UNIVERSITY OF MARYLAND

National Transportation Center

Project ID: NTC2014-SU-R-22

VALIDATION OF TRAVEL TIME RELIABILITY PREDICTION FROM PROBE DATA

Final Report

by

M. Shoaib Samandar (Graduate Research Assistant)
Institute for Transportation Research and Education
North Carolina State University

Billy M. Williams PhD, PE (PI)
Institute for Transportation Research and Education
North Carolina State University

Andy Wagner (Traffic Engineer)
HNTB

for

National Transportation Center at Maryland (NTC@Maryland)
1124 Glenn Martin Hall
University of Maryland
College Park, MD 20742

March, 2018

ACKNOWLEDGEMENTS

This project was funded by the National Transportation Center @ Maryland (NTC@Maryland), one of the five National Centers that were selected in this nationwide competition, by the Office of the Assistant Secretary for Research and Technology (OST-R), U.S. Department of Transportation (US DOT).

DISCLAIMER

The contents of this report reflect the views of the authors, who are solely responsible for the facts and the accuracy of the material and information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation University Transportation Centers Program in the interest of information exchange. The U.S. Government, University of Maryland and North Carolina State University assume no liability for the contents or use thereof. The contents do not necessarily reflect the official views of the U.S. Government, University of Maryland and North Carolina State University. This report does not constitute a standard, specification, or regulation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

SHRP 2 Project L08 – *Incorporation of Travel Time Reliability into the Highway Capacity Manual* – delivered implementable methodologies and tools for prediction of reliability performance on freeways and arterial streets. Although data sets employed in the development of those methodologies were chosen with robustness and generalizability, validation of the methodologies by applying them to multiple locations and conducting detailed assessment of the accuracy and usefulness of the reliability performance predictions were deemed necessary. This project made valuable contributions to this methodological validation by applying the methodologies to three freeways in the State of North Carolina. The central objective of this research was to validate the freeway reliability performance prediction methods developed under SHRP 2 Project L08 for incorporation into the Highway Capacity Manual. The inherent secondary objective was to identify and clearly define any modifications to the freeway reliability performance prediction methods that may be necessary or desirable.

The validation efforts in this project included validation of 1) facility travel time estimates under base-model, 2) facility performance modeling under non-recurring congestion, 3) scenario generation approach for travel time distribution estimation and 4) resulting travel time distributions.

The tool, FREEVAL, developed by the SHRP 2 L08 was capable of estimating travel times and the associated performance measures accurately under the validation of the base-model. For validation of non-recurring congestion on the freeways, three conditions were modeled and validated: presence of an incident, presence of an adverse weather event, and combination of the two. Performance of FREEVAL for modeling the impacts of incident (no adverse weather event present) on travel time reliability of freeway was at best questionable. However, the tool did a reasonable job modeling the impact of adverse weather events and combination of incident and adverse weather events on the travel time reliability and the associated performances of the freeways.

The scenario generation for estimation of travel time distribution was executed considering three types of data environment: data-poor (national default values), data-moderate (statewide values) and data-rich (facility specific values). It was found that the availability of data has an extremely significant impact on the scenario generation. With national default values the scenarios generated by FREEVAL were completely different from ground truth number of events and not representative of the facility. However, FREEVAL generated scenarios got better as national default values were substituted with statewide default values. With facility specific data, the gap between FREEVAL generated scenarios and ground truth number of events dropped significantly. The resulting scenarios were representative of the number of events that truly occurred on the facility.

Travel time index distributions generated by FREEVAL matched INRIX TTI's very well for all sites. However, FREEVAL's distribution were not quite aligned with Bluetooth TTI distributions for both directions of site 2 and one direction of site 3. The main reason for this difference was associated to the presence of weigh stations on both directions of sites 2 and 3. Apparently, FREEVAL has difficulties modeling the impact of weigh station on travel time reliability on these sites. As a result, its distribution does not match Bluetooth's very well. A systematic difference between TTI distributions of FREEVAL - INRIX and FREEVAL - Bluetooth was located on the left tail of the distributions. While INRIX and Bluetooth always had some data points with TTI less than one (translating to speeds greater than free flow speed), FREEVAL never predicted any travel time estimates that would translate to speeds greater than the free flow speed specified in the model, causing the repeating difference at the left tail of the distributions.

In summary, the findings of this project showed that the methodologies and tools developed for prediction of travel time reliability performance by SHRP 2 Project L08, predict freeway travel times and the associated reliability performance measures with sufficient accuracy.

1.0 INTRODUCTION

1.1 PROBLEM STATEMENT AND RESEARCH OBJECTIVES

1.1.1 Problem Statement

SHRP 2 Project L08 - *Incorporation of Travel Time Reliability into the Highway Capacity Manual* – delivered implementable methodologies and tools for predicting reliability performance on freeways and arterial streets. Although data sets used in the development of these methodologies were chosen with robustness and generalizability in mind, there is a pressing need to verify and validate the methods by applying the methods to multiple locations and conducting detailed assessment of the accuracy and usefulness of the reliability performance predictions. The focus for this research is on validation of the freeway reliability performance prediction methods.

1.1.2 Research Objectives

The central objective of this research is to validate the freeway reliability performance prediction methods developed under SHRP 2 L08 for incorporation into the Highway Capacity Manual. An inherent secondary objective is to identify and clearly define any modifications to the freeway reliability performance prediction methods that may be necessary prior to broad-based deployment of the methods.

1.2 RESEARCH APPROACH AND TASKS

The research makes extensive use of available data, both infrastructure based sensor data and probe data provided by traffic information companies such as INRIX. The L08 methods is applied using the data that was available for ongoing reliability monitoring and prediction. Reliability performance predictions was then assessed with follow on data. Temporary Bluetooth data collection was used to gather full route travel time samples to validate travel times derived from constituent route segments.

The L08 methods specifically assess the reliability impacts of inclement weather and traffic incidents. Along with the traffic data, weather and incident data was compiled from available sources. This data will allow congestion-induced reliability degradation to be attributed to the principal source of congestion. Following is the list of tasks that were executed under this project.

Task 1 – Selection of the validation sites

Task 2 – Development of data acquisition plan

Task 3 – Validation of the base-model facility travel time estimates

Task 4 – Validation of facility performance modeling under non-recurring congestion

Task 5 – Validation of the scenario generation approach for travel time distribution estimation

Task 6 – Validation of the resulting travel time distributions

Task 7 – Documentation of the validation tasks and recommendations for model enhancements (if enhancements are warranted)

1.3 BACKGROUND

1.3.1 Reliability and Travel Time Reliability

McGraw-Hill Dictionary of Scientific and Technical Terms define reliability in engineering and statistics applications as “The probability that a component part, equipment, or system will satisfactorily perform its intended function under given circumstances.” and “The amount of credence placed in a result. The precision of a measurement, as measured by the variance of repeated measurements of the same object” respectively (Parker, 2003).

From a systems engineering point of view, freeway facilities (and indeed, practically all other transportation facilities) possess an express need for assessment in terms of reliability to identify areas for improvement and more accurate quantification of performance. This is especially true as funding for freeway improvements becomes strained, particularly in urban environments where geometric growth is difficult or unfeasible. Travel time reliability continues to be an emerging field, as formal definitions for performance measures and overall goals are still being refined and evaluated. Bluetooth probe data (along with INRIX probe data) have been identified as critical for producing the distributions of travel times necessary to generate these performance measures.

Toppen and Wunderlich, in a 2003 report, presented the benefits of accurate reporting of travel time and possible methods of reaching these estimates. While travel time reliability had not yet reached maturity as a field, Toppen and Wunderlich outlined the need for accurate travel time reporting and information dissemination from a user-based perspective in terms of an Advanced Traveler Information System (ATIS). Moreover, value is identified for assessment of both accuracy of travel time and the variability within the traffic stream (Toppen and Wunderlich, 2003).

Sisiopiku and Islam focused on travel time reliability in their study along Interstate 65 in Alabama, specifically focusing on calculation of performance measures and comparison of INRIX data to the SHRP2 L03 modeling procedure. The selected reliability performance measures that were calculated and analyzed were 90th/95th percentile travel time, Buffer Time Index (BTI), and Planning Time Index (PTI). Significant effects from incident events were observed throughout the study. The L03 model was found to match quite well with the field data; however, particular errors were identified with the PTI values as they increased. Nevertheless, particular value was found in the reliability study and the L03 modeling potential (Sisiopiku and Islam, 2012).

Travel time reliability has been defined by a number of previous SHRP projects. As per the F-SHRP Reliability Research Program, travel time reliability is the variation in travel times over time; for example, hour-to-hour variability, or day-to-day variability in travel times (Cambridge Systematics Inc. et al., 2003). SHRP project L03 defines travel time reliability as “the level of consistency in travel conditions over time, [which] is measured by describing the distribution of travel times that occur over a substantial period of time" (Cambridge Systematics et al., 2013). SHRP 2 project L08, "Incorporating travel time reliability into the Highway Capacity Manual" proposed the following broad definition of travel time reliability to be included in the Highway Capacity Manual (Zegeer et al., 2014):

"Travel time reliability aims to quantify the variation of travel time. It is defined using the entire range of travel times for a given trip, for a selected time period (e.g., the p.m. peak hour during weekdays) over a selected horizon (e.g., a year). For the purpose of measuring reliability, a trip can be defined as occurring on a specific segment, facility (combination of multiple consecutive segments), or any subset of the transportation network; or the definition can be broadened to include a traveler's initial origin and final destination. Measuring travel time reliability requires that a sufficient history of travel times be present to track travel time performance. This history is described by the travel time distribution for a given trip."

Therefore, travel time distribution is the backbone of measuring travel time reliability. Upon generation of travel time distribution, performance measures can be obtained to establish reliability level of a segment or a facility. The latest edition of Highway Capacity Manual categorizes these measures into two groups: time-based and index-based reliability performance measures (*Highway Capacity Manual*, 2016). Following table shows the list of these measures.

Table 1: Reliability Performance Measures

No	Reliability Performance Measures	
	Time-Based	**Index-Based**
1	Planning Time	Travel Time Index
2	Buffer Time	Planning Time Index
3	Misery Time	80th Percentile Travel Time Index
4	On-time Percentage	50th Percentile Travel Time Index
5	Percentage of Trips Exceeding a Target Max. Travel Time	Misery Index
6	Standard Deviation	Reliability Rating
7	Semi-Standard Deviation	

1.3.2 Sources of Travel Time Variability

Highway Capacity Manual identifies following as the main sources for travel time variability:

- Demand fluctuation, these can be the variation in hourly, daily, monthly and seasonal demands;
- Severe weather that impacts capacity and free flow speed (e.g., heavy snow or rain, limited visibility);
- Incidents that impacts capacity and free flow speed of drivers (e.g., crashes, debris);
- Work zones that decreases capacity and impacts the demand; and
- Special events producing temporary intense increase in traffic demand (e.g., major sports events and festivals or concerts).

1.3.3 Freeway Travel Time Reliability Prediction Tools Developed under SHRP 2 Project L08

The SHRP 2 L08 framework was developed to work with HCM 2010 methodologies for evaluating freeway facilities. This framework is designed for operations analysis and planning applications where scarce data is available for estimation of current reliability or prediction of the impacts of demand changes, operational improvements, and design concepts on reliability. The framework uses scenarios to predict the impact of operational improvements on reliability or diagnose the causes of unreliable performance. Each scenario is a combination of demand, work zone, adverse weather, incident, and special events. HCM 2010 freeway methodologies are used to calculate the impact of each scenario on the performance of the facility. In addition, adjustment factors to be used with HCM methods and account for the effects of weather and incidents on capacities and speeds were developed by the research team. To obtain the statistics on facilities reliability the results of several scenarios are added up and weighted based on their probability of occurrence. Therefore, this framework is a deterministic approach that obtains the reliability statistics by applying probabilities of each event in a scenario at the end of the process. (Zegeer et al., 2014). Following figure shows the conceptual analysis framework of SHRP 2 Project L08.

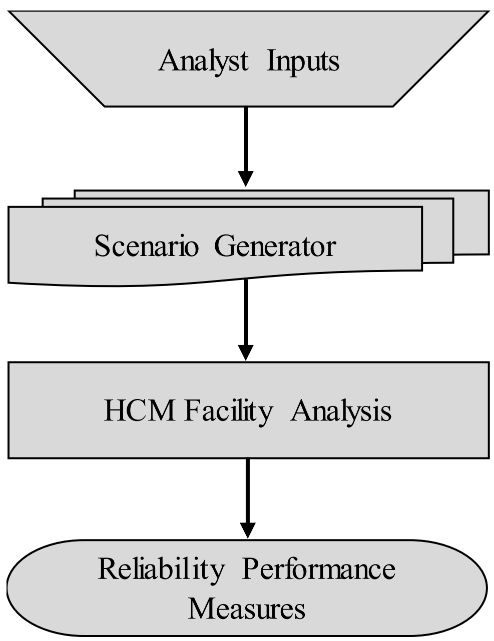

Figure 1: SHRP 2 Project L08 Conceptual Analysis Framework

The freeway facilities methodology consists of three main components: data depository, scenario generator, and core-computational procedure. The data depository houses all the necessary data to execute the methodology. These data in turn are entered into the scenario generator to obtain scenarios to run by the core computational engine in order to estimate travel times. The purpose of the scenario generator is to come up with a set of operational scenarios that a freeway facility may experience during its reliability reporting period along with their associated probabilities. The scenario generator creates the same number of input files for execution in the core computational procedure as there are scenarios to analyze.

FREEVAL is a computerized tool with the goal to execute operational analysis computations for both undersaturated and oversaturated freeway facilities. In addition, FREEVAL is the tool for reliability performance assessment of freeways. It contains all three components of the freeway facilities reliability methodologies. It houses the data depository, is able to generate and analyze reliability scenarios and contains the core computational module (also referred to as the seed file). "In the reliability method, various inputs and adjustment factors are applied to this seed file in scenarios representing impacts of demand variability, weather effects, or incident effects in a whole-year analysis context. The engine processes both the seed file and all reliability scenarios in an automated fashion, and reports summary statistics" (Trask et al., 2015).

1.4 ORGANIZATION OF THIS REPORT

This report is organized into 5 chapters. This chapter introduces the research questions and motivation as well as background information developed from past literature. Chapter 2 provides detail on the selected study areas for this research, and data collection efforts. Chapter 3 outlines data pre-processing and analysis processes for all datasets (Bluetooth, INRIX, weather, incident, work zone, and special events). Chapter 4 presents results generated from the data sources, and Chapter 5 puts forth an analysis of these results as well as commentary on lessons learned from this project.data acquisition, description and study locations

This chapter provides details on study location, reason for selection, the methodology on how data was collected/acquired from different sources, and description of the datasets used in this project.

2.0 SITE SELECTION, DESCRIPTIONS AND VISUALIZATIONS

Tools from the Vehicle Probe Project (vpp.ritis.org) were used to identify interstate candidate sites based on historic levels of congestion and delay. Study sites were selected at the outset of the project to include diverse population density and geographic area type, terrain, and prevailing facility use (e.g. commuter, freight, mixed). The selected study sites listed and detailed below represented what were interpreted to be the most useful locations for observation based on these criteria.

Three study areas from the state of North Carolina were selected to represent different facility trends, prevailing uses, and geographic conditions. Table 2 displays these study site locations along with some basic information on each. Analysis for this research was conducted in both directions for each study site facility.

Table 2: Study Site Locations and Durations

#	Facility	Metro Area	Length	Speed Limit
1	I-540	Raleigh, NC	12.7mi	70mph
2	I-95	Lumberton, NC	14.5mi	65mph
3	I-40	Asheville, NC	13.9mi	60mph

The three selected study sites are shown in the context of the state of North Carolina in Figure 2 on the next page as follows:

- •Site 1, I-540 in Raleigh, is shown in red.
- •Site 2, I-95 in Lumberton, is shown in blue.
- •Site 3, I-40 in Asheville, is shown in green.

