Impact of Data Source on Travel Time Reliability Assessment

Morgan State University
The Pennsylvania State University
University of Maryland
University of Virginia
Virginia Polytechnic Institute & State University
West Virginia University

The Pennsylvania State University
The Thomas D. Larson Pennsylvania Transportation Institute
Transportation Research Building ❖ University Park, PA 16802-4710
Phone: 814-865-1891 ❖ Fax: 814-863-3707
www.mautc.psu.edu

Impact of Data Source on Travel Time Reliability Assessment

Final Project report

Grant DTRT12-G-UTC03

Mid-Atlantic Universities Transportation Center

By:

Ali Haghani
Professor

Yanu Zhang
PhD Candidate

Masoud Hamedi
Research Scientist

Department of Civil & Environmental Engineering
University of Maryland
College Park, MD 20742

August 2014

1. Report No. UMD-2013-01	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle: Impact of Data Source on Travel Time Reliability Assessment	5. Report Date August 2014
	6. Performing Organization Code
7. Author(s) Ali Haghani, Yanru Zhang, Masoud Hamedi	8. Performing Organization Report No.
9. Performing Organization Name and Address University of Maryland College Park, MD	10. Work Unit No. (TRAIS)
	11. Contract or Grant No. DTRT12-G-UTC03
12. Sponsoring Agency Name and Address US Department of Transportation Research & Innovative Technology Admin UTC Program, RDT-30 1200 New Jersey Ave., SE Washington, DC 20590	13. Type of Report and Period Covered Final
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Travel time reliability measures are becoming an increasingly important input to the mobility and congestion management studies. In the case of Maryland State Highway Administration, reliability measures are key elements in the agency's Annual Mobility Report that is used for both transportation planning and operations. Recent advancements in vehicle tracking technologies have provided both private sectors and transportation agencies with multiple technologies for travel time data collection. This paper investigates the effect of data source selection on travel time reliability assessment. One year data from two independent sources, probe and Bluetooth, on two major freeway corridors in Maryland are used for travel time reliability analysis. Bluetooth sensors are permanently installed on multiple points of these two corridors and probe data are provided by INRIX. A statistical method is applied to compare reliability measures obtained from both sources. Several travel time reliability performance measures are studied. Results show that some reliability measures are more sensitive to the data source than the others. It is also shown that performance measures for HOV and general purpose lanes must be calculated separately.

17. Key Words Travel time reliability, data source, mobility, probe, Bluetooth	18. Distribution Statement No restrictions. This document is available from the National Technical Information Service, Springfield, VA 22161

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

Acknowledgments and Disclaimer

This research was funded by grants from the Maryland State Highway Administration (SHA) and the Mid-Atlantic Universities Transportation Center (MAUTC). The authors would like to thank INRIX and the Center for Advanced Transportation Technology Laboratory at the University of Maryland for providing the data. The authors are solely responsible for the content and conclusions presented in this paper. They do not constitute an endorsement or recommendation of the Maryland State Highway Administration.

Executive Summary

Travel time reliability, as one of the performance measurements for the Annual Urban Mobility Report, plays an important role in transportation planning and operation. The accuracy of measurements of reliability highly depends on the data collection technologies. This project supports Objective 2.1 (Travel Reliability) of the SHA Business Plan and investigates the effect of data sources on freeway travel time reliability assessment.

Traffic congestion and its associated impacts such as air pollution pose major concerns to the public. The concept of travel time reliability is proposed to measure the level of congestion. Travel time reliability better represents a commuter's experience than a simple average travel time, and it is becoming increasingly important as part of the planning, traffic management, and system maintenance. In recent years, researchers and practitioners have developed several measurements for travel time reliability. The annual Urban Mobility Report[1], prepared by Texas Transportation Institute, includes travel time reliability as one of the congestion measurements. A 2012 Maryland State Highway Mobility Report utilizes the planning time index and the travel time index to measure the travel time reliability of the highway system. The accuracy of reliability measurements highly depends on the collected data. Different technologies that researchers have used to collect traffic data could affect travel time reliability assessments. Thus, there is a need to thoroughly evaluate existing data sources and determine how to leverage them to support travel time reliability computations.

This project investigated the effects of data sources on travel time reliability performance measurements by comparing travel time reliability assessments based on travel time data obtained from permanently deployed Bluetooth sensors with INRX data. The studied area covers two freeway segments in Maryland: one from Interstate 95 (I-95) southbound between Route 100 and I-495, the other on I-270 northbound from I-495 junction to exit 18. Both INRIX and permanently installed Bluetooth sensors provided traffic data for the entire 2012 year. A statistical method was carefully selected to assess the impact of data sources on travel time reliability measurements. Several issues regarding data quality are worth mentioning below: equivalent INRIX path travel time and high-occupancy vehicle (HOV) lane travel time.

INRIX provides traffic information by utilizing a common industry convention known as "TMC location codes" to uniquely define road segments, and the Bluetooth technique provides segment travel time between two consecutive Bluetooth sensors, which comprised multiple individual TMCs. However, simply adding up the INRIX travel time measured at the same instant for each TMC would be problematic because, as traffic conditions change over time and space, a simple summation of instant travel time could not represent the real travel time experience of travelers. Instead, the Bluetooth technique detects a vehicle when it enters and exits the segment, and the real travel time experienced by individual travelers can be directly measured. To be able to compare INRIX speed feeds with Bluetooth data, there is a need to estimate equivalent INRIX path travel time. The research group developed an algorithm to estimate equivalent path travel time based on INRIX data.

An HOV lane is a traffic lane restricted to vehicles with a minimum number of occupants during peak hours or longer. HOV lanes are normally created to encourage higher vehicle occupancy with the goal of reducing traffic congestion. The entire segment of I-270 that has been studied has HOV lanes on it; these lanes operate during the afternoon peak hours from 3:30 p.m. to 6:30 p.m. on weekdays. Travel time between HOV lanes and general-purpose lanes could be different during HOV lane operation hours, so it is critical to treat the HOV and general purpose lanes separately and calculate separate travel time reliability measurements. Although the Bluetooth technique is able to detect individual vehicle travel time for different lanes, it cannot automatically separate vehicle travel time for each lane. Thus, the research group developed a classification technique to separate vehicle travel time on an HOV lane from that on general-purpose lanes based on the nature of the data. On the other hand, INRIX only provides one measure of aggregated traffic information at each time stamp, and it does not specify traffic information for different types of lanes. The estimated travel time reliability measurements on I-270 would be different for INRIX and Bluetooth data.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1. Problem Statement

Traffic congestion and its associated impacts pose major concerns to the public. While many travelers are used to daily traffic congestion and plan for it, the unexpected changes in travel time patterns cause widespread dissatisfaction. Travel time reliability is proposed to measure the extent of these unexpected delays. Although the concept of reliability is relatively new in transportation planning and operations compared to other engineering disciplines, it is becoming increasingly important as part of the planning, traffic management, and system maintenance processes. Among all reliability measures in the transportation system, travel time reliability poses an increasing concern to both travelers and transportation authorities. Travel time reliability better represents a commuter’s experience than a simple measure of average travel time. Empirical studies indicate that travel time reliability plays a vital role in travelers’ route choice decisions under circumstances in which they are uncertain about the outcome of their decisions [2]. Travelers prefer a reliable but on average longer route to a faster route with higher uncertainty [3]. The Federal Highway Administration (FHWA) defines travel time reliability as "the consistency or dependability in travel times, as measured from day-to-day and/or across different times of the day.” Several different criteria have been proposed to measure travel time reliability with no undisputed opinion on a single optimal measure. Most of these performance measures are related to the property of day-to-day travel time distribution. The wider the distribution, the more unreliable the travel time [3].

Travel time reliability has been studied through a significant amount of research sponsored by the Strategic Highway Research Program 2 and FHWA [4]. The applications of the travel time reliability concept range from traffic operation to long-term transportation planning procedures. The annual *Urban Mobility Report* published by the Texas Transportation Institute applies reliability measures to address the nation’s urban congestion problems [1]. In 2012, the Maryland State Highway Administration published its first *State Highway Mobility Report* [5]. Despite the important and broad applications of the travel time reliability concept, there remains a debate on how to select and calculate the appropriate measure for reliability. Some existing literature compares and evaluates different performance measures based on empirical data [6, 7]. The SHRP 2 L03 recommended several reliability measures for the purpose of determining the impacts of reliability mitigation strategies [8]. The SHRP 2 L03 project emphasizes that reliability measures should be derived from the distribution of travel time data. A minimum of six months of traffic data is required to develop travel time reliability measures in areas where winter weather is not a major factor. A full year of data is preferred to minimize all of the influences of random events. In the SHRP 2 L03 project, researchers also suggested specifying the peak hour and peak period for each study section [8].

Since calculation of travel time reliability measures is based on data distribution, it is necessary to first understand the existing data sources that can be used to support travel time computations. The traditional travel time detection method is the test vehicle or “floating car” technique [9], which is both costly and provides limited coverage. Traditional Intelligent Transportation Systems (ITS) detectors, such as loop, radar, and image processors, cannot directly measure travel time, since they instead measure speed and volume at specific points of the roadway. Recent technological advancements in vehicle tracking, Internet-enabled mobile devices, and in-vehicle electronic systems enable access to direct, inexpensive, and accurate travel time data at high resolution. Private-sector companies like INRIX take advantage of these resources to provide real-time travel time information on both arterials and freeways mainly by capturing, consolidating, and filtering GPS tracks of the probe vehicles. Bluetooth traffic detectors have emerged as a cost-effective source for travel time measurement and are being increasingly used due to their relative low cost and ease of installation. The I-95 Corridor Coalition has relied on Bluetooth-generated travel time as ground-truth for validating probe data in their Vehicle Probe Project (VPP) [10]. Electronic toll tag readings can be used to measure travel time where such data are available. As more technologies become available to collect traffic data, one concern remains: whether the selection of data sources affects a travel time reliability assessment. There is a need to thoroughly evaluate existing data sources and determine how to leverage them to support travel time reliability computations.

