These trips experienced delays during peak hours, work zone traffic, or due to crashes. Each period of congestion was verified by investigating crash reports and work zone reports. For example, many westbound trips on Friday, May 20, 2016 from 7pm to midnight were found to have an average travel time of 21 minutes, perhaps due to weekend traffic heading into Chicago. Figure 78 (a) illustrates a westbound trip experiencing congestion during the evening peak (travel time of 38.5 minutes (Figure 78 (b))). Figure 78 (c) shows an eastbound trip that experienced delay due to a crash around 1pm on June 6, 2016 near mile marker 9. This is well observed by the slow-downs after 6 miles from the origin, as shown in Figure 78 (d). Such trips with long travel times were kept in the data set.

6.3.2.6.2 Indeterminate

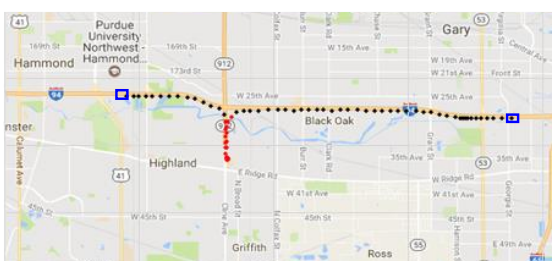
Similar to the earlier screening of outliers, some trips in the “non-outlier” data set had long travel times and no WOR, but the GPS data was too sparse to confirm whether they followed the study route. Such trips (a total of 19) were rejected from the data set.



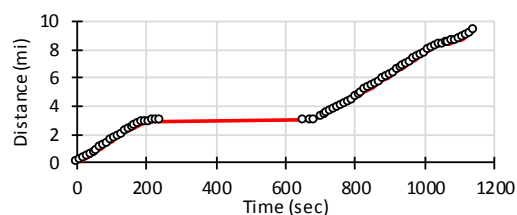
(a) Missed checkpoint trip



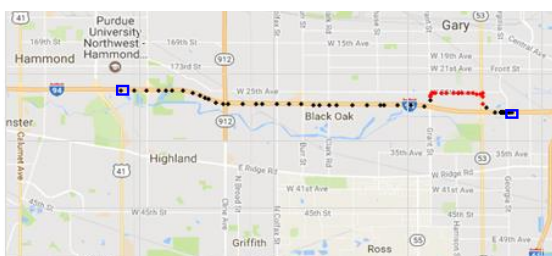
(b) Distance – time plot for (a)



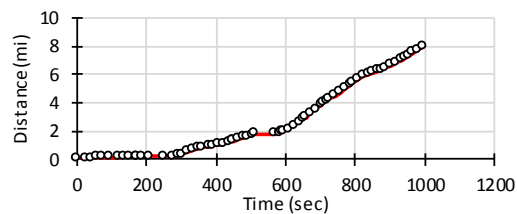
(c) WOR – Outlier (Trip chaining)



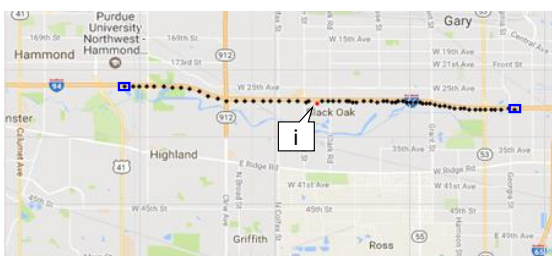
(d) Distance – time plot for (c)



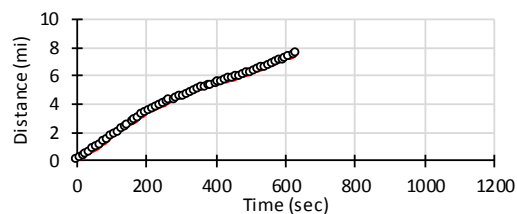
(e) WOR – Outlier (Alternate routes)



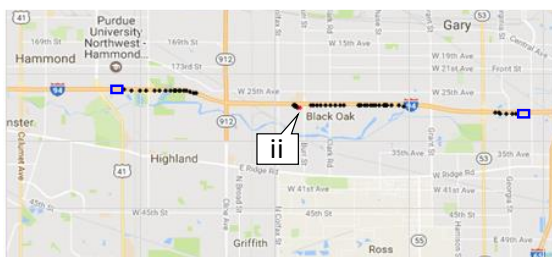
(f) Distance – time plot for (e)



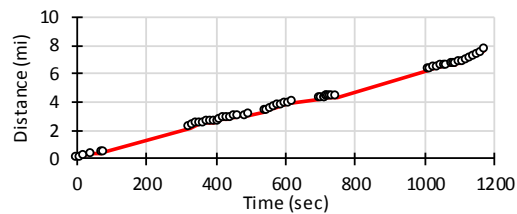
(g) WOR – Non-outlier (GPS error)



(h) Distance – time plot for (g)

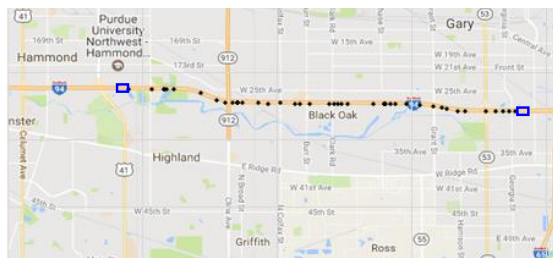


(i) WOR – Rejected (Indeterminate GPS)

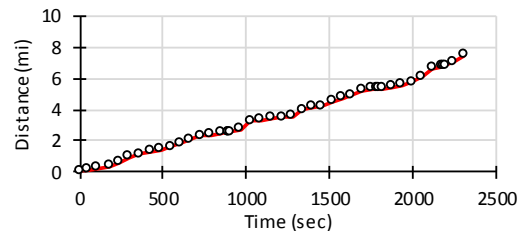


(j) Distance – time plot for (i)

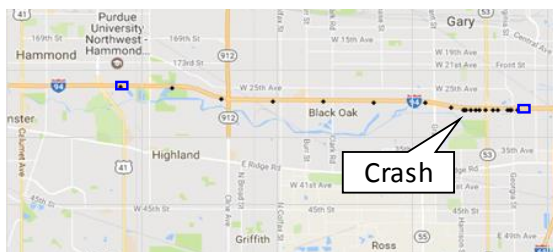
Figure 77. Missed checkpoint and waypoint outside roadway (WOR) trips



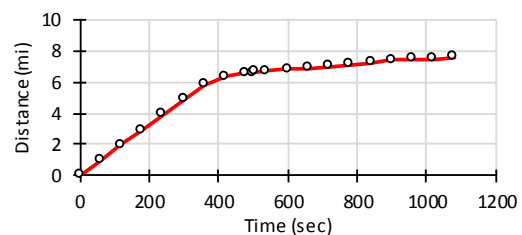
(a) Non-outlier trip experiencing congestion



(b) Travel time of 38.5 minutes for (a)



(c) Non-outlier trip – delay due to crash



(d) Slow down after 6 mi for (c)

Figure 78. Non-outlier trips with high travel time

6.3.2.7 Curated Data Set

A total of 257 trips (missed checkpoint and indeterminates) were removed to develop the curated data set, which consisted of 31,621 trips (15,487 eastbound and 16,134 westbound). The confirmed outliers were less than 1% of the total trips for each direction. Figure 79 shows the curated travel time data with confirmed outliers for both directions. A sample dataset is shown in Appendix C.

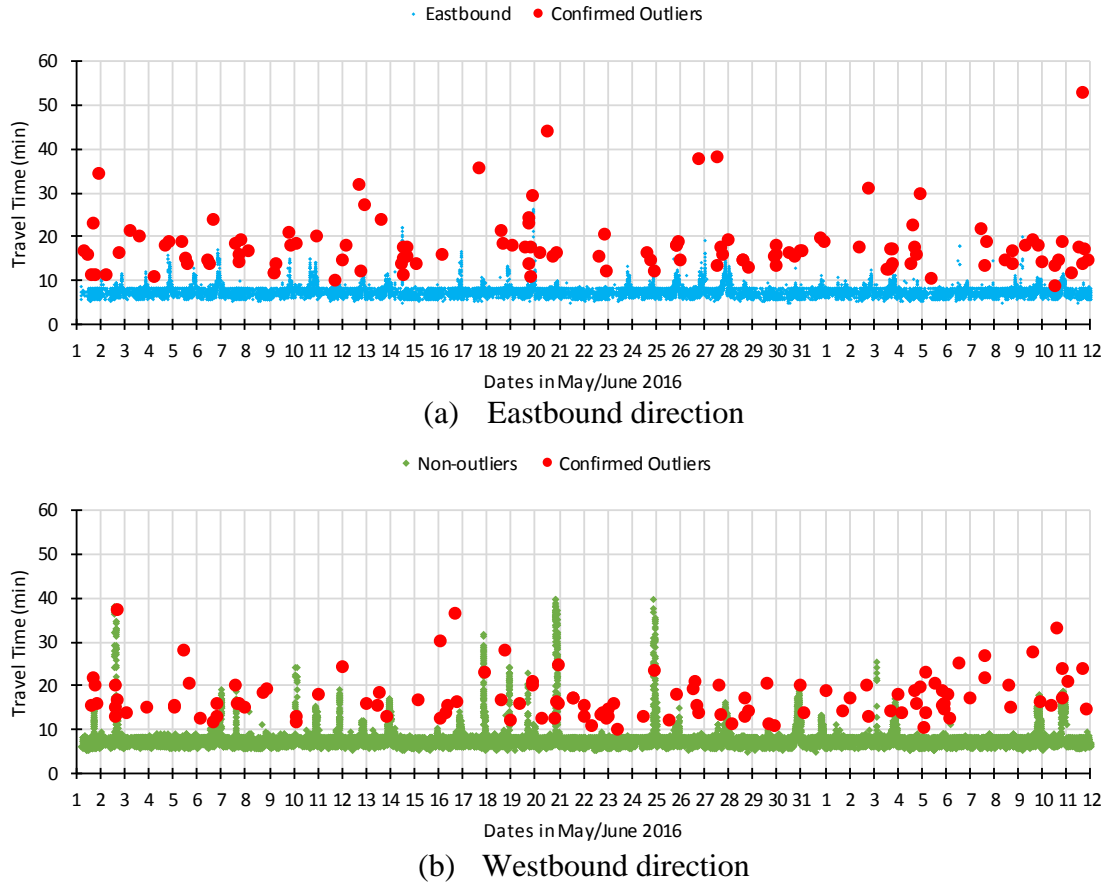


Figure 79. Curated dataset after removing missing checkpoint and indeterminate trips

6.3.3 Outlier filtering methods

The performance of three common outlier filtering algorithms is evaluated using the dataset: median absolute deviation (MAD), modified z-score and box plots.

6.3.3.1 Median absolute deviation (MAD)

The MAD is a statistical filter commonly used for identifying outliers [65]. The filter identifies outliers by comparing the travel time with the neighboring travel times within a period. For each travel time value X_i at time i , the filter compares all the values within time j (in this case, 5 minutes before and 5 minutes after the present value). MAD is computed as:

$$MAD = \text{median} (| X_i - \text{median}(X_j) |) \quad (6)$$

Data points are flagged as outliers if they are greater than the upper bound value (UBV), or lower than the lower bound value (LBV). The UBV and LBV are given by:

$$UBV = median + \sigma f \quad (7)$$

$$LBV = median - \sigma f \quad (8)$$

where σ is the standard deviation from MAD, in which a normally distributed data can be approximated as $\sigma = 1.4286 * MAD$ [124], [125], and σf represents the scatter of the data, where f is a scale factor. If f is small, the scatter (gap between UBV and LBV) will be small, and vice-versa [126]. The value of f has been suggested by researchers to be in the range of 1 to 5 [65]. For the present study, the value of f is assumed to be 2.

6.3.3.2 Modified Z-score (Mod Z)

The modified Z-score [127] is a standardized score that measures the strength of an outlier. It is a revised version of the Z-score method, which uses the sample mean and standard deviation to identify outliers. The standard deviation can be inflated by the presence of extreme values which present the problem of masking (less extreme outliers go undetected because of more extreme outliers). The modified Z-score addresses this problem by employing the median and MAD instead of the mean and standard deviation [128]. A window of 10 minutes (5 minutes before and 5 minutes after) was also used. The modified Z-score is computed as shown in equation (9).

$$M_i = \frac{(x_i - \tilde{x})}{1.4286 * MAD} \quad (9)$$

where \tilde{x} is the sample median for the 10-minute window and MAD is computed by equation (6). The data points with an absolute value of modified z-score greater than 3.5 are labelled as outliers.

6.3.3.3 Box and Whisker Plots

Box and whisker plots are frequently used to compare distributions across groups. A box and whisker plot consists of a rectangle with top and bottom as the first and third quartiles, a horizontal line at the median, and whiskers 1.5 times the interquartile range (IQR). The data points are classified as outliers if they fall outside the whiskers [129].

Figure 80 provides a graphical representation of the box plot. Outlier filtering using boxplots was also carried out across a 10-minute window (5 minutes before and 5 minutes after) for each data point.

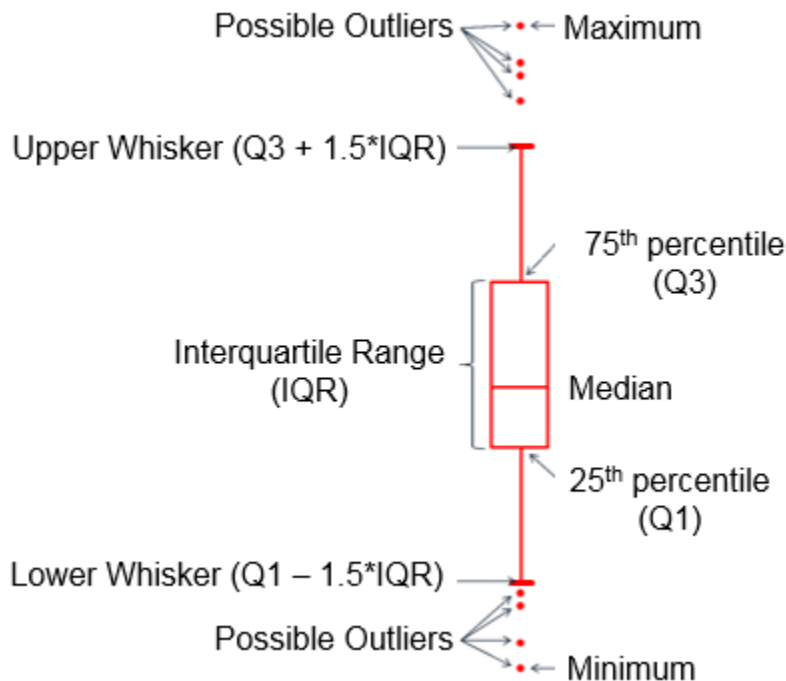


Figure 80. Graphical representation of box and whisker plot

6.4 Results

6.4.1 Comparison of outlier filtering algorithms

Two performance measures are identified for evaluating the outlier-filtering algorithms by the proportion of true samples rejected and proportion of outliers correctly identified. The true samples refers to confirmed non-outliers in the curated dataset. Table 9 shows a comprehensive comparison of the three outlier filtering algorithms on the curated dataset. All methods correctly removed some outliers, while also incorrectly removing some non-outliers (true samples). While the MAD and modified Z-score removed more than 70% of the confirmed outliers, the boxplot only removed about 55%. On examining the non-outliers, the MAD incorrectly removed nearly 10% of the samples (around 1500), followed by modified Z-score (around 5%) and boxplots (around 3.5%). If the objective is to strike a balance between correctly removing the confirmed outliers while avoiding incorrectly removing confirmed non-outliers, the modified z-score seems

to achieve a good compromise; it removes only 5% of the non-outliers while removing more than 70% of the confirmed outliers.

Table 9. Comparison of outlier filtering algorithms

Method	Eastbound				Westbound			
	Total data points = 15,487				Total data points = 16,134			
	Total outliers = 124				Total outliers = 119			
	Correctly identified outliers	% of correctly identified outliers	True samples rejected	% of true samples rejected	Correctly identified outliers	% of correctly identified outliers	True samples rejected	% of true samples rejected
MAD	92	74.19%	1496	9.74%	89	74.79%	1576	9.84%
Mod Z	87	70.16%	790	5.14%	86	72.27%	791	4.94%
Boxplot	64	51.61%	537	3.50%	70	58.82%	544	3.40%

6.4.2 Performance during real-time estimation/predictions

Further comparisons of the three methods along with the unfiltered (UF) data was carried out to study the deviation from the “ideally-filtered (IF)” dataset across 10-minute sampling windows. The unfiltered data is the curated dataset containing the confirmed outliers and confirmed non-outliers, without any filters. The ideally-filtered is the curated dataset with confirmed outliers removed. This is important for a number of intelligent transportation system (ITS) applications that rely on real-time data to provide accurate travel time estimates.

Figure 81 shows eastbound travel times for a 3-hour period on May 9, 2016, using data filtered by the three methods examined earlier, unfiltered, and ideally filtered. Travel times in each direction are plotted for the median and the 75th, 85th, and 95th percentiles. On examining the median plot (Figure 81 (a)), all four data series closely follow the ideally-filtered, except for a small period between 19:00 and 19:30. There is a minor spike (of 2.5 minutes shown by callout (i)) for the unfiltered and boxplot data, due to undetected outliers. The modified Z-score closely follows the ideally-filtered data during this period; however, the MAD underestimates the travel time by 0.8 minutes. The spikes become even more discernible at the 75th, 85th, and 95th percentile. While the modified Z-score followed the ideally-filtered data at the median, there was a negative spike of more than 4 minutes at the 95th percentile during the 19:00-19:30 period (Figure 81 (d)).

Another example is the boxplot during the 20:30 to 21:00 period, where there were differences of 0, 4, 6 and 8 minutes in travel time compared to the ideally-filtered data, at the median, 75th, 85th and 95th percentiles respectively.

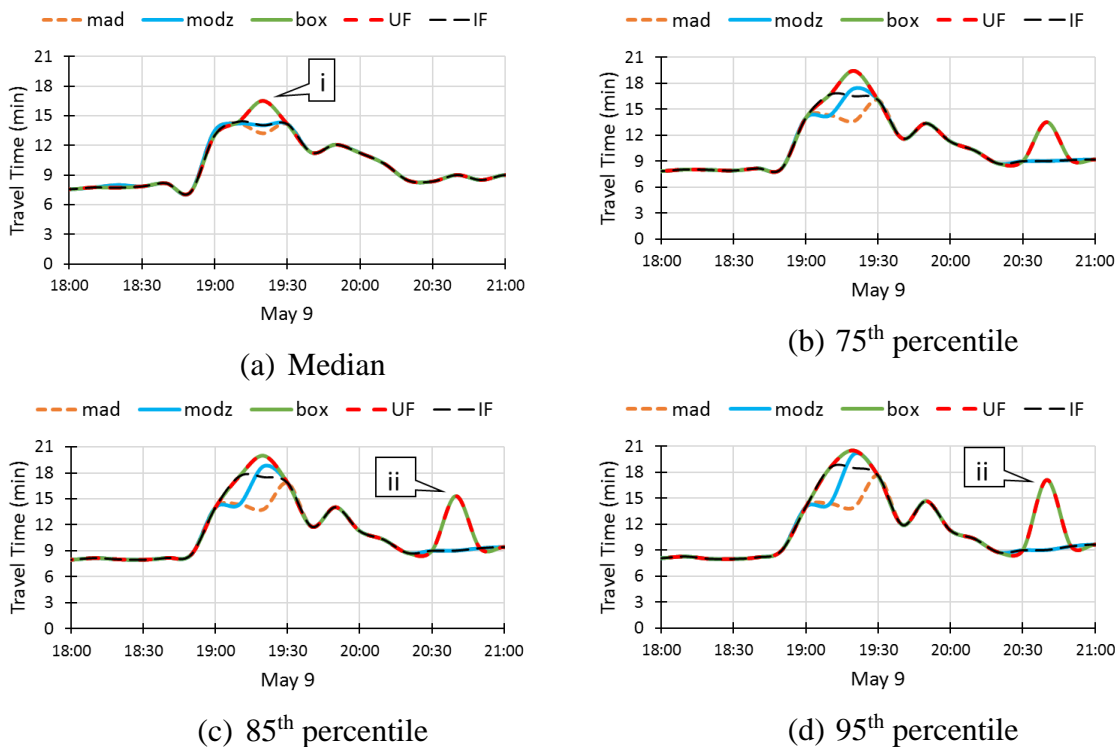


Figure 81. Performance of outlier filtering algorithms compared to ideally-filtered (IF) and unfiltered (UF) data

Spikes with an absolute difference in travel time more than 5 minutes (callout (ii) on Figure 81), when compared to the ideally-filtered data, were recorded as “false spikes.” Table 10 shows the number of false spikes reported by the three filtering methods for both directions over the entire study period. As expected, the number of false spikes increases for all the methods as we progress from the median to the 95th percentile. For MAD and modified Z-score, the false spikes increased by nearly 10, whereas for the boxplots, the number of spikes doubled from median to the 95th percentile. It is also interesting to note that the boxplots had the lowest number of false spikes in five out of the eight categories (one in eastbound and four in westbound), probably because this method removed the fewest number of outliers and non-outliers overall.