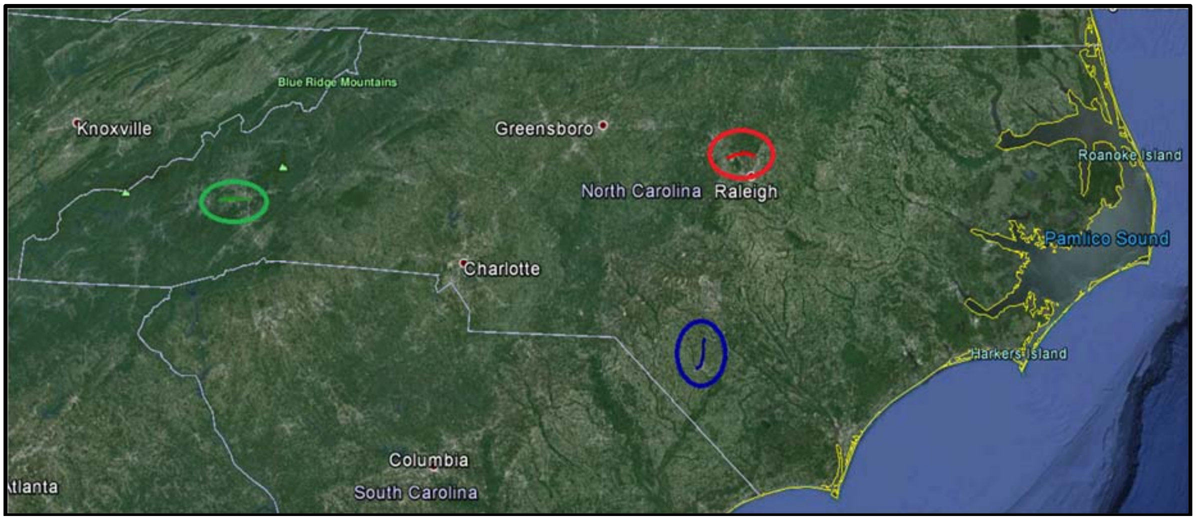

Figure 2: Study Site Locations within North Carolina

Site 1, a 12.7-mile section of Interstate 540 in the northern area of Raleigh, is primarily a commuter route to and from the Research Triangle Park (RTP), located to the west of the study area between Raleigh and Durham. In addition, this section of I-540 runs adjacent to Raleigh-Durham International Airport (RDU), a major source of trips in this area of Raleigh, at its western end. As such, Site 1 provided a study facility located in a large urban environment that showed consistent, reliable congestion from commuter traffic in weekday peaks. This congestion was at times exacerbated by the presence of incidents and weather events. Limited non-peak congestion events/incidents were also observed.

Site 2, a 14.5-mile section of Interstate 95 running through Lumberton and to the north to and from St. Pauls, is part of the major freight and tourism corridor running north-south along the eastern coast of the United States. While local traffic contributed to the traffic stream, the level of through traffic for Site 2 was expected to be higher compared to the other study sites; based on observed Bluetooth trips during the study, this expected higher through proportion was confirmed. This site was advantageous because of the high level of freight truck traffic, providing a useful contrast to Site 1, which consisted of a high percentage of passenger cars. In addition, Site 2 allowed for the inclusion of the effects of a weigh station inside the study area, about 2 miles from the northernmost point. The presence of a weigh station in the study area presented a trio of analytical challenges: first, increased Bluetooth travel times from freight vehicles not necessarily from deteriorating travel conditions; second, the heightened opportunity for discrepancies between Bluetooth and INRIX-derived observations; finally, the possibility of truck queue spillback from the weigh station into the main travel lanes.

Site 3, a 13.9-mile section of Interstate 40 in the southern area of Asheville, represented a mix between local traffic (including commuters) and through traffic (including freight). Asheville, while an urban area, is considerably smaller than Raleigh, offering less commuter traffic leading to less consistent congestion along this study site route. Site 3, like Site 2, contains a weigh station, providing another environment for observing effects from such a facility. Another potential source of delay and complexity for this study site was a work zone on Interstate 26 just south of the study facility; I-26 has a large interchange with I-40 inside of the study site, and that interchange was a primary source of complexity and congestion.

2.1 DATA COLLECTION

This project's data needs can be divided into two different categories. The first category depends to the data needed to predict travel time reliability performance measures through FREEVAL (i.e., data needed to run FREEVAL software) and the second category depends on the data needed to obtain true travel time reliability performance measures. Incident, weather, special event, work zone and facility demand are data needed to run FREEVAL and predict the reliability performance measures. However, to calculate the ground truth reliability performance measures, ground truth travel times should be available to generate travel time distribution and obtain reliability performance measures. Ground truth travel times were collected using Bluetooth devices and acquired through third party data provider (INRIX). Following sub-sections provide detailed description and method of collection of each dataset.

2.1.1 Bluetooth Data Collection

Each study site employed four commercial Bluetooth sensors, with one sensor at each endpoint and two intermittent sensors placed at roughly the one-third and two-third points along the study facilities; Bluetooth sensor spacing ranged between 3 and 6 miles. The Bluetooth sensors used were fifth-generation BlueMAC units, manufactured by Digiwest LLC. Figure 3 on the next page provides a representative setup of the sensor equipment from the Site 2 study in Lumberton (NCSU4).

For Site 1, all units were installed on gantry poles located in the wide median on Interstate 540. The geometry for Site 2 did not allow this, and 2 of the units were installed on infrastructure on the northbound roadside (the other 2 were installed in the median). Site 3 presented even more significant accessibility and geometric challenges, so all 4 devices were mounted on the roadside (1 on the westbound side, 3 on the eastbound side).

Figure 3 shows the basic elements of the Bluetooth sensor deployment used at the study sites. The top unit is the main operating structure, housing one 12V 12Ah battery, a GSM mobile radio, the Bluetooth radio and antenna proper, and a solar charging assembly. The solar charging panel is mounted on the front of this top unit and is adjusted in the field for maximum efficiency. The bottom unit is the supplementary battery pack, housing two additional 12V 12Ah batteries. The battery pack and the solar panel are connected to the top unit via umbilicals. Both units are mounted to an existing pole or other available infrastructure using metal brackets and straps as shown. Supplementary battery packs were not used for most of the Site 1 study, but were part of the sensor installations for the studies at Sites 2 and 3.

Figure 3: Example Bluetooth Sensor Installation

Figures 4, 5 and 6 show the sensor locations along each route along with a more detailed view of the study sites and the surrounding roads.

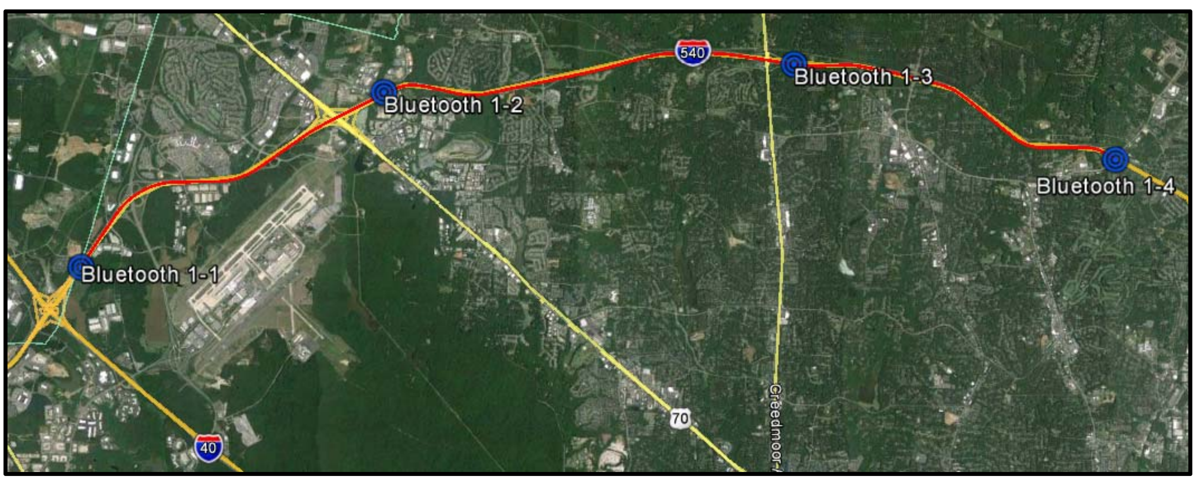

Figure 4: Site 1 (I-540, Raleigh) Bluetooth Sensor Locations

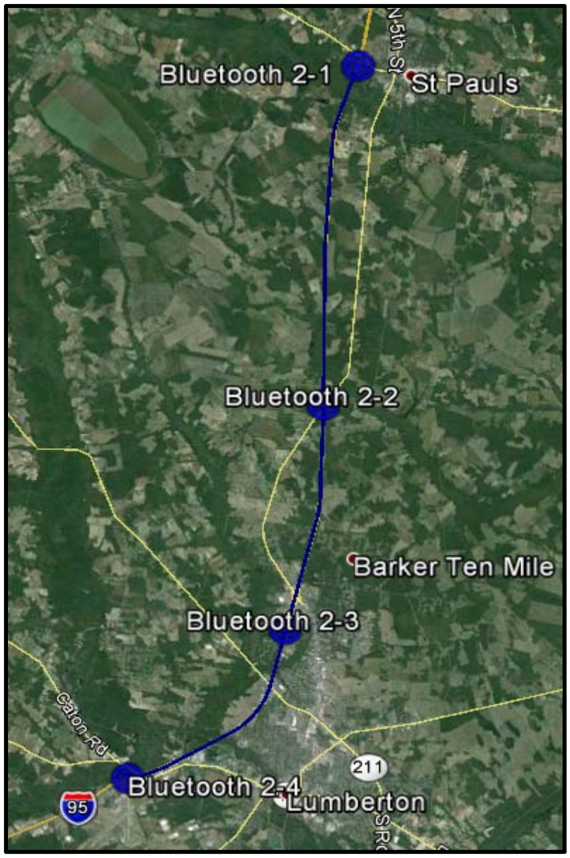

Figure 5: Site 2 (I-95, Lumberton) Bluetooth Sensor Locations

Figure 6: Site 3 (I-40, Asheville) Bluetooth Sensor Locations

The study sites were set up as close as possible to endpoints of INRIX TMC segments to minimize the need for adjustments when comparing Bluetooth and INRIX-derived travel times. Nevertheless, necessary infrastructure was not available to completely avoid such adjustments. It should be noted that for the duration of each study, all Bluetooth sensors operated in the 20dB (Class 1) mode. Following table provides a temporal perspective of Bluetooth sensor deployment at each study site.

Table 3: Bluetooth Sensor Deployment Temporal Information

#	Facility	Metro Area	Bluetooth Study Start/End Dates	Duration
1	I-540	Raleigh, NC	2014-01-09 to 2015-06-10	5 mo.
2	I-95	Lumberton, NC	2015-06-15 to 2015-09-01	2.5 mo.
3	I-40	Asheville, NC	2015-10-21 to 2016-02-21	4 mo.

2.1.2 INRIX Data Acquisition

INRIX probe data is a popular tool for agencies both public and private for speed and travel ime assessments because of its wide reach and non-invasive nature as probe data. INRIX uses a variety of data sources in its synthesis process leading to output speed and travel time data, including GPS, cellular networks, road sensors, and traffic cameras (INRIX, 2016). INRIX travel time data is also used extensively for research purposes in a variety of contexts typically involving major arterials or freeway facilities.

Each study site is composed of several INRIX traffic message channels (TMCs) each with their own speed and travel times. INRIX speed and travel time were downloaded for each study site through Vehicle Probe Project (vpp.ritis.org). The data downloaded for each study site covers the same duration for which Bluetooth data was collected. VPP provides different level of data aggregation. In order to provide consistency with FREEVAL analysis, data were downloaded at a 1-minute aggregation level (lowest aggregation level) and then averaged to 15-minute travel times.

2.1.3 Volume Data Acquisition

The most important set of data needed to run FREEVAL is the mainline demand for the facility. There are two different methods by which mainline demand can be entered into FREEVAL, direct entry (from available volume data) and indirect entry (from annual average daily traffic (AADT)). In the first method, the mainline demand is manually entered for each analysis period. There are several technologies available to obtain traffic volume on roadways (e.g., radar, video cameras, loop detectors). In this project, only volume data was available for site 1. This data was obtained from sidefire radar sensors deployed by Here.com at both directions of I-540. Volume data was not available for site 2 and 3 and AADT were used to estimate the demand on the facilities. AADTs were provided by the Traffic Survey Group of North Carolina Department of Transportation ("Traffic Survey Group, NCDOT," 2018).

2.1.4 Incident and Work Zone Data Acquisition

Incident and work zone data are necessary to run FREEVAL's scenario generator and obtain travel time reliability performance measures. To generate incident frequencies and durations, following characteristics of incidents should be known:

- Incident type;
- Incident severity;
- Incident start and end time; and

- Incident duration

Work zone information that needs to be available for reliability performance measurement all as follow:

- Type
- Severity
- Start and end time
- Duration
- Location
- Area type
- Barrier type
- Work zone speed limit
- Lateral distance between barriers and active traffic lane

Incident and work zone data for all three study sites were downloaded from Traveler Information Management System (TIMS) of NCDOT. The TIMS system gives information to travelers on the types of events that causes delay on the highway system. These events include accidents, work zones, and natural disaster that affect the condition of the road. Information on these events are entered by NCDOT field forces and include necessary information mentioned above under necessary characteristics of incidents and work zones.

2.1.5 Weather Data Acquisition

The last input for scenario generator and reliability analysis part of FREEVAL is the weather data. The inputs include probabilities, duration and adjustment factors for weather events. These weather events are medium rain, heavy rain, light snow, light-medium snow, heavy snow, severe cold, low visibility, very low visibility and minimum visibility. The probabilities of occurrence for each event type should be calculated and entered for the months included in the reliability reporting period. Weather data for each site was downloaded from Weather Underground website (“Weather Underground,” 2018).

To make sure weather data downloaded for each site represents the true weather of the study site, weather data from airport closest to the site was downloaded and used. Raleigh-Durham international airport (KRDU), Lumberton Municipal Airport (KLBT), and Asheville Regional Airport (KAVL) were selected for site 1, 2 and 3 respectively.

3.0 ANALYSIS

Data analysis for this research was conducted in two phases: the processing of the raw data from each source to either prepare it for use as input in FREEVAL or to create travel times and the comparison analysis of travel times predicted by FREEVAL to that obtained from probe data (Bluetooth, and INRIX). Process detailed in the following sections are applicable to all study sites and are for a general case. Only end-to-end trips were considered for travel time data processing. The main reason for this consideration was lack of overlap between Bluetooth internal segments and INRIX TMCs.

3.1 DATA PRE-PROCESSING

Data pre-processing for this project was divided into two parts: the pre-processing of data necessary to obtain true travel times for the study sites and pre-processing of data needed as input in FREEVAL. Bluetooth and INRIX data are the two datasets that were used to obtain true travel times for each of the study sites. Volume, weather, incident, and work zone data are the datasets that are used as inputs into FREEVAL. The first two following sections provide details of Bluetooth and INRIX data pre-processing followed by those needed for as FREEVAL inputs.

3.1.1 Bluetooth Data Pre-processing

For the duration of each study, each Bluetooth sensor unit recorded one line/record of raw data for each detection of a Bluetooth device in its vicinity. This record includes the date and time of the detection, a partial media access control (MAC) address identifying the device, and the received signal strength indication (RSSI). As a device passes by the sensor unit, it is typically detected many times; for a 100-meter Bluetooth detection range, a vehicle traveling at a free-flow speed of 65mph would take about 7 seconds to pass through a sensor's detection zone. All of the individual detections are written to file as separate records.

There are two main steps in distilling the raw Bluetooth device detections into travel time records: converting the individual Bluetooth detections into single-time records and matching those records based on MAC address.
To accomplish both of these steps, the analysis software program BluStats ("Traffax Inc." 2012) was used. Before beginning to use BluStats, however, the raw data from each BlueMAC sensor unit needed to be converted to a proper format for use in BluStats. The BlueMAC units use three fields, comma delimited, for each record (date and time, partial MAC address, and RSSI, as described earlier). BluStats expects five fields, tab delimited, made up of essentially the same information. To convert the BlueMAC raw data into BluStats-friendly format, a Python script was developed.

The first step of binding individual detections into single-time records (known within BluStats as the "station" phase) involves only one parameter: the station *gapout*. Station gapout specifies the amount of time between Bluetooth device detections for a specific device before a single-time record is created and the program "gaps out" to create a new detection for a device. For this research, the default station gapout value was left unchanged at 60 seconds. Theoretically, the station gapout can be adjusted much lower than 60 seconds with little change in results, but the decision was made after brief testing that the default value should be kept for simplicity and convenience in repeated studies.

The second step of deriving the Bluetooth travel times from these device detections is slightly more complex. In BluStats, this step is known as the "segments" phase. For this step, a number of options are presented. The first is the search window, for which a minimum travel time and maximum travel time must be set for each segment. Recall that only the end-to-end travel time segments (i.e. Segment 1-to-4 and Segment 4-to-1) are being considered. Bearing in mind the maximum levels of observed congestion and the length of the study facilities, the minimum time was kept at the default of 0 minutes and the maximum time was increased from the default to a value of 45 minutes. Secondly, the matching algorithm must be specified. At both ends of the study route (upstream and downstream stations), a setting must be defined for what BluStats refers to as the "time tag." There are nine total algorithm options, made up of three options for this setting for each end:

- First, which is simply the time of the first individual device detection (or the first detection after the gapout time has passed for a particular device)
- Middle, the average time of the First and Last options presented here
- Last, which is simply the time of the last individual device detection (or the last detection before the gapout time has passed for a particular device)

Each of these options was tested for each study, and upon observing the results it was determined that the First-First algorithm (using the First setting at both the upstream and downstream stations) was the most effective for accurately deriving Bluetooth travel times, rather than the default algorithm of Middle-Middle or the alternative of Last-Last. The differences between the methods were admittedly subtle, but the selection of the First-First algorithm originated from three characteristics:

- Algorithms with differing approaches for the beginning and ending stations (e.g. Last-First, Middle-Last) were not preferred due to the adjacent segment analysis to be conducted at Site 1.
- The "Middle" setting averaged two different measurements at each station, meaning that two sources of detection error were present as opposed to one (either the first or last hit)
- First-First allowed for slightly better travel time detection in congestion events where queue accumulation extended upstream of the first station; this is suspected to have happened multiple times at Site 1. The delay in the zone upstream of that first station is considered when using First-First, but not with Last-Last.