Travel time reliability, as one of the measurements for the annual *Urban Mobility Report*, plays an important role in transportation planning and operation. Travel time reliability better represents a commuter’s experience than a simple average travel time. The accuracy of reliability measures can be affected by the data source being used to collect the travel time data. Thus, there is a need to thoroughly evaluate the existing data sources and determine how to leverage them to support travel time reliability computations.

1.2. Research Approach

This project supports Objective 2.1 (Travel Reliability) of the SHA Business Plan: achieve an annual user cost savings of at least $1.1 billion as a result of congestion management. The objective of this project was to investigate the effects of different data sources (Bluetooth and INRIX) on freeway travel time reliability assessments. To achieve this objective, a comprehensive analysis was necessary to evaluate the effects of data sources on a travel time reliability assessment. The primary focus of this project was to investigate the effects of data sources on travel time reliability assessments by comparing travel time data obtained from both Bluetooth devices and probes as two independent sources. The studied area covers two freeway segments in Maryland: one from Interstate 95 (I-95) southbound between Route 100 and I-495, the other on I-270 northbound from I-495 junction to exit 18. Both INRIX and permanently installed Bluetooth sensors provided travel time data for the entire year of 2012 for these two segments. The major tasks for this study are summarized below:

- <u>Bluetooth data processing:</u> The Bluetooth sensors only report the identification number of the devices along with a timestamp. Converting sensor data into travel time and trajectory information requires matching, filtering and aggregation.
- <u>INRIX data processing:</u> INRIX reports data on segments of different lengths called TMC (Traffic Message Channel). In order to accurately obtain the travel time of a path that consists of several TMC segments, a backtracking algorithm must be applied to generate new records.
- <u>Reliability assessment:</u> This study relies on data from two independent sources for reliability assessment. A thorough comparison of two travel time reliability measures was conducted. By carefully selecting an appropriate statistical method, the effects of data sources on travel time reliability measurements can be assessed.
- <u>Final report:</u> Findings of the effects of data sources on travel time reliability assessment and analysis details are presented in this final report.

1.3. Organization of the Report

This final report summarizes the findings of this study led by the research group. First, this report provides a brief review about why travel time reliability is important and how it is defined, in addition to providing equations of existing reliability measures. Next, detailed explanations on how to obtain travel time data from different data sources are provided. This project focused on two different data sources: INTRIX and Bluetooth travel time data. INRIX data provides traffic speed for each TMC, while the Bluetooth travel time path is comprised of multiple consecutive TMCs. In order to make this comparison, researchers developed a methodology on how to convert INRIX speed data to travel time for each Bluetooth segment. Then, the methodology proposed a statistical method to compare performance measures based on different data sources. A special case study for HOV lanes on I-270 indicates that travel times on HOV lanes are different from that on general-purpose lanes. During HOV lane operating hours, the collected Bluetooth travel times on I-270 show two separate trends. Therefore, there is a need to calculate different performance measures for HOV and general purpose lanes. Finally, the last section summarizes the findings and makes conclusions for this study.

2. Literature Review

2.1. The Importance of Travel Time Reliability

Travel time reliability, as a critical congestion performance measure, gradually generates interest among researchers and practitioners. Traditionally, the transportation planning process compares “average travel times” to assess system performance, which formed the basis for comparison and prioritization of system investments. However, the use of “average travel times” does not tell the full story [11]. What travelers experience every day is very different from a simple average throughout the year. While many travelers are used to daily congestion and plan for it, the unexpected changes in traffic condition cause widespread dissatisfaction. The traditional approach failed to analyze the causes of those unexpected changes in travel time. For example, the “annual average travel time” does not detect bottlenecks, traffic incidents, work zones, weather, poor signal timing, and special events, which leads to a vast underestimation of actual congestion. Since many operational strategies, such as incident management systems, focused on reducing or mitigating the impact of those nonrecurring congestions, the traditional approach cannot evaluate the impact of these kinds of operational strategies. Likewise, ignoring the impacts of these non-typical day congestions when comparing and prioritizing system investments may lead to suboptimal investment decisions.

Travel time reliability, defined as the consistency or dependability in travel times, as measured from day-to-day and/or across different times of the day, measures the extent of those non-typical day delays. Chen et al. [12] stated that travel time reliability is “an important measure of service quality for travelers.” Among all reliability measures in transportation systems, travel time reliability is of increasing concern for both travelers and authorities.

The importance of measuring travel time reliability can be summarized as follows [8]:

“(a) From an economic perspective, reliability is highly important because travelers must either build in extra time in their trips to avoid arriving late or suffer the consequences of being late. This extra time has value beyond the average travel time used in traditional economic analyses. Recent work has documented the fact that reliability has value to travelers and that their behavior is influenced by it.

(b) Because of the extra time required in planning trips – and the uncertainty about what travel times will actually be for a trip – reliability influences decisions about where, when, and how travel is made.

(c) Due to the extra economic cost of unreliable travel on users, transportation planners and operators need to include these costs in the project planning, programming, and selection processes. This is particularly true of strategies that deal directly with roadway events (e.g., incidents). In the past, most assessments of these types of strategies have missed this important aspect of travel.”

2.2. Definitions of Travel Time Reliability

The definition of travel time reliability varies in different studies, and there is no universally accepted definition. Studying the existing definitions is helpful in clarifying the meaning of travel time reliability and its measurement. This section briefly reviews travel time reliability definitions in existing literature.

The Federal Highway Administration (FHWA) [11] defines travel time reliability as *“The consistency or dependability in travel times, as measured from day-to-day and/or across* *different times of the day.*" Florida DOT [13] defines travel time reliability as "*(a) the percent of trips that succeed in accordance with a predetermined performance standard for time or speed; (b) the variability of travel times that occur on a facility or a trip over a period of time.*" Three major components of reliability have been addressed; travel time, expected travel time, and acceptable additional travel time. In the work performed for NCHRP Project 3-68 [14], it is noted that continuous travel time data collected over a sufficiently long time period is required for a reliability study in order to capture the variations that occur due to both random and planned events that occur on the roadway system. This leads to a more general definition of reliability: "*Travel time reliability is defined as the level of consistency in travel conditions over time, and is measured by describing the distribution of travel times that occur over a substantial period of time.*"

The F-SHRP Reliability Research Program [15] proposed that reliability and variability should be used interchangeably, since they are related and useful in different purposes. Reliability is based on the notion of probability or the occurrence of failure, and it is commonly used in reference to the level of consistency in transportation services. Variability might be thought of as the amount of inconsistency in operating conditions. Variability measures how many times the travel-time threshold is exceeded, while on-time performance measures how many times the threshold is not exceeded.

From a practical standpoint, travel time reliability has a more "marketable" connotation for measurement purposes because it relates the quality of the service provided. The traveling public and a variety of companies or product sectors use the term reliability in their goal statements. Thus, the F-SHRP defines travel time reliability *"in terms of how travel times vary over time (e.g., hour-to-hour, day-to-day),"* which can be extended to any other travel time-based metrics such as average speeds and delay. The Texas Transportation Institute (TTI) *Urban Mobility Report* [16] made a distinction between variability and reliability of travel time: "*variability measures the amount of inconsistency in transportation operating conditions, reliability relates to the level of consistency in transportation service.*" Some non-U.S. reliability research focused on the unreliability as the probability of "failure" or "vulnerability," where "failure" is defined in terms of traffic flow breakdown and "vulnerability" is a measure of how vulnerable the network is to breakdown conditions, which could be applied at the link or network level.

HCM 2010 [17] proposed two widely accepted ways to define travel time reliability: “*(a) The variability in travel times that occur on a facility or a trip over the course of time, as expressed through metrics such as a 50th, 80th, or 95th percentile travel time; (b) The number of trips that fail or succeed in accordance with a pre-determined performance standard, as expressed through metrics such as on-time performance or percent failure based on a target minimum speed or travel time.”* Reliability is quantified from the distribution of travel times on the facility, and this distribution can be used to produce both variability and failure/success performance measures.” NCHRP 311 [18] summarized the definition of reliability in existing literature as: *“(a) The likelihood of a traveler’s expectations being met. Reliability is measured as the variability between the expected travel time (based on scheduled or average travel time) and the actual travel time (due to the effects of nonrecurrent congestion); (b) The range of travel times experienced during a large number of daily trips; and (c) The impact of nonrecurrent congestion on the transportation system, estimated as a function of the variation in the duration, extent, and intensity of traffic congestion on a system.*”

The second Strategic Highway Research Program (SHRP 2) released several reliability project reports, and each one gives different definitions of travel time reliability with different focuses.

Project L02 (Establishing Monitoring Programs for Travel Time Reliability) introduced travel time probability density functions (TT-PDFs), based on Ebeling’s [19] definition of reliability, which is: “*the probability that a component or system will perform a required function for a given period of time when used under stated operating conditions. It is the probability of a non-failure over time.*” The L02 definition “*allows agencies to portray the variation in travel time that exists between two locations (point-to-point, P2P) or areas (area-to-area, A2A) at a given point in time or across some time interval.”*

For the purpose of determining the effects of reliability mitigation strategies, project L03 research gives a functional definition of reliability, which is: “*the variability of travel times on an extended highway section over the course of six months to one year for different time slices of the day. This definition allows direct measurement with the available data and is consistent with the current state of the practice in performance measurement and economic analyses.”*

Project L04 (Incorporating Reliability Performance Measures in Operations and Planning Modeling Tools) used this definition: *"…models formulated in this research is based on the basic notion that transportation reliability is essentially a state of variation in expected (or repeated) travel times for a given facility or travel experience. The proposed approach is further grounded in a fundamental distinction between (a) systematic variation in travel times resulting from predictable seasonal, day-specific, or hour-specific factors that affect either travel demand or network capacity, and (b) random variation that stems from various sources of largely unpredictable (to the user) unreliability."*

Project L05 (Incorporating Reliability Performance Measures into the Transportation Planning and Programming Process) uses HCM 2010's definition and suggests incorporating the notion of predictability into the basic definition of reliability.

Project L07 (Identification and Evaluation of the Cost-Effectiveness of Highway Design Features to Reduce Nonrecurrent Congestion) used L03's definition.