Table 10. Number of false spikes reported for travel time difference above 5 minutes

Direction	Method	Median	75th percentile	85th percentile	95th percentile
Eastbound (EB)	MAD	37	40	43	48
	Modified Z Score	30	33	36	40
	Boxplots	23	33	41	54
Westbound (WB)	MAD	36	43	45	46
	Modified Z Score	29	36	37	40
	Boxplots	15	24	30	35

6.5 Summary

This chapter developed a framework using high-fidelity GPS trajectory data to distinguish outliers associated with trip chaining and route diversion from trips affected by incidents or weather. Outliers were verified by examining the waypoints of every outlier trip to confirm whether they diverted. Less than 1% of the data was removed to eliminate trips with inadequate GPS data or missed checkpoints. Trips experiencing high travel time (due to congestion or crashes) were verified against crash reports and work zone reports.

The curated dataset developed consisted of 31,621 trips with 243 confirmed outliers. This dataset was used to analyze the performance of three common outlier-filtering algorithms, median absolute deviation, modified Z-score and boxplots. Two performance measures, the proportion of true samples rejected and proportion of outliers correctly identified, were used to evaluate the algorithms. The modified Z-score had the best performance with successful removal of 70% of the confirmed outliers and incorrect removal of only 5% of the true samples. The MAD was more aggressive, removing a higher percentage of both confirmed outliers (75%) and true samples (10%), while the boxplot method was less aggressive, removing lower percentages (55% and 3.5%).

The performance of the outlier filtering algorithms over 10-minute sampling windows was also analyzed. The variation of these filters using the median, 75th, 85th and 95th percentile was evaluated. For each method, the number of false spikes (the difference between estimated and actual travel time > 5 minutes) increased while progressing from the median to the 95th percentile, indicating the tendency of the methods to remove extreme outliers.

A number of ITS applications rely on the real-time data collected by sensors to provide accurate estimates of the travel time. Outliers are an inherent part of any data collection technique, especially for travel time estimates. One of the challenges in each of these sampling techniques using probe data, is to employ filtering techniques to remove the outliers associated with trip chaining, but not remove important features in the data associated with incidents or traffic congestion. Numerous outlier filtering algorithms have been proposed in the literature, but their performance has never been evaluated. The first ever curated dataset developed by this research, where the outlier status of every data point is known, will be beneficial for researchers and other professionals to assess the performance of their outlier filtering algorithms. Moreover, this dataset published in open access [123] will be an essential data source for the development of robust and accurate outlier filtering algorithms in the future. It is to be noted that the confirmed outliers in this study are the ones that left the roadway of interest for trip chaining activities. These trips are different from the “statistical outliers” where the outlier status of a data point is defined on the basis of statistical computations.

CHAPTER 7. TRAFFIC SIGNAL PERFORMANCE MEASURES FOR CONNECTED VEHICLES

This chapter introduces performance measures vehicle vendors can use to characterize the accuracy of traffic signal state predictions for connected vehicle applications.

7.1 Introduction

In 2011, the USDOT announced the “Connected Vehicle Program” that allows vehicles to communicate with other vehicles (V2V), infrastructure (V2I) and other devices (V2X) to improve safety, mobility and reduce environmental impacts like fuel consumption and emissions [12], [13]. These V2I and V2V technologies are components of the connected vehicle program that allows the wireless communication between vehicles and infrastructure. The vehicle-generated data can be used to provide information such as advisories from the infrastructure to the drivers alerting them of safety and mobility conditions on the roadway. Recently, the American Association of State Highway and Transportation Officials (AASHTO) announced the Signal Phase and Timing (SPaT) challenge to state and local agencies to kick start infrastructure deployments for V2I communications [130]. The challenge involved the deployment of Dedicated Short Range Communication (DSRC) infrastructure with SPaT broadcasts (current intersection signal light phase) on at least 20 signalized intersections in all of the 50 states by 2020. The DSRC refers to a two-way communication on the 5.9GHz band set aside for V2V and V2I applications. Figure 82 shows an overview of the DSRC communication standard. The traffic signal controller (callout 1) is connected to a road side unit (RSU) (callout 2) which transmits (such as SPaT) and receives data wirelessly from the on-board units (OBU) and high fidelity GPS units (callout 3) inside the vehicles. The DSRC technology is capable of interacting with fast moving vehicles and prioritizing safety messages with the dedicated wireless transmission [12]. It operates on a short range, meaning that the communication between the vehicles and radios are limited to a certain radius (typically 1000 – 1500 ft).

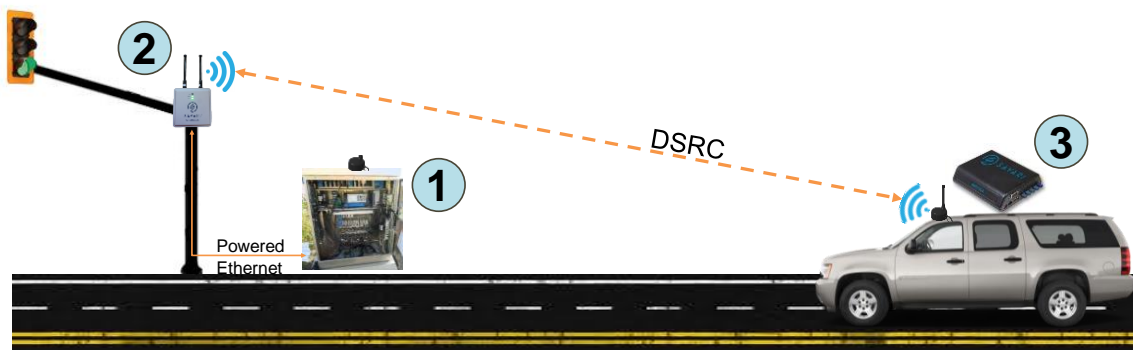


Figure 82. DSRC mode of communication for connected vehicles

An alternative architecture that has been gaining popularity is the cellular mode of communication (Figure 83). In this method, the original equipment manufacturers (OEM) send vehicle data directly to the cloud using 4G/5G cellular technology. The data from the signal controllers and vehicle undergo integration in a cloud-based system, which then communicates it back and forth with the signals and the vehicles. The advantage of this method is that vehicles can communicate with the cloud as long as the cellular network is available. With the OEM's sending data to the cloud, there is less roadway equipment and maintenance required by the local agencies and state DOTs

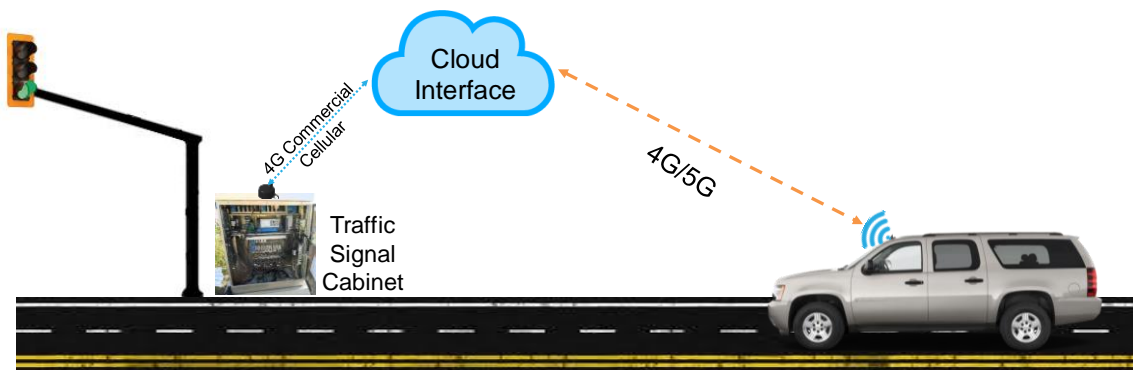


Figure 83. Cellular mode of communication for connected vehicles

In October 2017, Volkswagen of America entered into a partnership with Purdue University, INDOT and other industry partners to develop a working prototype of the V2I communications using the cellular standard. Seven intersections along the US 231 corridor in West Lafayette, IN, were prepared for V2I communications. Data from the

Volkswagen research vehicle and the INDOT signal controllers were used to deploy the traffic signal status indications in Volkswagen vehicles. As the vehicle approaches an intersection, the dashboard displays the current indication of the traffic signal for the vehicle movement. If the vehicle arrives during the red phase or if the algorithm computes that the vehicle will not make the green, the application displays a countdown timer with the time remaining for the next green (Figure 84). The early applications of this technology include eco-driving and dilemma zone reduction.

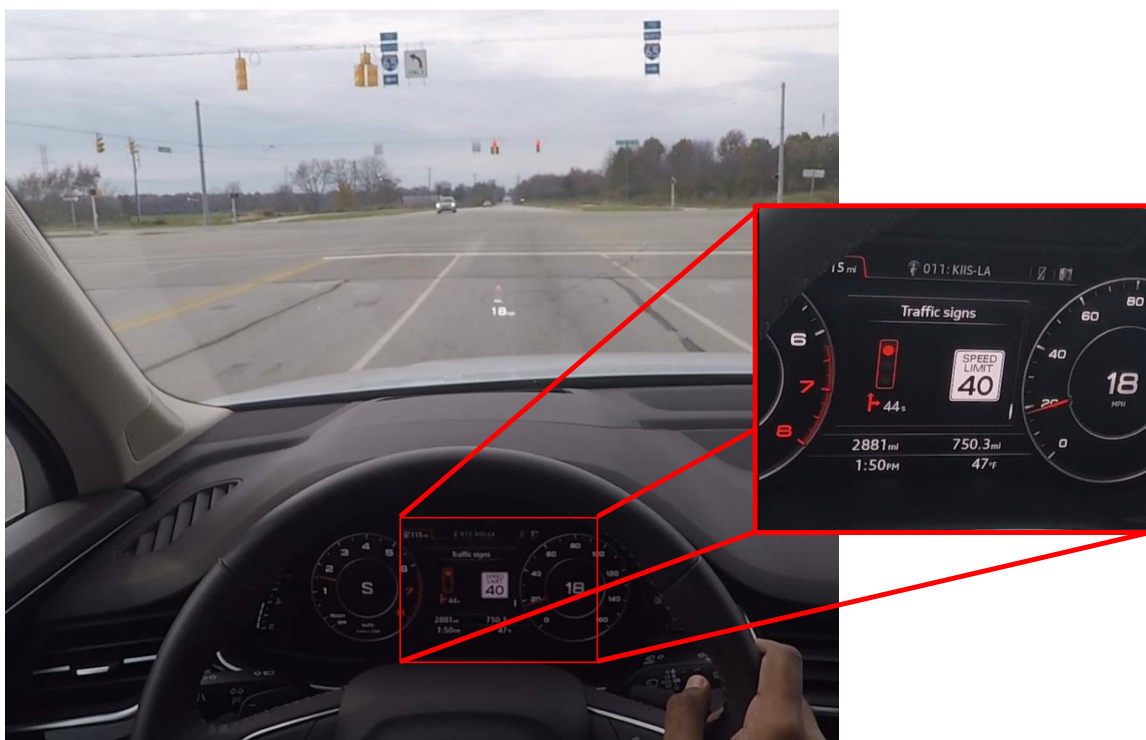


Figure 84. Dashboard with countdown timer for the next green

Currently, New York City, Tampa, and Wyoming are undergoing real-world deployments of the connected vehicle program for testing and research purposes. As connected vehicle technologies gain popularity, it is necessary to have performance measures to better communicate information about the system. Signal performance measures have been widely discussed in the literature and in Chapter 3. The Federal Highway Administration (FHWA) has been promoting performance-based signal re-timing to monitor and optimize the traffic signals which could lead to improved mobility, safety and targeted maintenance [131]. With advancements in high-resolution signal

controller data [132], a number of performance measures have been developed [8], [133]–[135] to improve progression and signal timing and provide feedback on system operation. In 2013, the Utah Department of Transportation (UDOT) developed automated visualization tools using the performance measures [136] which are currently implemented by 12 states. Researchers have also refined these performance measures to provide better estimates of field conditions [137]. Studies have found that performance measures such as vehicle arrivals on green and percent on green and visualization tools such as flow profiles [8] have helped agencies to understand the system better and improve system performance [138]–[140].

7.2 Motivation and Scope

During the USDOT Public Listening Summit on Automated Vehicle Policy held in March 2018, the Secretary of Transportation listed out three main visions: 1) safety, 2) infrastructure and 3) preparing for the future. Safety is one of the primary concerns in transportation and the evolution of connected and automated vehicles will have a prominent role in reducing the number of fatalities and crashes. This requires sound, robust and efficient infrastructure that can facilitate proper V2V and V2I communications. The traffic light status information is one recent development in this emerging area. Two early applications of this technology are eco-driving and dilemma zone reduction, which need an adaptive and dynamic algorithm to provide accurate predictions. The stochastic variation of signal phases that have detectors and communication latency can impact the accuracy of the traffic signal prediction. This chapter evaluates the performance of this technology and proposes two performance measures, probability of phase splits and distribution of green for a phase, to improve the efficiency of the predictive algorithm.

7.3 Methodology and Data Collection

The first part of this study involved collecting field data to compare the accuracy of the red signal countdown prediction. Residual plots are prepared to study the accuracy of the prediction.

The southbound left turn movement at the US 231 & River Road intersection was selected as the study location. This is a protected movement with “left turn only on green”. Figure 85 illustrates the phasing and sequencing diagram at this intersection, which runs in actuated coordination [8] from 6AM to 9PM.

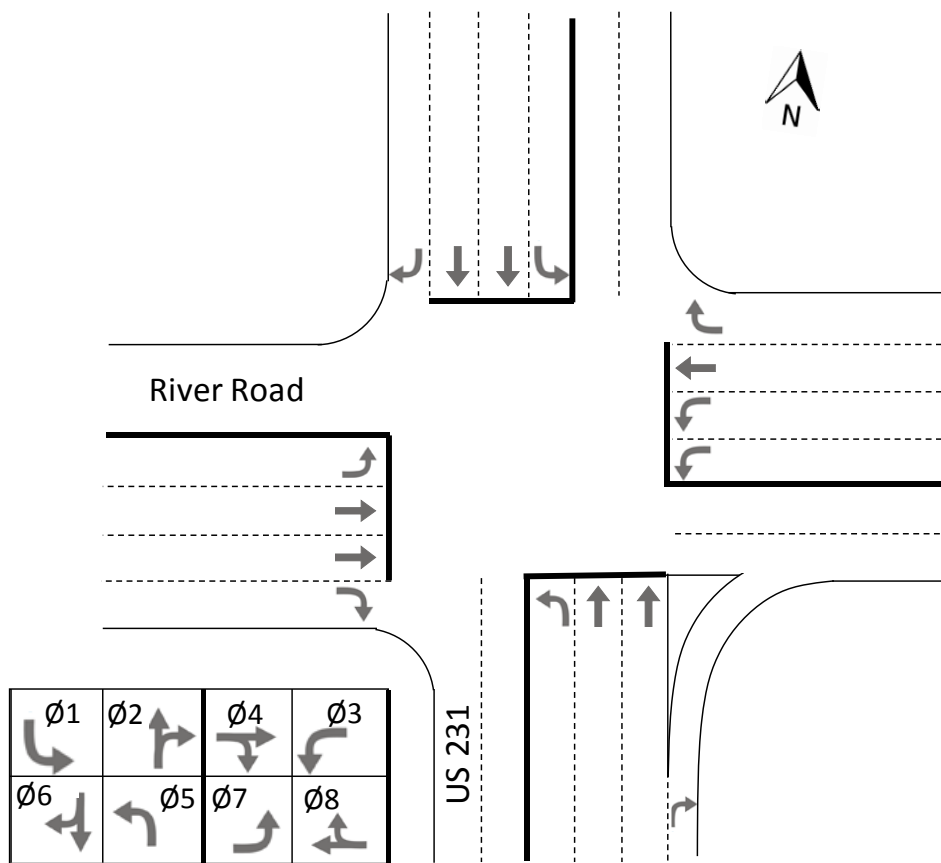


Figure 85. US 231 & River Road intersection diagram

Video data collected at 30fps from two time-synchronized cameras was used to compare the actual event and prediction. Figure 86 shows the data collection setup inside the research vehicle. The “dashboard cam” recorded the traffic signal countdown prediction (similar to Figure 84) on the dashboard and the “traffic signal cam” captured the actual traffic signal events as the vehicle approached the intersection.



Figure 86. Video camera setup for data collection

Figure 87 shows a screenshot of the data extraction process during which the two videos were compared to each other. The timestamp (shown on the top left corner in Figure 87 (a)) of the final red signal countdown timer (usually stops at 4s to avoid driver distraction before the start of green) was extracted using a video editing tool. The predicted time of green was then estimated by adding this time to the timestamp. For example, Figure 87 (a) shows the red signal countdown to be “10s” at a timestamp of 12:25:59.500 as the vehicle reaches the stop bar. The timer disappears at “4s” to avoid driver distraction before the green signal. This timestamp is noted as 12:26:05.467 (Figure 87 (b)). The predicted green time will then be 12:26:09.467 after the addition of 4s. This is compared with the timestamp when the signal actually turns green in the field (12:26:10.367 from Figure 87 (c)), to estimate whether the prediction was early, on time or late. In this case, the prediction was early by 0.9s. Data was collected for 15 runs to generate a reasonable sample size for the accuracy tests.



(a) 10s red signal countdown at 12:25:59.500



(a) Final red signal countdown time (4s) before it disappears at 12:26:05.467



(b) Actual start of green at 12:26:10.367

Figure 87. Extracting timestamps for comparison.

7.4 Results

Figure 88 shows the residual plot comparing the time between prediction of green and actual green from the 15 test runs on the southbound left turn movement. On an average, the difference was 2.69 sec with a standard deviation of 5.61 sec. Half of the predictions were late by more than 4 seconds. Figure 89 illustrates conditions associated with the 10.9s late prediction (callout i) shown in Figure 88. The late predictions occurred in cases where there was no demand on phases 3 and 8 (Figure 85). As a result, there were no detector calls and hence phases 3 and 8 were skipped. This led to an early green for the southbound left movement (phase 1) which was not captured by the prediction algorithm.

There were two runs in which the prediction was early as a result of force-off on the westbound movement. Force-off occurs when the split-timer based on the background coordination cycle expires, thereby resulting in a termination of the phase [8]. Figure 90 illustrates the conditions associated with the 8s early prediction (callout ii) shown in Figure 88 callout ii. As seen from Figure 90 (c), the westbound movement (phase 3) is still active during the predicted green time. The prediction system did not account for variation in demand and detector actuations, which resulted in an early prediction of green for the southbound movement (phase 1). If the prediction algorithm is unable to capture these variations, there could be potential safety risks.