Finally, BluStats allows for the implementation of a simple statistical filter to screen for "outlier" travel times, which is referred to as IQR4. According to the BluStats manual ("Traffax Inc." 2012), any travel time record that is three or more standard deviations away from the mean of the thirty most adjacent travel time records is flagged as an outlier travel time. As noted later, there are many possible reasons for outlying travel times, from unexpected vehicle departure from the facility and subsequent re-entry to vehicle speed significantly lower than the traffic stream. To approximate the standard deviation, the IQR4 filter uses the inter-quartile range (the difference between the 25th and 75th percentile readings) as an estimate of 0.75 times the standard deviation, which is near what such value would be on a truly normal distribution. This inter-quartile range is then multiplied by four (thus resulting in the name IQR4) to arrive at the buffer value equal to three standard deviations used in screening.

Finally, these matched Bluetooth travel times are written to an output file, one per line, comma delimited, with eight fields: the MAC address, an ID number, the beginning date, the beginning time, the ending date, the ending time, the travel time in minutes, and an "outlier flag" (set to 0 if not an outlier and 1 if an outlier).

3.1.2 INRIX Data Pre-processing

INRIX raw data processing was considerably simpler when compared to that of the Bluetooth sensor stations. Bi-directional data for each route was downloaded from the Vehicle Probe Project (VPP) with a one-minute resolution. In the downloaded data, individual records are given for each TMC segment with the average space mean speed for each minute in the requested dataset. Additionally, a "confidence score" and "c-value" are provided. The "confidence score" is the more important value here, as it provides information concerning the basis of the reported speed in a TMC. The "confidence score" can take three values (13):

- 30: The reported speed is wholly based on real-time probe data
- 20: The reported speed is based partially on real-time probe data and partially on historical speed data
- 10: The reported speed is wholly based on historical speed data

The "c-value" is an additional percentage-based value reporting confidence in those readings with a "confidence score" of 30 – those based entirely on real-time probe data. In general, "confidence scores" of 20 are rare, and values of 10 typically appear in low-demand off-peak hours only.

To generate synthetic travel times from the INRIX speed values, a stitching algorithm macro-enabled spreadsheet was used. The end product of this stitching algorithm is a synthetic travel time through the study route composed of the target TMC segments for a theoretical vehicle beginning travel through the route (in a specified direction) beginning at a specified minute.

This process requires a pre-processing step for the INRIX space mean speed records: using the list of TMCs in one direction, a script using data processing and statistical analysis software program SAS re-formats the raw data (with one line per minute per TMC segment) into a more compact format, providing the date and time in the leftmost column and space mean speed for each TMC segment along one direction of the study route (in travel order) in each subsequent column moving left to right.

Coupling these re-formatted space mean speeds from INRIX with the segment lengths and free flow speeds for the study area, the stitching algorithm calculates travel times for each segment based on the segment lengths and the set of space mean speeds per minute and then adds these together to arrive at the target set of minute-by-minute theoretical travel times.

An important distinction in the INRIX data processing procedures is between the concepts of a "simultaneous" travel time and a truly "stitched" travel time. A simultaneous travel time developed from INRIX data simply involves calculating travel times for each TMC segment based on a single 1-minute interval and adding them together. A stitched travel time is developed using the process described above; this process is significantly more realistic and the theoretical travel times for each study facility for this research were developed using this process.

3.1.3 Incident and Work Zone Data Pre-processing

North Carolina Department of Transportation's Traveler Information Management System (TIMS) provides information on the types of events that most often cause delays on the highway system. This information includes major accidents, construction or maintenance projects, and natural disaster that affect road conditions. This information is usually entered by NCDOT field forces with the goal of having the information be as timely, accurate and helpful as possible. This project used TIMS to download incident and work zone data for use in the scenario generator of FREEVAL.

Data was downloaded for the entire length of the facilities initially. Subsequently, mile markers were used to segregate events that occurred within the boundary of the study area from those that happened outside the boundary of the study area. As mentioned above, TIMS contain all events that cause delay on the highway system, however, this project used Highway Capacity Manual's definition of incidents and removed data points that did not comply with the definition of Highway Capacity Manual. As an example, TIMS logs recurring congestion as an incident in their system. However, recurring congestion is not considered an incident in the Highway Capacity Manual and thus needed to be removed from the analysis dataset.

While most of the incidents recorded by TIMS were representative of the true incidents on the facilities, some events were outliers and did not make sense in the context of this project. These events were either extremely long in duration (hundreds of days) or had negative duration. These events were considered outliers and removed from the analysis dataset.

In order to run FREEVAL's reliability analysis two distributions pertaining to incidents need to be entered into the software. These two distributions are temporal distribution of incident event frequency and distribution of incident duration given severity. Following tables show the incident duration given severity distributions for each direction of each study site. The unit of mean, standard deviation, maximum and minimum durations in the tables below is minute.

Table 4: Distribution of Incidents Given Severity for Eastbound of Site 1

Site 1 (I-540) - Eastbound					
Incident Severity	**Distribution %**	**Mean Duration**	**Std. Dev**	**Min. Duration**	**Max. Duration**
Shoulder Closure	64.4	47.6	29.4	6.3	122.4
1 Lane Closure	29.8	40.0	27.6	6.8	118.9
2 Lane Closure	4.8	17.5	11.4	6.8	36.2
3 Lane Closure	1.0	70.1	0.0	70.1	70.1
4+ Lane Closure	-	-	-	-	-

Table 5: Distribution of Incidents Given Severity for Westbound of Site 1

Site 1 (I-540) - Westbound					
Incident Severity	**Distribution %**	**Mean Duration**	**Std. Dev**	**Min. Duration**	**Max. Duration**
Shoulder Closure	67.7	54.1	36.7	8.0	176.8
1 Lane Closure	27.4	58.4	79.0	5.6	476.4
2 Lane Closure	4.0	128.1	196.2	5.6	476.4
3 Lane Closure	0.8	107.5	0.0	107.5	107.5
4+ Lane Closure	-	-	-	-	-

Table 6: Distribution of Incidents Given Severity for Northbound of Site 2

Site 2 (I-95) - Northbound					
Incident Severity	**Distribution %**	**Mean Duration**	**Std. Dev**	**Min. Duration**	**Max. Duration**
Shoulder Closure	0.0	0.0	0.0	0.0	0.0
1 Lane Closure	100.0	422.1	517.3	15.8	720.0
2 Lane Closure	0.0	0.0	0.0	0.0	0.0
3 Lane Closure	0.0	0.0	0.0	0.0	0.0
4+ Lane Closure	0.0	0.0	0.0	-	-

Table 7: Distribution of Incidents Given Severity for Southbound of Site 2

Site 2 (I-95) - Southbound					
Incident Severity	**Distribution %**	**Mean Duration**	**Std. Dev**	**Min. Duration**	**Max. Duration**
Shoulder Closure	14.3	56.0	0.0	56.0	56.0
1 Lane Closure	85.7	109.7	61.2	43.9	217.3
2 Lane Closure	0.0	0.0	0.0	0.0	0.0
3 Lane Closure	0.0	0.0	0.0	0.0	0.0
4+ Lane Closure	0.0	0.0	0.0	-	-

Table 8: Distribution of Incidents Given Severity for Eastbound of Site 3

Site 3 (I-40) - Eastbound					
Incident Severity	**Distribution %**	**Mean Duration**	**Std. Dev**	**Min. Duration**	**Max. Duration**
Shoulder Closure	15.4	47.1	19.2	19.4	60.0
1 Lane Closure	50.0	56.2	39.1	8.9	149.3
2 Lane Closure	23.1	34.4	20.6	9.6	66.7
3 Lane Closure	11.5	79.2	95.4	15.9	189.0
4+ Lane Closure	-	-	-	-	-

Table 9: Distribution of Incidents Given Severity for Westbound of Site 3

Site 3 (I-40) - Westbound					
Incident Severity	**Distribution %**	**Mean Duration**	**Std. Dev**	**Min. Duration**	**Max. Duration**
Shoulder Closure	23.3	46.9	16.8	16.5	62.5
1 Lane Closure	53.3	61.0	68.7	8.7	292.6
2 Lane Closure	23.3	73.4	39.4	23.8	148.0
3 Lane Closure	-	-	-	-	-
4+ Lane Closure	-	-	-	-	-

Work zone data was also downloaded from TIMS for the entire facility. Consequently, events located outside the study area on the facility were excluded from the analysis dataset and process similar to incident outlier detection was taken to exclude outliers from the work zone analysis dataset. Work zone events need to be entered individually in FREEVAL with information on their temporal (start date, end date, start and end analysis periods), spatial (start segment and end segment) and physical (severity, area type, barrier type, work zone speed limit, and lateral distance) characteristics.

3.1.4 Weather Data Pre-processing

The ideal weather data set would include 15-minute weather records collected near the facility of interest for at least one full year. However, 15-minute weather data was not available for the study sites. As a result, hourly weather reports by Weather Underground were used to obtain weather probabilities. Weather Underground's historical hourly weather reports can be downloaded free of charge from their website. These datasets rely on meteorological aviation reports to archive all the necessary matrices for each airport in the United States. These reports were also used by the SHRP 2 L08 project freeway methodology to develop a 10-year averages of weather occurrence probabilities for 101 metropolitan areas in the United States.

FREEVAL software contains these databases which can be an extremely handful tool for data-poor agencies that do not want to compile and analyze weather reports for their projects.

FREEVAL-RL allows analyst to custom input the weather probabilities into the software. For this purpose, weather data must be collected and classified into one of the following weather types:

- Medium rain;
- Heavy rain;
- Light snow;
- Light-medium snow;
- Medium-heavy snow;
- Severe cold;
- Low visibility;
- Very low visibility;
- Minimal visibility; or
- Nonsevere weather.

Classifying weather reports into these categories is a time-consuming process, but it can be done easily with the use of spreadsheet. To conduct this stage of data acquisition, the same tools used to develop the weather averages for SHRP 2 L08 were employed. First, a python script is used to download the daily weather data from Weather Underground servers for the three airports for each day of 2015 and store it in a folder. Second, another python script is run to create a single csv file for each airport will readings. After the second stage, data is ready to be entered into the Excel workbook. Third, a developed Excel workbook calculates average storm durations, frequencies and each weather type probability. In 1 year, there should be 8,760 (365*24) hourly observations. The probability of occurrence of a weather type is simply the ratio of the number of hourly observations of that weather type to 8,760. Following tables show each weather type's probability and duration (in minute) for the three study sites for the year 2015.

Table 10: Weather Probability and Duration for Site 1

	Med Rain	Heavy Rain	Ligh Snow	LM Snow	MH Snow	Heavy Snow	Severe Cold	Low Vis	Very Low Vis	Min Vis	Normal Weather
January	0.9	0.5	0.6	0.0	0.0	0.0	0.0	1.2	0.0	3.1	93.7
February	0.5	0.2	3.0	0.5	1.2	0.1	0.0	0.7	0.0	0.3	93.5
March	0.6	0.6	0.0	0.0	0.0	0.0	0.0	1.4	0.0	3.2	94.2
April	0.7	0.6	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0	97.3
May	0.6	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.1
June	1.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.2
July	0.3	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.2
August	0.2	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.0
September	0.9	0.8	0.0	0.0	0.0	0.0	0.0	0.5	0.0	1.8	96.0
October	1.1	0.4	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	98.4
November	2.4	0.9	0.0	0.0	0.0	0.0	0.0	0.8	0.0	2.2	93.7
December	1.7	1.1	0.0	0.0	0.0	0.0	0.0	2.0	0.0	4.9	90.3
Avg Dur (min)	37.5	35.3	110.8	26.7	35.2	7.5	0	67.9	0	244.6	

Table 11: Weather Probability and Duration for Site 2

	Med Rain	Heavy Rain	Ligh Snow	LM Snow	MH Snow	Heavy Snow	Severe Cold	Low Vis	Very Low Vis	Min Vis	Normal Weather
January	0.8	1.2	0.1	0.0	0.0	0.0	0.0	0.4	0.0	1.7	95.8
February	1.9	1.8	0.7	0.1	0.1	0.0	0.0	0.5	0.0	1.1	93.8
March	1.0	1.2	0.0	0.0	0.0	0.0	0.0	0.5	0.0	1.6	95.6
April	0.3	0.5	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.2	98.8
May	0.3	0.3	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	99.2
June	0.6	1.3	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.3	97.2
July	0.3	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	99.3
August	0.5	1.2	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.3	97.9
September	0.9	0.6	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.2	98.2
October	1.7	2.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	1.0	94.9
November	0.8	1.4	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.1	96.8
December	1.4	1.8	0.0	0.0	0.0	0.0	0.0	3.5	0.0	3.0	90.3
Avg Dur (min)	23.7	29.2	44.9	14.3	15	12	0	29.9	0	80.2	

Table 12: Weather Probability and Duration for Site 3

	Med Rain	Heavy Rain	Ligh Snow	LM Snow	MH Snow	Heavy Snow	Severe Cold	Low Vis	Very Low Vis	Min Vis	Normal Weather
January	1.5	0.4	0.4	0.0	0.0	0.0	0.0	1.0	0.0	0.1	96.5
February	0.7	0.1	3.5	0.5	1.1	0.1	0.0	0.1	0.0	0.1	93.8
March	0.5	0.2	0.5	0.0	0.1	0.0	0.0	0.4	0.0	0.3	98.1
April	1.5	0.8	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	97.4
May	0.1	0.4	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.3	98.2
June	0.8	0.7	0.0	0.0	0.0	0.0	0.0	1.3	0.0	0.6	96.6
July	0.4	0.5	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.1	98.7
August	0.2	0.6	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.2	98.7
September	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.0
October	2.9	1.5	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.2	93.4
November	2.8	1.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	95.6
December	3.1	1.8	0.0	0.0	0.0	0.0	0.0	1.3	0.0	1.5	92.3
Avg Dur (min)	45	15.9	56.5	16.1	32.9	3.8	0	22.5	0	15	

3.2 ANALYSIS FRAMEWORK

To verify and validate the freeway reliability performance prediction methods developed under SHRP 2 L08 and to identify and clearly define any modifications to the freeway reliability performance prediction methods that may be necessary prior to broad-based deployment of the methods, following aspects of travel time reliability prediction needed to be validated:

- Base-model facility travel time estimates;
- Facility performance modeling under non-recurring congestion;
- Scenario generation approach for travel time distribution estimation; and
- Travel time distributions.

Further detail on the framework and analysis method of above mentioned tasks are provided below. First each task is explained in detail as to what it means and why it is important to be validated. Second, the methodology and framework for validation is discussed. The first two tasks outlined above have periods equal to or less than 24 hours. As a result, the travel time performance measures and travel time index distribution are very sensitive to model inputs. Thus, these two tasks are validated only for site 1 where detailed model inputs (such as ramps volumes and mainline volumes) are available.

3.2.1 Validation of the base-model facility travel time estimates

For the purpose of this project base-model facility means the normal conditions on the freeway (i.e. absence of all the factors that impact travel time reliability – incidents, adverse weather, work zone, recurring and non-recurring congestions). To validate the base-model facility travel time estimates following steps were followed:

a) Periods that represented base-model facility had to be determined for each study site. To do this a number of sources were consulted to determine which period could represent base-model facility. First, the congestion scan feature of INRIX was used to see if the candidate period was congested or not. Second, data downloaded from TIMS was looked up for any incidents for the period. Third, Weather Underground data was checked to make sure there was no adverse weather event present for the period. If a period conforms to all the requirements, it can be used for this task. These periods can span anywhere between 1-24 hours.

b) Consequently, these periods are modeled and run in FREEVAL-RL.
c) 15-minute INRIX data was downloaded for the same periods and processed to obtain travel times. Similarly, Bluetooth data for the same periods was acquired and processed to obtain travel times for the periods.
d) Travel time index distribution and travel time reliability matrices were obtained using FREEVAL's output and INRIX and Bluetooth travel time data.
e) Finally, to validate the base-model facility, travel time index distributions obtained using FREEVAL's output, INRIX and Bluetooth were compared to each other visually. In addition, travel time reliability performance measures obtained from FREEVAL's output were also compared to those obtained from Bluetooth and INRIX data.

This task was executed only for eastbound of site 1 (I-540 at Raleigh). Following figure shows the location of the modeled portion of I-540.