Project L11 aimed to "*identify and evaluate strategies and tactics to satisfy the travel-time reliability requirements of users of the roadway network—those engaged in both freight and person transport in urban and rural areas.*" This project defines travel time reliability as follows: "*Travel-time reliability is related to the uncertainty in travel times. It is defined as the variation in travel time for the same trip from day to day (same trip implies the same purpose, from the same origin, to the same destination, at the same time of the day, using the same mode, and by the same route). If there is large variability, then the travel time is considered unreliable. If there is little or no variability, then the travel time is considered reliable.*"

2.3. Travel Time Reliability Measurements

Although a significant amount of research has focused on travel time reliability and has proposed several reliability measures in the literature, there exists a debate on which method is the most effective. A study by Lomax et al. provides a comprehensive review of existing travel time reliability measures [20]. A report by the FHWA indicates that although a reliability measure must have technical merit, it also must be easily understood by non-technical audiences [11]. Finding an effective measure that provides the best understanding of existing traffic conditions is critical to travelers and transportation decision makers alike. This section briefly reviews some common reliability measures. Based on a study by Van Lint et al. [3], travel time reliability measures are classified into four categories: (1) statistical range methods, (2) buffer time methods, (3) "tardy-trip" indicators, and (4) probabilistic measures.

Statistical Range Measures

Statistical range methods interpret travel time reliability using some basic statistical concepts: standard deviation, percent variation, skew statistic, and width statistic. These methods are easy to understand for statisticians, but most are difficult to explain to nontechnical audiences.

Standard deviation (STD) represents the variation or dispersion around the average.

Percent variation (PV) is the normalized standard deviation. It is the ratio of the standard deviation to the mean μ. This measure is useful when comparing the degree of variation among different datasets.

$$PV = STD/\mu \times 100\% \qquad \text{EQ 1}$$

Skew statistic (λ^{skew}) and width statistic (λ^{var}) describe the size and shape of the travel-time distribution. The skew statistic relates the range of the 40th percentile travel time above the median to the range of the 40th percentile below the median. It therefore indicates the level and direction of the skewness of the data [21].

$$\lambda^{skew} = \frac{TT_{90th} - TT_{50th}}{TT_{50th} - TT_{10th}} \qquad \text{EQ 2}$$

$$\lambda^{var} = \frac{TT_{90th} - TT_{10th}}{TT_{50th}} \qquad \text{EQ 3}$$

where TT_{90th}, TT_{50th}, and TT_{10th} represent the 90^{th}, 50^{th}, and 10^{th} percentile travel time.

Buffer Time Methods

Different from statistical methods, an average traveler can comprehend the buffer time concept, as it often relates to the way travelers make route choice decisions. It measures the extra percentage of travel time travelers should plan for their trip to ensure they will arrive on time.

Buffer Index (BI) represents the extra percentage of time a driver should add to the average travel time to ensure he or she will arrive on time. The Buffer Index is the difference of the 95th percentile travel time and average (or median) travel time divided by the average (or median) travel time.

$$BI = \frac{BT}{\mu} = \frac{TT_{95th} - \mu}{\mu} \quad \text{EQ 4}$$

Travel Time Index (TTI) compares the average time that a traveler would experience during a trip to travel time during free-flow conditions. It is the ratio of average travel time across the entire year to travel time at free-flow conditions.

$$TTI = \frac{TT_{mean}}{TT_{free\ flow}} \quad \text{EQ 5}$$

where TT_{mean} is the mean travel time and $TT_{free\ flow}$ is the free-flow travel time.

Planning Time Index (PTI) denotes the total percentage of time, compared with the free-flow travel time that a traveler should allow to ensure meeting 95th percentile on-time arrival. The planning time index is calculated as the 95th percentile travel time divided by the free-flow travel time.

$$PTI = \frac{TT_{95th}}{TT_{free\ flow}} \quad \text{EQ 6}$$

Similar to the planning time index, there are 80th, 85th, and 90th percentile Travel Rate. They are calculated as the 80th, 85^{th}, and 90th percentile travel time, respectively, divided by the free-flow travel time.

Tardy Trip Indicators

Tardy trip indicators emphasize unreliability by using the amount of trips that lead to late arrivals. The Misery Index (MI), for example, is the average travel time of the 20 percent worst trips minus average travel time of the total trips and divided by the average travel time. This method focuses on the extra delays during the worst day of the week.

$$MI = \frac{\text{Average travel time for the longest 20\% of trips} - \text{Average Travel Time}}{\text{Average Travel Time}} \quad \text{EQ 7}$$

Probabilistic Measures

Probabilistic measures calculate the probability that the observed travel times are larger than *n* times for some predefined travel time threshold (for example, median travel time on a given time of day, or day of the week).

$$PR(a) = P(TT_i \geq \alpha \cdot TT_{50th}) \quad \text{EQ 8}$$

The parameter α used here can be chosen by the analyst. A value of 1.2 denotes the probability that travel time is 20 percent greater than the median travel time [3].

Travel time reliability measurements, as an important indicator for roadway system performance, have been utilized by several transportation agencies as a portion of their mobility measurement in their performance evaluations. Different agencies chose different reliability measurements according to their needs. Table 1 summarizes some selected agencies' choices of reliability measures. Maryland SHA included travel time index and planning time index in the 2012 *Maryland State Highway Mobility Report*.

Table 1. Transportation Agencies Use Reliability Measures as a Portion of Their Mobility Measurement in Their Performance Evaluations

Agency	Reliability Metrics Used
Georgia Regional Transportation Authority and Georgia DOT	Buffer Index Planning Time Index
Florida DOT	Buffer Index On-Time Arrival
Southern California Association of Governments	Buffer Index
Washington State DOT	95th Percentile Travel Time
National Transportation Operations Coalition (NTOC)	Buffer Index
Maryland SHA	Travel Time Index Planning Time Index

3. Data Collection and Preprocessing

The travel time reliability index is measured based on the underlying distribution of travel time on a segment over time. Effective calculation of this index requires accurate, high-quality data. This study focused on the impact of selecting INRIX and Bluetooth data on travel time reliability assessments. Two major freeway segments were selected for the case study: one from Interstate 95 southbound and the other from Interstate 270 northbound (Figure 1). I-95 is one of the most heavily traveled Interstate highways in Maryland, and runs diagonally from northeast to southwest. I-270 is an auxiliary Interstate highway in the State of Maryland. The studied portion of I-270 consists of a local-express lane configuration as well as a high-occupancy vehicle (HOV) lane that is in operation during afternoon peak hours (3:30 p.m. – 6:30 p.m.). Both selected segments have uninterrupted coverage of both Bluetooth and INRIX data 24 hours per day and 365 days a year.

Figure 1 shows the locations of detectors and the configuration of segments on the studied corridors. The red pushpins denote the locations of the Bluetooth sensors. Two Bluetooth sensors are required for each corridor to obtain corresponding travel time, where "Bluetooth_Start" and "Bluetooth_End" indicates the start and end points of the studied section. As shown in the upper panel of Figure 1, the selected I-95 southbound section is between the Route 100 and I-495 junction with a total length of 15.94 miles. For I-270, a 17.89-miles-long corridor was selected, covering the portion of I-270 northbound starting from the I-495 junction and ending at I-270 exit 18, as shown in the lower panel of Figure 1. To compare travel time reliability measures, INRIX data are also captured for both corridors. The INRIX data are reported on Traffic Message Channel (TMC) location codes. TMC is an industry standard developed and maintained by the leading electronic mapping vendors to uniquely define road segments [15]. Each of the studied corridors is covered by multiple TMCs with different lengths. Individual TMC portions of the corridors are marked by green color and their identification code (e.g., 110N04420) in Figure 1. For the I-95 corridor, the start Bluetooth sensor is located 0.91 miles north of the southern end of the TMC (110-04421) and the end Bluetooth sensor is located 0.28 miles south of the northern end of the TMC (110-04260). On the I-270 corridor, the first Bluetooth sensor is located 0.6 miles north of the start of the southern end of the TMC

(110+4103), while the last Bluetooth sensor is located 0.86 miles south of the northern end of the TMC (110+04115). The collected INRIX travel time is adjusted to match with the Bluetooth