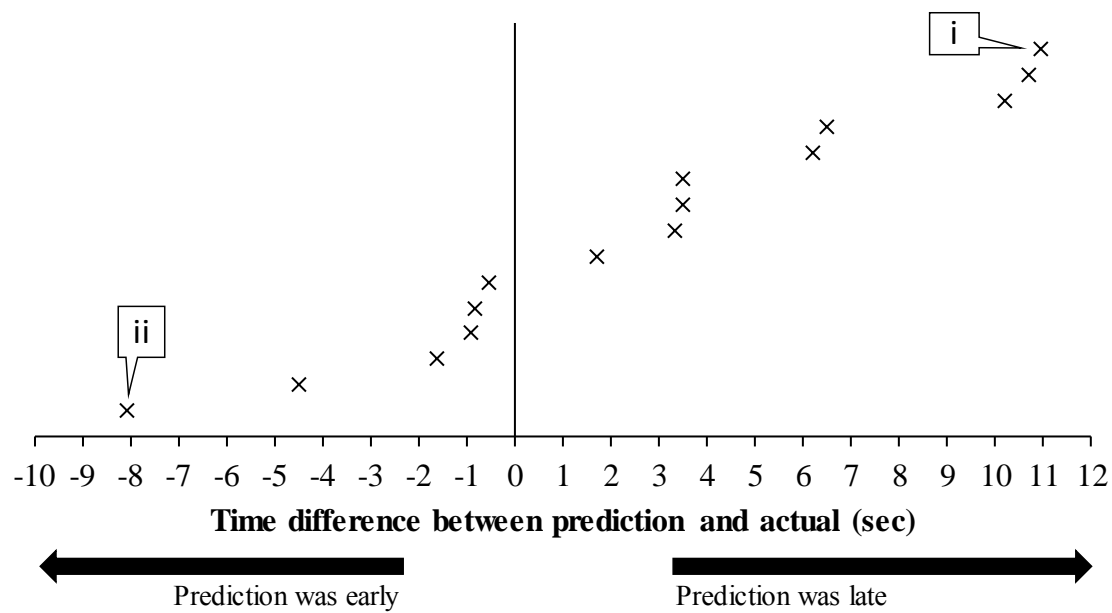
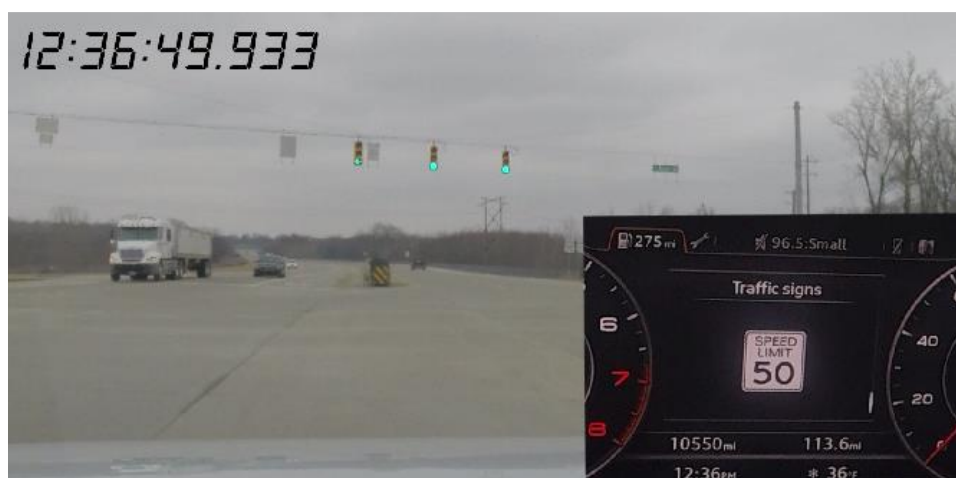


Figure 88. Residual plot comparing prediction and actual



(a) Final red countdown step displayed was at “14s” at 12:36:46.900



(b) Signal turns green at 12:36:49.933 resulting in a 10.9s late prediction

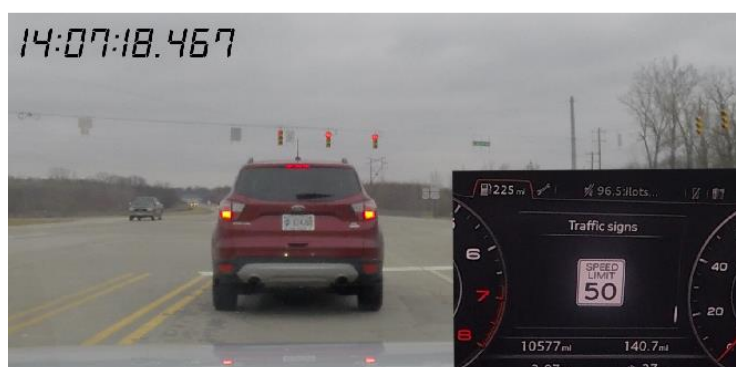
Figure 89. Late prediction



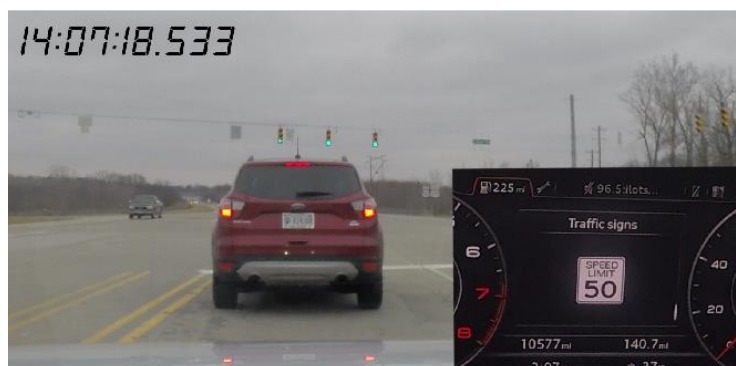
(a) Final red countdown step at "4s" at 14:07:06.467



(b) Signal is still red after 4s at 14:07:10.467.



(c) Signal remains red until 14:07:18.467



(d) Signal turns green at 14:07:18.533, 8s after prediction time

Figure 90. Early prediction

7.5 Performance Measures to Improve Prediction

As seen in the previous section, the prediction system performance degrades during phase omits and force-offs. With actuated coordinated systems, there is a significant variation in the activation of a phase during a cycle. To improve the accuracy of the predictions, the following sub-sections propose a series of performance measures that characterize the stochastic variation of individual phases.

High resolution signal controller data was utilized to develop these performance measures. This data with a temporal fidelity of 0.1s was recorded using a data logger software that captures all the detection and phase events at a given intersection [141].

7.5.1 Split diagrams and probability of expected clearance time

The concept of the split diagram is very similar to the coordination diagram [8], [59] without the vehicle detections. In conventional coordination, a split is the amount of cycle time assigned to a phase by the controller [8]. Splits are inclusive of the green time and the clearance time associated with a phase. In this chapter, the clearance time for a phase is defined as the sum of yellow and all-red intervals. The split diagram for one cycle (ring 1 from Figure 85) at the US 231 & River Road intersection is shown in Figure 91. The two axes are both time; the horizontal axis represents the time of the day and the vertical axis shows the time in cycle. The dashed yellow line depicts the start of clearance time associated with the phase. The cycle length is the sum of the preceding effective red and the subsequent effective green. For coordinated systems, the split times for all phases in a ring must sum to the cycle length.

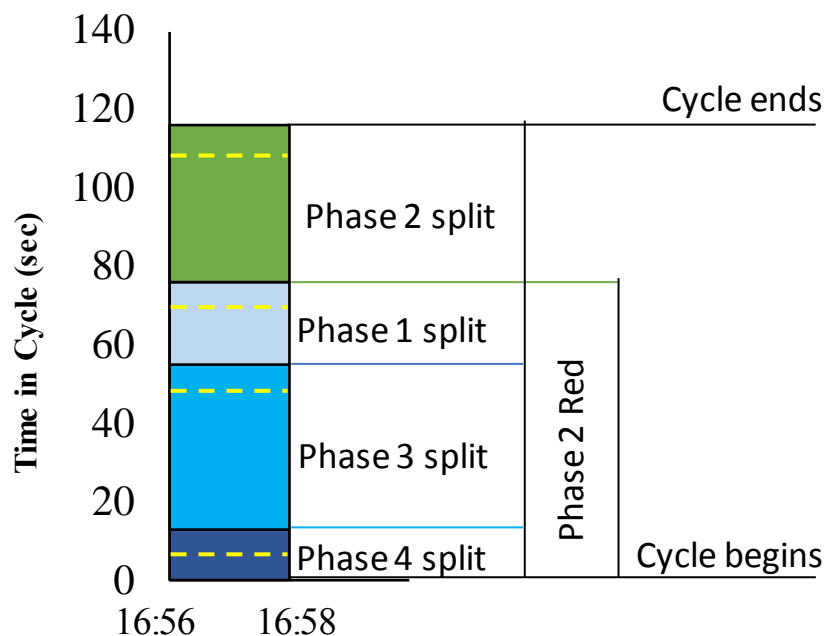


Figure 91. State change event diagram for one cycle

When multiple cycles are shown in succession, we can see the variations in the split times due to the actuated coordination. The split times and cycle length for each cycle during this period are shown in Table 11. Figure 92 shows the split diagram for the same data. The cycle time during this signal timing plan is programmed for 116s, but we can see variations when the cycle time ends early (callout ii) or extends (callout i) due to actuated coordination.

Table 11.Split times for all phases from high resolution data

Time	Phase 4 Split			Phase 3 Split			Phase 1 Split			Phase 2 Split			Cycle Length
	Green	Clearance		Green	Clearance		Green	Clearance		Green	Clearance		
		Y	AR		Y	AR		Y	AR		Y	AR	
16:56	7	4	2	36	4	2	15	4	2	33	5	2	116
16:58	10	4	2	20	4	2	18	4	2	43	5	2	116
17:00	8	4	2	34	4	2	6	4	2	42	5	2	114
17:02	2	4	2	30	4	2	0	4	2	65	5	2	121
17:04	0	0	0	43	4	2	16	4	2	31	5	2	109
17:06	10	4	2	33	4	2	14	4	2	33	5	2	115
17:07	7	4	2	33	4	2	17	4	2	34	5	2	116
17:09	7	4	2	36	4	2	13	4	2	34	5	2	115
17:11	9	4	2	34	4	2	10	4	2	37	5	2	115
17:13	9	4	2	34	4	2	16	4	2	37	5	2	121
17:15	0	0	0	43	4	2	16	4	2	31	5	2	109
17:17	8	4	2	35	4	2	16	4	2	31	5	2	115
17:19	4	4	2	34	4	2	16	4	2	36	5	2	121
17:21	9	4	2	34	4	2	16	4	2	31	5	2	110
17:23	10	4	2	33	4	2	13	4	2	35	5	2	116
17:25	8	4	2	30	4	2	20	4	2	32	5	2	115
17:27	4	4	2	21	4	2	15	4	2	50	5	2	115
17:29	10	4	2	33	4	2	16	4	2	31	5	2	115
17:31	7	4	2	35	4	2	18	4	2	37	5	2	122
17:33	0	0	0	20	4	2	20	4	2	51	5	2	110
17:35	10	4	2	33	4	2	17	4	2	31	5	2	116

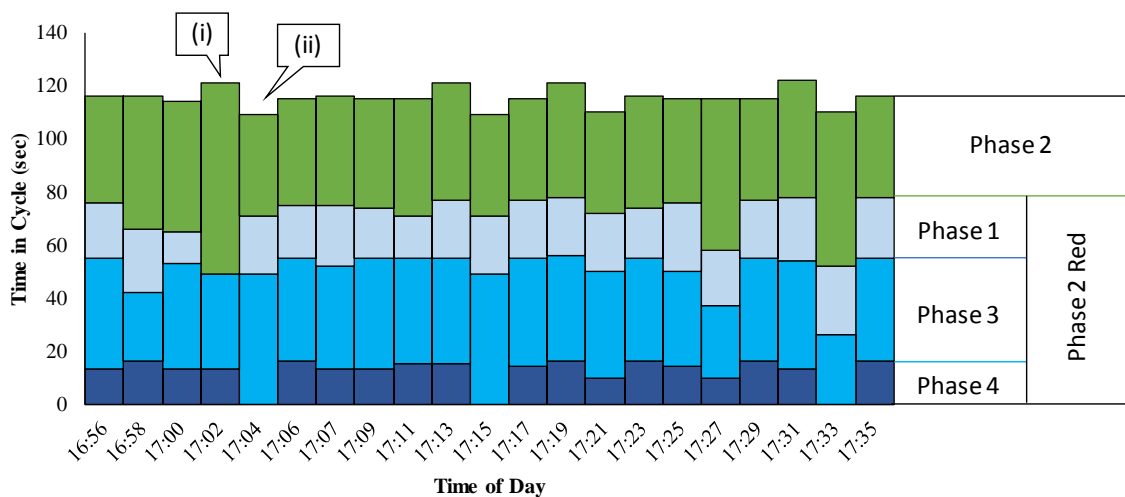


Figure 92. Split diagram for ring 1

The split diagram can be used to develop the probability distribution of the split times for each phase over the cycle length. The probability distribution is estimated for each 1-second bin of the cycle length. This is very similar to the green distribution in the flow profiles data generated by TRANSYT [142]. Figure 93 shows the distribution of the

splits for the median cycle length during the time period. The probability of all distributions at any time in the cycle should add up to 1. For example, if a vehicle arrives at the sixth second of a cycle, the probability of phase 4 being active is 85% and the probability of phase 3 being active is 15%.

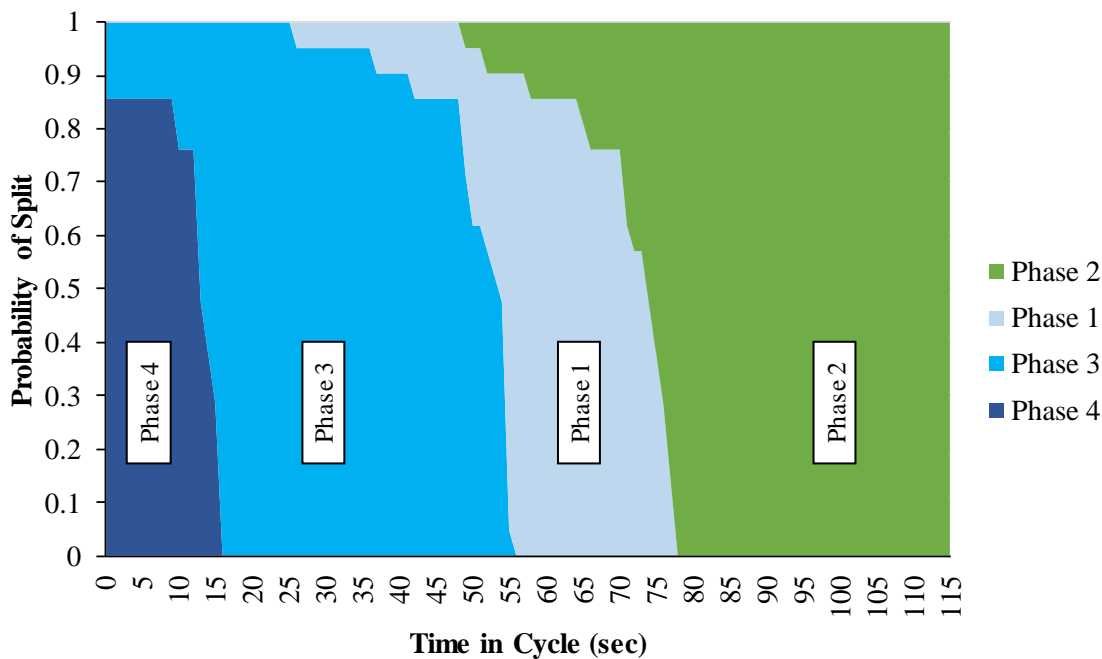
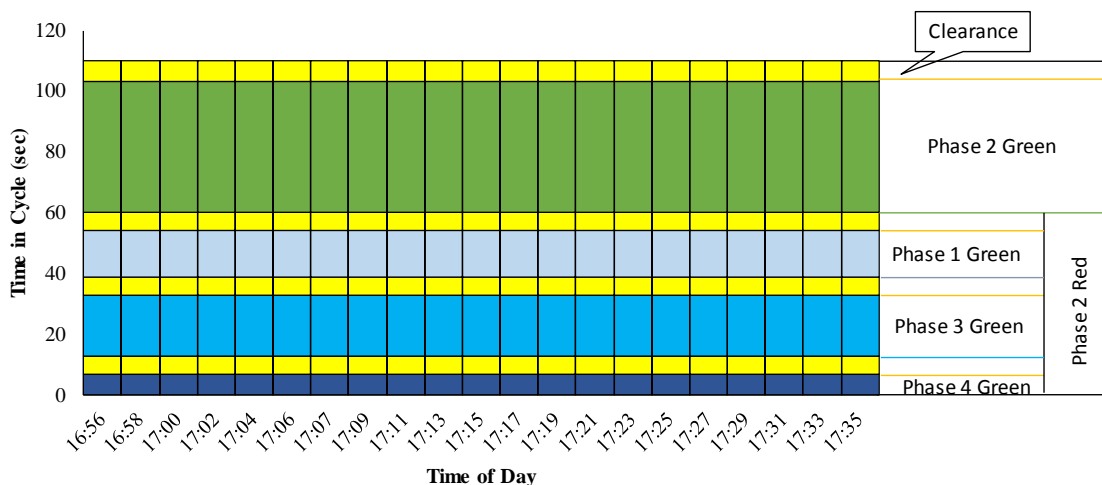


Figure 93. Probability of split for each phase

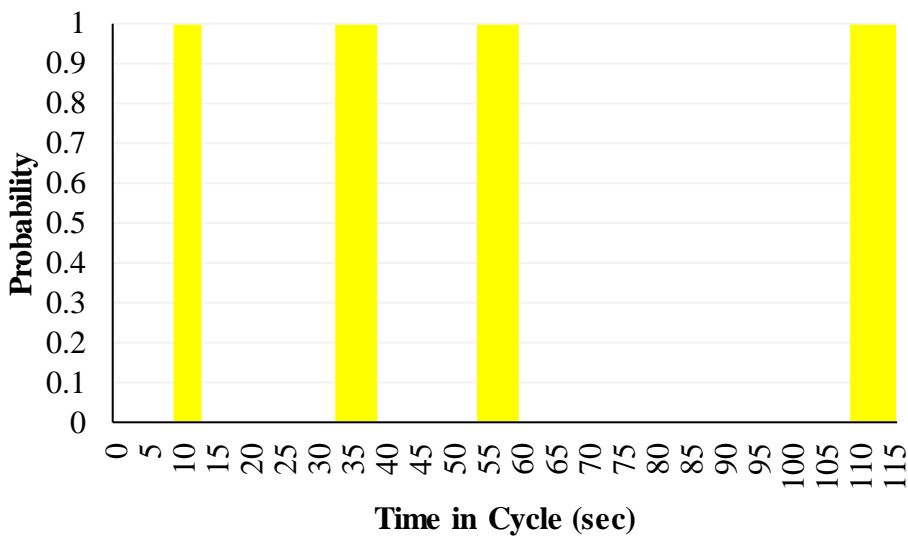
Controllers typically make “phase next” decisions at the onset of clearance state, usually 5 to 7 seconds (for high speed approaches) before phase termination. Clearance states are those periods within the cycle when a phase transitions from green to red. If the probability of clearance states in a cycle is taken into consideration, it may provide a better prediction for the start of the next phase. The probability of the expected clearance state during 1-second bins in the cycle can be estimated in a similar way the probability of phase split was calculated. For signal controllers running a fixed-time [143], clearance states occur at fixed periods within the cycle. However, signal controllers running actuation coordination, there is considerable stochastic variations within each phase due to detector actuations.

For the fixed-time system, as seen in Figure 94 (a), the clearance state occurs at fixed periods (7-12, 33-38, 54-59 and 109-116 seconds) for all cycles, and the probability of the expected clearance state is 100% during these periods (Figure 94 (b)). For the

actuated coordinated system (Figure 95 (a)), there is considerable the variation in the start of the clearance states and Figure 95 (b) shows the corresponding probability distribution of clearance states across the 1-second bins in the cycle. For example, the probability of expected clearance state during the 7-12 second period for fixed system is 100%; the same period in the actuated coordinated system varies between 38% and 76%. Accounting for these variations in the prediction algorithm may improve the signal predictions.

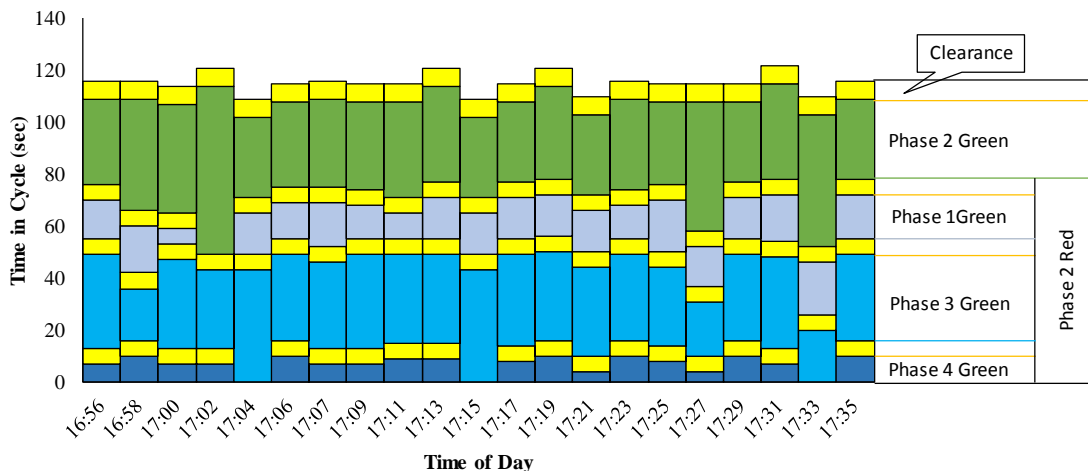


(a) Split diagram with clearance

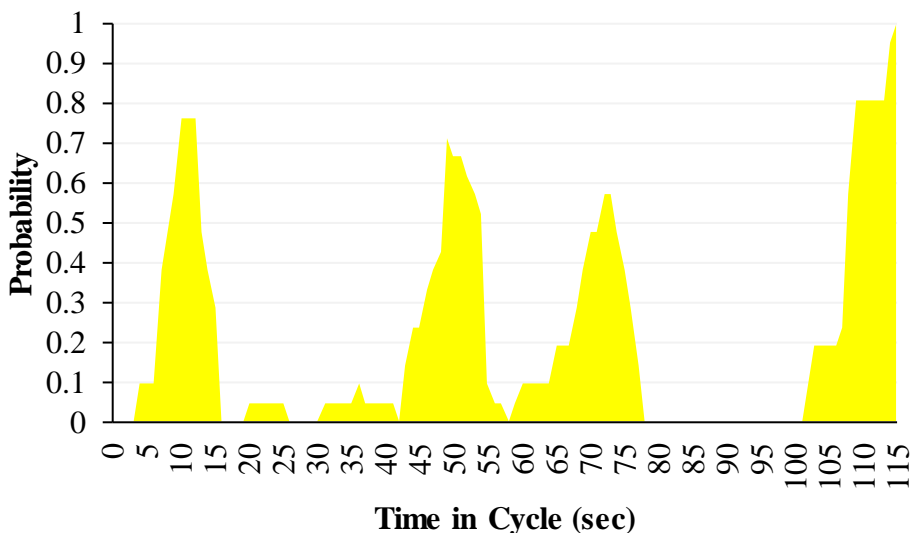


(b) Probability distribution of clearance

Figure 94. Fixed time with clearance



(a) Split diagram with clearance



(b) Probability distribution of clearance

Figure 95. Actuated coordination with clearance

7.5.2 Distribution of green for a phase

Similar to estimating the probability of the expected phase splits, the probability of expected green for each phase within the cycle can also be estimated. Figure 96 illustrates the probability of expected green for the phase 2, northbound through movement on US 231 & River Road. Based on the coordination, if the vehicle arrives between 52s and 67s period in the cycle, the probability of receiving green is 93.5%. For periods before the 50s (callout i), there is a significant variation in the probability of

expected green due to detector actuations. As we approach the force-off at 50s, the probability of accurately predicting the green is very hard.