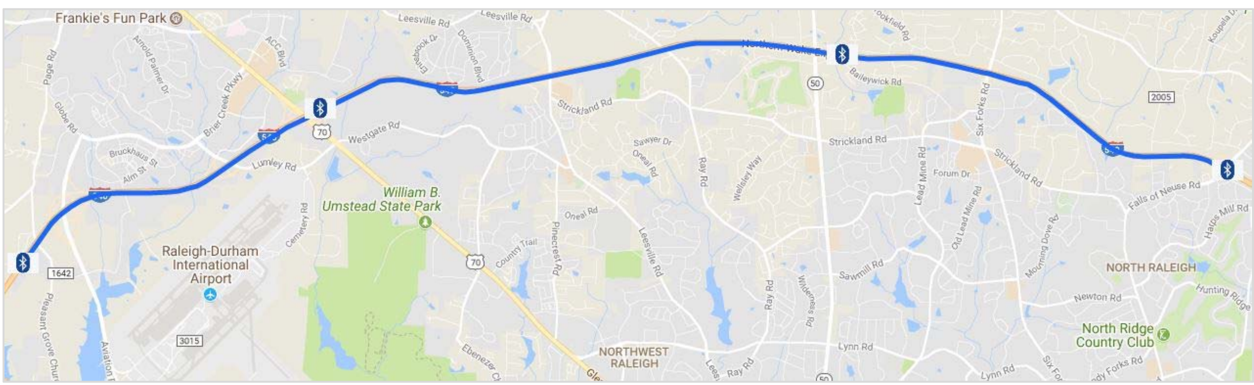

Figure 7: Modeled Section of I-540 in FREEVAL for Validation of Base-Model and Non-Recurring Congestion

3.2.2 Validation of facility performance modeling under non-recurring congestion

Federal Highway Administration (FHWA) defines a non-recurring congestion as a congestion caused by non-recurring causes, such as crashes, disabled vehicles, work zones, adverse weather events, and planned special events. To validate whether FREEVAL is capable of predicting accurate travel times under non-recurring congestion the following framework was developed and tested.

a) TIMS and Weather Underground databases were used to obtain periods with non-recurring congestion events (incident events and adverse weather events, respectively). For the purpose of this task, non-recurring congestion events that were modeled in FREEVAL were boiled down to incidents and adverse weather events on the facility.

b) Following the identification of non-recurring congestion using TIMS and Weather Underground databases, INRIX congestion scan was used to verify that the event was truly a non-recurring and that it impacted the travel times on the facility.
c) Periods where non-recurring congestion happened are then modeled in FREEVAL. A total of three periods were selected for site 1 for this task. The detail of these periods are as follows:
    I. The first period is April 13, 2015 where an accident occurs at 15:35 that continues until 18:22. This accident causes one lane of the facility to close down. The accident occurs on mile marker 5 eastbound on I-540.
    II. The second period is March 19, 2015. This day was selected because there is light rain and light drizzle that start from 12:51 pm and continues till 7 in the morning the next day.
    III. The third day is March 3, 2015. This day includes both an accident and adverse weather event. The accident occurs on mile marker 11 and starts at 18:26 and continues until 19:20. The accident closes one lane of the facility. The weather was overcast until 2 pm and a mix of light drizzle, and light rain started from 2:33 pm until midnight.
d) Bluetooth data for the same periods were obtained and processed for travel times. In addition, fifteen-minute INRIX speeds were downloaded from VPP suite and processed for travel times.
e) Travel time index distribution and travel time reliability matrices were obtained from FREEVAL outputs, INRIX and Bluetooth travel time data.
f) Finally, to validate the facility performance modeling under non-recurring congestion, travel time index distributions obtained using FREEVAL's output, INRIX and Bluetooth were compared to each other visually. In addition, travel time reliability performance measures obtained from FREEVAL were also compared to those obtained from Bluetooth and INRIX data.

3.2.3 Validation of the scenario generation approach for travel time distribution estimation

The freeway scenario generation process is potentially the most complex part of FREEVAL analysis tool. The freeway scenario generation (FSG) generates scenarios from factors that affect travel time variability: traffic demand, weather, and incidents. It converts these factors into an aggregated set of operational conditions on the facility, each with a predetermined probability. The FSG first develops base scenario followed by study period and detailed scenarios. Each base scenario is a combination of events that occur within a given time period (a weekday, or a few hours of day). These base scenario probabilities are provided as the portion of time a specific combination of events occur during the study period of interest. The methodology tries to mix the states of freeway operation to model weather events and incidents more realistically.

The FSG works both in data-rich and data-poor environments. It can also work well in environments with moderate data availability. If extensive data on the facility is available (data-rich facility), the user is asked to input as much local data on the facility as possible. If local data is not available, the tool relies on national default values to generate scenarios. In addition, the

FSG also includes the determination of the required number of events that need to be modeled in all the study periods.

For further detail on freeway scenario generator methodology, readers are advised to refer to SHRP 2 L08 final report (Zegeer et al., 2014). It is worthwhile mentioning that the FSG methodology is completely automated in the new version of FREEVAL. For the purpose of validation of the freeway scenario generation, this project has sufficed on validating the required number of events that need to be modeled in all the study periods. The impact of data availability was also explored on the scenario generation methodology. Three levels of data availability were explored: data-rich, data-poor and data-neutral environments. Following chart shows the workflow for this task.

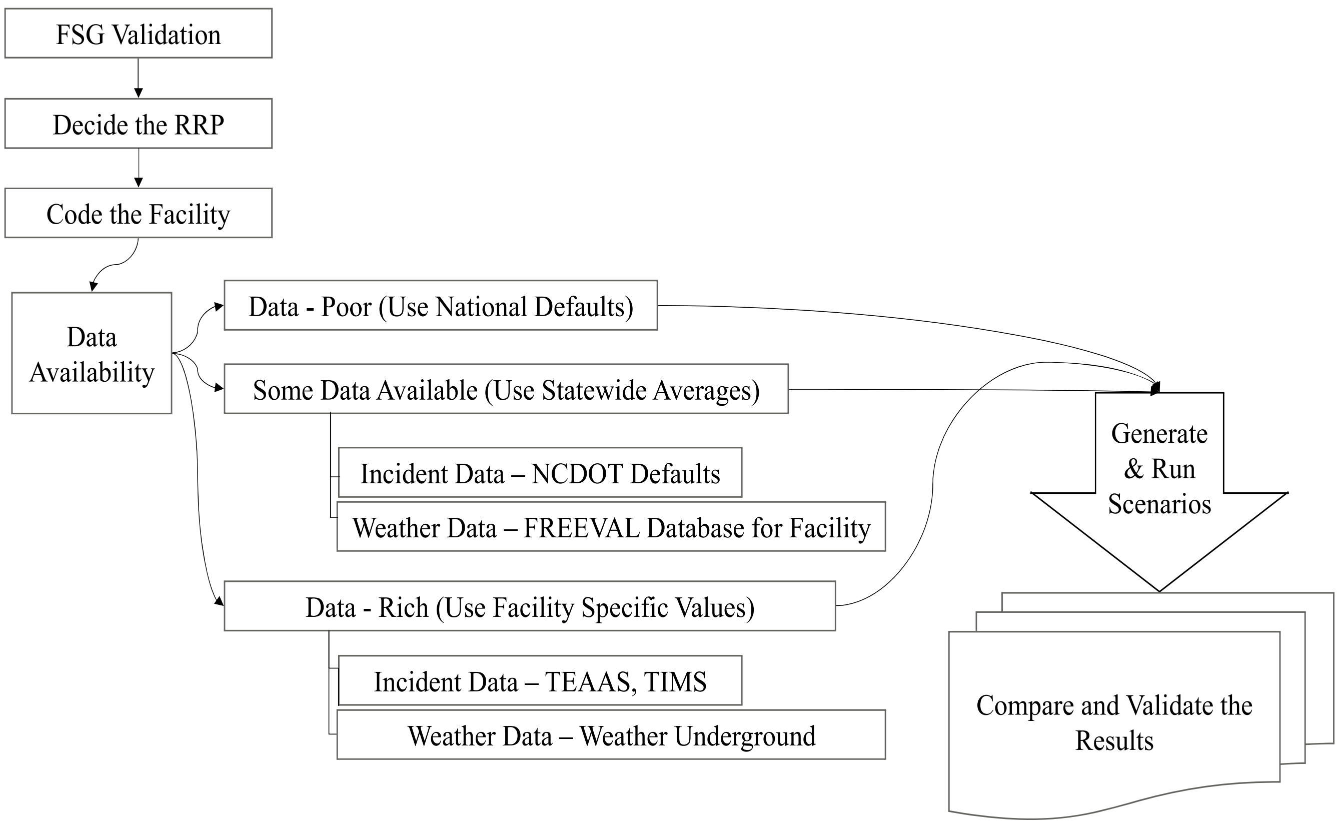

FIGURE 8: Flowchart for Freeway Scenario Generation Validation

3.2.4 Validation of the resulting travel time distributions

Travel time distribution is the starting and backbone of travel time reliability assessment for any facility because from a measurement point of view, reliability is quantified from the distribution of travel times. Usually, one year is considered sufficient to capture almost all of the variability caused by disruptions. Upon generation of travel time distribution, performance measures can be established to capture reliability. These measures can be standard statistical measures, percentile-based measures, on-time measures, and failure measures.

Both SHRP 2 L08 and FREEVAL used travel time index (TTI) as the variable of interest because of the need to normalize travel time. As a result, the fundamental distribution is actually based on the TTI distribution, instead of travel time. For the purpose of this project, travel time index distributions for all three sites were developed from Bluetooth and INRIX data and compared to that given by FREEVAL. The flowchart for this task is provided by the following figure.

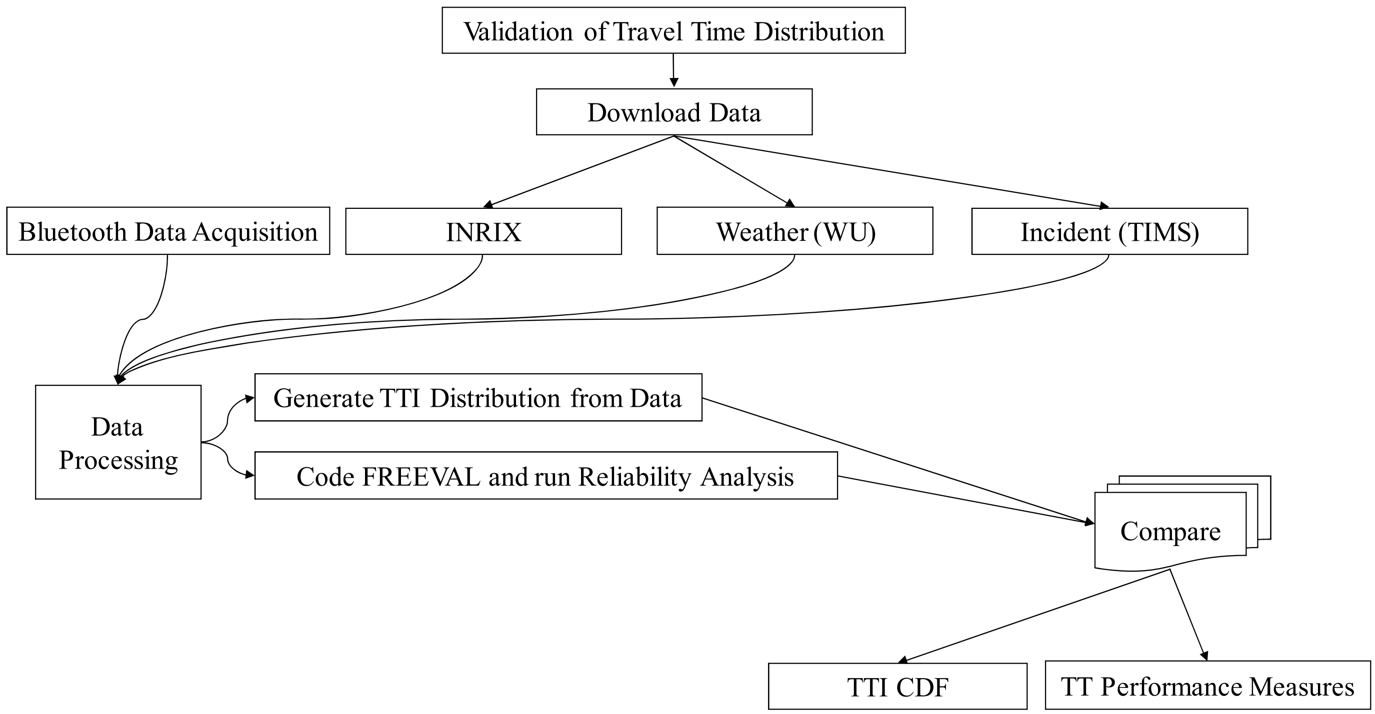

FIGURE 9: Flowchart for Validation of Travel Time Distribution

4.0 STUDY RESULTS

This chapter outlines the analytical findings of the analyses conducted under the methodologies described in chapter 3.

4.1.1 Validation of the base-model facility travel time estimates

As mentioned in earlier sections of this project, a base-model facility refers to a condition of facility where there are no incidents, work zones or adverse weather present. Therefore, the only source of travel time variability in the base-model facility is the demand fluctuation. After careful consideration of the characteristics of base-model facility, consultation of INRIX's congestion scan, Weather Underground dataset and TIMS dataset, January 28, 2015 was selected as the date for which the base-model facility travel time estimates needed to be validated. The INRIX congestion scan for the facility is shown in the following figure. Observation of the figure reveals that the facility is completely uncongested until 4:45pm. However, at 4:45 pm the facility starts to get congested which lasts until around 7 pm. This congestion is not caused because of any incidents, weather or work zone. This is a recurring congestion due to increase in pm peak demand. Similar patterns were observed for other days (without incident, adverse weather and work zone events) at the facility.

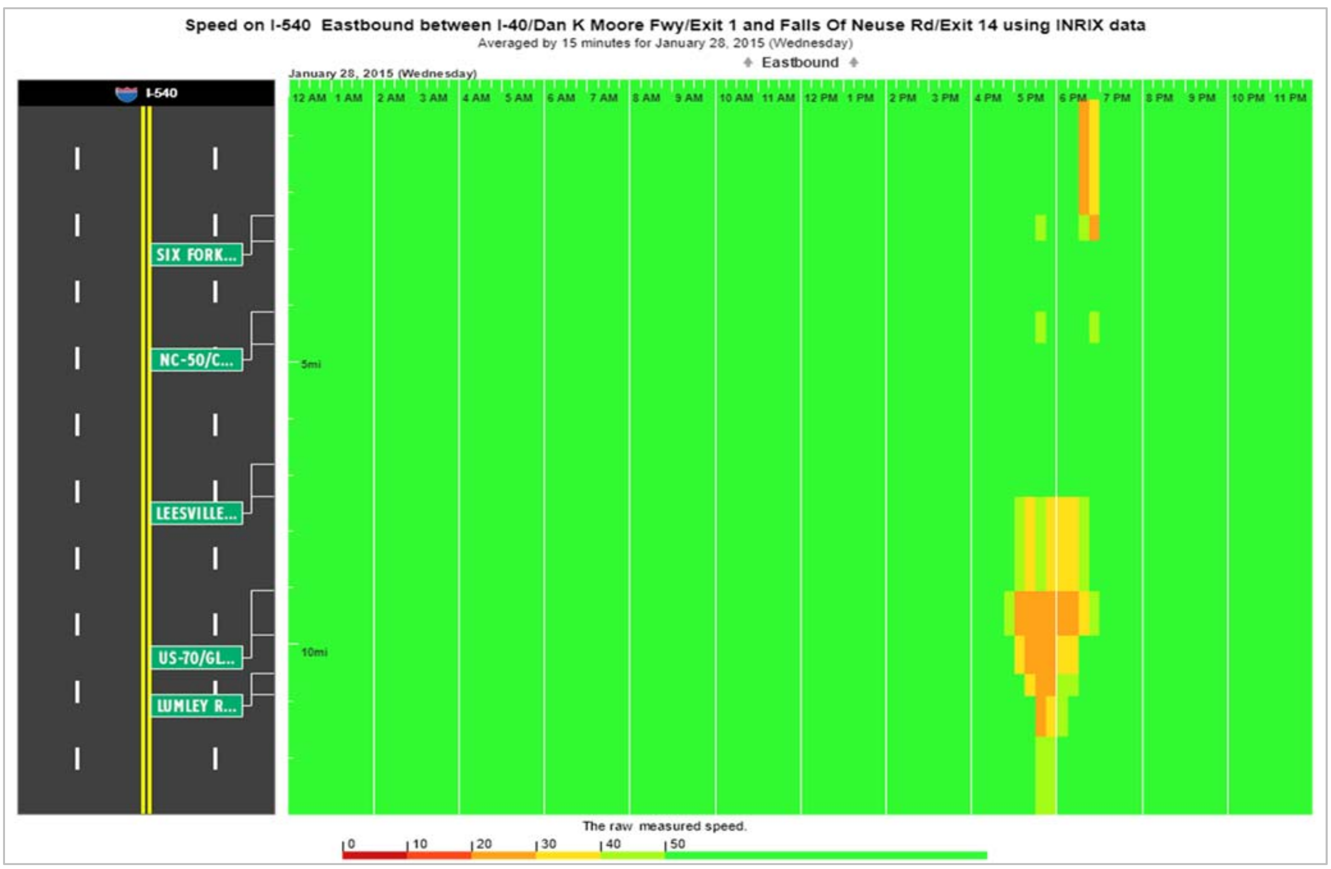

Figure 10: INRIX Congestion Scan for 01/28/2015

Two types of study periods were validated in this part of the project: short and long study periods. Short study periods refer to a study period where the duration is less than an entire day (for example AM or PM peak periods), where long term study period for the purpose of this task refers to an entire day. For the short study period, PM peak was selected for validation where recurring congestion is present on the facility. In addition, the impact of data availability was explored under the curtain of this task to see how much it can impact the final results. As discussed above, in a base-model environment, the only factor that impacts travel time reliability is the fluctuation of demand on the facility. Thus, mainline and on/off ramps demand are the most important input into the FREEVAL computational engine for this task. As a result, the impact of the following scenarios of mainline and on/ramps demand availability were explored.