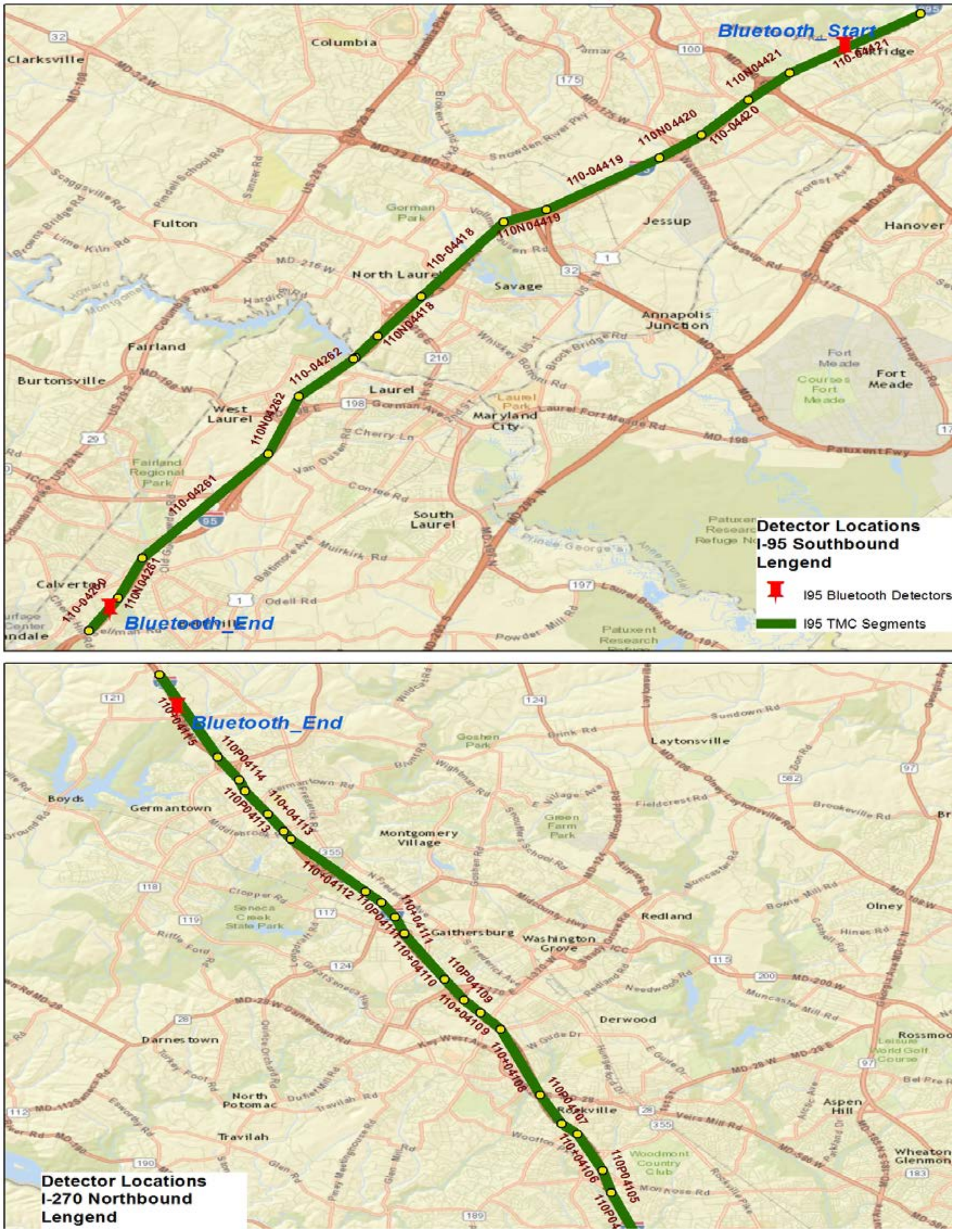


Figure 1. Case study: Bluetooth detector and INTRIX TMC locations.

data based on the path length difference. The following briefly describes the procedure to collect and process the Bluetooth and INRIX data.

3.1. Bluetooth Data

Bluetooth is an open, wireless communication platform enabling digital devices to connect with each other using short-range wireless communications. Many computers, mobile phones, car radios and dashboard systems, PDAs, headsets, and other personal devices come equipped with Bluetooth wireless capabilities to communicate with other Bluetooth-enabled devices anywhere from 1 m to about 100 m, or 300 ft. In the context of vehicle travel time collection, the Bluetooth detector captures the electronic identifier, or tag, called the Machine Access Control (MAC) address, in each Bluetooth-enabled device and places a timestamp when the vehicle enters the detection range of the sensor. When the same vehicle passes subsequent detectors, the detected MAC address can be matched. The difference of the detection time for the same vehicle is the time used to traverse this segment between the two detectors. The Bluetooth technique requires at least two detectors to obtain travel time data. The Bluetooth detector provides more accurate traffic data with a relatively low cost and easy installation, but since only a small percentage of drivers have Bluetooth-enabled devices in their vehicles, it leads to lower detection rates.

There are some issues that should be considered when filtering traffic data obtained from Bluetooth devices. First, the Bluetooth sensors record the same wireless network ID more than once, especially when the speed of the vehicle is low. During the data preprocessing process, there is a need to filter out redundant information to guarantee unique information for each vehicle. Secondly, the Bluetooth sensor has a wide detection range and collects undesirable vehicle IDs. For example, a Bluetooth sensor desirable for highway traffic detection may also collect information on vehicles at a ramp. The undesirable vehicle information would compromise the accuracy of the collected freeway traffic information, and it must be filtered. Both issues should be carefully addressed during the data preprocessing process. The raw data are MAC addresses along with their collection time. A four-step filtering algorithm proposed by Haghani et al. [23] is utilized to extract travel time from the pool of Bluetooth observations. The first two steps are designed to identify and discard outliers among single observations in each time interval. The third and fourth steps are designed to exclude time intervals during which we either do not have enough observations (proxy for low-volume traffic conditions) or when there are large variations among individual observations within the time interval, even if there is a sufficient number of observations to consider.

Step 1: Abnormal travel times typically indicate a match between two observations at consecutive Bluetooth sensors in which the same MAC address has been registered during different vehicle trips between these locations. This is mainly due to the sampling nature of the technology utilized to measure the “ground truth.” For example, consider the case that during a multi-day travel time study a commuting vehicle gets detected by the first sensor on the first day but is not detected by the second sensor. If, in the second day, the same vehicle is not detected by the first sensor but is detected by the second sensor, the recorded MAC ID of day one will be matched with the same MAC ID of day two, resulting in a very large travel time. To filter out these obvious outliers, a method based on the assumption of the smooth transition of travel times is adopted. In this method, the histogram of observed speeds in a past time window is formed. Then, a moving average of the speed frequencies is calculated, which establishes the basis for identification of lower and upper cut-off points. Observations falling beyond these lower and upper limits are automatically flagged as outliers and therefore will not be part of the ensuing calculations.

Step 2: In this step, the variations in speed observations are considered to identify additional outlier speed observations. To that end, all observations corresponding to each of the time intervals for which we have Bluetooth observations are identified, and the average and standard deviation of the speeds in those time intervals are calculated. Observations that correspond to speeds falling within ±1.5 times the standard deviation are kept, and the rest are discarded. Assuming a normal distribution for the observations around the mean, this approach translates into keeping nearly 87 percent of the data. At this point, the number of observations, the averages and the standard deviations of the speed in all time intervals for which we have Bluetooth data are updated and kept ready for future use in the analysis.

Step 3: This step is designed to exclude all time intervals with a small number of observations that will render ground truth estimation in those time intervals impossible. Similarly, this step can serve as a surrogate method for identification of time intervals with extremely low traffic volumes (typically less than 500 vehicles per hour). Generally speaking, this step is designed to ensure that the "ground truth" speed estimations are reliable under low sampling rate conditions. The following equation specifies the minimum number of required observations in each time interval.

$$n_{min} = \frac{V_{tresh} \times T \times PSR}{60}$$

where

n_{min} is the minimum number of observations required per time interval,

V_{tresh} is the threshold hourly volume of traffic, below which ground truth estimation with low sampling rate may not be reliable (vehicles per hour),

T is the length of time interval for which estimation is being performed (minutes), and

PSR is the percentage of sampling rate that can be reasonably sustained throughout the analysis period.

Step 4: Finally, to ensure that the variability among speed observations inside a given time interval is within a reasonable level, the coefficient of variations (COV) of Bluetooth speed observations in each time interval that survives the previous three steps is estimated, time intervals that have a COV greater than 1 are excluded, and their corresponding observations are discarded from further consideration in the ground truth estimation process.

3.2. INRIX Data

The company INRIX is a private-sector travel time data provider that derives travel times from its smart driver network, which aggregates traffic data from probe vehicles and traditional sensor sources. The probe vehicles utilized include taxis, airport shuttles, service delivery vans, long-haul trucks, consumer vehicles, GPS-enabled consumer smartphones, and so on. The data fusion methods are proprietary, and travel times are reported on TMC segments.

In order to calculate path travel times in different intervals that are directly comparable with Bluetooth data points, data from all TMCs must be consolidated. Bluetooth observations are time tagged at the end sensor, so a backtracking algorithm is required to generate an equivalent INRIX record per each Bluetooth observation. The INRIX path records are then aggregated in desired intervals. Details of this method can be found in Hamedi et al. [24]. A brief description of this method follows:

In order to compare INRIX speed feeds with Bluetooth travel time, an equivalent travel time should be estimated based on INRIX data. The segment between each pair of Bluetooth detectors is comprised of multiple individual TMC segments. The estimation should be based on either the observed entrance or exit time of the vehicle to/from the path. For the sake of simplicity in illustration, the initial example given herein and the proposed algorithm are based on the exit times. However, this study applies to both forward and backward procedures and reports the result.

First, the time period to which the end time of a Bluetooth match belongs should be identified. Then, given the distances and speeds of each segment in that time period, the average travel time for each segment is calculated. Based on these travel times, we can determine where the vehicle has been at the beginning of the current time period by backtracking. Obviously, in the case that vehicle has been midway on the path, the procedure can be repeated with data from the previous period until it is determined that the vehicle has entered the path during the current time period. The concept is illustrated through an example.

0.1	45(0.13)	30(0.2)	40(0.15)	⇩ 25(0.24)
0.2	30(0.4)	40(0.3)	25(0.48)	45(0.27)
0.3	15(1.2)	30(0.6)	45(0.4)	25(0.72)
0.4	25(0.96)	15(1.6)	30(0.8)	45(0.53)
	8:00	8:05	8:10	8:15

The table above shows an aggregated space-time diagram of a 1-mile path comprised of four smaller segments, or TMCs, which are 0.4, 0.3, 0.2, and 0.1 miles long (shown on the Y-axis). The sample period starts from 8:00 a.m. to 8:15 a.m. with 5-minute intervals (X-axis). Thus, each cell of the table corresponds to one segment during a certain time interval. The value in each cell represents sample INRIX speeds, measured in mph, for each segment during each time period. Values within parentheses correspond to average INRIX travel times in minutes. Two Bluetooth sensors are placed at both ends of the one-mile segment. For example, a

Bluetooth match with exit time 8:11 a.m. (indicated with the arrow in the figure) belongs to the fourth time period above. Based on estimated segment travel time during each time interval, the trajectory of this vehicle can be derived. This vehicle spent 1 minute in the fourth time period. During this one minute, it travels through the last two segments and parts of the second one:

1-0.24>=0, our vehicle has totally covered the fourth segment during this period,

1-0.24-0.27>=0, our vehicle has totally covered the third segment during this period,

1-0.24-0.27-0.72<0, our vehicle has spent 1-0.24-0.27=0.49 min on the second segment during this period. This translates to 0.3*0.49/0.72=0.2 miles of the second segment.

In other words, at the beginning of this period, our vehicle has already covered 0.3-0.2=0.1 mile of the second segment.

Given the average speed on the second segment in the third time period, our vehicle would have required 0.1/0.3*0.4=0.13 min to cover 0.1 mile on the second segment:

5-0.13>=0, our vehicle has covered the 0.1 mile on the second segment during the previous period,

5-0.13-0.8>=0, our vehicle has totally covered the first segment during the previous period.