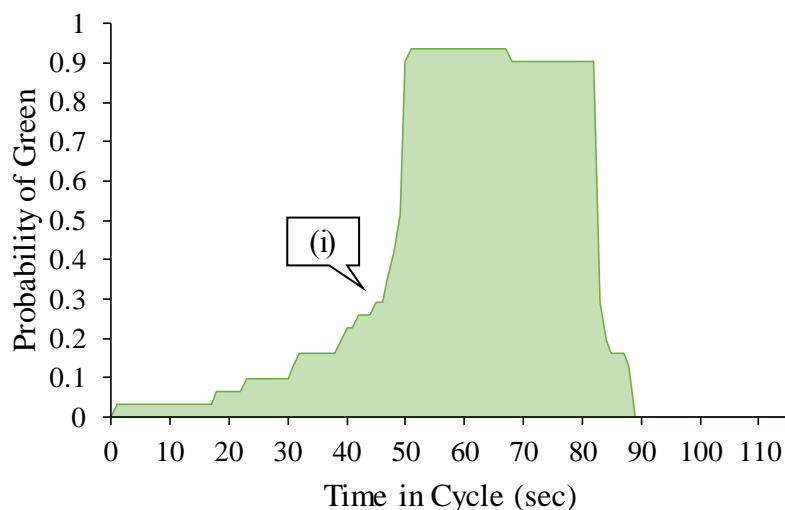


Figure 96. Green distribution for phase 2 (northbound through) movement

Similarly, the green distribution for all the phases in the cycle can be estimated (Figure 97). The sum of all the distributions across a ring at any period in the cycle should equal to one. The minor road movements, especially phases 4 and 7 are highly consistent with more than 80% probability of receiving green between 90 and 100 seconds in the cycle. As for the major movements phases 1, 2 and 6) we can see the variations in the green probability as a result of the detector actuations. The accuracy of the prediction system can be improved by accounting for these probabilities in the prediction algorithm.

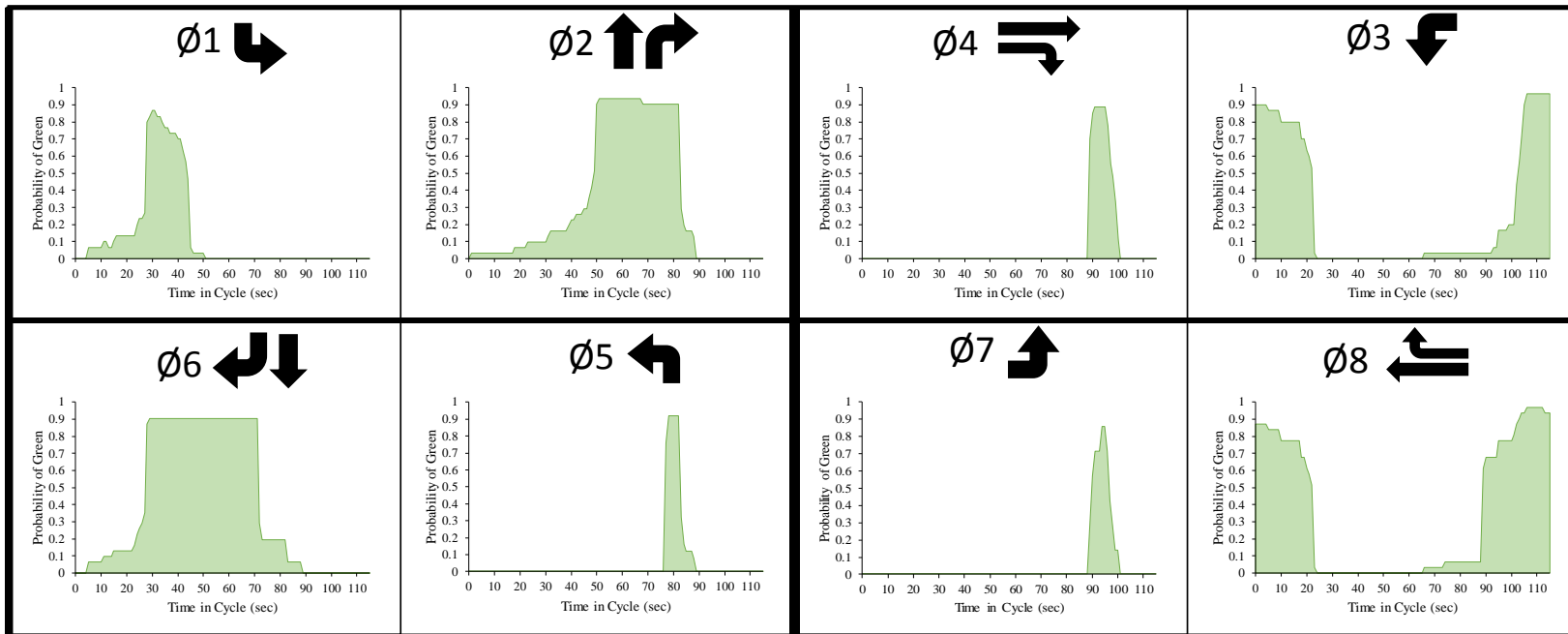


Figure 97. Green distribution on all phases

While actuated coordinated systems are more efficient than the fixed time systems along a corridor, it is difficult to predict the distribution of green at each cycle. As connected vehicle research gains popularity, it is vital to have a technology that provides an accurate estimation of the signal status. In order to use this technology for improving dilemma zone and eco-driving, the V2I communication and predictions need to be accurate. The proposed measures outlined in this chapter including the probability of clearance states and the expected green for each phase will be helpful in improving the accuracy of signal status prediction algorithms.

7.6 Summary

The accuracy and efficiency of V2I communications is important for a variety of connected vehicle applications. This chapter proposes performance measures for evaluating and improving the traffic signal state prediction for connected vehicle applications. Residual plots were prepared as performance measures to evaluate the accuracy of the signal prediction algorithm using video data collected from the field and a data set of 15 trials. The results found considerable variation in the prediction accuracy of the onset of green for the southbound left movement because of the actuated signal coordination. The two measures proposed, including the probability of clearance states and the expected green for each phase, accounts for the variations in an actuated coordinated signal environment and integration into the signal prediction will potentially help improve the efficiency of connected vehicle applications.

Fixed time signal deployments provide very accurate phase status indication (Figure 94). Prior to deployment, figures similar to Figure 93 and Figure 97 should be constructed to identify which phases may have variation that could make it challenging to provide accurate phase prediction for the onset of green. Once systems are deployed, developing residual plots similar to Figure 7 can provide valuable feedback on how effective the phase status algorithms are operating. Finally, a controller typically makes a “Phase Next” decision at the moment a yellow indication starts. For high speed approaches, the time to start of next green may be on the order of 7s. If there is sufficient latency in the cloud based communication, the point the phase next decision can provide an effective point to predict the start time of the next phase.

Advancements in communication technologies has resulted in connected vehicle technologies which may improve safety and mobility and reduce vehicle emissions and fuel consumption. DSRC and cellular technologies allow drivers, vehicles, pedestrians, infrastructure, and transportation management professionals to share data and form a connected transportation network. As the connected vehicle technology develops, it is necessary to have a robust and effective infrastructure that can facilitate proper V2V and V2I communications. The adoption of the measures outlined in this chapter will assist in providing better estimations of predicted onset of green for connected vehicle applications.

CHAPTER 8. CONCLUSIONS

The transportation system is poised for change as new technologies that provide real-time data are increasingly used for a variety of applications. In 2016, the USDOT announced an additional \$65 million in grants on the success of the Smart City Challenge to support IoT transportation projects in four different cities to reduce congestion, increase connectivity and improve opportunity. New technologies and devices provide additional data regarding vehicles, passengers, bikes, traffic signal infrastructure, and overall user mobility. Developing techniques for agencies, transportation system managers, and manufacturers to mine this new emerging “big data” is important to make design, operation, and maintenance decisions that will improve overall mobility and safety.

This dissertation documents performance measures to support analysis techniques for a range of IoT applications. Selected case studies demonstrate the impact of these new data sources on design, operation, and maintenance decisions.

8.1 Rumble Strips

Roadway rumble strips have been proved to be effective in reducing crashes but exterior noise is usually considered a nuisance by nearby residents. The use of vibration, noise level and retroreflectivity as performance measures were demonstrated through a case study of three different sinusoidal rumble strip configurations. The results indicated the 12 in wavelength satisfies the NCHRP recommendations (Table 1) and demonstrated the usefulness of the performance measures developed. The promising results from this research has allowed INDOT and other agencies to develop quantitative data to incorporate 12 in wavelength rumble strips into their design manual.

While rumble strips are potential safety countermeasures on roadways, their application may not be appropriate for all modes. This was demonstrated in a controlled field test that evaluated rumble strips on airfield taxiways as a tool to improve pilot situational awareness and potentially reduce runway incursions. The performance measures included aircraft acceleration and durability. The results indicated that although

the performance measures were useful, long-term deployment of aviation rumble strips is not recommended because the resulting forces may cause undesirable airframe fatigue (Figure 25).

8.2 Arterial Probe Data Performance Measures

Highway monitoring and performance measure requirements have been increasingly emphasized for federal transportation funding mandates. Visualization tools and performance dashboards are actively used by agencies to identify opportunities for mobility improvements in the network. Commercial probe vehicle data from the greater Philadelphia, PA area (around 75 billion records for a period of 3 years) was used to analyze traffic conditions on 138 arterials in the Commonwealth of Pennsylvania. Three web dashboards were developed using performance measures such as CFDs of travel time, normalized median and IQR travel time, and cumulative miles of a corridor operating at different speeds. These dashboards allowed PennDOT to evaluate the impacts of signal system upgrades, signal retiming and maintenance activities on arterials. For five corridors analyzed before and after signal upgradations, the study found a reduction in 1.2 million veh-hours of delay, 10,000 tons of CO₂ (Table 6) and an economic benefit of \$32 million dollars (Figure 52). This analysis has been widely shared by PennDOT and FHWA as a good case study on before/after analysis to quantify signal system infrastructure upgrades and re-timing.

8.3 Airport Security Wait Times

Several billion dollars per year is expended upon security checkpoint screening at airports. Historically, the performance of checkpoints was measured by handing out time stamped paper cards to passengers and collecting them at the x-ray machine. Using anonymized wait time data from consumer electronic devices over a one-year period, performance dashboards were developed that identified periods of the day with high median wait times (Figure 60). Systematically analyzing these irregularities can provide the framework for identifying opportunities to examine operational procedures and resource allocations to improve the service. This has led airports, such as Cincinnati

(CVG) to invest in consolidating checkpoint operations so they would not have to shift staff between physical piers and for the federal government to institute various expedited screening procedures such as TSA PreCheck. This study also provided the first ever systematic evaluation documenting the differential wait time of expedited screening and standard screening using median wait times and reliability as performance measures (Figure 57). As participation evolves in the TSA PreCheck program and inclusion parameters are adjusted, this methodology provides a scalable technique for real-time monitoring and management of our nation's multi-billion dollar investment in airport screening staff.

8.4 Bike share Programs

Bike sharing programs are an eco-friendly mode of transportation gaining immense popularity all over the world. With the growth of bike share programs, there is little in the way of objective performance measures to monitor these programs. Several performance measures were discussed that provided light on the usage patterns, user behaviors and effect of weather on a bike sharing program initiated at Purdue University. A preliminary result from this research (Figure 69) helped the stakeholders to implement policy revisions, which curtailed the users from renting the bikes over 24 hours. These performance measures are a valuable tool for decision makers and agencies to develop new models that monitor the progress of the system and improve the asset management strategies. Since the first study was performed, additional bike share programs have appeared on campus. The metrics proposed in this dissertation provide a framework for organizations (such as Purdue) to monitor how this new transportation mode is evolving, particularly as we enter a period where dock less bike sharing is emerging.

8.5 Outlier Filtering Algorithms and Curated Dataset

Travel time is a key indicator of traffic system performance. Agencies collect or purchase travel time data based on vehicle location data, which requires filtering techniques to remove outliers associated with trip chaining while retaining data linked with incidents or congestion. Although a number of outlier filtering algorithms have been

proposed in the literature, their performance has never been evaluated. In order to so, this study proposed a methodology to develop a first ever curated dataset (Figure 79), where the outlier status of every data point is confirmed. The dataset consisted of 31,621 data points with 243 confirmed outliers. This is published in open access [123] which is beneficial for agencies and researchers to evaluate the performance of their algorithms. Additionally, two performance measures, the proportion of true samples rejected and proportion of outliers correctly identified, are used to analyze three common outlier-filtering algorithms: median absolute deviation, modified z-score and box plots. The modified z-score was found to have the best performance with a successful removal of 70% of the confirmed outliers and incorrect removal of only 5% of the true samples from the curated dataset (Table 9). A number of ITS systems rely on real-time data to provide accurate travel time estimates. Outlier filtering of this data is an essential step and evaluating the accuracy of various algorithms will help agencies to deploy the best ones that provide accurate estimates.

8.6 Traffic Signal State Prediction

The connected vehicles research is an emerging technology that has gained popularity in the recent years. A number of V2V, V2I and V2X applications are being proposed to improve the safety, mobility and reliability of the transportation system. The accuracy of V2I communication is an important metric for connected vehicle applications. Traffic signal state indication is an early data element that vehicle manufacturers are incorporating into vehicles (Figure 84). For fixed time systems (Figure 94) it is easy to predict the signal state. However, when actuated control is used, there is significant stochastic variation on the initiation of a phase (Figure 95), which are influenced by cycle-by-cycle variations in vehicles. This chapter proposed a simple residual plot as performance measure (Figure 88) for analyzing outcome, as well as plots indicating phase-by-phase variations to provide vehicle manufacturers and agencies with data to understand the conditions under which phase predictions work well and degrade (Figure 93). Over the longer term, it is anticipated that these quantitative graphics will stimulate a dialog for additional data elements such as “Phase-next” controller flags to be provided by agency signal controllers. This communication will allow vehicles to have

deterministic near term data to update their phase predictions 5-7s prior to the start of the next phase.

8.7 Summary of Transportation Modes Analyzed and Contribution

Table 12 summarizes the impact of this research and illustrates the application of the performance measures across three transportation modes and the transportation focus areas of safety, mobility and operations.

Table 12. Summary of transportation areas and modes analyzed

Topic	Transportation Areas			Mode	Partner
	Safety	Mobility	Operations		
Sinusoidal rumble strips	X			Roadway	INDOT
Aviation rumble strips	X			Aviation	FAA
Arterial probe data performance measures	X	X	X	Roadway	PennDOT
Airport security wait time			X	Aviation	TSA
Bike share programs			X	Bikes	Zagster/ Purdue University
Outlier filtering and curated data set		X		Roadway	Researchers
Connected Vehicle	X	X	X	Roadway	Vendors

The contributions of this research include:

- The development of performance measures to evaluate sinusoidal rumble strips alternatives.
- The development of performance measures to evaluate aviation rumble strips as a potential safety countermeasure at airports.

- The development of web dashboards and performance measures to evaluate the mobility impact of signal retiming, maintenance, traffic diversion and construction activities on arterials.
- The development of performance measures to identify operational performance at airport security checkpoints and identify opportunities to improve resource allocation.
- The development of performance measures to monitor operations, policies and asset management for bike share programs.
- The development of a curated data set and performance measures to evaluate the accuracy of outlier filtering algorithms for improving travel time estimates.
- The development of performance measures to assess in-vehicle signal predictions for connected vehicle applications.

As we transition into a technology-driven era, the performance measures discussed in this dissertation will provide a framework for agencies and transportation professionals to assess the performance of system components and support investment decisions.

APPENDIX A. STATISTICAL MODELLING OF RUNWAY INCURSION OCCURRENCES IN THE UNITED STATES

INTRODUCTION

Although aviation is the safest mode of transportation [144], the FAA continues to examine additional ways to improve aviation safety and one of their highest priorities is runway safety [145] and the reduction of runway incursions. The FAA defines a runway incursion as, “any occurrence at an aerodrome involving the incorrect presence of an aircraft, vehicle or person on the protected area of a surface designated for the landing and take-off of aircraft” [45].

The annual number of runway incursions reported through the FAA’s Aviation Safety Information Analysis (ASIAS) database has increased by nearly 80% since 2002; this is especially noteworthy since there has been a decrease in the total operations over this period (Figure A1). Over the years, as total operations declined, the incursion rate per 100,000 operations increased from 1.5 in 2002 to more than 3.5 in 2015. According to the FAA Runway Safety Report 2011-12, this increase in the number of incursions is a direct result of the safety culture enhancements adopted by the FAA which encourage the reporting of the incursions through the Air Traffic Organization’s (ATO) Safety Management System [43].

In the 2015 National Runway Safety Plan, the FAA states, “the goal for runway safety is to improve safety by decreasing the number and severity of runway incursions” [146]. There are thousands of airports in the United States, ranging from small general aviation (GA) airports to large hub airports. In order to better identify the different characteristics associated with incursions at these different size airports, this research models incursions by grouping airports based on their National Plan of Integrated Airport Systems (NPIAS) classification.

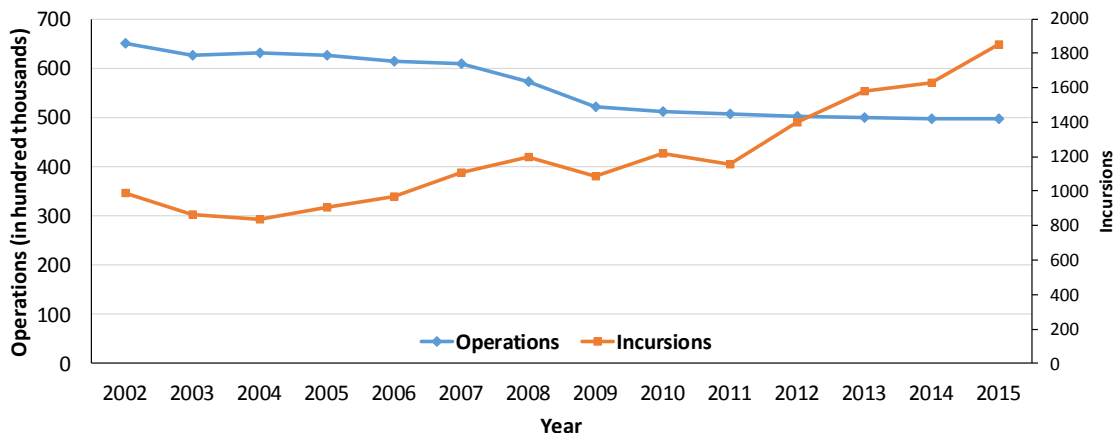


Figure A1. Runway incursions and operations from 2002 to 2015

The purpose of this paper is to use statistical methods to examine the factors that correlate with runway incursions and severity for different airport categories. This paper provides an updated analysis from previous work [147], incorporating additional 2015 data, additional literature, and addition of a new variable (local time when the incursion took place).

BACKGROUND

There are 3,331 airports in the NPIAS [148]. For commercial service airports, NPIAS categorizes airports based on the annual passenger enplanements; there are 389 primary airports (large hub, medium hub, small hub and non-hub) and 2,942 non-primary airports (GA, reliever and non-primary commercial service). Operating characteristics, financial resources, infrastructure and the deployment of technologies can vary across and within these airport categories. Between 2002 and 2015, there were 16,785 runway incursions reported at United States airports. Incursions occurred at airports of every size, from GA to large hubs [149].

FAA categorizes incursions based on severity and causation. Severity designations reflect four major categories, with A being the most severe and D being the least severe [43]:

- *Category A*: an incident in which a collision is narrowly avoided.
- *Category B*: an incident in which the separation decreases and there is a significant potential for collision, which may result in a time critical corrective/evasive response to avoid a collision.
- *Category C*: an incident characterized by ample time and/or distance to avoid the collision.
- *Category D*: an incident which meets the definition of runway incursion, such as the incorrect presence of a vehicle/person/aircraft on the protected area of a surface, designated for landing and take-off of the aircraft, but with no immediate safety consequences.

FAA also categorizes incursions based on the cause, reflecting the following three incident types [43]:

- *Operational incident (OI)*: the action of an air traffic controller resulting in less than required minimum separation between two or more aircraft or between an aircraft and obstacle.
- *Pilot deviation (PD)*: the action of a pilot that violated Federal Aviation Regulation, such as entering the runway without permission.
- *Vehicle/pedestrian deviation (V/PD)*: entry of vehicles or pedestrians into the airport movement areas without air traffic controller (ATC) authorization.