- Only AADT data available;
- AADT and its profile available (using Bluetooth detection rate for the day);
- Detailed and accurate 15-minute mainline volume data available.

Following figure shows the cumulative distribution function (CDF) or travel time index (TTI) with only AADT data available. Mainline demand was estimated from AADT of the facility. The accuracy of these mainline demands entered in FREEVAL is at best questionable.

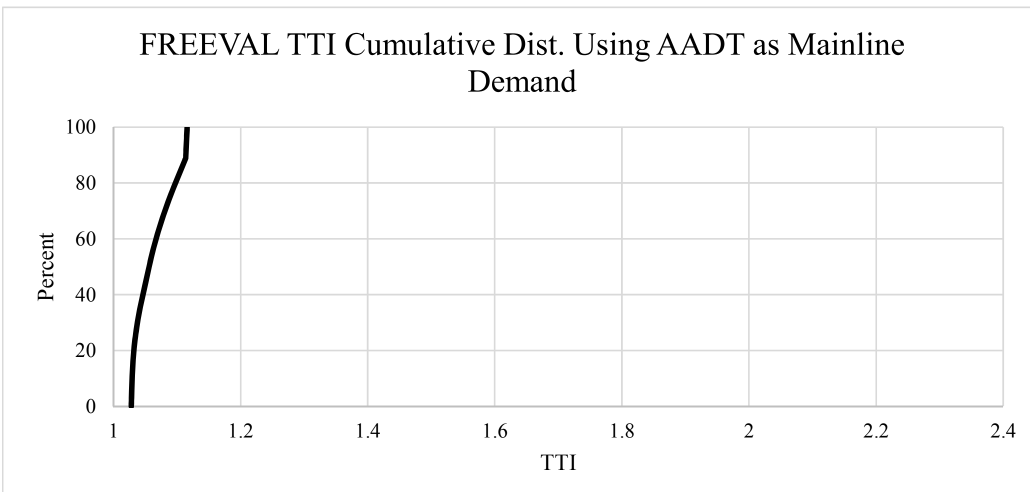

Figure 11: TTI CDF Using AADT for Mainline Demand

The next level, moderate-data availability, uses the detection pattern of Bluetooth sensors at the facility to determine the demand profile for the day and come up with 15-minute mainline demands on the bases of the detection pattern. The demand for each analysis period is calculated proportional to the number of devices detected by the sensor. Following figure shows the detection pattern of Bluetooth sensor NCSU 1 located at the start of eastbound direction of the facility.

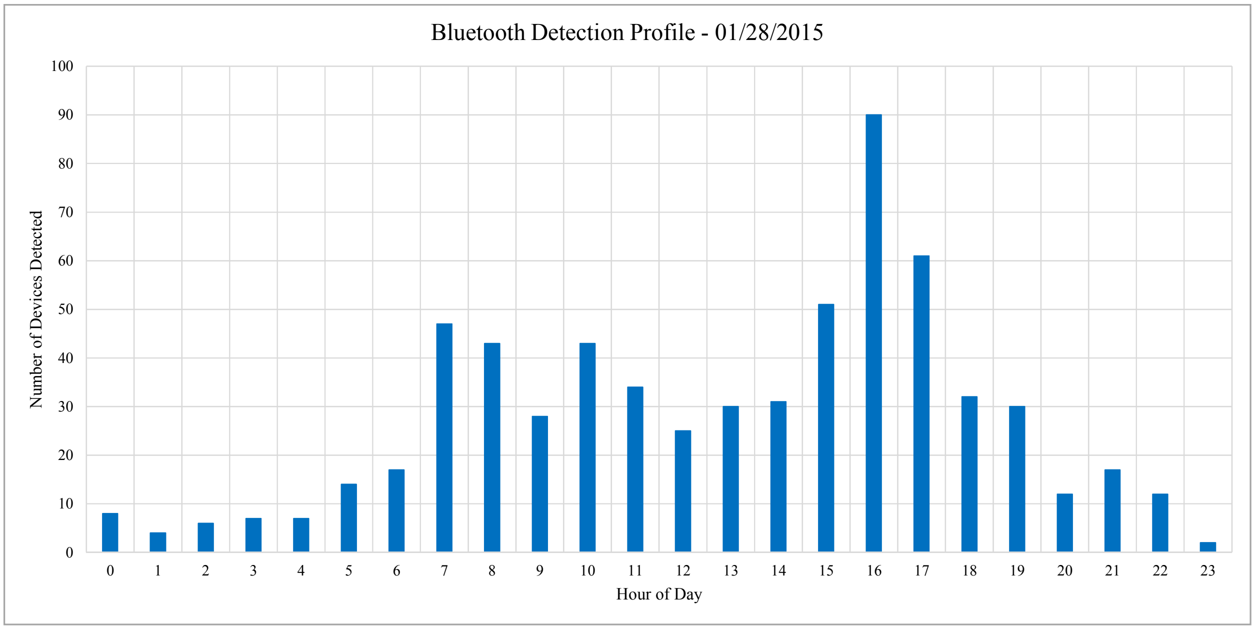

Figure 12: Bluetooth Sensor Device Detection Profile for 01/28/2015

Bluetooth sensor detection profile shown above was used to proportionally adjust the demand on the mainline of the facility. These adjusted demands were entered in FREEVAL, the resulting TTI cumulative distribution is shown in the figure below.

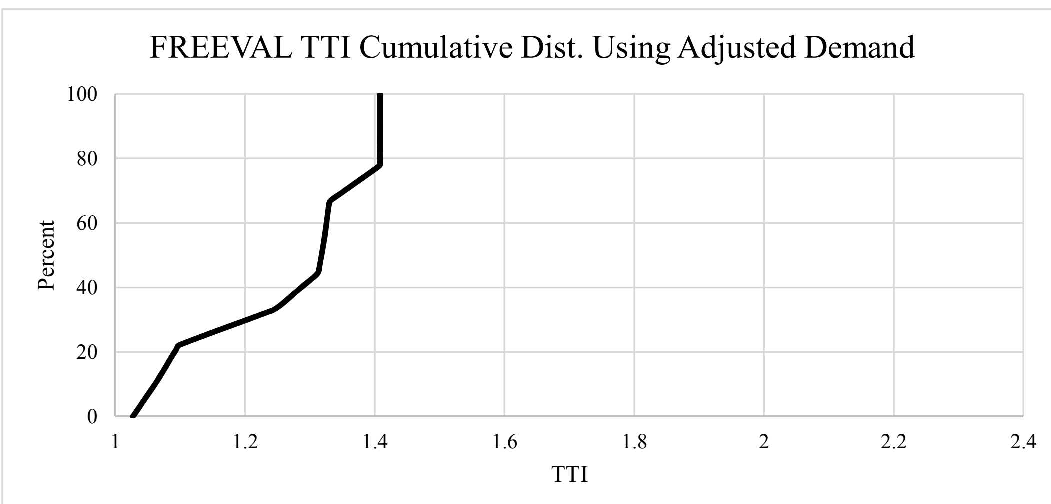

Figure 13: TTI CDF Using Demand Adjusted Based on Bluetooth Sensor Detection Profile

The last level, data-rich environment, uses the real data collected on the facility as the input in FREEVAL model. Volume data collected via Sidefire radar detectors at the facility was used as the mainline demand. As a result, following TTI distribution was generated by FREEVAL.

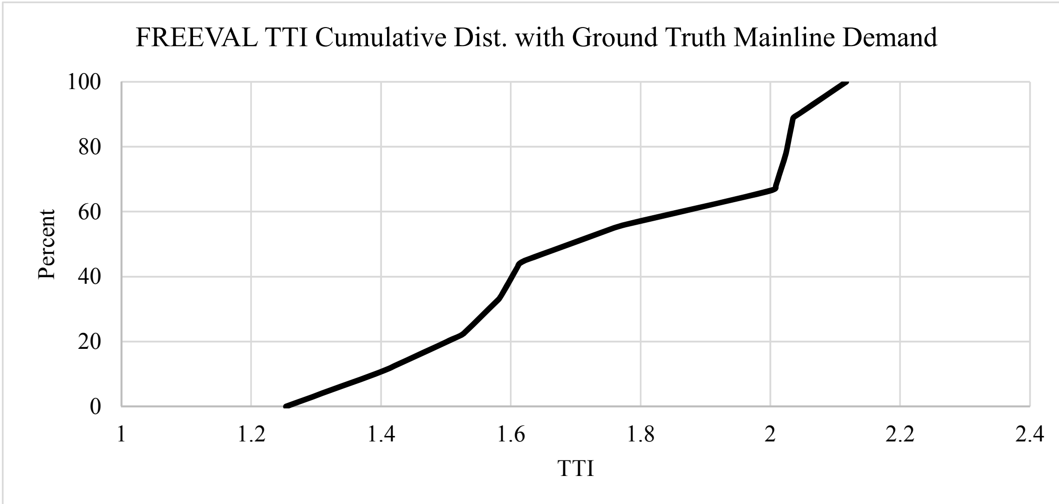

Figure 14: TTI CDF Using True Mainline Demand Collected via Sidefire Radar Detectors

Bluetooth data for the same period was aggregated to 15-minutes and plotted in the following figure.

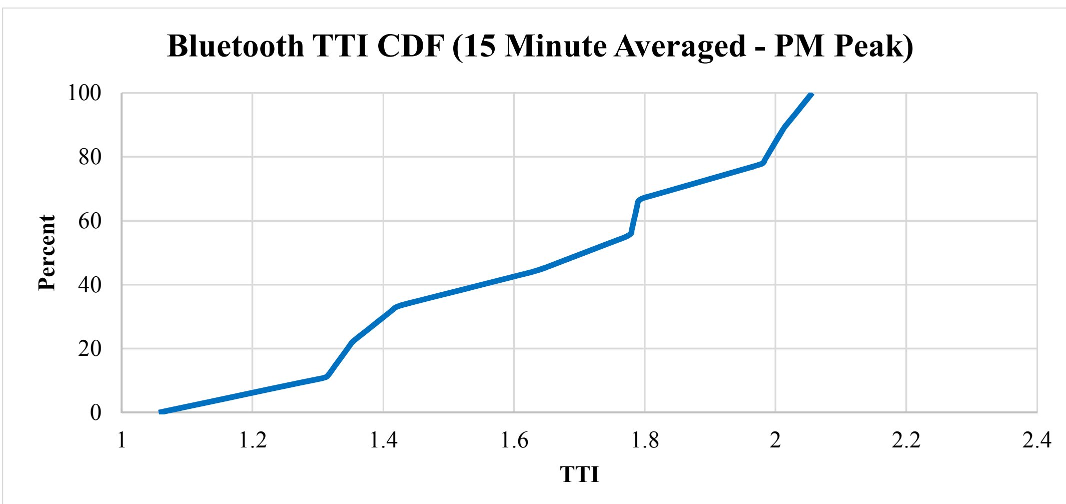

Figure 15: TTI CDF Using Bluetooth 15-Minutes Aggregated Data

Finally, INRIX's TTI distribution for the same period is provided in the following figure followed by the figure that compares these three distributions to each other (data-rich FREEVAL travel time index distribution to those of INRIX and Bluetooth).

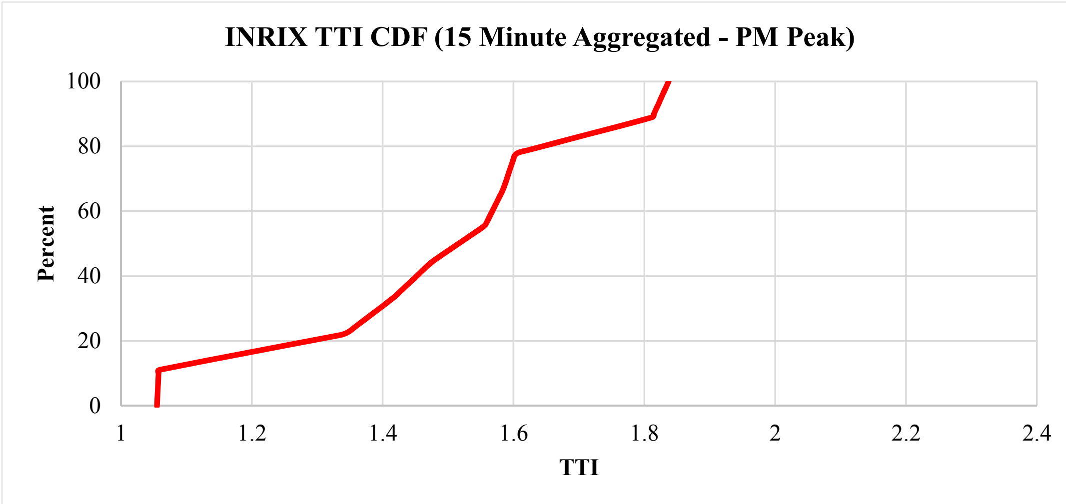

Figure 16: INRIX CDF based on 15-Minutes Aggregated Data

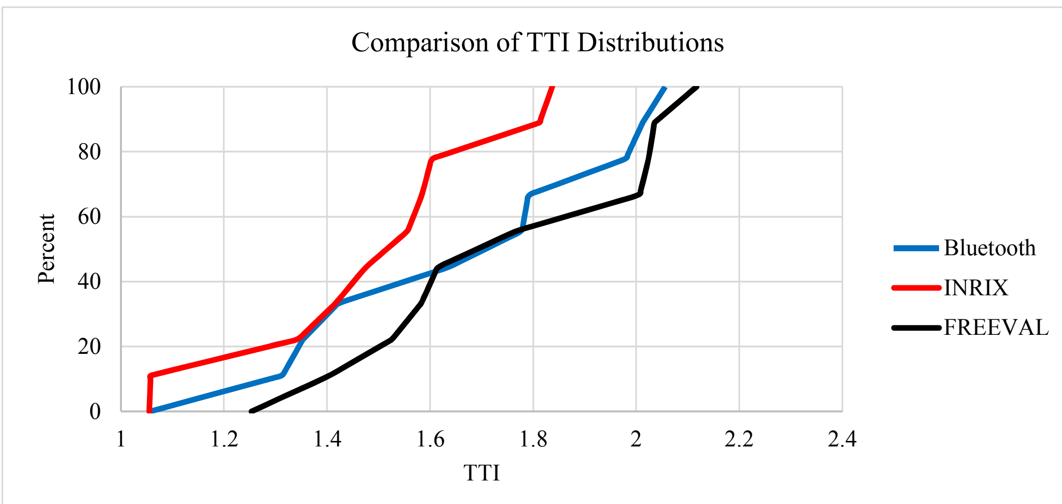

Figure 17: Comparison of FREEVAL to INRIX and Bluetooth Distributions for Eastbound of I-540

The preceding figure shows interesting results. First, both INRIX and Bluetooth distributions have values less than one, indicating speeds greater than the free flow speed (in this case the speed limit of the facility). However, the lowest TTI value that FREEVAL has is greater than 1.2. This shows that FREEVAL speeds do not exceed the free flow speed entered into the software. Second, the distributions of INRIX and Bluetooth are quite similar until about 40 percent, they start to take slightly different directions passing the 40 percent line. Interestingly, the gap between FREEVAL and Bluetooth distributions get lesser after passing the 40^{th} percent line. Following table shows the facility reliability performance measures for the PM peak period.

Table 13: Reliability Performance Measures for PM Peak (1/28/2015) for Eastbound of I-

Reliability Measures	Bluetooth	INRIX	FREEVAL	Abs. Dif. (BT - FVL)	Abs. Dif. (INRIX - FVL)
Mean TTI	1.640	1.474	1.733	0.093	0.259
Misery Index	2.056	1.837	2.117	0.061	0.280
50th % TTI	1.707	1.516	1.691	0.017	0.175
80th % TTI	1.986	1.645	2.026	0.040	0.382
Reliability Rating (%)	80	80	90	10	10
95th % TTI (PTI)	2.036	1.826	2.080	0.044	0.254

The table above shows that the difference in reliability performance measures of Bluetooth and FREEVAL are considerably lower than that of INRIX and FREEVAL. The reason why INRIX is so different compared to Bluetooth and FREEVAL are believed to be: data smoothing, potential capping of high speed data points, lower market penetration rates and removing potential non-outlier data points with low speeds. Based on the comparison of the distribution and reliability measures, it can be concluded that FREEVAL does estimate the base-model facility travel time estimates accurately for short study periods.