In total, our vehicle has spent 1+0.13+0.8=1.93 min to travel through the segments. This is the equivalent travel time estimated using average reported speeds on each segment, which results in an average travel speed equal to (0.4+0.3+0.2+0.1)/1.93*60=31.0 mph for our sample vehicle.

Accordingly, for the purpose of estimating equivalent travel times, the following algorithm is proposed:

Given a sequence of segments $i = 1, \dots, N$ forming a path, each with length $l(i)$, and speed $S(i, t)$ at time period t:

For each valid Bluetooth observation pair $(StTime, EndTime)$ ending at the same time period $\{T: (T-1) * 5 < EndTime \leq T * 5\}$, let: $TE = mod(EndTime, 5); n = N; t = T; PTT = 0;$

```
        while (Ture)
                If $TE \geq \sum_{i=1}^{n} TT(i,t)$ then
                        $PTT = PTT + \sum_{i=1}^{n} TT(i,t)$ ; Break.
                Find segment $j$ for which, $TE \geq \sum_{i=j+1}^{n} TT(i,t)$ and $TE < \sum_{i=j}^{n} TT(i,t)$
                $PTT = PTT + TE$;
                $TT(j,t-1) = TT(j,t-1) * \frac{\sum_{i=j}^{n} TT(i,t) - TE}{TT(j,t)}$
                $TE = 5min; n = j; t = t-1$;
        Loop while,
Repeat.
Compute the equivalent average path travel time, $\overline{\mathrm{PTT}}$
Compute the equivalent average speed for comparison with Bluetooth speed:
```

$$S_{INRIX}(T) = \frac{\sum_{i=1}^{N} l(i)}{\overline{PTT}}$$

3.3. Travel Time Data Distribution

Since travel time patterns during weekends are significantly different from weekdays, this study only focused on travel time reliability measures during weekdays. As a result, a total of 260 days of travel time data in 2012, including 24,960 data points for each segment, were selected. Figure 2 depicts weekday scatter plots and boxplots of every 15-minute Bluetooth and INRIX travel time for the entire year on both segments. Comparing scatter plots of data in Figure 2(a) and Figure 2(b) shows that both Bluetooth and INRIX illustrate both morning and afternoon peak hours for the I-95 southbound segment with similar patterns. However, similar plots for the I-270 segment presented in Figure 2 (e) and Figure 2 (f) display only afternoon peak hour and the Bluetooth scatter plot exhibits higher variation in travel time compared to the INRIX data. To better understand the underlying travel time distribution, box-and-whisker plots based on both Bluetooth and INRIX data are presented in Figure 2 (c, d, g, f) for both segments. These plots provide a convenient way to graphically compare the distribution of the data. On the vertical axis, the bottom and top of the blue box are the first and third quartiles of the travel time data for any given time of the day, and the black band inside the box is the median travel time. During non-peak hours, the first and third quartiles are close to the median with relatively smaller variations of travel time. During peak hours, distance between the 1st and 3rd quartile increases to approximately 10 minutes, which is an indicator of travel time variability. A comparison between Bluetooth and INRIX boxplots on I-270 shown is Figure 2 (g) and Figure 2 (h) suggests that both mean and variance of travel time are significantly different. The INRIX data tend to exhibit smaller variation and lower mean during peak hours. The cause of this phenomenon is discussed later in this report.

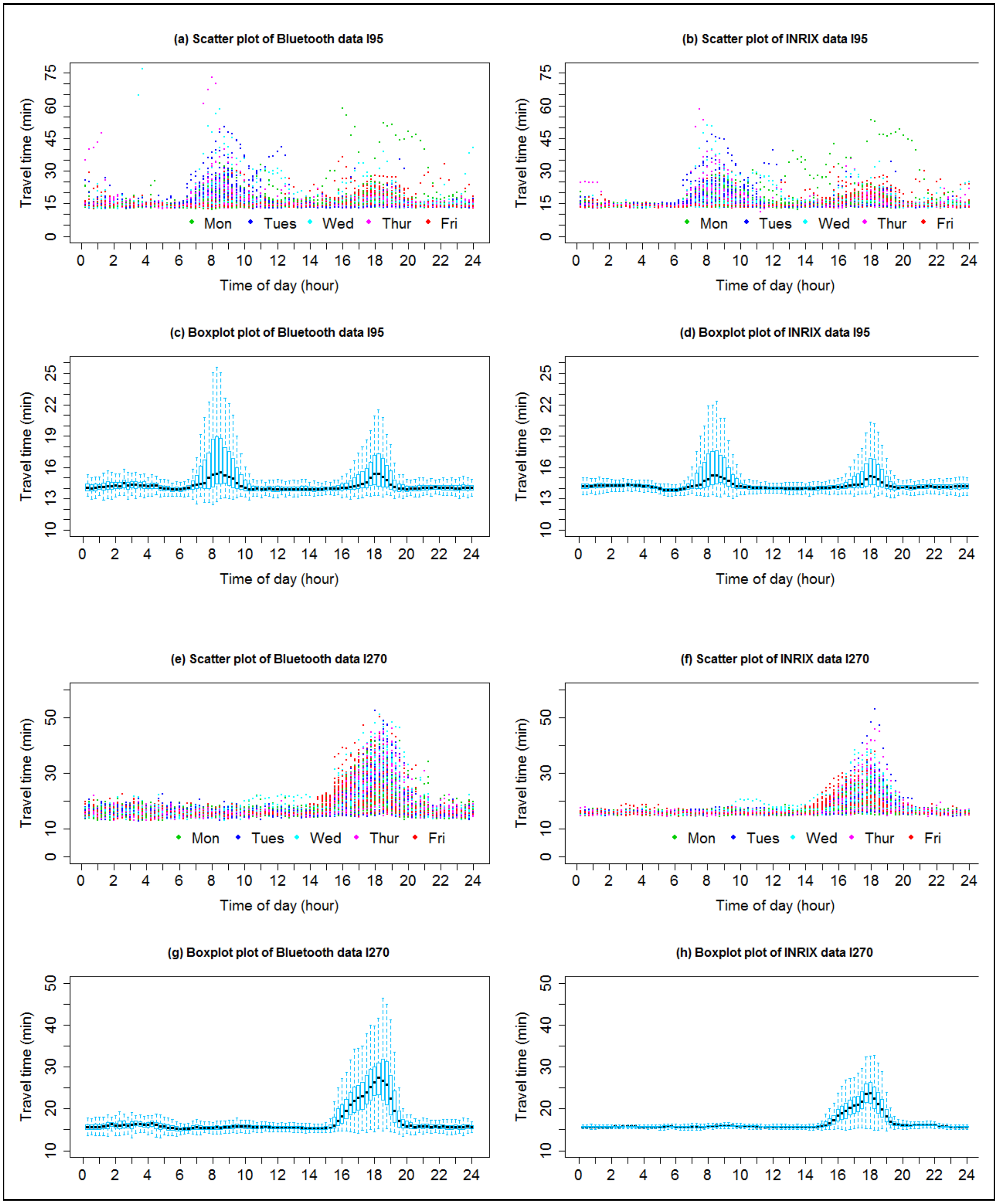


Figure 2. Scatter and boxplots of Bluetooth and INRIX travel time on I-95 and I-270 in 15 minute intervals

4. Methodology

Real-world data are subject to measurement errors, and the errors associated with different data collection methods could be different due to the uniqueness of data collection and processing procedures. As indicated in Figure 2, the travel time collected from Bluetooth sensors are more sparsely distributed compared to the INRIX data. Therefore, it is critical to test if measures based on different data sources would be different. This section first examines whether the correlations of different measures would be different based on these two data sources and then utilizes the Welch's t-test statistical method to assess the level of difference between two data sources.

4.1. Correlation among Different Measures

Based on Bluetooth data from the I-95 segment, Table 2 shows the correlations coefficient of reliability measures. Table 3 shows the difference between the correlation coefficient of the same reliability measures obtained from both Bluetooth and INRIX. The correlation matrix of Table 2 is a pairwise comparison of reliability measures, with higher values indicating higher correlations between the two corresponding measures. Each cell is colored according to the coefficient value, with red representing the highest value, green the lowest, and yellow denoting value in between. It seems that the reliability measures of Width, BI, Planning Time Index (PTI), and 90th, 85th, and 80th Percentile Travel Rate are highly correlated with each other, while Skewness is significantly different from other measures. This finding is on a par with the results from a previous study [8].

Table 2 summarizes the difference of correlation coefficient matrix between INRIX and Bluetooth (correlation coefficient of INRIX minus corresponding value of Bluetooth data). In general, correlation coefficients based on Bluetooth and INRIX are similar, with Skewness relatively more sensitive to the data source.

Another tool to represent correlations of different performance measures is the correlation plot, which provides a straightforward illustration. It not only indicates the level of correlations between each pair of measures, but also shows how they are correlated. Each cell of the correlation plot in Figure 3 shows the correlation between a pair of measures from the same data source. The closer the points to the line of equality (a nonexistent line with $y = x$), the higher the correlation between these two measures. In Figure 3, all plots in a row share the common y axis, while all plots in a column share the common x axis. For example, the y axis for the first row represents the Skew Statistic, and the x axis for the second column denotes the probability measure PR(1.2). The cell located at the first row and second column is the plot of Skew Statistic against the probability measure PR(1.2), which visualizes the correlation between these two measurements. The measurements are arranged in such a way that those with higher correlations are closer to the principal diagonal. The cell is color coded according to the value of the correlations, with purple representing the highest correlation, yellow denoting the lowest correlation, and blue indicating the ones with modest correlation. It is evident from this plot that there are strong linear correlations between reliability measures Width, BI, Planning Time Index, and 90th, 85th,80th Percentile Travel Rate. Based on information from one of the measures, it is possible to derive values for the other measures by utilizing this linear correlation.