Runway incursions occur for a number of reasons. Previous research reveals that more than 70% of all aviation accidents are caused by human error [49], [50]. According to the FAA, 65% of the 1,264 runway incursions in 2014 were due to PD [46]. Lack of situational awareness is one of the most common factors leading to PD [150]. Some of the factors that lead to an OI incursion include call sign confusion, poor read-back procedure, and incorrect phraseology [56].

Cardosi and Yost (2001) synthesized previous research regarding runway incursions. Their findings generally reflect a human factors perspective and include:

- “Failure to anticipate the required separation or miscalculation of the impending separation,
- Forgetting about an aircraft, the closure of a runway, a vehicle on the runway, and/or a clearance that was issued,
- Communication errors, readback/hearback errors, issuing an instruction other than the one the controller intended to use, and
- Lack of, or incomplete, coordination between controller” [151].

In recent years, statistical modelling has been used to identify factors linked to runway incursions. Green (2014) developed a Bayesian Belief Network (BBN) to examine the contributing factors to a runway incursion, including airport issues, weather issues, operational issues, and communication issues. Wilke et al., (2015) examined the “airport surface system architecture” based on regression analysis, and concluded that the geometric characteristics of an airport affect the runway incursion severity. Johnson et al., (2016) confirmed the impact of airfield geometrics on incursions, reporting that airports with runway intersections (taxiway or another runway) have a higher occurrence of incursions than airports without runway intersections.

Biernbaum and Hagemann (2012) used a multinomial logit model to examine 8,812 incursions from 2001 to 2010 and found that OI incidents were less likely to result in a more severe incursion at the 35 busiest airports, but more likely to result in severe incursions (A and B) at other airports. This finding highlights that incursion characteristics vary depending on airport size.

STUDY SCOPE

While previous research has focused on factors that increase the likelihood of a runway incursion, no research has examined the factors that influence the likelihood of an incursion based on the category of airport using a multinomial logit model with random parameters. Figure A2 shows the top 50 airports in terms of the number of runway incursions in the past 5 years. This list includes airports from all six NPIAS categories (Large, Medium, Small, General Aviation, Reliever, and non-hub). Analysis of runway incursions by airport size allows a better understanding of the characteristics associated with incursions at different kinds of airports. This analysis may help identify appropriate

mitigation measures for incursions at different sizes of airports. Different size airports may serve different aircraft fleets, serve aircraft operating under different regulations, serve different operating volumes, and have different resources available (both funds and technologies) for incursion mitigation. The increasing number and rate of runway incursions, and the importance of reducing runway incursions to ensure aviation safety, warrant new ways to examine incursion data. Although factors such as airfield geometry and crossing runways [154] as well as hotspots play an important role in runway incursions, the scope of this paper was limited to the modelling factors affecting the severity and incident types at airports grouped by airport category.

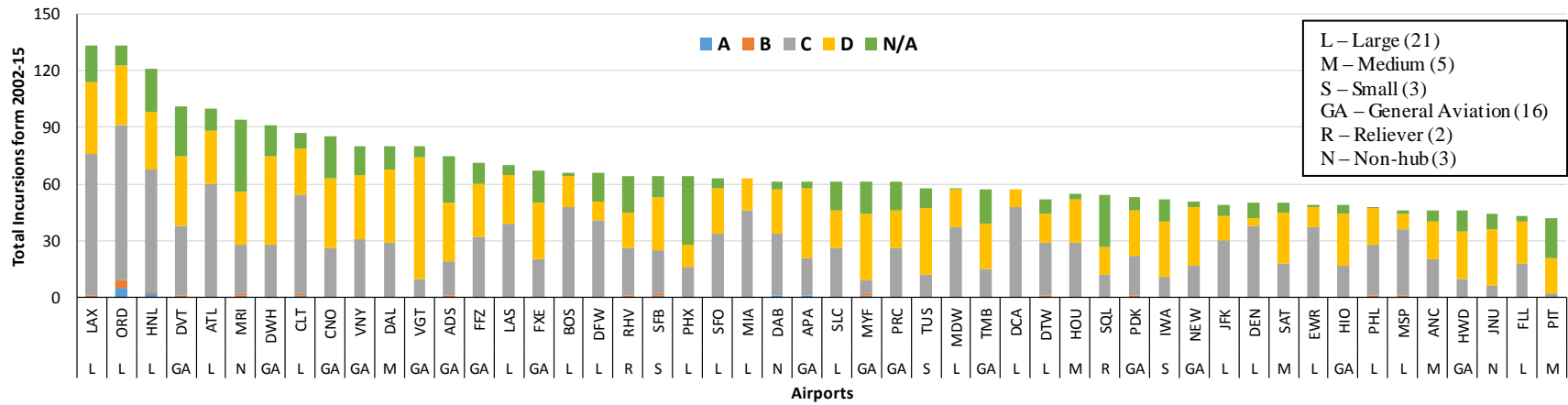


Figure A2. Top 50 airports with highest number of runway incursions, color coded by reported severity (2010-2015)

DATA DESCRIPTION

The data for this study was obtained from the FAA Runway Safety Office – Runway Incursions (RWS) database [149] which is part of the FAA Aviation Safety Information Analysis and Sharing (ASIAS) system. The database contains the records for all runway incursions that occurred in the United States since 2001. This analysis is based on 16,785 observations from 2002 through 2015. Table A1 shows the variables recorded in the RWS database.

Table A1. Variable description

Variables	Description
Incursion Severity	Severity category - A/B/C/D
Incident Type	OI/PD/VPD
Event Date	Date of event
Airport Location	Location of airport
Airport Code	3 or 4 letter alphanumeric airport code
Aircraft Flight Conduct Code	FAR part under which the aircraft was operated
Aircraft Code	Type of aircraft (e.g., A320)
Weather	Weather condition in METAR format
Local Time	Local time when the incursion occurred
Runway	Runway on which the incident occurred
Event description	Detailed narrative of the incident

This incursion data was divided into three categories, reflecting large, medium/small and GA/non-hub/reliever/non-primary commercial service, as defined by the NPIAS categories shown in Table A2. Medium and small hub airports were combined into one category to satisfy minimum sample size requirements. Incursions of severity A and B are rare and contribute to less than 4% of all incursions, however, these incursions are the most severe and may be most important to mitigate. At medium and small hub airports, analysis of severity A and B incursions were combined to ensure an adequate minimum sample.

Table A2. Summary Statistics for NPIAS Airport Categories Analyzed

Airport Category	Number of Airports Reporting Incursions	Number of NPIAS Airports in 2017	Total 2011-2015 Enplanements (in millions)	Total 2011-2015 Operations (in millions)	Example Airport From Figure 2
Large	30	30	2706.4	63.06	Los Angeles (LAX)
Medium	33	33	606.2	25.03	Houston Hobby (HOU)
Small	71	76	307.4	32.53	Orlando Sanford (SFB)
GA	200	2,553	2.3	79.01	Phoenix Deer Valley (DVT)
Reliever	12	264	-	5.92	San Carlos (SQL)
Non-hubs	160	251	105.1	40.73	Daytona Beach (DAB)

This research utilized several variables from the RWS database. In addition to incursion severity and incident type, operational characteristics such as the Federal Aviation Regulation (FAR) under which the aircraft operated (e.g., part 91 for GA and part 121 for domestic commercial operations), event date, and local time were also examined. Adverse weather conditions can be a potential factor affecting incursions; however, the METAR data in the RWS database was reported inconsistently, making it difficult to perform analysis regarding weather (e.g., instrument or visual flight rules).

After removing records with missing information, the final data set had 11,262 observations (2,747 large hub, 2,952 medium/small hub, and 5,563 GA/non-hub). All variables were tested for each of the three airport categories. Table A3 shows the descriptive statistics of all the variables for each of the airport categories. In some cases, findings may be constrained by the limited sample. For example, there were only 107 OI incursions of severity A or B; which makes evidence of correlation more challenging. This may be addressed by using the methodology demonstrated in this paper with additional data reflecting incursions over a longer period of time in the future.

Table A3. Descriptive statistics of the variables

Variable Description	Large		Medium/Small		GA/Non-Hub	
Total Observations (number of incursions)	2,747		2,952		5,563	
Dependent variable	Percentage		Percentage		Percentage	
Incursion severity (A/B/C/D)	2/2/55/41 %		1/1/35/63 %		2/2/32/64 %	
Indicator variables	Percentage		Percentage		Percentage	
Pilot deviation (PD), 1 if yes else 0	48.08%		63.98%		66.17%	
Operational incident (OI), 1 if yes else 0	36.69%		20.14%		12.69%	
Vehicle/Pedestrian deviation (VPD), 1 if yes else 0	15.14%		15.70%		20.96%	
Commercial aircraft involved, 1 if yes else 0	73.98%		32.41%		2.91%	
GA aircraft involved, 1 if yes else 0	-		64.63%		-	
Incident between 2 commercial aircraft, 1 if yes else 0	36.67%		-		-	
Incident between commercial and GA aircraft, 1 if yes else 0	-		13.15%		1.42%	
Incident between 2 GA aircraft, 1 if yes else 0	-		-		31.40%	
Charter aircraft involved, 1 if yes else 0	10.41%		11.22%		4.00%	
Foreign aircraft involved, 1 if yes else 0	9.57%		-		-	
Incident during months of December, January and February, 1 if yes else 0	-		-		19.39%	
During peak operating hours, 1 if yes, else 0	91.85% (06:00 – 22:00)		60.73% (09:00 – 16:00)		83.57% (08:00 – 17:00)	
Continuous variable	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
Number of years since 2002	8.48	3.65	8.04	3.62	8.69	3.49

MULTINOMIAL LOGIT MODEL

A multinomial logit model was used to model the runway incursions for each airport size category. This model was selected since the dependent variable, runway incursion severity, is ordered discrete choice data. Although the incursion severities are ordered (ranging from the most severe A, to the least severe D), and an ordered model would also be a possible choice, it was not selected due to the inability of an ordered model to capture the effect of interior category variables [83].

There are a number of factors that could affect the severity of the incursions, but are difficult to measure. A random parameter (also known as a mixed logit model or MNL with random parameters) model can be used to quantify the heterogeneity due to the individual differences associated with the inconsistent effects of the independent variables on the dependent variable [156]. The mixed logit model accounts for the influences of the unobserved heterogeneity by allowing the estimated parameters to vary across the observations according to a specific distribution.

To develop the modelling approach [83], the severity function with discrete outcome probabilities is defined as

$$T_{in} = \beta_i \mathbf{X}_{in} + \varepsilon_{in} \quad (1)$$

where T_{in} is the severity function determining the severity category i (A, B, C and D) for observation n ; β_i is a vector of the estimable parameters; \mathbf{X}_{in} is a vector of the explanatory variables that determine the discrete outcome for observation n and ε_{in} is the error term. The explanatory variables shown in Table 1 (PD, OI, Incident between 2 GA aircraft, etc.) are all binomial variables; the variable year since 2002, which is a continuous variable, was also used. If the error terms are assumed to be generalized extreme value distributed, then this gives rise to the standard MNL, given by

$$P_n(i) = \frac{EXP[\beta_i \mathbf{X}_{in}]}{\sum_{\forall I} EXP(\beta_i \mathbf{X}_{in})} \quad (2)$$

where $P_n(i)$ is the probability of observation n having discrete severity outcome i (from the set of all categories I). For the mixed logit model, in order to allow for the parameter variations across observations, a mixing distribution m is defined for the outcome probabilities, where

$$P_n^m(i) = \int_{\mathbf{x}} P_n(i) f(\boldsymbol{\beta} | \varphi) d\boldsymbol{\beta} \quad (3)$$

where $f(\boldsymbol{\beta} | \varphi)$ is the density function of $\boldsymbol{\beta}$ with φ referring to a vector of parameters of the density function (mean and variance). Substituting (2) in (3) gives the mixed logit model,

$$P_n^m(i) = \int_{\mathbf{x}} \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_l EXP(\boldsymbol{\beta}_l \mathbf{X}_{in})} f(\boldsymbol{\beta} | \varphi) d\boldsymbol{\beta} \quad (4)$$

For model estimations, $\boldsymbol{\beta}$ can account for the variations of \mathbf{X} across the incursion categories, with the density function used to estimate $\boldsymbol{\beta}$. The mixed logit probabilities are a weighted average of the different values of $\boldsymbol{\beta}$ across the observations where some may be fixed and some random. For the functional form of parameter density function, distributions such as normal, log normal, uniform and triangular were considered.

Maximum likelihood estimations of the multinomial logit models are computationally intensive and hence simulation based maximum likelihood method was employed using Halton draws to estimate the model using Limdep (NLOGIT 4.0).

MODEL ESTIMATION RESULTS

The estimation results of the three mixed logit models developed using 200 Halton draws (Milton et al., 2008) are shown in Table A4, Table A5 and Table A6 for GA/non-hub, medium/small and large airports, respectively. All the estimated parameters in the model are statistically significant (90% confidence level) and the signs are plausible. The parameters found to be random were those which produced a statistically significant (90% confidence level) standard error for their assumed distribution. In each case, the normal distribution provided the best statistical fit for the random parameters. The parameters with estimated standard errors that were not significantly different from zero were treated as fixed parameters. The McFadden pseudo R – squared (ρ^2) shown in Tables 4, 5 and 6 is a common measure that indicates the overall model fit. The adjusted ρ^2 statistic is given by

$$\rho^2 = 1 - \frac{LL(\boldsymbol{\beta}) - K}{LL(0)} \quad (14)$$

where $LL(\boldsymbol{\beta})$ is the log-likelihood at convergence with parameter vector $\boldsymbol{\beta}$, $LL(0)$ is the restricted log – likelihood at zero and K is the number of parameters estimated in the

model. The ρ^2 statistics lies between 0 and 1; a statistic close to one indicates a better fit which suggests that the model predicts the outcomes with near certainty [83]. For practical data sets, ρ^2 values above 0.40 is considered a good fit.

Pilot Deviation

The impact of PD varies depending on the size of the airport. Consider first, GA and non-hub airports. Looking at Table A4, the variable pilot deviation was significant for severity C and D incursions with a fixed parameter estimate of 1.3 and 1.67, respectively. The positive sign here implies that pilot deviations are more likely to result in severity C and D incursions. PD may be due to a loss of situational awareness, resulting in wrong turns and failure to stop at the hold short line, and may be more likely for pilots at unfamiliar airports. Many airports in this category do not have an air traffic control tower, in which case all incursions would be either a PD or a V/PD.

PD also increased the likelihood of a severity D incursion at medium/small airports (parameter estimate of 1.75) (Table A5). These airports also have a significant number of GA operations and thus a loss of situational awareness among the pilots could lead to the incursion. At large airports (Table A6), pilot deviations were significant for severity B incursions with a parameter estimate of -1.28. This means that PD incursions are less likely for severity B at large airports. This could be the direct impact of the steps taken by the FAA to reduce PD through awareness (advisory circulars and hotspots), education (video tapes to enhance the recognition of signs and markings), procedures and technologies, such as runway status lights (RWSL) and enhanced taxiway markings [157]. The runway safety report from 2008 also confirms that the FAA implemented some of these measures at higher volume airports to improve runway safety [158].

Operational Incidents

At GA/non hub airports, OI were more likely to result in severity C incursions (Table A4); at these airports the OI variable was statistically significant but not random. Inadequate tower height has been reported to cause OI at selected airports [159]. At airports such as Sarasota–Bradenton International Airport (SRQ) (small hub airport in Florida) and Flying Cloud Airport (FCM) (GA airport in Minnesota), the limited view of

the air traffic controller has resulted in OI [159]. Low aircraft volumes at GA/non-hub airports may also result in lower severity incursions.

The OI for severity C was statistically significant and varied randomly at medium/small hub airports (Table A5). The parameter for this indicator variable was normally distributed with a mean of 0.26 and a standard deviation of 2.68. Given these estimates, the parameter was less than zero for 54% and greater than zero for 46% of incursions at medium hub airports. In other words, 54% of the OI at the medium/small hub airports increases the likelihood of a severity C incursion.

OI for severity C at large hub airports (Table A6) was also statistically significant as a random parameter, with a mean of 2.05 and standard deviation of 1.76. With these parameter estimates, 89% of the OI at large hub airports were more likely to result in an incursion of severity C and the remaining 11% are less likely to result in an incursion of severity C. Large airports are very busy with many take-off and landings. The resulting heavy workload for the air traffic controllers could increase the likelihood of an OI event at the majority of large hub airports.

Table A4. Estimation results for NPIAS GA/Reliever/Non-Hub airports

Variable	Parameter estimate	t-stat	p-value
Severity A			
Number of years since 2002	-0.25	-7.02*	<0.00001
Incident between 2 GA aircraft	1.52	4.69*	<0.00001
Incident between GA and Commercial aircraft	2.00	2.96*	0.0031
Incident during Dec or Jan or Feb (standard deviation of parameter distribution)	-7.23 (5.29)	-1.36 (2.01)*	0.208 (0.068)
Severity B			
Constant	-1.78	-4.68	<0.00001
Incident between 2 GA aircraft	1.77	6.03*	<0.00001
Charter involved	2.00	4.86*	<0.00001
Severity C			
Constant	1.10	3.19	0.0014
Operational incident	1.26	4.19*	<0.00001
Pilot deviation	1.35	4.67*	<0.00001
Local time (08:00 to 17:00)	0.22	2.65*	0.0082
Severity D			
Constant	1.79	4.85	<0.00001
Pilot deviation	1.67	7.84*	0.0027
Vehicle/Pedestrian deviation	0.89	3.00*	<0.00001
Number of observations	5563		
Restricted log-likelihood (at zero)	-7711.959		
Log-likelihood at convergence	-3957.482		
McFadden Psuedo R-squared	0.48		

*significantly different from zero at more than 90% confidence

Vehicle Pedestrian Deviations

V/PD were more likely to result in severity D incursions at GA/non-hub (Table A4) and medium/small airports (Table A5). Low aircraft volumes at GA/non-hub may result in lower severity incursions. Vehicles and pedestrians may not have separate access roads at these airports and are more likely to enter an active runway without ATC clearance (since many smaller airports do not have an air traffic control tower) and thus, could result in an incursion. GA airports vary widely and local conditions may present specific challenges. According to the Runway Incursion Airport Assessment Report,

inadequate fences/signs to provide warning of secured areas could result in pedestrians and vehicles entering these areas without proper authorization [159]. V/PD at large hub airports were also found to result in severity C as well as severity D incursions (Table A6).

Table A5. Estimation results for NPIAS Medium/Small Hub airports

Variable	Parameter estimate	t-stat	p-value
Severity A and B			
Number of years since 2002	-0.18	-4.67	<0.00001
GA aircraft involved	0.67	2.03*	0.0423
Severity C			
Constant	1.69	4.23	<0.00001
Commercial aircraft involved	0.66	5.57	<0.00001
Operational incident (standard deviation of parameter distribution)	0.26 (2.68)	0.57 (2.36)*	0.5706 (0.0182)
Charter involved	0.68	4.63*	<0.00001
Incident between GA and Commercial aircraft	1.05	6.45*	<0.00001
Local time (09:00 to 16:00)	0.25	2.53*	0.0113
Severity D			
Constant	1.5	3.98	0.0001
Pilot Deviation	1.75	6.21*	<0.00001
Vehicle/Pedestrian Deviation	1.64	5.46	<0.00001
Number of observations	2952		
Restricted log-likelihood (at zero)	-3239.808		
Log-likelihood at convergence	-1895.743		
McFadden Pseudo R-squared	0.41		

*significantly different from zero at more than 90% confidence

Interaction Between GA and Commercial Aircraft

Incidents between a GA and commercial aircraft were more likely to result in an incursion of severity A at the GA airports (Table A4). The GA community may be unfamiliar with the operational needs of commercial aircraft (e.g. turbine aircraft), especially when commercial aircraft utilize GA airports. Turbine aircraft may require greater separation than GA aircraft due to speed. Unless given proper training, GA pilots may not be familiar with the operating characteristics of commercial aircraft and

misjudge the distance required, resulting with inadequate separation distance between aircraft and resulting in a severe incursion. At medium/small airports, this variable was more likely to result in severity C incursions. This may be due to fewer GA operations at medium/small airports relative to GA/non-hub airports. This variable was not significant in the large hub model, which may reflect that GA aircraft are less likely to fly into large airports.