The second part of this task validates a long study period with duration of 24 hours. It analyzes the entire day of January 28, 2015. Following figure shows comparison of the TTI distribution for this day.

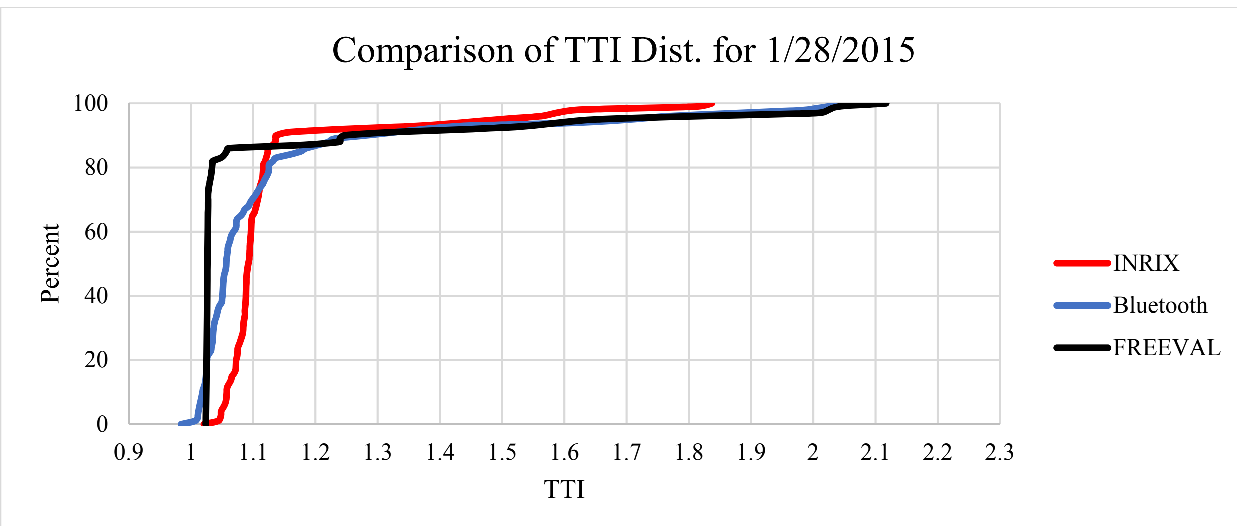

Figure 18: Comparison of TTI Dist. for the Entire Day of 01/28/2015

The figure above shows that majority of FREEVAL travel times correspond to speeds close to the free flow speed on the facility (TTI close to 1), where Bluetooth and INRIX have a limited number of speeds corresponding to the free flow speeds. There is quite a significant difference between the distributions on the left tail. However, the right tail of the distributions tends to get closer above 80 percent line. Bluetooth and FREEVAL overlap each other above the 85th percent line. Following table shows the reliability performance measures and the difference between FREEVAL-Bluetooth and FREEVAL-INRIX performance measures for the entire day of 1/28/2015.

Table 14: Reliability Performance Measures for Entire Day of (1/28/2015) for Eastbound of I-

Reliability Measures	Bluetooth	INRIX	FREEVAL	Abs. Dif. (BT - FVL)	Abs. Dif. (INRIX - FVL)
Mean TTI	1.132	1.042	1.060	0.073	0.018
Misery Index	1.914	1.679	1.990	0.076	0.311
50th % TTI	1.057	1.092	1.026	0.030	0.066
80th % TTI	1.125	1.116	1.034	0.091	0.082
Reliability Rating (%)	91	94	96	5	2
95th % TTI (PTI)	1.706	1.496	1.653	0.054	0.157

The table above shows that the difference between FREEVAL and Bluetooth reliability measures are negligible. Similarly, the difference between the reliability measures of INRIX and FREEVAL are not significant but are larger than those of Bluetooth and FREEVAL.

4.1.2 Validation of facility performance modeling under non-recurring congestion

The first scenario considered under modeling of non-recurring congestion is the presence of only an accident on the modeled section of the facility. After consultation of the TIMS database, Weather Underground database and INRIX's congestion scan tool, April 13, 2015 was selected for the first scenario. On April 13, 2015 an accident occurs at 15:35 that continued until 18:22. This accident caused one lane of the facility to close down. The accident occurred on mile marker 5 eastbound on I-540. Following figure shows the INRIX congestion scan for the day.

The figure shows that the facility is not congested for the entire morning and afternoon until around 15:35 where the accident happens. The accident causes significant drop in travel speed on the facility as low as 10 mph and even lower.

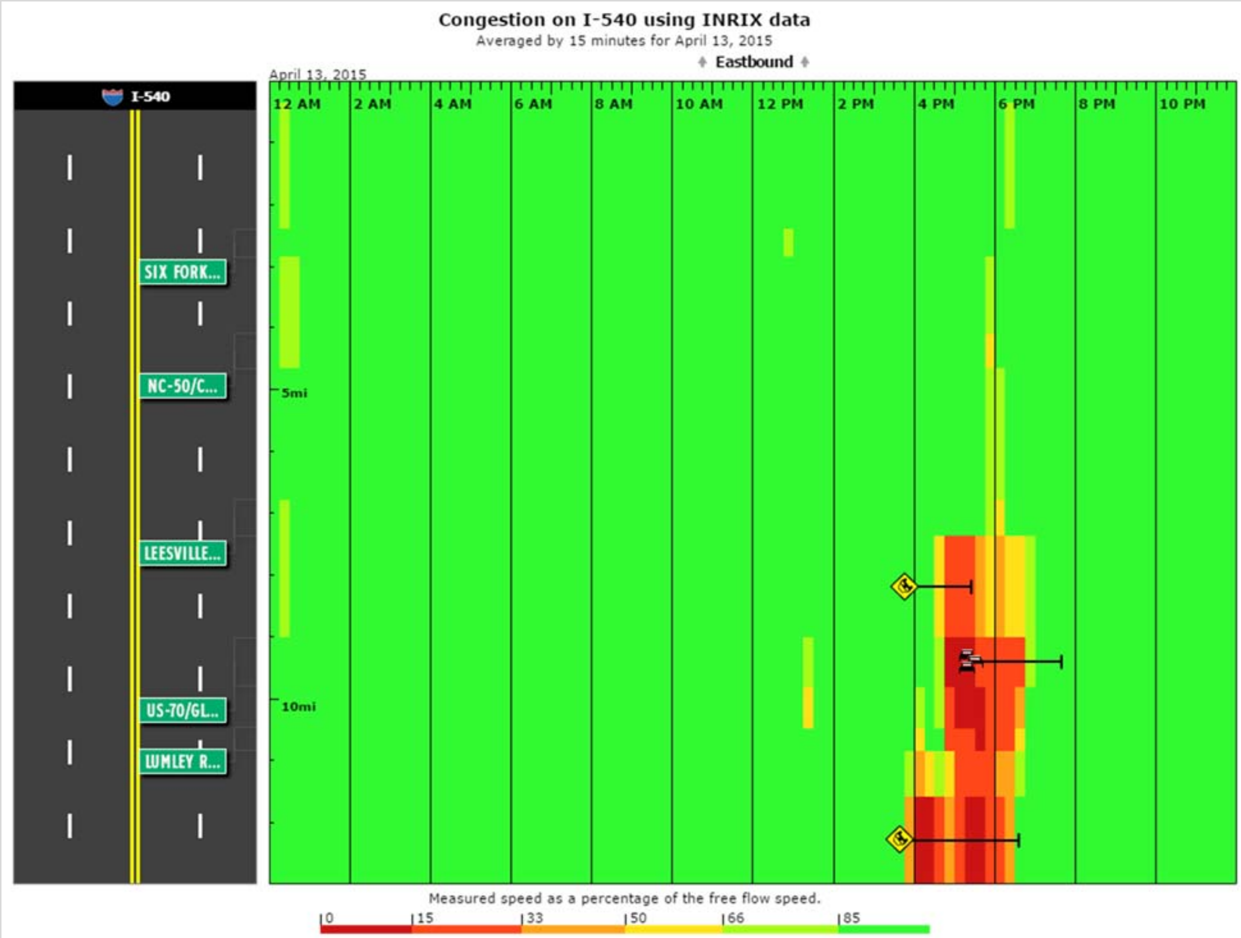

Figure 19: INRIX Congestion Scan for 04/13/2015

The figure below compares the travel time index distribution of FREEVAL to that of INRIX and Bluetooth.

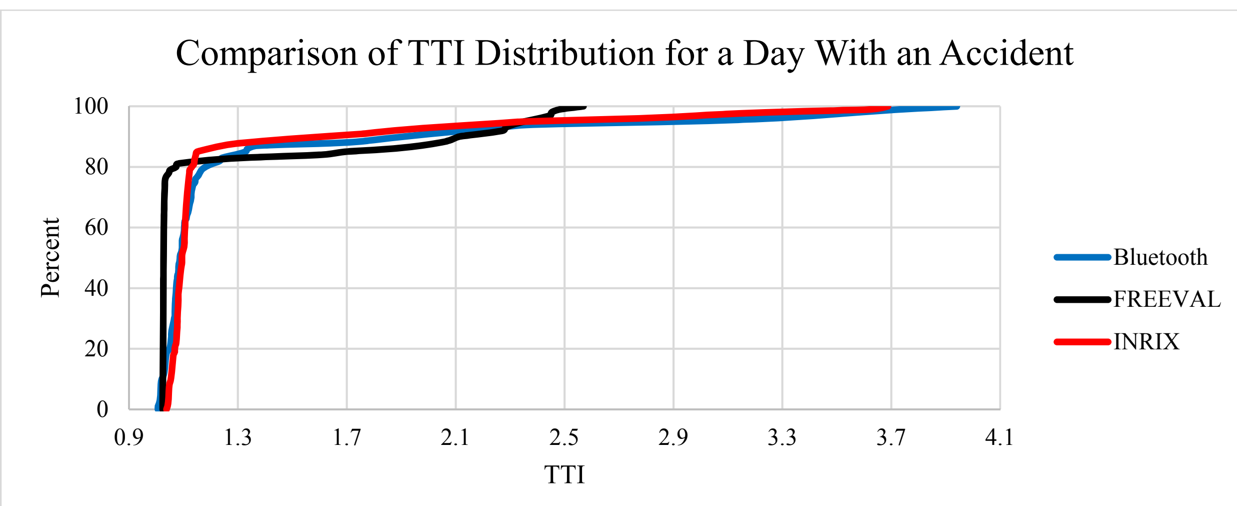

Figure 20: Comparison of TTI Dist. for 4/13/2015 (A day with accident)

Visual observations of the figure above show that INRIX and Bluetooth distributions do not have significant differences and follow the same trend. However, FREEVAL's distribution does not follow the two distributions and peels off of the two distributions at the beginning of the distribution. There are a significant number of travel times in the lower tail of the distribution related to travel speeds during the uncongested regimes. About 70 percent of trips have TTI's slightly higher than one showing no degradation of travel speeds on the facility for majority of the study period. However, a small percent of INRIX and Bluetooth TTIs are below one showing speeds greater than the free flow speed on the facility. FREEVAL's TTI never gets less than one for the entire period. The difference between FREEVAL and the other two distributions gets significant on the right tail of the distribution passing the 90^{th} percentile travel time index. Following table shows the reliability performance measures obtained from FREEVAL, INRIX and Bluetooth and their comparison.

Table 15: Reliability Performance Measures for 4/13/2015

Reliability Measures	Bluetooth	INRIX	FREEVAL	Abs. Dif. (BT - FVL)	Abs. Dif. (INRIX - FVL)
Mean TTI	1.302	1.160	1.171	0.132	0.011
Misery Index	3.518	3.241	2.468	1.050	0.773
50th % TTI	1.087	1.095	1.027	0.060	0.068
80th % TTI	1.181	1.131	1.073	0.107	0.058
Reliability Rating (%)	84.706	88.542	82.292	2	6
95th % TTI (PTI)	2.936	2.364	2.352	0.583	0.011

As noted above, the difference of FREEVAL - INRIX, and FREEVAL - Bluetooth is very significant on the right side of the distribution. This impact is also visible from the measures located on the right side of the distribution, planning time index and misery index, these measures between FREEVAL and the two other data sources are very pronounced compared to other measures.

The second scenario considered under modeling of non-recurring congestion is the presence of only an adverse weather event on the facility. After consultation of TIMS database, Weather Underground database and INRIX's congestion scan tool, March 19, 2015 was selected for the second scenario. On March 19, 2015 there was a light rain and light drizzle that started from 12:51 pm and continued till 7 in the morning the next day. Following figure shows the INRIX congestion scan for the day. The figure shows that the facility is not congested for the entire morning period and only experiences congestion a little passed 4 pm. The cause of the congestion is mainly the surge in demand due to the PM peak, but same weekdays for the facility show a lighter congestion during the PM peak and that the one occurring on March 19 is believed to have worsened because of the presence of rain and drizzle.

Figure 22 compares the travel time index distribution of FREEVAL to that of INRIX and Bluetooth for March 19, 2015 where an adverse weather is present.