Table 2. Correlation Coefficient of Reliability Measures (Bluetooth on I-95)

Correlation	Mean	STD	PV	SKEW	WIDTH	BI	PTI	90th%	85th%	80th%	MI	PR(1.2)
Mean	1	0.88	0.83	0.65	0.98	0.97	0.97	0.99	0.99	0.99	0.97	0.97
STD	0.88	1	0.99	0.68	0.86	0.83	0.87	0.86	0.84	0.82	0.94	0.88
PV	0.83	0.99	1	0.68	0.81	0.77	0.82	0.8	0.78	0.76	0.9	0.84
SKEW	0.65	0.68	0.68	1	0.72	0.74	0.74	0.7	0.64	0.59	0.78	0.78
WIDTH	0.98	0.86	0.81	0.72	1	1	0.99	1	0.99	0.97	0.97	0.98
BI	0.97	0.83	0.77	0.74	1	1	0.98	0.99	0.98	0.96	0.96	0.97
PTI	0.97	0.87	0.82	0.74	0.99	0.98	1	0.98	0.97	0.95	0.98	0.98
90th%	0.99	0.86	0.8	0.7	1	0.99	0.98	1	0.99	0.98	0.96	0.97
85th%	0.99	0.84	0.78	0.64	0.99	0.98	0.97	0.99	1	0.99	0.95	0.96
80th%	0.99	0.82	0.76	0.59	0.97	0.96	0.95	0.98	0.99	1	0.93	0.94
MI	0.97	0.94	0.9	0.78	0.97	0.96	0.98	0.96	0.95	0.93	1	0.98
PR(1.2)	0.97	0.88	0.84	0.78	0.98	0.97	0.98	0.97	0.96	0.94	0.98	1

In general, we can conclude that although different measures are developed for different purposes, most of them have some correlations. The only measure that is significantly different from the others is the Skewness. Based on the definition, it is the only measure that indicates the level and direction of the distribution. The scatter plot matrix is a good way to visualize the level of correlation between different measures and how they are correlated.

Table 3 Difference Between Correlations Coefficient between INRIX and Bluetooth (INRIX-Bluetooth on I-95)

Difference	Mean	STD	PV	SKEW	WIDTH	BI	PTI	90th%	85th%	80th%	MI	PR(1.2)
Mean	0	0.01	0.02	0.19	0	0	0	0	0	0	0	0.01
STD	0.01	0	0	0.14	0.02	0.03	0.02	0.02	0.02	0.01	0.02	0.04
PV	0.02	0	0	0.13	0.04	0.04	0.04	0.03	0.04	0.02	0.03	0.05
SKEW	0.19	0.14	0.13	0	0.17	0.16	0.17	0.18	0.2	0.2	0.13	0.14
WIDTH	0	0.02	0.04	0.17	0	0	0	0	0	0	0	0.01
BI	0	0.03	0.04	0.16	0	0	0	0	0	0	0	0.01
PTI	0	0.02	0.04	0.17	0	0	0	0	0	-0.01	0	0.01
90th%	0	0.02	0.03	0.18	0	0	0	0	0	0	0	0.01
85th%	0	0.02	0.04	0.2	0	0	0	0	0	0	0	0.01
80th%	0	0.01	0.02	0.2	0	0	-0.01	0	0	0	-0.01	0
MI	0	0.02	0.03	0.13	0	0	0	0	0	-0.01	0	0.01
PR(1.2)	0.01	0.04	0.05	0.14	0.01	0.01	0.01	0.01	0.01	0	0.01	0

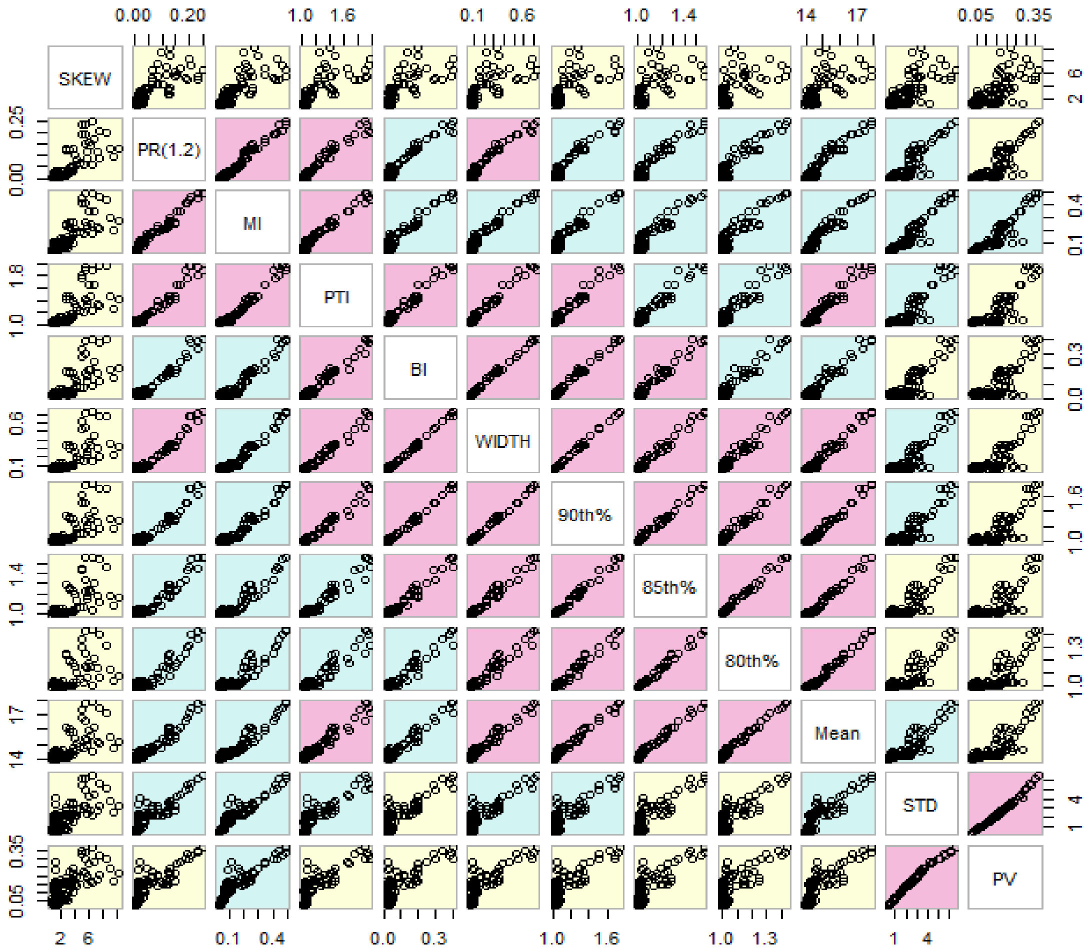


Figure 3. Reliability measures correlation plot (Bluetooth Data)

4.2. Welch’s T-test

The purpose of this study was to assess the effect of data selection on travel time reliability measurements. A comparison was necessary to test whether measurements based on different data sources agree with each other. Differences within a small range would be acceptable and will not cause problems in representing traffic conditions or affect transportation managers’ decision making processes. Since the “ground truth” value remains unknown, the difference between INRIX and Bluetooth measurements becomes the objective of interest. In other words, we would want to test whether reliability measures derived from these two data sources are statically equal.

This study implements the statistical method Welch's t test. The Welch's t test is an adaptation of the student’s t-test, which is used to determine if there is a significant difference between the mean of two populations of observations that have possibly unequal variances. If there is no significant difference, the mean of the differences between these two populations should be zero. Otherwise, the mean of the differences is not equal to zero. Instead of collecting travel time of all vehicles on a certain segment, both Bluetooth and INRIX collect a sample of vehicle travel time. In other words, travel time collected by these two detectors can follow different distributions with possibly unequal mean and variance. Therefore, the Welch's t test is selected here to compare travel time information obtained by the two detectors. Let $\{X_{11}, X_{12}, X_{13}, \ldots, X_{1m}\}$ and $\{X_{21}, X_{22}, X_{23}, \ldots, X_{2n}\}$ be the collected sample from two populations. The first sample has a sample size of m with mean μ_1 and variance σ_1^2. The second population has a sample size of n with mean μ_2 and variance σ_2^2. The goal is to test if these two population means are equal. The test statistic for the Welch's t-test is as follows:

Null hypothesis: $H_0: \mu_1 = \mu_2$

Alternative hypotheses: $H_1: \mu_1 \neq \mu_2$

Test statistic: $T_0 = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$

P-value: Probability above $|t_0|$ and below $-|t_0|$

Rejection region: $t_0 > t_{0.5\alpha,v}$ or $t_0 < -t_{0.5\alpha,v}$

Degree of freedom: $\upsilon \approx \frac{(\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2})^2}{\frac{s_1^4}{N_1^2\upsilon_1}+\frac{s_2^4}{N_2^2\upsilon_2}}$

where $\overline{X}_1$ and $\overline{X}_2$ are sample mean, s_1^2 and s_2^2 are sample variance, m and n are sample size for the first and second population, $\mu_1$and μ_2 are true population mean, $v_1 = m - 1$ and $v_2 = n - 1$ are the degrees of freedom associated with the first and the second variance estimate. T_0 is t value and $(1 - \alpha)100\%$ gives the confidence level for the test.

4.3. Experiment Results

For the purpose of comparing reliability measures, measures based on Bluetooth and INRIX data at each time interval are the samples. The test is performed for measures in every 15 minute intervals, for all days and peak hours as shown in Table 4. The null hypothesis of the test is that there is no significant difference between the mean travel time reliability measures based on Bluetooth and INRIX data. The alternative hypothesis is that the mean travel time reliability measures based on Bluetooth and INRIX data are different. For 95 percentage confidence, if the p value is higher than 0.05, then there is no significant evidence to reject the hypothesis that the measures obtained from these two data sources are different from each other. On the other hand, if the p value is less than 0.05, we would reject the null hypothesis with 95 percentage confidence and conclude that there is significant difference between the two measures.