Table A6. Estimation results for NPIAS Large Hub airports

Variable	Parameter estimate	t-stat	p-value
Severity A			
Number of years since 2002	-0.37	-3.65*	0.0003
Commercial aircraft involved (standard deviation of parameter distribution)	-9.81 (6.75)	-1.82(2.49)*	0.069 (0.0128)
Operational incident	2.02	2.87*	0.0041
Severity B			
Constant	-0.43	-0.59	0.5586
Pilot deviation	-1.28	-3.43*	0.0006
Severity C			
Constant	1.22	1.75	0.0804
Operational incident (standard deviation of parameter distribution)	2.11 (1.76)	5.58(2.96)*	<0.00001 (0.0031)
Incident between 2 commercial aircraft	1.03	2.70*	0.0069
Vehicle/Pedestrian deviation	1.53	2.89*	0.0038
Charter involved	0.68	4.22*	<0.00001
Foreign involved (standard deviation of parameter distribution)	0.89 (2.06)	3.49 (1.62)*	0.0005 (0.1050)
Local time (06:00 to 22:00)	0.31	2.97*	0.0029
Severity D			
Constant	2.33	3.355	0.0008
Incident between 2 commercial aircraft	-0.87	-2.39*	0.0171
Vehicle/Pedestrian deviation	1.00	1.94*	0.0530
Number of observations	2747		
Restricted log-likelihood (at zero)	-3808.151		
Log-likelihood at convergence	-1863.514		
McFadden Pseudo R-squared	0.51		

*significantly different from zero at more than 90% confidence

Commercial Aircraft Involvement

Involvement of a commercial aircraft was statistically significant as a random parameter for severity A incursions at large hub (Table A6) airports. The parameter for this indicator variable was normally distributed with a mean of -9.81 and a standard deviation of 6.75. Given these estimates, 93% of the incursions at large hub airports involving a commercial aircraft were less likely to result in an incursion of severity A, and more likely to result in incursions of severity B, C or D. This implies that the impact varies over the sample, which may be due to the wide range of airports within the category. The largest large hub airports are Chicago O'Hare International Airport (ORD) and the Hartsfield – Jackson Atlanta International Airport (ATL), each with more than 882,000 operations in 2015; and the smallest large hub airport is Tampa International Airport (TPA) with more than 189,000 operations in 2015. Commercial aircraft at large hub airports have highly trained and experienced pilots as well as ATC support for guidance, both factors which are likely contribute to the reduced probability of an incursion. For the remaining 7% of the incursions at large hub airports, the parameter is positive, implying that the incursions involving a commercial aircraft are more likely to result in severity A incursions. These incursions could be due to the higher frequency of aircraft, reflecting more commercial operations that occur at large hub airports.

Number of years since 2002

The continuous variable, “number of years since 2002”, was found to be significant for all the three airport categories for severity A incursions. This variable has a negative co-efficient indicating that the incidence of severity A incursions has decreased since 2002. Severity A incursions are very rare and account for less than 2% of incursions at all airport categories (Table A3). Although specific causes cannot be identified from this data set, perhaps strong continuous improvement for safety in the aviation community and the introduction of new technologies at larger airports such as Airport Surface Detection Equipment, Model-X (ASDE-X) [160], which allows the controllers to track the surface movement of aircraft and vehicles, have been contributors. ASDE-X was first implemented in 2003 with the specific motive of reducing severity A and B incursions. At GA/reliever/non-hub airports, an increased focus by FAA on

publicizing hotspots (Federal Aviation Administration, 2014d) may be one reason for the reduction in severe incursions.

Time of Day

The peak incursion hours were identified by mapping histograms of the incursions based on time of day, using one hour bins. At GA/non-hub (Table A4) and medium/small airports (Table A5), incursions during the 08:00 to 16:00 and 09:00 to 17:00 time periods respectively, were more likely to result in a severity of type C. These time periods presumably represent the hours when most operations occur. At large hub airports, incursions during the 06:00 – 22:00 time period (Table A6), were also found to increase the likelihood of severity C. Large hub airports primarily serve commercial and international operations which extend all throughout the day.

Winter Periods

Incursions that occurred during the months of December, January and February at GA/non-hub airports (Table A4) were found to vary randomly with a statistically significant mean (-7.23) and standard deviation (5.29). This means that 91.4% of the incursions at GA/non-hub airports were less likely to be severity A and were more likely to be a less severe incursion. This is probably due to the reduced number of operations during the winter periods (Figure A3) at GA/non-hub airports in many parts of the country. The remaining 8.6% of the incursions during this period were more likely to result in severity A incursions.

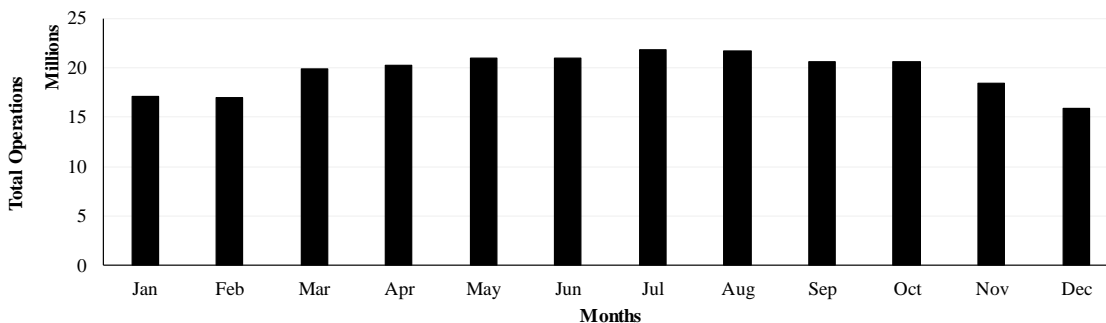


Figure A3. Reduced GA operations during the winter periods (2002 – 15)

INCURSION CHARACTERISTICS BY AIRPORT CATEGORY

Table A7 shows a synthesis of the most important variables for each of the three models. Discussion on the incursion characteristics by airport category is provided below.

GA/Non – Hub/Reliever

Incursion characteristics varied depending on the category of airport. At GA/non-hub airports, more than 95% of the incursions belonged to severity C and D categories, these incursions were more likely to be PD. Incidents involving a GA aircraft were more likely to result in a severity A or B incursion (Table A4). As a result, mitigation initiatives should enhance the situational awareness of the pilots, and targeted to GA pilots and aircraft.

Medium/Small Hub

At medium and small hub airports, a variety of factors were correlated with incursions. PD and V/PD were positively correlated with severity D incursions. Incidents involving GA aircraft were likely to result in severity A/B incursions, while incidents involving a commercial aircraft were more likely to result in a severity C incursion. At medium and small hub airports, air traffic control becomes a factor affecting the probability of incursions. For more than half of the incursions, OI had an increased likelihood of resulting in a severity C incursion. Since medium and small hub airports comprise a wide range of airports with respect to both operations and physical layout, a variety of incursion countermeasures may be appropriate. The range of operations in this category is exemplified by Cincinnati/Northern KY International airport (CVG), a medium airport with 133,500 operations in 2014, 4% of which were GA [161], and Long Beach airport (LGB), a small hub airport with 251,957 operations in 2014, 86% of which were GA [162]. While low cost countermeasures may be appropriate at smaller airports with substantial GA operations, technology based countermeasures including those focused on ATC may be more appropriate at airports with more operations and more commercial service.

Table A7.Overall model results

Variable	GA/Non-Hub		Medium/Small		Large	
	Severity	β	Severity	β	Severity	β
Pilot Deviation	C	1.35	D	1.75	B	-1.28
	D	1.67				
Operational Incident	C	1.26	C	0.26 (2.68)	A	2.02
					C	2.11 (1.76)
Vehicle/Pedestrian Deviation	D	0.89	D	1.64	C	1.53
					D	1.00
Commercial aircraft involved	-	-	C	0.66	A	-9.81 (6.75)
Incident between 2 commercial aircraft	-	-	-	-	C	1.03
					D	-0.87
Incident between commercial and GA aircraft	A	2.00	C	1.05	-	-
Incident between 2 GA aircraft	A	1.52	-	-	-	-
	B	1.77				
GA aircraft involved	-	-	A/B	0.67	-	-
Time of Day	C	0.22	C	0.25	C	0.31
Number of years since 2002	A	-0.25	A/B	-0.18	A	-0.37
Incident during months of December, January and February	A	-7.23 (5.29)	-	-	-	-

Large Hub

Large hub airports serve significant volumes of commercial service and ATC plays a critical role in the maintenance of safe operations. OI was a significant factor at these airports and had an increased probability of resulting in a severity A and C incursion. Although PD were less likely to result in severity B incursions (PDs were not

significant in A, C, and D categories), V/PD were significant for both severity C and D incursions. Since OI errors are more prevalent at large hub airports, solutions to support ATC activities are probably most appropriate. Countermeasures that utilize advanced technologies are most appropriate at large hub airports, where the volume of commercial aircraft and high cost associated with even minor delays provide justification for the substantial financial outlay required. Furthermore, large hub airports are more likely to have a technology platform that can be leveraged to implement NextGen and advanced technology based solutions, and the commercial aircraft that utilize these large-hub airports are more likely to have the equipment required to utilize them.

CONCLUSIONS

Over 16,000 incursions occurred at 506 airports between 2002 and 2015. This research developed statistical models to evaluate factors that correlate with incursions at large, medium/small, and non-hub/GA airports. Runway incursion characteristics varied amongst airport sizes.

- At GA and non-hub airports, incursions due to PD were more common, as were severe incursions involving a GA aircraft.
- At medium and small hub airports, a more diverse set of factors were correlated with incursions including PD, OI and V/PD.
- At the largest airports, OI had an increased probability of resulting in severe incursions.
- The only variable that was statistically significant for all categories of airports was “number of years since 2002”, which was significant for severity A incursions. The parameter estimate was negative, indicating that the likelihood of the most severe incursions has diminished since 2002. Although specific causes for this reduction cannot be identified from this analysis, it is likely due to the strong continuous improvement in safety in the aviation community and the introduction of new technology at larger airports such as ASDE-X.

The statistical models presented in this paper provide a mathematical explanation for incursions based on the data reported by the FAA. As mentioned previously, there are limited data for severity A and B incursions; these two categories together contribute to

less than 4% of all incursions. As a result, there are fewer factors that correlate with statistical significance for category A and B incursions. When subsets are examined (e.g., severity A incursions at a large hub airport attributable to OI that includes a GA aircraft), the likelihood of statistical significance decreases further. Since there are more C and D incursions, the models are more likely to identify factors that are statistically significant for C and D incursions. The proposed models cannot provide statistical confirmation of all causes of runway incursions for all severity levels.

The intent of this paper is to provide and demonstrate a method to assess factors contributing to incursions at airports based on airport category, which is valuable since operations at different airports differ significantly, and as a result, incursions and appropriate mitigation measures also differ. The findings illustrate that the factors contributing to runway incursions do vary depending on the size of the airport and the severity of the incursion. When examining the top 50 airports by total incursions, large hubs accounted for 21 airports and general aviation accounted for 16 airports (Figure A2). For this reason, there is no single best solution to prevent runway incursions across all airports, and in fact, the most appropriate countermeasures should vary depending on the airport category. Although the rate of severe incursions (A/B) has slightly increased over the past 3 years, factors contributing to this increase may not be reflected in this research since this analysis examines the 14 years from 2002 to 2015. The methodology used in this paper can be applied in the future, as more data becomes available, to provide an increasing understanding of the factors that contribute to incursions. Future research could include alternative modelling approaches with different grouping strategies to increase the sample size for the severe incursion categories. It would also be worthwhile to model incursions at airports in other countries as well, provided that the airports use the ICAO definitions for runway incursion severity and appropriate data is available.

ACKNOWLEDGEMENTS

This work was supported by the Federal Aviation Administration through the PEGASAS Research Center. Although the FAA has sponsored this project, it neither endorses nor rejects the findings of this research. The presentation of this information is in the interest of invoking technical community comment on the results and conclusions of the research.

APPENDIX B. METHODOLOGY FOR BENEFIT COST ESTIMATION

The calculation of user benefits was achieved through the use of existing annual average daily traffic (AADT) data provided for the corridors in combination with information found in the 2014 Pennsylvania Traffic Data Report [163]. This report provided data on hourly vehicle and truck traffic percentages. These numbers were determined through 92 study locations across the state. Each of these locations was categorized into 10 traffic pattern groups. The five corridors in this study are defined as “TPG 3, Urban – Other Principle Arterials.” Additional user benefit design values were adapted from the *2015 Urban Mobility Scorecard* [164], including vehicle occupancy, commercial vehicle operating cost, and average cost of time. The time period selected for the analysis was from 6AM to 8PM.

The hourly volumes were estimated by equation (15)

$$vol_i = AADT * k_i * d \quad (15)$$

where,

vol_i = estimated volume for hour i

AADT = annual average daily traffic

k_i = hourly vehicle percentages from [163]

d = directional distribution (assumed to be 0.5)

The difference in travel time for each hour, before and after the adaptive signal deployment was calculated using the following equation

$$\Delta TT_i = TT_{before,i} - TT_{after,i} \quad (16)$$

where,

$TT_{before,i}$ = median travel time during the before period for hour i

$TT_{after,i}$ = median travel time during the after period for hour i

The user benefit for trucks during each hour was then calculated using equation (17)

$$user_{truck,i} = vol_i * \Delta TT_i * \%T_i * PPV_t * VOT_t \quad (17)$$

where,

$\%T_i$ = percentage of truck traffic for hour i, from [163]

PPV_t = number of passengers for commercial vehicles (1 for trucks)

VOT_t = time value of money for commercial vehicles, \$94.04/vehicle-hr from [164]

Similarly, the user benefits for passenger cars were computed using equation (18)

$$user_{car,i} = vol_i * \Delta TT_i * \%C_i * PPV_c * VOT_c \quad (18)$$

where,

$\%C_i$ = percentage of car traffic for hour i , assumed as $(1 - \%T_i)$

PPV_c = number of passengers for commercial vehicles, 1.25 for cars from [164]

VOT_c = time value of money for passenger cars, \$17.67/person-hr from [164]

The hourly user costs for passenger cars and commercial trucks in both the directions were aggregated to compute the daily user cost for each corridor. The annual user costs were then computed by using weekly and yearly multipliers. Analysis was carried out separately for weekdays and weekends, and the results are shown in Figure B1.

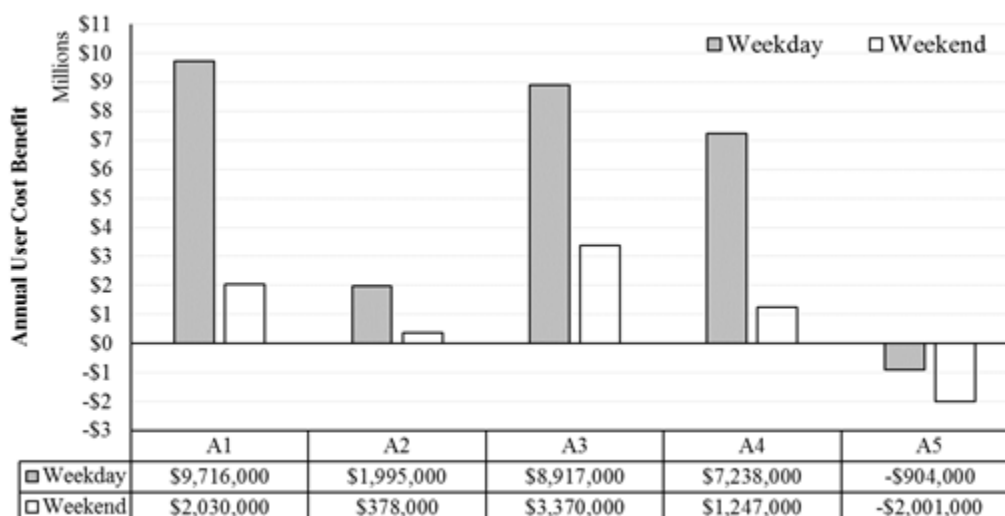


Figure B1. Annual user cost benefits for the five corridors

In addition to the user costs, changes in carbon dioxide (CO₂) emissions were computed using the method adopted by [165]. Using conversion factors from the Argonne National Laboratory, a passenger car is expected to consume 0.87 gal of gasoline per hour. This number was conservatively used to determine the fuel consumption, as given by equation (19)

$$fuel = \Delta TT * vol * \frac{0.87 \text{ gal}}{\text{hour}} \quad (19)$$

Equation (20) computes the CO₂ emissions in tons. According to the U.S. Environmental Protection Agency, the amount of CO₂ emitted when a gallon of gasoline burns is approximately 19.6 lb/gal [166].

$$CO_{2\text{ emissions}} = fuel * \frac{19.6\text{ lb}}{\text{gal}} * \frac{1\text{ ton}}{2,000\text{ lb}} \quad (20)$$

The USEPA also estimates the social cost of CO₂ as \$36/ton [167] and the cost of CO₂ was determined using equation (21).

$$CC = CO_{2\text{ emissions}} * \frac{\$36}{\text{ton}} \quad (21)$$

The weekday and weekend savings as well as emissions for CO₂ is given in Table B1.

Table B1. Emission reductions

Corridor	Weekday CO ₂ Savings		Weekend CO ₂ Savings	
	Tons	Dollars	Tons	Dollars
A1	3120	\$112,000	650	\$23,000
A2	640	\$23,000	120	\$4,000
A3	2890	\$104,000	1080	\$39,000
A4	2320	\$84,000	400	\$14,000
A5	-310	-\$11,000	-650	-\$23,000
Total	8660	\$213,000	1610	\$58,000

As shown, all corridors saw improvements in annual user benefits and CO₂ savings, except corridor A5. Altogether, these corridors accounted for an \$32.0 million annual user benefit and a \$369,000 CO₂ yearly savings. The greatest improvements are reflected by weekend totals of \$5 million and \$58,000, compared to weekday totals of \$27 million and \$312,000. Of these five corridors, A1 and A3 had the highest user benefit savings and CO₂ savings, totaling \$24 million and \$277,000. As mentioned previously, corridor A2 had reliable travel times before the adaptive installation, and, therefore, had the least amount of travel time impact and cost savings. Corridor A4 had a similar weekend impacts as A1 and A3, but was not as effective for weekdays. Negative user benefit and CO₂ savings from corridor A5 can be related to the increased travel times.

APPENDIX C. SAMPLE CURATED DATASET

trip_ID	start_time	end_time	outlier
0xEC78D084865FCE6567EAD39EC88F895C	5/1/2016 4:51:50	5/1/2016 4:57:37	No
0x95EA2041CABE75BCFEE47227FAABB8E8	5/1/2016 5:12:26	5/1/2016 5:19:39	No
0x8AF930EA70326385A86A0B869A567E7B	5/1/2016 5:13:32	5/1/2016 5:22:07	No
0x75B6D957A9B130DE80D20D4624DC854C	5/1/2016 5:26:32	5/1/2016 5:33:03	No
0xECF829D6644EE95DBC320FD5B221161A	5/1/2016 6:03:39	5/1/2016 6:10:38	No
0x5E295B6F036ACE6FAEDA818199BBEA20	5/1/2016 6:08:16	5/1/2016 6:15:56	No
0x5E33F7EB128351BA8C1165A0A30BAD84	5/1/2016 6:21:22	5/1/2016 6:28:12	No
0x782DDB99824ECF1B409C07A4C22BC4FF	5/1/2016 6:45:03	5/1/2016 6:52:30	No
0xFE6A100C92CA7ECC9FB397DC1C6AD9BA	5/1/2016 7:00:30	5/1/2016 7:07:20	No
0x738B6EDF82640B5A9BAFB6C7298E2F04	5/1/2016 7:14:10	5/1/2016 7:31:00	Yes
0x21970A5BD18DDFD4C2C93BA4847382AE	5/1/2016 7:35:43	5/1/2016 7:43:10	No
0x31AC35B54D2F3871505D1F6183F5E943	5/1/2016 7:44:40	5/1/2016 7:52:10	No
0xA45A79DDF49AF28D659D95CE8C5ED4B5	5/1/2016 8:14:47	5/1/2016 8:21:17	No
0x101D43A56E2AD5BD146E1CA5269D2DF5	5/1/2016 8:17:53	5/1/2016 8:25:08	No
0x92E83832550CED4C0358A9B5154DC166	5/1/2016 8:35:02	5/1/2016 8:41:37	No
0xEA7DDAD2ED8D2E055002BE591306F635	5/1/2016 10:02:54	5/1/2016 10:09:19	No
0x5EFA87829390B6E612CAED5137EB64B2	5/1/2016 10:39:18	5/1/2016 10:46:44	No
0x0AA5902AE845DEA424A1A38EC2E89974	5/1/2016 10:43:12	5/1/2016 10:58:59	Yes
0x2545EC205F8EEEECE4581A3AE96D6C9C0	5/1/2016 11:19:54	5/1/2016 11:25:47	No
0x8322EECD1E52A51B633FA21AF6E87697	5/1/2016 11:20:37	5/1/2016 11:26:47	No
0x40D6463E3D3F659EC2AFB68D118ABFA9	5/1/2016 11:21:43	5/1/2016 11:27:48	No
0xEFAD3C8269C3D79E403F5C18311A1F34	5/1/2016 11:28:20	5/1/2016 11:35:20	No
0xB3376BBA9BB3796E06C82B72FECEC831	5/1/2016 11:29:31	5/1/2016 11:36:48	No
0x8B4C1446F4C1583C75A836714C596F86	5/1/2016 11:41:56	5/1/2016 11:49:38	No
0x15128E8C7E8E014A8AFC7AB364E17C02	5/1/2016 11:45:57	5/1/2016 11:52:31	No
0x4D2A632E7728AEC25DA5D1500FA26F9C	5/1/2016 11:52:15	5/1/2016 11:58:55	No
0xD8D3FC583116006DF16097C28501DCAF	5/1/2016 11:55:29	5/1/2016 12:01:07	No
0xA0365CA0BE9F7E36435805E0406BEF9D	5/1/2016 12:02:36	5/1/2016 12:09:12	No
0xF1103C13FCC449BB459110D1EB1C2E1D	5/1/2016 12:04:21	5/1/2016 12:12:03	No
0x4EB9F0720A32E404D09CB9CFC9E14C52	5/1/2016 12:08:18	5/1/2016 12:14:03	No
0xAB36FF96F8DDFC650F8079CB85A7E7C8	5/1/2016 12:17:27	5/1/2016 12:23:18	No
0xF983D224AD9B11DDFBF0F71EB6749265	5/1/2016 12:21:17	5/1/2016 12:29:30	No
0xD9C033D26E4038429F0143202FCAEBAD	5/1/2016 12:37:47	5/1/2016 12:45:07	No

GLOSSARY OF TERMS

Term	Description
3-axis accelerometers	Sensors that measure the acceleration on the X, Y and Z axes
Annual Average Daily Traffic (AADT)	The average 24-hour volume at a given location over a year
AASHTO	American Association of State Highway and Transportation Officials
Actuated Coordinated Operation	In this operation, a set of signal systems coordinate and operate under the same cycle length by controlling offsets between adjacent intersection greens to keep the vehicles moving
Air Traffic Control (ATC)	Service provided by the ground-based air traffic controllers to direct the aircraft through ground and controlled air-space
AMASS	Airport Movement Area Safety System
AOG	Aircraft on Ground
Application Programming Interface (API)	A set of definitions, protocols and tools for building computer programming application software
Arterial Congestion Ticker	A performance dashboard which produces a chart of speed distributions on selected arterial routes over time
Arterial Ranking Tool	A performance dashboard which enables the user to view performance of several corridors and rank the corridors according to their normalized median and IQR travel times
Arterial Travel Time Comparison tool	A performance dashboard that compares the travel times on an arterial during “before” and “after” conditions, using filtering by date, day of week and time of day.
ASDE-X	Airport Surface Detection Equipment Model X
Bikes on Ground	Cumulative number of days for which a bike was unused
Bike sharing	An environmental friendly form of public transportation that provides the users with the flexibility of accessing public bikes at unattended stations.