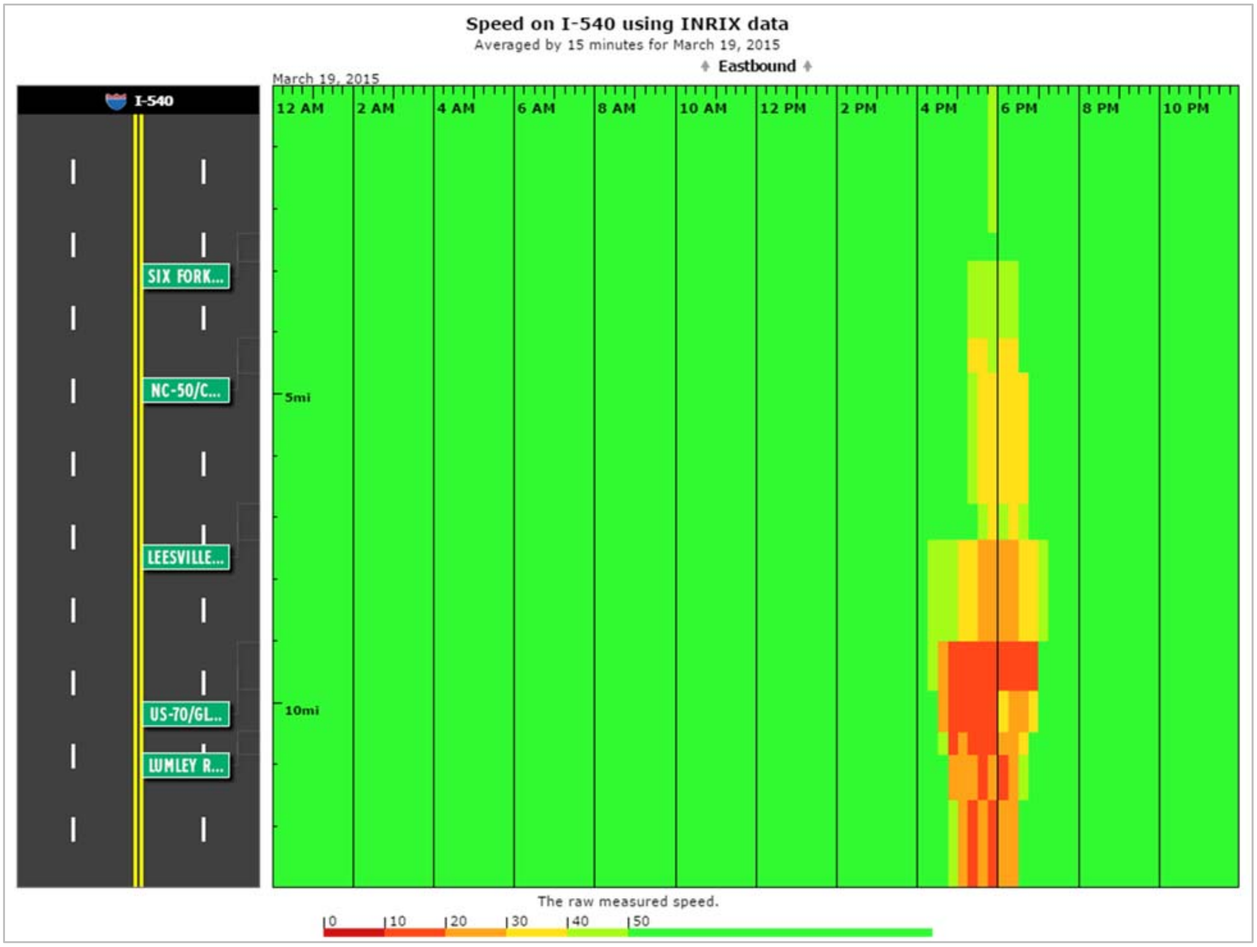

Figure 21: INRIX Congestion Scan for 03/19/2015

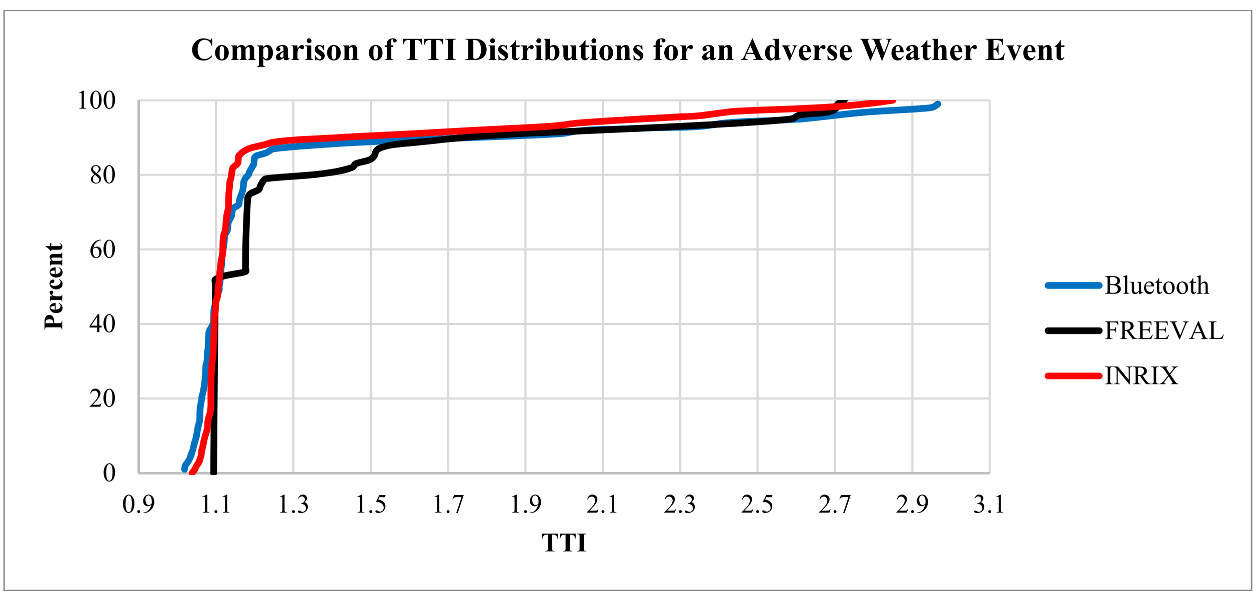

Figure 22: Comparison of TTI Dist. for 3/19/2015 (A day with an adverse weather event)

The figure above shows that INRIX and Bluetooth distributions are very similar and follow the same trend. The only difference between the two are either located at the left tail or right tail of the distribution. As mentioned above, it is believed that this difference between INRIX and Bluetooth data is data smoothing and speed capping by INRIX. FREEVAL's distribution has a stark difference in several regions. First the left tail of the distribution until about 50th percentile has a flat value of about 1.1. The value of TTI for FREEVAL never drops below 1, indicating restriction of travel speeds by FREEVAL never surpassing the free flow speed specified in the model. Second, the TTI has a significant drop right at about 50-55 percentile value and then picking back up. From 50 to about 85 percentile, the distribution of FREEVAL is significantly different from that of INRIX and FREEVAL. However, the difference between he distributions decrease as the percentage passes the 85th percent. Following table shows the reliability performance measures and their comparison for March 19.

Table 16: Reliability Performance Measures for 3/19/2015 (a day with an adverse weather event)

Reliability Measures	Bluetooth	INRIX	FREEVAL	Abs. Dif. (BT - FVL)	Abs. Dif. (INRIX - FVL)
Mean TTI	1.210	1.224	1.168	0.042	0.057
Misery Index	2.875	2.603	2.686	0.189	0.083
50th % TTI	1.108	1.106	1.098	0.010	0.007
80th % TTI	1.183	1.140	1.347	0.164	0.208
Reliability Rating (%)	87	90	88	0	2
95th % TTI (PTI)	2.615	2.199	2.586	0.029	0.387

Overall the table above shows that the reliability performance measures between different sources of data are not very significantly different. However, the reasoning provided above are validated by looking at the 80th percentile TTI, which shows a big difference of FREEVAL with Bluetooth and INRIX distributions. Apart from that, the performance measures obtained from the FREEVAL model can predict the impact of weather event reasonably.

The third and last scenario considered under modeling of non-recurring congestion is the presence of both an accident and adverse weather event on the facility. After consultation of TIMS database, Weather Underground database and INRIX's congestion scan tool, March 3, 2015 was selected for the third scenario. On this day an accident occurred on mile marker 11 at 18:26 and continued until 19:20. The accident caused one lane of the freeway to close down. The weather on the day was overcast until 14:00 and a mix of light drizzle and light rain governed the afternoon starting from 14:23 and continuing until midnight. Following figure shows the INRIX congestion scan for the day. The figure shows that the facility is uncongested before 4 pm, and PM peak congestion starts a little after 4 pm. The figure also shows the location of the accident relative to the facility and shows that almost the entire length of the facility is congested between 5:30 pm – 7:30 pm.

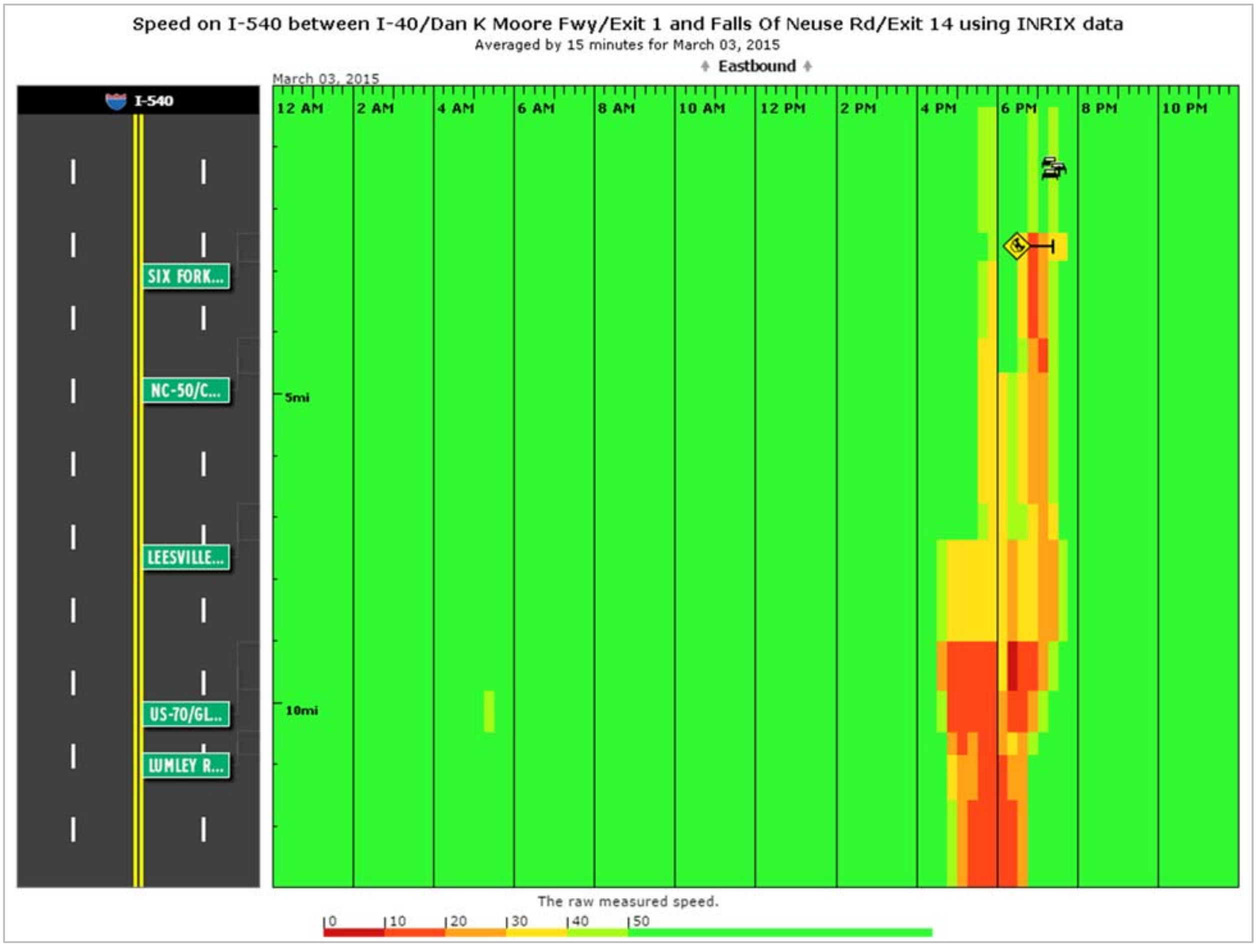

Figure 23: INRIX Congestion Scan for 03/3/2015

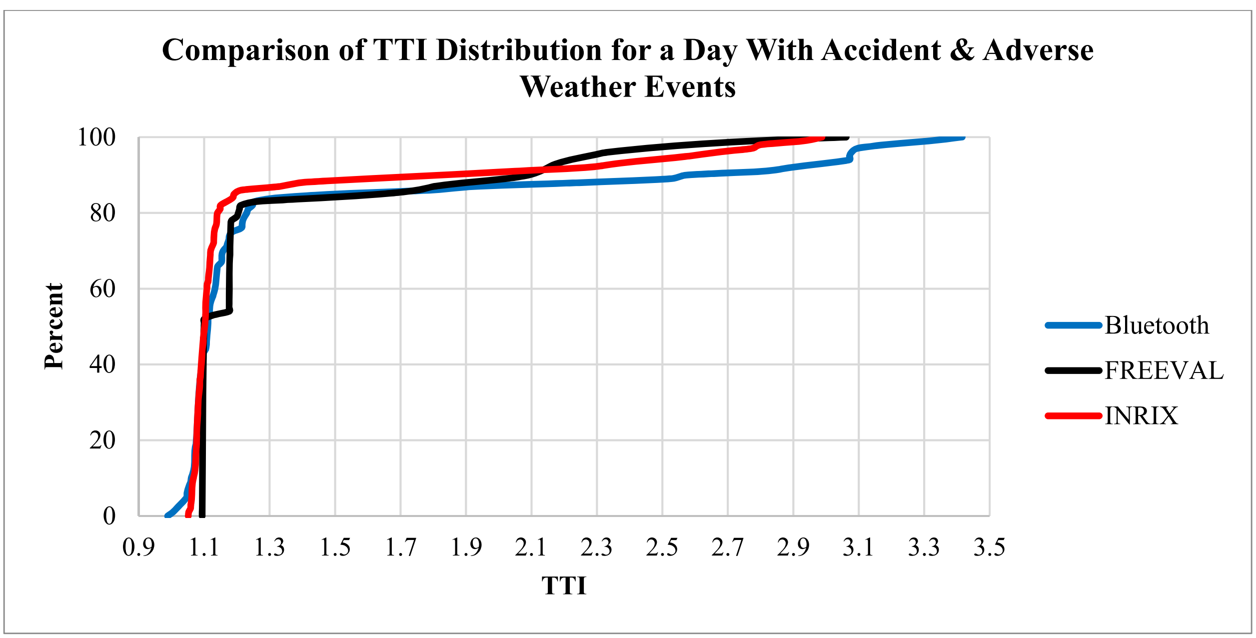

Figure 24: Comparison of TTI Dist. for 3/3/2015 (adverse weather + accident)

Interestingly the figure above shows that INRIX and FREEVAL distributions are very similar to each other at the extremes of the right tail. However, Bluetooth stays together with them for the most part, but peels off once the percent is above 85 percent. Again, similar pattern can be seen on the left tail of the distributions (Bluetooth with a good number of readings lower than TTI of one, INRIX with a limited number below TTI one and FREEVAL with no TTI less than one). Following table shows the comparison of the travel time reliability performance measures for the specific day.

Table 17: Reliability Performance Measures for 3/3/2015 (adverse weather event + accident)

Reliability Measures	Bluetooth	INRIX	FREEVAL	Abs. Dif. (BT - FVL)	Abs. Dif. (INRIX - FVL)
Mean TTI	1.275	1.265	1.169	0.106	0.095
Misery Index	3.222	2.823	2.625	0.597	0.198
50th % TTI	1.111	1.101	1.100	0.012	0.002
80th % TTI	1.230	1.141	1.204	0.026	0.063
Reliability Rating (%)	83.750	86.458	84.375	1	2
95th % TTI (PTI)	3.072	2.583	2.274	0.797	0.308

The table above shows that measures located on the right tail of the distributions (misery index, 95th percentile TTI) are different for Bluetooth and FREEVAL significantly, where these differences between INRIX and FREEVAL are negligible. It can be concluded from this section that FREEVAL reasonably predicts travel times in presence of non-recurring congestions.

4.1.3 Validation of the scenario generation approach for travel time distribution estimation

Since the necessary data available in FREEVAL for running the scenario generator is weather and incident, this task could be executed for the entire duration of 2015, but it was rather done for the same duration where Bluetooth data was available. The impact of data availability was also studied on scenario generation. Incidents can be created using three options: national default values, statewide default values, and facility specific values. However, there are two tracks of data availability for modeling of weather data: long-term regional weather data and facility specific weather for the duration of the study. FREEVAL in its database has the long-term regional weather data of more than 100 cities in the United States. These long-term weather probabilities are based on 10 years of weather data on each city. Following tables show the number of incident and weather events by TIMS, Weather Underground and FREEVAL using national default values for incident and Raleigh's long-term regional weather data for weather probabilities.

Table 18: Number of Incident Events Generated by FREEVAL Using National Default Values vs. Ground Truth Incidents for Eastbound of I-540

Incident Severity	FREEVAL	TIMS	Dif. (FREEVAL - TIMS)
	Number of Incidents	Number of Incidents	
Shoulder Closure	217	67	150
1 Lane Closure	100	31	69
2 Lane Closure	16	5	11
3 Lane Closure	3	1	2
4+ Lane Closure	0	0	0
Total	336	104	232

The table above shows significant difference between events generated by FREEVAL and ground truth events. FREEVAL overestimates the number of events in almost all incident severities.

Table 19: Generated Number of Weather Events by FREEVAL Using Raleigh's Long-term Weather Probabilities vs. Ground Truth Weather Events

Weather Severity	Description	FREEVAL	WU	Dif. (FREEVAL - WU)
		Number of Events	Number of Events	
Medium Rain	> 0.10 - 0.25 in/h	66	71	-5
Heavy Rain	> 0.25 in/h	69	38	31
Light Snow	>0.00 - 0.05 in/h	11	12	-1
Light/Medium Snow	>0.05 - 0.10 in/h	6	2	4
Medium/Heavy Snow	>0.10 - 0.50 in/h	7	4	3
Heavy Snow	>0.50 in/h	-	2	-2
Severe Cold	<-4 F	-	-	-
Low Visibility	0.50 – 0.99 mi	35	57	-22
Very Low Visibility	0.25 – 0.49 mi	-	-	-
Minimum Visibility	< 0.25 mi	18	29	-11
Total		212	215	-3

The table shown above indicates that there are some weather events that FREEVAL overestimates and there are some events that FREEVAL underestimates. Those that are overestimated are: heavy rain, light/medium snow and medium/heavy snow. The rest of the weather types are underestimated by the FREEVAL. However, the total difference is negligible between generated events and ground truth events. Following table show the number of ground truth incidents and incidents generated by FREEVAL using statewide default values.

Table 20: Number of Incident Events Generated by FREEVAL Using Statewide Default Values vs. Ground Truth Incidents for Eastbound of I-540

Incident Severity	FREEVAL	TIMS	Dif. (FREEVAL - TIMS)
	Number of Incidents	Number of Incidents	
Shoulder Closure	88	67	21
1 Lane Closure	40	31	9
2 Lane Closure	7	5	2
3 Lane Closure	1	1	0
4+ Lane Closure	0	0	0
Total	136	104	32

The table shown above indicate that the difference between the number of incidents generated by FREEVAL and TIMS get smaller as we move from national default values to statewide default values. However, the scenarios generated by FREEVAL still overestimates the number of incidents on the facility. Following tables show the number of incident and weather events by TIMS, Weather Underground and FREEVAL using facility specific values for incident and same year weather data for weather probabilities (the year for which reliability analysis is run).