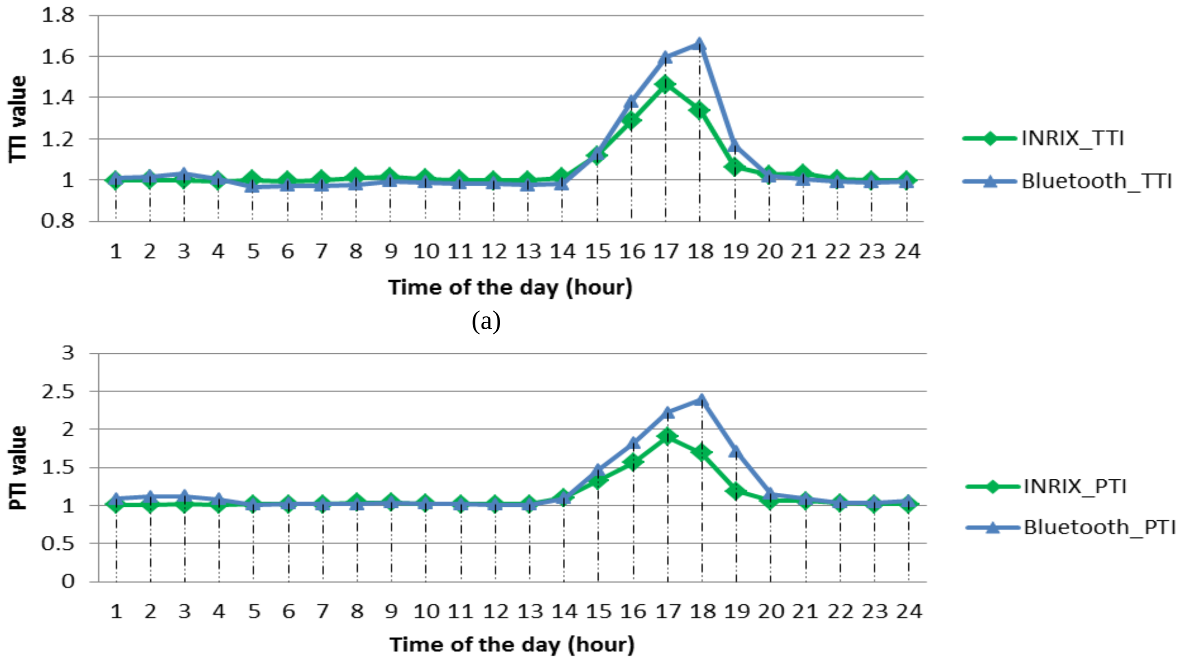
Table 4 summarizes the Welch's t test results for all day and peak hour measures of the study sites. For the I-95 corridor, except for standard deviation and performance variation measures, p values for all the other performance measures based on every 15 minute data are higher than 0.05. This is to say, with 95 percentage confidence, there is no significant evidence to reject the null hypothesis that there is a difference between the Bluetooth and INRIX performance measures, regardless of the time period of the day (except standard deviation and performance variation). Both standard deviation and performance variation measure the dispersion of the data from their mean. The higher value indicates higher variation or dispersion. By looking at the confidence intervals of both STD and PV for all day and peak hour measures, it can be observed that they assume values higher than zero. This indicates that STD and PV value estimated by Bluetooth data are higher than their counterparts estimated by INRIX data. The result is consistent with Figure 2, as the Bluetooth data show more variations or larger dispersions from the mean compared with the INRIX data. This is can be attributed to the fact that INRIX data source simply is not as granular and tends to make the data smooth, while Bluetooth takes the average of all valid observations. On the other hand, looking at Welch's t-test results on I-270 shows most p-values are lower than 0.05, which is to say that most performance measures estimated by Bluetooth data are different from the one estimated by INRIX. The selected I-270 segment covers HOV lanes, which can significantly save travel time during effective hours. It is possible that the different ways Bluetooth and INRIX collect travel time data on HOV lanes results in the difference between performance measures. In general, based on the I-95 segment results, it is safe to say that there is no significant difference between travel time reliability performance measures using either Bluetooth or INRIX data (with exception of standard deviation and performance variation measures). As a special segment that contains HOV lanes and shows different behavior, I-270 segment is further studied for more details.

Table 4. T-test of Different Performance Measures on Fifteen Minute Intervals (Bluetooth - INRIX)

	All Day	Mean	STD	PV	SKEW	WIDTH	BI	PTI	90th%	85th%	80th%	MI	PR(1.2)
I-95	P-value	0.33	0.02	0.01	0.06	0.06	0.07	0.19	0.2	0.32	0.36	0.18	0.5
	T-value	0.97	2.34	2.48	1.9	1.91	1.84	1.32	1.29	1	0.92	1.34	0.67
	C.I.	(-0.12, 0.36)	(0.07, 0.83)	(0.01, 0.05)	(-0.02, 1.28)	[0, 0.08)	[0, 0.05)	(-0.02, 0.11)	(-0.02, 0.08)	(-0.02, 0.05)	(-0.01, 0.04)	(-0.01, 0.05)	(-0.01, 0.02)
	D.F.	183.05	183.49	185.55	188.53	182.83	185.05	186.25	182.88	183.12	179.83	189.74	189.85
	Mean	14.67	1.85	0.12	2.75	0.15	0.07	1.18	1.11	1.08	1.05	0.13	0.05
	Peak	Mean	STD	PV	SKEW	WIDTH	BI	PTI	90th%	85th%	80th%	MI	PR(1.2)
	P-value	0.21	0.04	0.03	0.93	0.11	0.16	0.23	0.2	0.25	0.2	0.38	0.58
	T-value	1.25	2.14	2.3	-0.09	1.61	1.43	1.23	1.29	1.16	1.3	0.88	0.55
	C.I.	(-0.18, 0.78)	(0.04, 1.23)	(0, 0.07)	(-1.14, 1.05)	(-0.02, 0.16)	(-0.01, 0.09)	(-0.05, 0.2)	(-0.03, 0.16)	(-0.03, 0.12)	(-0.02, 0.1)	(-0.03, 0.08)	(-0.02, 0.04)
	D.F.	55.63	53.94	54.98	57.46	55.01	55.91	56.12	55.21	55.57	55.2	57.66	58
	Mean	15.63	3.34	0.21	5.32	0.31	0.17	1.44	1.29	1.22	1.16	0.27	0.12
I-270	All Day	Mean	STD	PV	SKEW	WIDTH	BI	PTI	90th%	85th%	80th%	MI	PR(1.2)
	P-value	0.21	<0.01	<0.01	0.07	<0.01	<0.01	0.01	0.03	0.06	0.11	<0.01	0.01
	T-value	1.26	4.25	5.3	1.83	4.89	4.71	2.57	2.14	1.87	1.62	4.12	2.64
	C.I.	(-0.29, 1.32)	(0.53, 1.46)	(0.03, 0.06)	(-0.02, 0.51)	(0.06, 0.15)	(0.03, 0.08)	(0.03, 0.22)	(0.01, 0.17)	(0, 0.14)	(-0.01, 0.12)	(0.03, 0.08)	(0.01, 0.04)
	D.F.	161.93	158.87	171.1	167.27	164.49	162.43	155.6	158.21	157.73	159.01	169.49	155.39
	Mean	18.36	1.42	0.07	1.39	0.15	0.08	1.21	1.16	1.13	1.11	0.09	0.03
	Peak	Mean	STD	PV	SKEW	WIDTH	BI	PTI	90th%	85th%	80th%	MI	PR(1.2)
	P-value	0.03	<0.01	<0.01	0.16	<0.01	<0.01	<0.01	<0.01	0.01	0.02	<0.01	<0.01
	T-value	2.25	4.98	6.11	1.43	5.57	6.67	3.88	3.3	2.87	2.51	5.91	5.88
	C.I.	(0.24, 4.81)	(1.41, 3.35)	(0.06, 0.11)	(-0.29, 1.62)	(0.14, 0.3)	(0.08, 0.16)	(0.2, 0.64)	(0.12, 0.52)	(0.08, 0.46)	(0.04, 0.41)	(0.08, 0.16)	(0.05, 0.11)
	D.F.	33.37	35.6	37.98	29.72	36.45	37.32	33.69	33.79	32.76	32.44	37.82	37.41
	Mean	22.75	4.16	0.18	2.32	0.42	0.22	1.76	1.62	1.54	1.47	0.26	0.14

4.4. Special Case for HOV Lane

To better understand the difference between reliability measures derived from Bluetooth and INRIX data, the focus has been put on TTI and PTI measures on I-270 segment. The Maryland State Highway Administration (SHA) uses TTI and PTI to quantify congestions in their annual mobility report [6]. The TTI denotes the ratio of actual travel time to the ideal or free flow travel time, while the PTI represents the percentage of time compared with free-flow travel time a traveler should allow to make sure he or she arrives on time by considering the worst conditions. Figure 5 depicts the hourly TTI and PTI measures based on both Bluetooth and INRIX data. Values for both measures begin to increase at 3:00 pm and retract to normal after 7:00 pm. The SHA Mobility Report defines congestion when TTI value is higher than 1.15. According to this criterion, Figure 4 shows highest level of congestion for the afternoon peak hour between 5:00 pm and 6:00 pm. In addition, it can be seen that both INRIX and Bluetooth TTI and PTI values are very similar during the non-congested periods. The differences begin to increase as congestion begins and reaches the highest at 6:00pm then gradually fade away



(a)

(b)

Figure 4. Plot of INRIX and Bluetooth Hourly Based Travel Time Index (a) and Planning Time Index (b) for the Entire Year

The reason behind the differences during the peak hour is that HOV regulations are in effect on this segment during afternoon peak period. The northbound HOV lane operates from 3:30 p.m. to 6:30 p.m, and HOV users enjoy major travel time savings compared to the drivers on general purpose lanes. INRIX only provides one value for each time point and does not distinguish between lane types at least on this segment. However since individual travel time samples are provided by Bluetooth detectors, the HOV impact can be captured by this data as shown in Figure 6. Although eventually all Bluetooth individual samples are aggregated to provide a single travel time value, they can reveal underlying travel time patterns resulting from non-homogenous lane operations. In sum, it is clear that aggregating travel time data from HOV and general purpose lanes would lead to questionable reliability measures, as travel time on HOV and general purpose lanes can be different during HOV lane operation hours.