Term	Description
Clearance time	Sum of yellow and all-red interval for a phase
Connected Vehicles	A program announced by the USDOT that allows communication between vehicles, infrastructure and others to improve safety, mobility and reduce emissions
Cumulative Frequency Diagram (CFD)	Plots showing the cumulative frequency of a dataset
Curated Dataset	A travel time dataset where the outlier status of every data point is known
CVG	FAA airport code for Cincinnati/Northern Kentucky airport
Cycle	A signal cycle is one complete rotation through all the indications provided
Cycle Length	The time in seconds it take to complete one full cycle of indication
dB	Decibel
Dedicated Short Range Communication (DSRC)	Two-way short- to- medium-range wireless communications capable of very high data transmission which is critical in communications-based active safety applications
DOT	Department of Transportation
Early Prediction	When the start of green time predicted by the connected vehicle system was before the actual start of green in the field
EB	Eastbound
EFB	Electronic Flight Bag
Federal Aviation Administration (FAA)	National authority in the United States with powers to regulate all aspects of civil aviation.
FAR	Federal Aviation Regulation
Fixed or Pretimed Operation	In this operation the cycle length, phase sequence and timing of each interval are constant

Term	Description
Foreign Object Debris (FOD)	Any object, live or not, located in an inappropriate location in the airport environment that has the capacity to injure airport or air carrier personnel and damage aircraft
Fully Actuated Operation	In this operation, the cycle length, phase sequence and timing of each interval vary depending on programmed rules established on the signal controller
General Aviation (GA)	One of the six airport categories listed by the National Plan of Integrated Airport Systems which includes all civil aviation operations other than scheduled air services and non-scheduled air transport operations for remuneration or hire
GPS	Global Positioning System
HCM	Highway Capacity Manual
IEC	International Electrotechnical Commission
INDOT	Indiana Department of Transportation
Internet of Things (IoT)	A network of smart sensors and computing devices connected via Internet that are capable of receiving and sending data
Interquartile Range (IQR)	Difference between the 75 th and 25 th percentile
Interval	Period of time during which no signal indication changes
ITS	Intelligent Transportation Systems
KLAF	FAA airport code for Purdue University airport
Late Prediction	When the start of green time predicted by the connected vehicle system was after the actual start of green in the field
LOS	Level of Service
MAD	Median Absolute Deviation
Managed Inclusions	A program which allowed passengers that were not on known traveler lists or given TSA PreCheck based on automated risks assessments to utilize expedited screening under certain circumstances

Term	Description
MAP-21	Moving Ahead for Progress in the 21 st century
Mod Z	Modified Z-score
ms	Millisecond
National Association of City Transportation Officials (NACTO)	A coalition of DOTs in North America committed to raising the state of practice for street design and transportation
National Cooperative Highway Research Program (NCHRP)	A forum for coordinated and collaborative research, the NCHRP addresses issues integral to the state Departments of Transportation (DOTs) and transportation professionals at all levels of government and the private sector.
NDSU	North Dakota State University
Operational Incident (OI)	Type of runway incursion that can occur due to the action(s) of an air traffic controller that results in less than required minimum separation between two or more aircraft, or between an aircraft and obstacles, or providing clearance to incorrect/unsafe operations
Outlier (in Chapter 6)	All trips that left the study route for trip chaining purposes or due to route diversions
Pilot Deviation (PD)	Type of runway incursion that can occur when the action(s) of a pilot violates any federal aviation regulation
PennDOT	Pennsylvania Department of Transportation
Performance Measures	Quantitative metrics to monitor the progress of a system towards defined organizational tasks and objectives
Phase/Split	Amount of cycle time assigned to a phase by a controller. This consists of the green interval and the yellow and all-red intervals that follow it
Protected Left Turn Movement	A left turn movement made without an opposing vehicular flow
Retroreflectivity	Optical phenomenon in which reflected rays of light are preferentially returned in directions opposite to the source
RMS	Root Mean Squared

Term	Description
Roadway Rumble Strips	Safety countermeasures that use tactile vibration and audible rumbling to alert inattentive drivers of potential danger.
RPM	Raised Pavement Marker
RSU	Road Side Unit
Rumble Stripes	Rumble strips painted with a retroreflective coating to increase the visibility of the pavement edges and centerline, at night and during adverse weather conditions
Runway incursion	Any occurrence at an aerodrome involving the incorrect presence of an aircraft, vehicle or person on the protected area of a surface designated for the landing and takeoff of aircraft
Runway Incursion Mitigation (RIM)	A safety program initiated by FAA to identify airport risk factors that might contribute to a runway incursion and develop strategies to help airport stakeholders mitigate those risks
RWSL	Runway Status Light Program
SHRP 2	Strategic Highway Research Program 2
Signal Phase and Timing (SPaT)	Current intersection signal light phase
Sinusoidal Rumble Strips	Roadway rumble strips designed and constructed in the form of a sine wave
SPF	Safety Performance Function
SPL	Sound Power Level
Sound Level Meter	Sensors that record sound in decibels (dB)
Traffic Signal Controller	A signal controller determines the various signal indications
Transportation Security Administration (TSA)	Agency of the U.S. Department of Homeland Security that has authority over the security of the traveling public in the United States.
TSA PreCheck	An expedited screening that uses a risk-based strategy, allowing low-risk passengers to undergo expedited security screening with faster moving lines and improved traveler experience

Term	Description
TT	Travel Time
UDOT	Utah Department of Transportation
USDOT	United States Department of Transportation
V/C	Volume-to-capacity ratio
V2I	Vehicle-to-infrastructure communication
V2V	Vehicle-to-vehicle communication
V2X	Vehicle-to-everything communication
Vehicle/Pedestrian Deviation	Type of runway incursion that can occur when pedestrians or vehicles enter any portion of the airport movement areas (runways/taxiways) without authorization from air traffic controller
Vibration	Movement or mechanical oscillation about an equilibrium position of a machine or component
WB	Westbound
WOR	Waypoints Outside Roadway

REFERENCES

- [1] S. Pickrell, "Transportation Performance Measures for Outcome Based System Management and Monitoring," Oregon Department of Transportation, Salem, OR, FHWA-OR-RD-15-03, 2014.
- [2] Florida DOT, "2013 Performance Report," 2013. [Online]. Available: <http://www.dot.state.fl.us/planning/performance/2013/2013PerformanceReport.pdf>. [Accessed: 12-Jun-2016].
- [3] HRB, "Highway Capacity Manual, second ed. Special Report 87," National Research Council, Washington D.C., 1965.
- [4] JHK & Associates, "Performance Measures and Levels of Service in the Year 2000 Highway Capacity Manual. NCHRP Project 3-55(4)," National Research Council, Washington D.C., 1996.
- [5] T. J. Lomax *et al.*, "Quantifying Congestion. Volume 1: Final Report," TRB, Washington D.C., 1997.
- [6] AASHTO, "Highway Safety Manual," American Association of State Highway & Transportation Officials, Washington D.C., 2010.
- [7] US Department of Transportation, "Combined Performance Plan and Report," 2002. [Online]. Available: <https://www.ncua.gov/Legal/Documents/Combined2001PerformanceReport2003InitialPlan.pdf>. [Accessed: 13-Jun-2016].
- [8] C. M. Day *et al.*, "Performance Measures for Traffic Signal Systems: An Outcome-Oriented Approach," Purdue University, West Lafayette, IN, Apr. 2014.
- [9] M. Jin and H. Wang, "System Performance Measures For Intermodal Transportation with A case study and industrial application," Mississippi State University, Mississippi State, MS, 2004.
- [10] R. H. Pratt and T. J. Lomax, "Performance Measures for Multimodal Transportation Systems," *Transportation Research Record*, no. 1518, pp. 85–93, 1996.

- [11] Federal Highway Administration, “National Performance Management Measures; Assessing Performance of the National Highway System, Freight Movement on the Interstate System, and Congestion Mitigation and Air Quality Improvement Program,” Federal Highway Administration, Washington, D.C., Rule 82 FR 5970, 2017.
- [12] US Department of Transportation, “Connected Vehicles,” 2016. [Online]. Available: <https://www.pcb.its.dot.gov/eprimer/module13.aspx>. [Accessed: 13-Mar-2018].
- [13] US Department of Transportation, “Connected Vehicle Basics,” *Intelligent Transportation Systems Joint Program Office*, 2018. [Online]. Available: https://www.its.dot.gov/cv_basics/cv_basics_what.htm. [Accessed: 13-Mar-2018].
- [14] US Department of Transportation, “Smart City Challenge,” 2016. [Online]. Available: <https://www.transportation.gov/smartcity>. [Accessed: 17-Mar-2018].
- [15] A. Balmos, J. Mathew, D. Plattner, J. Krogmeier, and D. Bullock, “Evaluation of Sinusoidal Rumble Strip Noise Levels,” in *Transportation Research Board 2017 Annual Meeting Compendium of Papers*, 2017.
- [16] D. M. Bullock, S. M. L. Hubbard, J. K. Mathew, W. L. Major, C. Furr, and D. W. Gallagher, “Evaluation of Aviation Rumble Strips,” Washington D.C., DOT/FAA/TC-14, in press, 2016.
- [17] R. Srinivasan, J. Baek, and F. Council, “Safety evaluation of transverse rumble strips on approaches to stop-controlled intersections in rural areas,” *Journal of Transportation Safety and Security*, vol. 2, no. 3, pp. 261–278, 2010.
- [18] Federal Highway Administration, “Center Line Rumble Strips,” Washington D.C., Technical Advisory T 5040.40 Revision 1, 2011.
- [19] Federal Highway Administration, “Shoulder and Edge Line Rumble Strips,” Washington D.C., Technical Advisory T 5040.39, Revision 1, 2011.
- [20] Federal Highway Administration, “Rumble Strip Types,” 2011. [Online]. Available: https://safety.fhwa.dot.gov/roadway_dept/pavement/rumble_strips/rumble_types/. [Accessed: 24-Jan-2018].

- [21] Federal Highway Administration, "Rumble Strips and Rumble Stripes: Design and Construction," 2015. [Online]. Available: https://safety.fhwa.dot.gov/roadway_dept/pavement/rumble_strips/design-and-construction.cfm. [Accessed: 25-Jan-2018].
- [22] T. Bucko, "Evaluation of Milled-In Rumble Strips, Rolled-In Rumble Strips and Audible Edge Stripes," California DOT, Sacramento, CA, May 2001.
- [23] J. Kragh and B. Andersen, "Traffic noise at rumble strips on roads: A pilot study," *Transport Research Arena Europe*, 2008.
- [24] E. Terhaar and D. Braslau, "Rumble Strip Noise Evaluation," Minnesota Department of Transportation, St. Paul, MN, MN/RC 2015-07, Feb. 2015.
- [25] Federal Highway Administration, "Rumble Strips and Rumble Stripes: General Information," 2015. [Online]. Available: https://safety.fhwa.dot.gov/roadway_dept/pavement/rumble_strips/general-information.cfm. [Accessed: 25-Jan-2018].
- [26] S. R. Mitkey *et al.*, "Comparison of the Retroreflective Durability Between Rumble Stripes and Painted Lines," *ITE Journal of Transportation*, vol. 3, no. 1, pp. 33–50, 2012.
- [27] R. B. Gibbons and B. M. Williams, "Assessment of the Durability of Wet Night Visible Pavement Markings: Wet Visibility Project Phase IV," Virginia Department of Transportation, Blacksburg, VA, VCTIR 12-R13, Jun. 2012.
- [28] N. Hawkins, O. Smadi, S. Kinckerbocker, and P. Carlson, "Rumble Stripe: Evaluation of Retroreflectivity and Installation Practices," Minnesota Department of Transportation, St. Paul, MN, MN/RC 2016-13, Apr. 2016.
- [29] T. M. Brennan, S. R. Mitkey, and D. M. Bullock, "Alternatives to Raised Pavement Markers (RPMs)," Joint Transportation Research Program, West Lafayette, IN, Purdue University, FHWA/IN/JTRP-2014/01, Feb. 2014.
- [30] D. E. Karkle, "Effects of Centerline Rumble Strips on Safety, Exterior Noise, and Operational Use of the Travel Lane," Kansas State University, 2011.
- [31] M. D. Finley and J. D. Miles, "Exterior Noise Created by Vehicles Traveling over Rumble Strips," *Transportation Research Board 86th Annual Meeting*, 2007.

- [32] DelDOT, “SR 24 Longitudinal Edge Line Rumble Strip Noise Study,” 2012.
- [33] D. J. Torbic *et al.*, “NCHRP Report 641 – Guidance for the Design and Application of Shoulder and Centerline Rumble Strips,” Transportation Research Board, Washington D.C., 2009.
- [34] T. V Sexton, “Evaluation of Current Centerline Rumble Strip Design (s) to Reduce Roadside Noise and Promote Safety,” Washington State Department of Transportation, WA-RD 835.1, Sep. 2014.
- [35] J. S. Lamancusa, “Outdoor Sound Propagation,” 2009. [Online]. Available: http://www.mne.psu.edu/lamancusa/me458/10_osp.pdf. [Accessed: 15-Jul-2016].
- [36] Gulf Coast Data Concepts, “High Sensitivity 2g USB Accelerometer X2-2,” 2016. [Online]. Available: <http://www.gcdataconcepts.com/x2-1.html>. [Accessed: 01-Jun-2016].
- [37] S. W. Smith, “Audio Processing,” in *The Scientist and Engineer’s Guide to Digital Signal Processing*, no. 1, 1999, pp. 631–642.
- [38] A. J. Hudspeth, “How the ear’s works work.,” *Nature*, vol. 341, no. 6241, pp. 397–404, 1989.
- [39] “Electroacoustics Sound Level Meters Part 1: Specifications.” IEC 61672-1.
- [40] DELTA, “LTL-M mobile retroreflectometer for road markings,” 2018. [Online]. Available: <https://roadsensors.madebydelta.com/products/ltl-m-mobile-retroreflectometer-road-markings/>. [Accessed: 25-Jan-2018].
- [41] INDOT, “Standard Specifications,” vol. 808.07, Indianapolis, IN, 2018, pp. 833–834.
- [42] Federal Aviation Administration, “Focus on Hot Spots, Prevent Runway Incursions,” 2014. [Online]. Available: [http://www.faa.gov/airports/runway_safety/publications/media/Focus on Hot Spots Brochure.pdf](http://www.faa.gov/airports/runway_safety/publications/media/Focus%20on%20Hot%20Spots%20Brochure.pdf). [Accessed: 05-Jan-2015].
- [43] Federal Aviation Administration, “Runway Safety Report 2011 – 2012,” 2012. [Online]. Available: https://www.faa.gov/airports/runway_safety/publications/media/2012-AJS-475-FY2011-Runway-Safety-Annual-Report.pdf. [Accessed: 05-Jan-2015].

- [44] Federal Aviation Administration, “National Runway Safety Plan 2015-17,” 2015. [Online]. Available: http://www.faa.gov/airports/runway_safety/publications/media/2015_ATO_Safety_National_Runway_Safety_Plan.pdf. [Accessed: 20-Jul-2015].
- [45] Federal Aviation Administration, “Runway Incursions,” 2015. [Online]. Available: http://www.faa.gov/airports/runway_safety/news/runway_incursions/. [Accessed: 25-Mar-2015].
- [46] Federal Aviation Administration, “Runway Incursion Totals for FY 2014,” 2014. [Online]. Available: http://www.faa.gov/airports/runway_safety/statistics/regional/?fy=2014. [Accessed: 05-Jan-2015].
- [47] Federal Aviation Administration, “Runway Incursions Totals for FY 2012,” 2013. [Online]. Available: http://www.faa.gov/airports/runway_safety/statistics/regional/?fy=2012. [Accessed: 02-Jun-2016].
- [48] Federal Aviation Administration, “Runway Incursion Totals by quarter FY2010 vs. FY2009,” 2014. .
- [49] Bureau of Air Safety Investigation, “Human Factors in Fatal Aircraft Accidents,” 1996. [Online]. Available: https://www.atsb.gov.au/media/28363/sir199604_001.pdf. [Accessed: 06-Jan-2015].
- [50] G. Li, S. P. Baker, M. W. Lamb, J. G. Grabowski, and G. W. Rebok, “Human factors in aviation crashes involving older pilots.,” *Aviation, space, and environmental medicine*, vol. 73, no. 2, pp. 134–138, Feb. 2002.
- [51] Federal Aviation Administration, “NextGen Implementation Plan,” 2010.
- [52] D. Olson, B. Manchas, R. Glad, and M. Sujka, “Performance Analysis of Centerline Rumble Strips in Washington State,” 2011.
- [53] Federal Aviation Administration, “FAA Implements New Airport Safety Program,” 2015. [Online]. Available: <https://www.faa.gov/news/updates/?newsId=83046>. [Accessed: 22-Jul-2015].

- [54] J. K. Mathew, W. L. Major, S. M. Hubbard, and D. M. Bullock, "Statistical modelling of runway incursion occurrences in the United States," *Journal of Air Transport Management*, vol. 65, pp. 54–62, 2017.
- [55] Federal Aviation Administration, "FAA Foreign Object Debris Program," 2015. [Online]. Available: http://www.faa.gov/airports/airport_safety/fod/. [Accessed: 07-Feb-2016].
- [56] Eurocontrol, "European Action Plan for the Prevention of Runway Incursions," 2012.
- [57] J. Mathew, D. Krohn, H. Li, C. Day, and D. Bullock, "Implementation of Probe Data Performance Measures," Harrisburg, PA, PA-2017-001-PU WO 001, Apr. 2017.
- [58] C. A. Quiroga and D. Bullock, "Measuring Control Delay at Signalized Intersections," *Journal of Transportation Engineering*, vol. 125, no. 4, pp. 271–280, Jul. 1999.
- [59] C. Day *et al.*, "Evaluation of Arterial Signal Coordination," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2192, pp. 37–49, Dec. 2010.
- [60] S. Quayle, P. Koonce, D. DePencier, and D. Bullock, "Arterial Performance Measures with Media Access Control Readers," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2192, pp. 185–193, Dec. 2010.
- [61] Y. Wang, B. A. Namaki, Y. Malinovskiy, J. Corey, and T. Cheng, "Error Assessment for Emerging Traffic Data Collection Devices," Pacific Northwest Transportation Consortium, University of Washington, Report WA-RD 810.1, 2014.
- [62] J. Hu, M. D. Fontaine, and J. Ma, "Quality of Private Sector Travel-Time Data on Arterials," *Journal of Transportation Engineering*, vol. 142, no. 4, p. 4016010, Apr. 2016.
- [63] E. Sharifi, S. E. Young, S. Eshragh, M. Hamed, R. M. Juster, and K. Kaushik, "Quality Assessment of Outsourced Probe Data on Signalized Arterials: Nine Case Studies in Mid-Atlantic Region," in *Transportation Research Board Annual Meeting*, 2016.