Figure 6 shows a sample plot consisting of Bluetooth travel time observations, average Bluetooth travel times and INRIX travel time on Wednesday, May 16, 2012, in the study area a typical weekday. As the graph shows, two separate travel time patterns emerge during the HOV hours (3:30 p.m. - 6:30 p.m.) presented by individual Bluetooth observations. Since there is only one HOV lane, aggregate Bluetooth travel time is close to the travel time of general purpose lanes. However, during the last hour of HOV operations, INRIX data is biased toward HOV travel times, creating differences as high as 20 minutes between the two sources. So, if Bluetooth is selected as a data source, the impact of HOV lanes on reliability measures can be calculated by separating the trends using pattern recognition algorithms. This is not possible using INRIX.

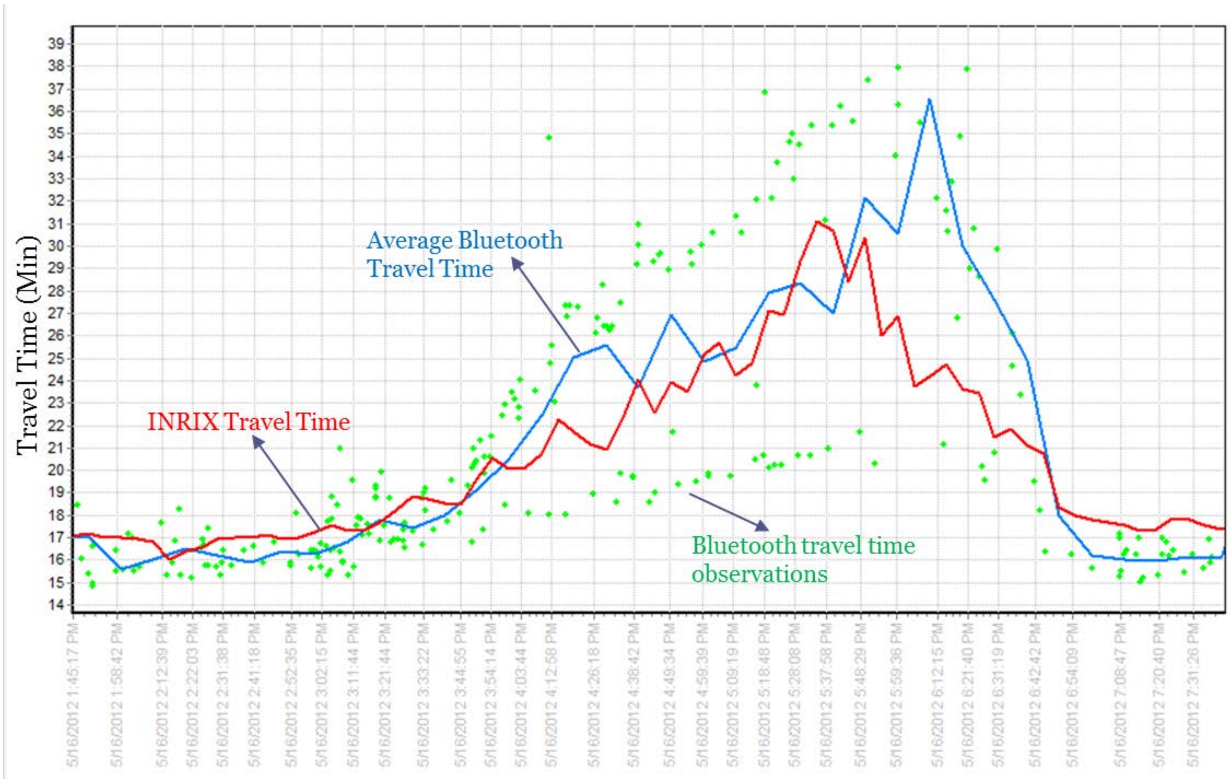


Figure 5. Plot of INRIX and Bluetooth Data for Peak Hours of May 16, 2012

5. Findings and Conclusions

This project investigated the effects of data sources on travel time reliability measures by utilizing two independent sources, Bluetooth and INRIX, to derive different performance measures. By applying the Welch's t statistical method, the statistical differences between reliability performances measures derived from two sources were tested. The experiment is performed on two freeway segments I-95 and I-270 in Maryland where I-270 contains one HOV lane. The test results indicate that there is no significant evidence that reliability performance measures derived from the Bluetooth and INRIX data are different from each other on the I-95 segment. However, reliability measures obtained from these sources are statistically significantly different from each other on I-270, due to presence of HOV lanes. Results show that it is necessary to calculate separate performance measures for HOV and general purpose lanes in such facilities, as travel times on HOV and general purpose lanes are different. Because INRIX data on I-270 segment does not differentiate between lanes, the reliability measures obtained from INRIX and Bluetooth are different from each other. In order to calculate the effect of HOV operations on reliability index, Bluetooth data may be used. Authors are developing pattern recognition algorithms to separate the travel time patterns between the HOV and general purpose lanes which is a direction for future research. Another finding is among all the reliability performance measures, standard deviation and performance variation measures are more sensitive to the data source.

6. Recommendations

In summary, different data sources may provide different values for the same travel time reliability index. Thus, it is critical to test the accuracy of the data and select the appropriate data collection methodology for the target application before calculating reliability measures. It appears both Bluetooth and INRIX data are suitable for calculating reliability measures for segments with homogeneous lanes. Although the INRIX data is sufficient for performance reporting, there is a value in maintaining the Bluetooth data for validation purposes. For non-homogeneous lanes, HOV lane for example, it is necessary to calculate reliability measures for each type of lanes separately. To further explore this problem, including arterial segments in future studies will offer critical insights to travel time reliability assessment. The methodology developed in this paper is general and can be extended to other travel time data sources and different segments of interest.

References

[1] D. Schrank, B. Eisele, and T. Lomax, "TTI's 2012 Urban Mobility Report," *Texas A&M Transportation Institute. The Texas A&M University System*, 2012.

[2] A. Chen, Z. Ji, and W. Recker, "Travel time reliability with risk-sensitive travelers," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1783, pp. 27-33, 2002.

[3] J. Van Lint, H. J. van Zuylen, and H. Tu, "Travel time unreliability on freeways: Why measures based on variance tell only half the story," *Transportation Research Part A: Policy and Practice*, vol. 42, pp. 258-277, 2008.

[4] SHRP2. *Reliability* Available: http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/Pages/Reliability_159.aspx

[5] S. Mahapatra, M. Wolniak, K. F. Sadabadi, T. Jacobs, and E. Beckett, "2012 Maryland State Highway Mobility Report," *State Highway Administration; Maryland Department of Transportation*, 2012.

[6] L. Jin and D. S. McLeod, "Comparison of Travel Time Indexes and Other Travel Time Reliability Measures Using Florida Freeway Spot Speed Data," in *Transportation Research Board 92nd Annual Meeting*, 2013.

[7] R. T. Chase Jr, B. M. Williams, and N. M. Rouphail, "Detailed Analysis of Travel Time Reliability Performance Measures from Empirical Data," in *Transportation Research Board 92nd Annual Meeting*, 2013.

[8] C. S. Inc, *Analytical procedures for determining the impacts of reliability mitigation strategies*. Washington, DC: Transportation Research Board, 2013.

[9] S. M. Turner, "Advanced techniques for travel time data collection," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1551, pp. 51-58, 1996.

[10] A. Haghani, M. Hamedi, and K. F. Sadabadi, "I-95 Corridor Coalition Vehicle Probe Project: Validation of Inrix Data July-September 2008," *Transportation Program, Civil Engineering Department, University of Maryland, College Park,* 2009.

[11] FHA. (February). *Travel Time Reliability: Making It There On Time, All the Time*. Available: www.ops.fhwa.dot.gov/publications/tt_reliability/TTR_Report.htm

[12] C. Chen, A. Skabardonis, and P. Varaiya, "Travel-time reliability as a measure of service," *Transportation Data Research,* pp. 74-79, 2003.

[13] F. D. O. Transportation. (2012, 3). *Mobility Performance Measures Definitions*. Available: http://www.dot.state.fl.us/planning/statistics/mobilitymeasures/mpmdfns4.pdf

[14] "Guide to Effective Freeway Performance Measurement: Final Report and Guidebook," 2006.

[15] Future Strategic Highway Research Program (U.S.), National Research Council (U.S.). Transportation Research Board., and Cambridge Systematics. (2003). *Providing a highway system with reliable travel times study 3 : reliability*. Available: http://trb.org/publications/f-shrp/f-shrp%5Fwebdoc%5F3.pdf

[16] D. Watling, Sumalee, A., Connors, R., and C. Balijepalli, "Advancing Methods For Evaluating Network Reliability," University of Leeds,Leeds2004.

[17] National Research Council (U.S.). Transportation Research Board., *HCM 2010 : highway capacity manual*, 5th ed. Washington, D.C.: Transportation Research Board, 2010.

[18] T. Shaw and K. Fisher, *Performance measures of operational effectiveness for highway segments and systems : a synthesis of highway practice*. Washington, D.C: Transport Research Board, 2003.

[19] C. E. Ebeling, *An introduction to reliability and maintainability engineering*. New York: McGraw Hill, 1997.

[20] T. Lomax, D. Schrank, S. Turner, and R. Margiotta, "Selecting travel reliability measures," *Texas Transportation Institute monograph (May 2003),* 2003.

[21] M. A. P. Taylor and Susilawati, "Modelling Travel Time Reliability with the Burr Distribution," *Procedia - Social and Behavioral Sciences,* vol. 54, pp. 75-83, 2012.

[22] INRIX. *Scorecard*. Available: http://www.inrix.com/scorecard/methodology.asp

[23] A. Haghani, M. Hamedi, K. F. 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, 2010.

[24] M. H. Ali Haghani, Kaveh Farokhi Sadabadi. (2010). *Estimation of Travel Times for Multiple TMC Segments*. Available: http://www.i95coalition.org/i95/Portals/0/Public_Files/uploaded/Vehicle-Probe/I-95%20CC-Estimation%20of%20Travel%20Times%20for%20Multiple%20TMC%20Segments%20-%20FINAL2.pdf

[25] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*: John Wiley & Sons, 2007.