- [64] C. M. Day *et al.*, “Performance Ranking of Arterial Corridors Using Travel Time and Travel Time Reliability Metrics,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2487, pp. 44–54, Aug. 2015.
- [65] R. Pearson, “Outliers in process modeling and identification,” *Control Systems Technology, IEEE Transactions*, vol. 10, no. 1, pp. 55–63, 2002.
- [66] G. List *et al.*, “Establishing Monitoring Programs for Travel Time Reliability,” Transportation Research Board, Washington, D.C., Report S2-L02-RR-1, Sep. 2014.
- [67] Federal Highway Administration, “Adaptive Signal Control Technology,” 2017. [Online]. Available: <https://www.fhwa.dot.gov/innovation/everydaycounts/edc-1/asct.cfm>. [Accessed: 26-Feb-2018].
- [68] J. K. Mathew, S. M. Hubbard, and D. M. Bullock, “Management Oriented Performance Measures for Airport Security Checkpoints,” in *12th ITS European Congress, Strasbourg, France, 2017*.
- [69] R. W. Poole and J. J. Carafano, “Time to Rethink Airport Security,” 2006. [Online]. Available: <http://www.heritage.org/research/reports/2006/07/time-to-rethink-airport-security>. [Accessed: 16-Jan-2017].
- [70] D. Bullock, R. Haseman, J. Wasson, and R. Spitler, “Automated Measurement of Wait Times at Airport Security,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2177, pp. 60–68, Dec. 2010.
- [71] US Customs and Border Protection, “Key Regulatory Changes in 2008,” 2008. [Online]. Available: https://www.cbp.gov/sites/default/files/documents/reg_change.pdf. [Accessed: 16-Jan-2017].
- [72] U.S. Government Accountability Office, “Aviation Security: TSA’s Managed Inclusion Process Expands Passenger Expedited Screening, But TSA Has Not Tested Its Security Effectiveness,” no. GAO-15-465T. 2015.
- [73] J. Fishel, P. Thomas, M. Levine, and J. Date, “Undercover DHS Tests Find Security Failures at US Airports,” 01-Jun-2015.
- [74] K. Philipovitch, “Testimony of Kerry Philipovitch,” *House Homeland Security Committee’s Subcommittee on Transportation Security*, 26-May-2016.

- [75] M. Schlangenstein, "Major airlines spend millions in bid to cut TSA lines in record summer," *Chicago Tribune*, 2016. [Online]. Available: <http://www.chicagotribune.com/business/ct-airlines-tsa-lines-20160525-story.html>.
- [76] B. Mutzabaugh, "American Airlines paying \$4 million to hire help for TSA," *USA Today*, 2016. [Online]. Available: <https://www.usatoday.com/story/travel/flights/todayinthesky/2016/05/19/american-airlines-paying-4-million-hire-help-tsa/84583016/>. [Accessed: 17-Jan-2017].
- [77] Homeland Security, "Statement By Secretary Johnson On Additional Congressional Support For Airport Security | Homeland Security," 2016. [Online]. Available: <https://www.dhs.gov/news/2016/06/13/statement-secretary-johnson-additional-congressional-support-airport-security>. [Accessed: 17-Jan-2017].
- [78] C. Unger, "Congress OKs Funding to Ease TSA Wait Times - SmarterTravel," 2016. [Online]. Available: <https://www.smartertravel.com/2016/06/23/congress-oks-funding-tsa-screeners/>. [Accessed: 17-Jan-2017].
- [79] S. M. Remias, A. M. Hainen, and D. M. Bullock, "Leveraging Probe Data to Assess Security Checkpoint Wait Times," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2325, no. 1, pp. 63–75, Dec. 2013.
- [80] J. K. Mathew, S. M. Hubbard, and D. M. Bullock, "Impact of Security Reconfiguration and Expedited Security at CVG on Reducing Passenger Wait Time," *Transportation Research Board 95th Annual Meeting*. 2016.
- [81] Transportation Security Administration, "TSA Pre[√]™ Begins at Cincinnati/Northern Kentucky International Airport," 2012. [Online]. Available: <http://www.tsa.gov/press/releases/2012/09/05/tsa-pre-√-begins-cincinnati-northern-kentucky-international-airport>. [Accessed: 10-Feb-2015].
- [82] Bureau of Transportation Statistics, "Cincinnati, OH: Cincinnati/Northern Kentucky International (CVG)," *US Department of Transportation*, 2018. [Online]. Available: https://www.transtats.bts.gov/airports.asp?pn=1&Airport=CVG&Airport_Name=Cincinnati,OH:Cincinnati/NorthernKentuckyInternational&carrier=FACTS. [Accessed: 19-Feb-2018].

- [83] S. Washington, M. Karlaftis, and F. Mannering, *Statistical and econometric methods for transportation data analysis*, Second. Boca Raton: Chapman & Hall/CRC, 2011.
- [84] C. Unger, “Long TSA Lines Are Gone, But for Good? - SmarterTravel,” 2016. [Online]. Available: <https://www.smartertravel.com/2016/09/22/long-tsa-lines-are-gone/>. [Accessed: 17-Jan-2017].
- [85] Cincinnati/Northern KY Airport, “CVG Airport,” 2018. [Online]. Available: <http://www.cvgairport.com/>.
- [86] Transportation Security Administration, “MyTSA App,” 2018. [Online]. Available: <https://www.tsa.gov/mobile>.
- [87] USDOT Bureau of Transportation Statistics, “Bike share Stations in the United States,” 2016. [Online]. Available: https://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/bts_technical_report/april_2016. [Accessed: 08-Mar-2018].
- [88] D. Malouff, “All 119 US bikeshare systems, ranked by size,” *Greater Greater Washington*, 2017. [Online]. Available: <https://gwwash.org/view/62137/all-119-us-bikeshare-systems-ranked-by-size>. [Accessed: 08-Mar-2018].
- [89] National Association of City Transportation Officials, “Bike Share in the US: 2010-2016,” 2017. [Online]. Available: <https://nacto.org/bike-share-statistics-2016/>.
- [90] S. Shaheen, E. Martin, and A. Cohen, “Public Bikesharing and Modal Shift Behavior: A Comparative Study of Early Bikesharing Systems in North America,” *International Journal of Transportation*, vol. 1, no. 1, pp. 35–54, Dec. 2013.
- [91] Zagster, “How to build better communities: Start with bike sharing,” 2017. [Online]. Available: <https://www.zagster.com/blog/how-to-build-better-communities-start-with-bike-sharing>. [Accessed: 09-Mar-2018].
- [92] S. L. Handy, Y. Xing, and T. J. Buehler, “Factors associated with bicycle ownership and use: a study of six small U.S. cities,” *Transportation*, vol. 37, no. 6, pp. 967–985, Nov. 2010.
- [93] R. Godavarthy, J. Mattson, and A. R. Taleqani, “Evaluation Study of the Bike Share Program in Fargo, North Dakota,” SURLC 17-005, 2017.

- [94] T. L. Hamilton and C. J. Wichman, "Bicycle Infrastructure and Traffic Congestion: Evidence from DC's Capital Bikeshare." 2015.
- [95] E. Fishman, S. Washington, and N. Haworth, "Bike share's impact on car use: Evidence from the United States, Great Britain, and Australia," *TRANSPORTATION RESEARCH PART D*, vol. 31, pp. 13–20, 2014.
- [96] M. L. Grabow *et al.*, "Air quality and exercise-related health benefits from reduced car travel in the midwestern United States.," *Environmental health perspectives*, vol. 120, no. 1, pp. 68–76, Jan. 2012.
- [97] J. Woodcock, M. Tainio, J. Cheshire, O. O'Brien, and A. Goodman, "Health effects of the London bicycle sharing system: health impact modelling study.," *BMJ (Clinical research ed.)*, vol. 348, p. g425, Feb. 2014.
- [98] D. Rojas-Rueda, A. de Nazelle, M. Tainio, and M. J. Nieuwenhuijsen, "The health risks and benefits of cycling in urban environments compared with car use: health impact assessment study.," *BMJ (Clinical research ed.)*, vol. 343, p. d4521, Aug. 2011.
- [99] S. Shaheen, E. Martin, A. Cohen, and R. Finson, "Public Bikesharing in North America: Early Operator and User Understanding," San Jose, CA, CA-MTI-12-1029, 2012.
- [100] Weather Underground, "Weather History for KLAF," 2018. [Online]. Available: https://www.wunderground.com/history/airport/KLAF/2016/1/1/CustomHistory.html?dayend=31&monthend=5&yearend=2016&req_city=&req_state=&req_statename=&reqdb.zip=&reqdb.magic=&reqdb.wmo=. [Accessed: 09-Mar-2018].
- [101] W. El-Assi, M. Salah Mahmoud, and K. Nurul Habib, "Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto," *Transportation*, vol. 44, no. 3, pp. 589–613, May 2017.
- [102] K. Gebhart and R. B. Noland, "The impact of weather conditions on bikeshare trips in Washington, DC," *Transportation*, vol. 41, no. 6, pp. 1205–1225, Nov. 2014.

- [103] J. K. Mathew, C. Day, H. Li, and D. Bullock, "Curating Automatic Vehicle Location Data to Compare the Performance of Outlier Filtering Methods," *97th Annual Meeting Transportation Research Board*. Washington D.C., 2018.
- [104] J. Kennedy, C. R. Cantrell, M. D. Varney, Z. Czyzewski, and B. D. V. Smith, "Highway travel time analysis using license plate image capture techniques," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 2004, vol. 5272, pp. 294–303.
- [105] B. Coifman, "Estimating travel times and vehicle trajectories on freeways using dual loop detectors," *Transportation Research Part A: Policy and Practice*, vol. 36, no. 4, pp. 351–364, May 2002.
- [106] M. Hunter, S. Wu, and H. Kim, "Practical Procedure to Collect Arterial Travel Time Data Using GPS-Instrumented Test Vehicles," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1978, no. 1, pp. 160–168, Jan. 2006.
- [107] H. D. Robertson and J. E. Hummer, *Manual of transportation engineering studies*. Prentice Hall, 1994.
- [108] C. A. Quiroga and D. Bullock, "Travel time studies with global positioning and geographic information systems: an integrated methodology," *Transportation Research Part C: Emerging Technologies*, vol. 6, no. 1, pp. 101–127, 1998.
- [109] L. Vanajakshi, S. C. Subramanian, and R. Sivanandan, "Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses," *IET Intelligent Transport Systems*, vol. 3, no. 1, pp. 1–9, 2009.
- [110] P. Murphy, E. Welsh, and J. P. Frantz, "Using Bluetooth for short-term ad hoc connections between moving vehicles: a feasibility study," *Vehicular Technology Conference. IEEE 55th Vehicular Technology Conference. VTC Spring 2002 (Cat. No.02CH37367)*, vol. 1, pp. 414–418, 2002.
- [111] A. Haghani, M. Hamed, K. Sadabadi, S. Young, and P. Tarnoff, "Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2160, pp. 60–68, Dec. 2010.

- [112] A. Janecek, K. A. Hummel, D. Valerio, F. Ricciato, and H. Hlavacs, "Cellular data meet vehicular traffic theory: Location area updates and cell transitions for travel time estimation," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*, 2012, p. 361.
- [113] Y. Liu, F. Dion, and S. Biswas, "Dedicated Short-Range Wireless Communications for Intelligent Transportation System Applications: State of the Art," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1910, pp. 29–37, Jan. 2005.
- [114] A. Bhaskar, L. M. Kieu, M. Qu, A. Nantes, M. Miska, and E. Chung, "On the use of Bluetooth MAC Scanners for live reporting of the transport network," *10th International Conference of Eastern Asia Society for Transportation Studies*. Taipei, Taiwan, 2013.
- [115] L. Minh, L. Minh Kieu, A. Bhaskar, and E. Chung, "Bus and Car Travel Time on Urban Networks: Integrating Bluetooth and Bus Vehicle Identification Data," *25th ARRB Conference : Shaping the future: Linking Policy, Research and Outcomes*. Perth, WA, 2012.
- [116] J. Jang, "Outlier filtering algorithm for travel time estimation using dedicated short-range communications probes on rural highways," *IET Intelligent Transport Systems*, vol. 10, no. 6, pp. 453–460, Aug. 2016.
- [117] J. K. Mathew, V. L. Devi, D. M. Bullock, and A. Sharma, "Investigation of the Use of Bluetooth Sensors for Travel Time Studies under Indian Conditions," *Transportation Research Procedia*, vol. 17, pp. 213–222, 2016.
- [118] S. Clark, S. Grant-Muller, and H. Chen, "Cleaning of Matched License Plate Data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1804, pp. 1–7, Jan. 2002.
- [119] J. M. Roth, "A Time Series Approach to Removing Outlying Data Points from Bluetooth Vehicle Speed Data," University of Akron, 2010.
- [120] X. J. Ban, Y. Li, A. Skabardonis, and J. D. Margulici, "Performance Evaluation of Travel-Time Estimation Methods for Real-Time Traffic Applications," *Journal of Intelligent Transportation Systems*, vol. 14, no. 2, pp. 54–67, Apr. 2010.

- [121] National Academies of Sciences Engineering Medicine, “Guide to Establishing Monitoring Programs for Travel Time Reliability,” Nov. 2014.
- [122] S. Moghaddam and B. Hellenga, “Evaluating the Performance of Algorithms for the Detection of Travel Time Outliers,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2338, pp. 67–77, Dec. 2013.
- [123] J. Mathew, C. Day, H. Li, and D. Bullock, “Automatic Vehicle Location Data Set,” *Purdue University Research Repository*, 2017. [Online]. Available: <http://dx.doi.org/10.4231/R7XS5SJJ>. [Accessed: 27-Jul-2017].
- [124] R. K. Pearson, “Exploring process data,” *Journal of Process Control*, vol. 11, no. 2, pp. 179–194, Apr. 2001.
- [125] P. J. Huber and E. Ronchetti, *Robust statistics*. Wiley, 2009.
- [126] L. Davies and U. Gather, “The Identification of Multiple Outliers,” *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 782–792, 1993.
- [127] B. Iglewicz and D. C. Hoaglin, *How to detect and handle outliers*. Milwaukee, Wisconsin: ASQC Quality Press, 1993.
- [128] S. Seo, “A review and comparison of methods for detecting outliers in univariate data sets,” University of Pittsburgh, 2006.
- [129] J. W. Tukey, *Exploratory data analysis*. Boston, MA: Addison-Wesley Pub. Co, 1977.
- [130] AASHTO, “The National Connected Vehicle SPaT Deployment Challenge,” 2017. [Online]. Available: <http://www.transportationops.org/spatchallenge>. [Accessed: 13-Mar-2018].
- [131] Federal Highway Administration, “EDC-4: Automated Traffic Signal Performance Measures (ATSPMs),” 2017. [Online]. Available: https://www.fhwa.dot.gov/innovation/everydaycounts/edc_4/atspm.cfm. [Accessed: 13-Mar-2018].
- [132] E. Smaglik, A. Sharma, D. Bullock, J. Sturdevant, and G. Duncan, “Event-Based Data Collection for Generating Actuated Controller Performance Measures,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2035, pp. 97–106, Dec. 2007.

- [133] S. M. Lavrenz, W. B. Smith, and D. M. Bullock, "Assessing Longitudinal Arterial Performance and Traffic Signal Retiming Outcomes," *Transportation Research Board 95th Annual Meeting*, vol. 47906, no. 16, pp. 1–19, 2016.
- [134] H. Li, S. Lavrenz, C. Day, A. Stevens, and D. Bullock, "Quantifying Benefits of Signal Timing Maintenance and Optimization using both Travel Time and Travel Time Reliability Measures," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2487, no. January, pp. 55–68, 2015.
- [135] R. Freije *et al.*, "Graphical Performance Measures for Practitioners to Triage Split Failure Trouble Calls," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2439, pp. 27–40, Dec. 2014.
- [136] UDOT, "Automated Traffic Signal Performance Measures - Automated Traffic Signal Performance Metrics," 2017. [Online]. Available: <http://udottraffic.utah.gov/atspm/>. [Accessed: 13-Mar-2018].
- [137] I. Dakic, A. Stevanovic, M. Zlatkovic, and C. Kergaye, "Refinement of Performance Measures based on High-Resolution Signal and Detection Data," *Transportation Research Procedia*, vol. 22, pp. 372–381, Jan. 2017.
- [138] C. M. Day, "Performance-based management of arterial traffic signal systems," Purdue University, 2010.
- [139] A. M. Hainen *et al.*, "Performance Measures for Optimizing Diverging Interchanges and Outcome Assessment with Drone Video," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2487, pp. 31–43, Aug. 2015.
- [140] J. Zheng, "A Performance Visualization and Fine-tuning Tool for Arterial Traffic Signal Systems," University of Minnesota, 2014.
- [141] J. Sturdevant *et al.*, "Indiana Traffic Signal Hi Resolution Data Logger Enumerations," *Indiana Department of Transportation and Purdue University*, Nov-2012. [Online]. Available: <https://docs.lib.purdue.edu/jtrpdata/3>. [Accessed: 11-Mar-2018].
- [142] D. I. Robertson, "'TRANSYT' Method for Area Traffic Control," *Traffic Engineering and Control*, vol. 11, pp. 276–281, 1969.

- [143] R. Roess, E. Prassas, and W. McShane, *Traffic engineering*, Third. Pearson Prentice Hall, 2004.
- [144] US Department of Transportation, “Transportation Accidents by Mode,” 2015. [Online]. Available: http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/html/table_02_03.html. [Accessed: 19-Jul-2015].
- [145] Federal Aviation Administration, “Runway Safety,” 2016. [Online]. Available: http://www.faa.gov/airports/runway_safety/. [Accessed: 20-Jul-2015].
- [146] Federal Aviation Administration, “National Runway Safety Plan 2015 – 2017,” 2015. [Online]. Available: http://www.faa.gov/airports/runway_safety/publications/media/2015_ATO_Safety_National_Runway_Safety_Plan.pdf. [Accessed: 06-Mar-2015].
- [147] J. K. Mathew, W. L. Major, S. M. Hubbard, and D. M. Bullock, “Statistical Modelling of Runway Incursions,” in *Transportation Research Board 95th Annual Meeting*, 2016.
- [148] Federal Aviation Administration, “National Plan of Integrated Airport Systems (NPIAS) 2015-2019,” 2014.
- [149] Federal Aviation Administration, “FAA Runway Safety Office - Runway Incursions (RWS),” 2015. [Online]. Available: <http://www.asias.faa.gov/pls/apex/f?p=100:28:0::NO:28::> [Accessed: 20-Jul-2015].
- [150] Y.-H. Chang and K.-M. Wong, “Human risk factors associated with runway incursions,” *Journal of Air Transport Management*, vol. 24, pp. 25–30, 2012.
- [151] K. Cardosi and A. Yost, “Controller and Pilot Error in Airport Operations: A Review of Previous Research and Analysis of Safety Data,” 2001.
- [152] L. Green, “Development of a Bayesian Belief Network Runway Incursion Model,” *AIAA Aviation Technology, Integration and Operations (ATIO) Conference*, vol. 14, no. June, pp. 1–11, 2014.
- [153] S. Wilke, A. Majumdar, and W. Y. Ochieng, “Modelling runway incursion severity,” *Accident Analysis & Prevention*, vol. 79, pp. 88–99, 2015.

- [154] M. E. Johnson, X. Zhao, B. Faulkner, and J. P. Young, “Statistical Models of Runway Incursions Based on Runway Intersections and Taxiways,” *Journal of Aviation Technology and Engineering*, vol. 5, no. 2, May 2016.
- [155] L. Biernbaum and G. Hagemann, “Runway Incursion Severity Analysis,” Volpe, Cambridge, MA, DOT-VNTSC-FAA-12-13, Sep. 2012.
- [156] J. C. Milton, V. N. Shankar, and F. L. Mannering, “Highway accident severities and the mixed logit model: An exploratory empirical analysis,” *Accident Analysis and Prevention*, vol. 40, no. 1, pp. 260–266, 2008.
- [157] Federal Aviation Administration, “FAA Runway Safety Report,” 2003.
- [158] Federal Aviation Administration, “Runway Safety Report 2008,” 2008.
- [159] Federal Aviation Administration, “Runway Incursion Airport Assessment Report,” Dec-2002. [Online]. Available: http://www.cast-safety.org/pdf/runway_incursion_assessment.pdf. [Accessed: 21-Nov-2016].
- [160] Federal Aviation Administration, “Airport Surface Detection Equipment, Model X (ASDE-X),” 2014. [Online]. Available: https://www.faa.gov/air_traffic/technology/asde-x/. [Accessed: 15-Jul-2015].
- [161] Cincinnati/Northern KY Airport, “Air Traffic Statistics-2014,” 2014. [Online]. Available: <http://www.cvgairport.com/docs/default-source/stats/cvg-stats-14.pdf?sfvrsn=8>. [Accessed: 08-Feb-2015].
- [162] Long Beach Airport, “Air Traffic Statistics,” 2014. [Online]. Available: <http://www.lgb.org/civicax/filebank/blobdload.aspx?BlobID=2914>. [Accessed: 17-Jul-2015].
- [163] Pennsylvania Department of Transportation, “2014 Pennsylvania Traffic Data.” 2014.
- [164] David Schrank, Bill Eisele, Tim Lomax, and Jim Bak, “2015 Urban Mobility Scorecard,” 2015.
- [165] C. Day *et al.*, “Reliability, Flexibility, and Environmental Impact of Alternative Objective Functions for Arterial Offset Optimization,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2259, pp. 8–22, Dec. 2011.

- [166] U.S. Environmental Protection Agency, “Greenhouse Gas Equivalencies Calculator.” [Online]. Available: <https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator>. [Accessed: 18-Jul-2016].
- [167] U.S. Environmental Protection Agency, “Social Cost of Carbon,” 2015. [Online]. Available: <https://www3.epa.gov/climatechange/Downloads/EPAactivities/social-cost-carbon.pdf>. [Accessed: 18-Jul-2017].