Table 21: Number of Incident Events Generated by FREEVAL Using Facility Specific Values vs. Ground Truth Incidents for Eastbound of I-540

Incident Severity	FREEVAL	TIMS	Dif. (FREEVAL - TIMS)
	Number of Incidents	Number of Incidents	
Shoulder Closure	71	67	4
1 Lane Closure	33	31	2
2 Lane Closure	5	5	0
3 Lane Closure	1	1	0
4+ Lane Closure	0	0	0
Total	110	104	6

The table above shows that FREEVAL still overestimates the number of incidents when using facility specific values. However, the difference is very minimal and negligible. Studying these tables, it can be observed that data availability does have a significant impact on the results that will be obtained from this analysis tool.

Table 22: Generated Number of Weather Events by FREEVAL Using Same-Year (2015) Weather Probabilities vs. Ground Truth Weather Events

Weather Severity	Description	FREEVAL	WU	Dif. (FREEVAL - WU)
		Number of Events	Number of Events	
Medium Rain	> 0.10 - 0.25 in/h	69	71	-2
Heavy Rain	> 0.25 in/h	76	38	38
Light Snow	>0.00 - 0.05 in/h	10	12	-2
Light/Medium Snow	>0.05 - 0.10 in/h	5	2	3
Medium/Heavy Snow	>0.10 - 0.50 in/h	12	4	8
Heavy Snow	>0.50 in/h	2	2	-2
Severe Cold	<-4 F	-	-	-
Low Visibility	0.50 – 0.99 mi	30	57	-27
Very Low Visibility	0.25 – 0.49 mi	-	-	-
Minimum Visibility	< 0.25 mi	19	29	-10
Total		223	215	8

The figure above shows that the difference gap between generated scenarios by FREEVAL and ground truth events slightly widen as we use same-year weather probabilities. This small change in difference between the two can be neglected. To get an overall sense of how data impacts the output of the analysis engine, following figure shows all three data availability options and the difference between generated number of events by FREEVAL and TIMS for each data availability option.

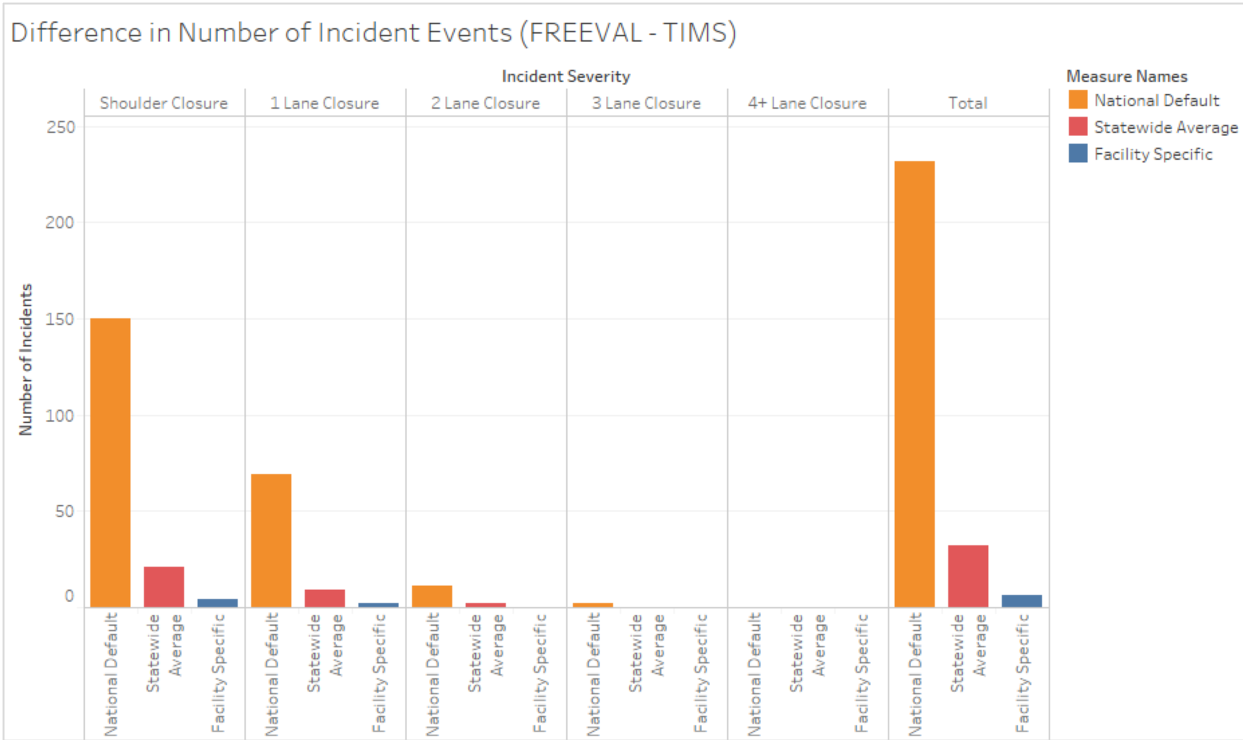

Figure 25: Impact of Data Availability on FREEVAL Output (Difference in Number of Generated Events by FREEVAL and Ground Truth Number of Events by TIMS)

The figure above shows that as the data is customized for the facility, the outputs of FREEVAL gets closer and closer to the ground truth number of events. The scenario where national default value is used for generation of incidents, is the scenario with the largest difference in all types of incident severity. Similarly, employing statewide default values for incident frequencies and duration result in moderate differences between the two extremes. Finally, using the facility specific data significantly diminishes the difference between FREEVAL output and ground truth number of incidents. Following figure compares the difference between FREEVAL generated weather events and Weather Underground events using long-term regional weather data and reliability reporting period weather data.

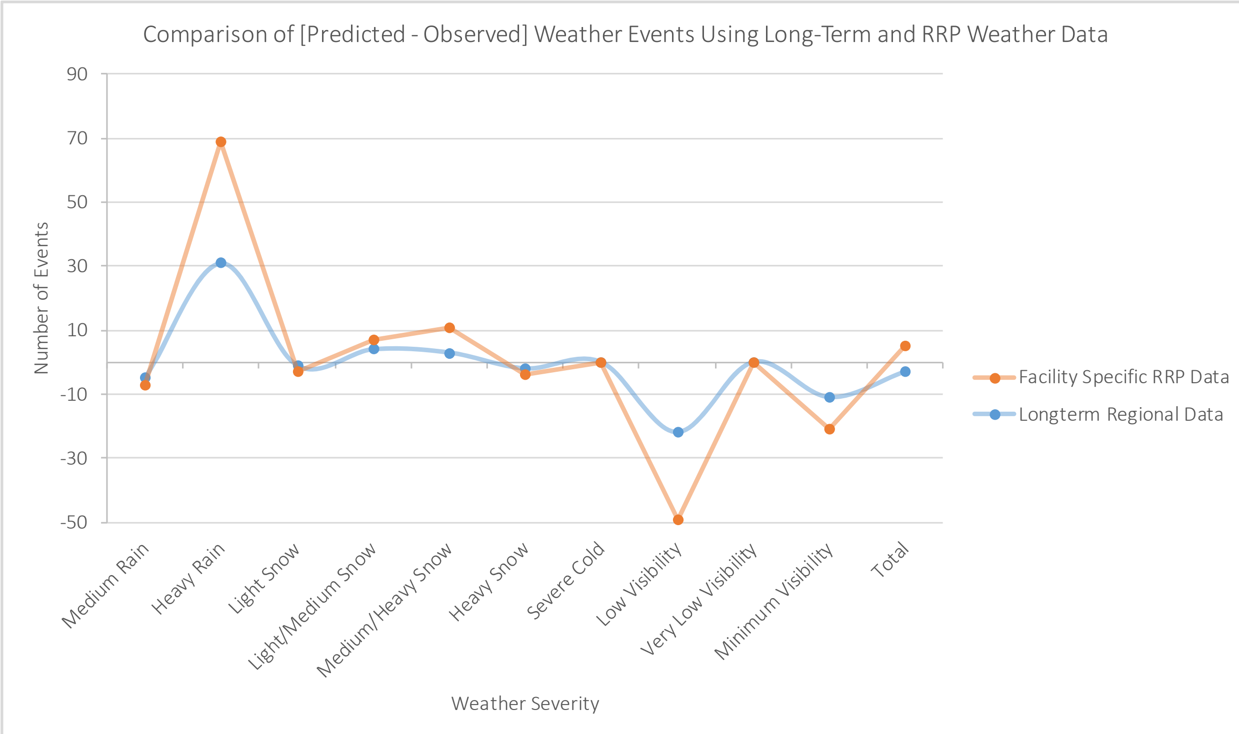

Figure 26: Comparison of (FREEVAL-WU) Weather Events Obtained Using Long-term and RRP Weather Data

The figure above shows that when long-term regional weather data is used, FREEVAL does a better job of predicting the number of weather events compared to when facility specific weather data is used for just the reliability reporting period. The main reason that comes to mind for this behavior of FREEVAL is related to the fact that the long-term regional data is the average of facility weather for ten years. Therefore, there is no surprise as to why FREEVAL's output is closer than the average of several years than the reliability reporting period.

In conclusion, feeding facility specific incident data to FREEVAL predicts the closest number of weather event compared to national and statewide default values. However, feeding long-term regional weather data for weather event prediction gives better results compared to feeding facility specific weather data for the reliability reporting period.

4.1.4 Validation of the resulting travel time distributions

The resulting travel time distributions for both directions of site 1 are presented below. The reliability reporting period started on January 9, 2015 and finished on June 6, 2015. There were 99 weekdays with Bluetooth data. The INRIX travel time index distribution shown in the following figures (six figures for all sites) were constructed from data points from periods where Bluetooth data was available. For this purpose the VLOOKUP() function in Microsoft Excel was used.

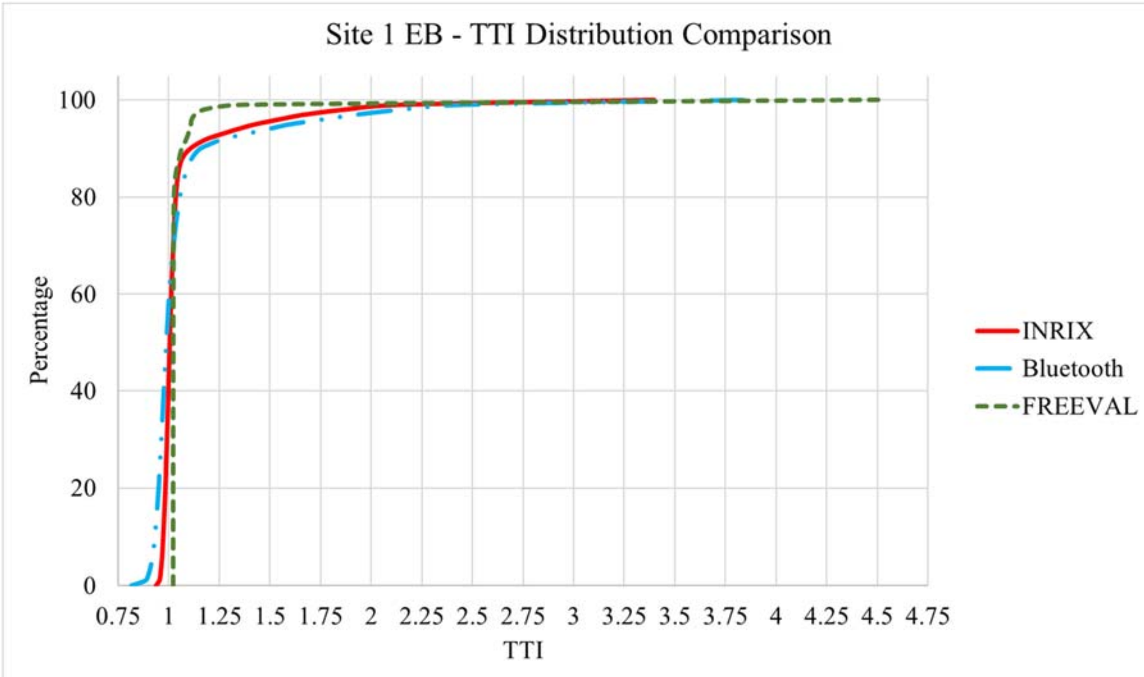

Figure 27: Travel Time Index Distribution Comparison for Site 1 – Eastbound

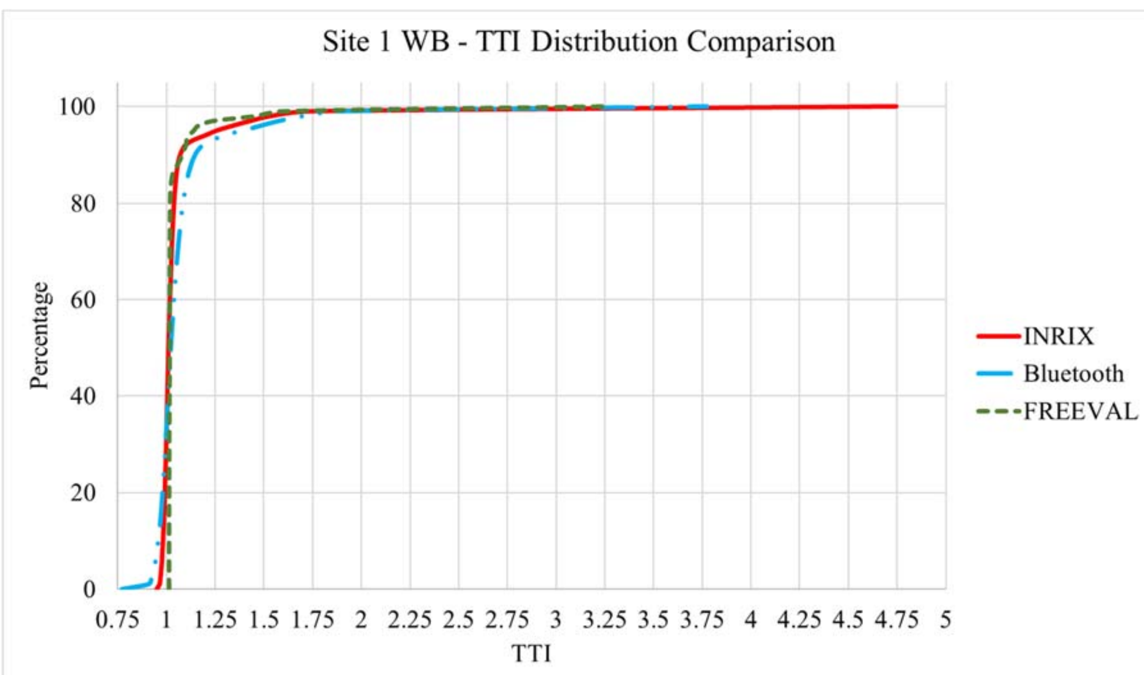

Figure 28: Travel Time Index Distribution Comparison for Site 1 - Westbound

The two figures above have a lot of similarities. The figure for the eastbound direction of the facility show that there are a significant number of analysis periods with Bluetooth data that translate to speeds greater than the free flow speed on the facility. This conclusion is obtained from observation of the left tail of Bluetooth distribution. INRIX distribution shows similar pattern. However, the number of data points from INRIX is not as significant as Bluetooth's and the reason for that is believed to be smoothing and capping of data by INRIX algorithms. To no surprise, FREEVAL does not have any data points that would indicate travel speed greater than the free flow speed specified in the model.

The eastbound figure shows that while the right tail of the Bluetooth and INRIX distributions remain the same, FREEVAL's distribution peels off a little bit from the two right around 90th percentile TTI. However, it overlaps the two other distributions back at around 95th percentile. The westbound figure, however, shows consistent similarity between the three distributions on the right tails of the distributions. Following two tables show the reliability performance measures for site 1's eastbound and westbound directions.

Table 23: Comparison of Reliability Performance Measures for Eastbound of I-540

Reliability Measures	Bluetooth	INRIX	FREEVAL	Abs. Dif. (BT - FVL)	Abs. Dif. (INRIX - FVL)
Mean TTI	1.036	1.064	1.039	0.003	0.025
50th % TTI	0.990	1.007	1.023	0.033	0.017
80th % TTI	1.057	1.037	1.026	0.031	0.011
Reliability Rating (%)	93	94	99	6	5
95th % TTI (PTI)	1.601	1.430	1.110	0.491	0.320

Table 24: Comparison of Reliability Performance Measures for Westbound of I-540

Reliability Measures	Bluetooth	INRIX	FREEVAL	Abs. Dif. (BT - FVL)	Abs. Dif. (INRIX - FVL)
Mean TTI	1.068	1.047	1.042	0.026	0.005
50th % TTI	1.024	1.012	1.016	0.008	0.004
80th % TTI	1.087	1.040	1.019	0.068	0.021
Reliability Rating (%)	94	96	97	3	2
95th % TTI (PTI)	1.396	1.259	1.241	0.154	0.018

The two tables above show that most of the reliability measures are similar for the three distributions. However, for the eastbound FREEVAL's right tail measure (PTI) is different compared to Bluetooth and INRIX indicating that the right of the distribution does not quite overlap with INRIX and Bluetooth distributions. Other measures are very close to each other, especially for the westbound direction of the facility. Following figure show travel time index distributions for northbound directions of site 2.

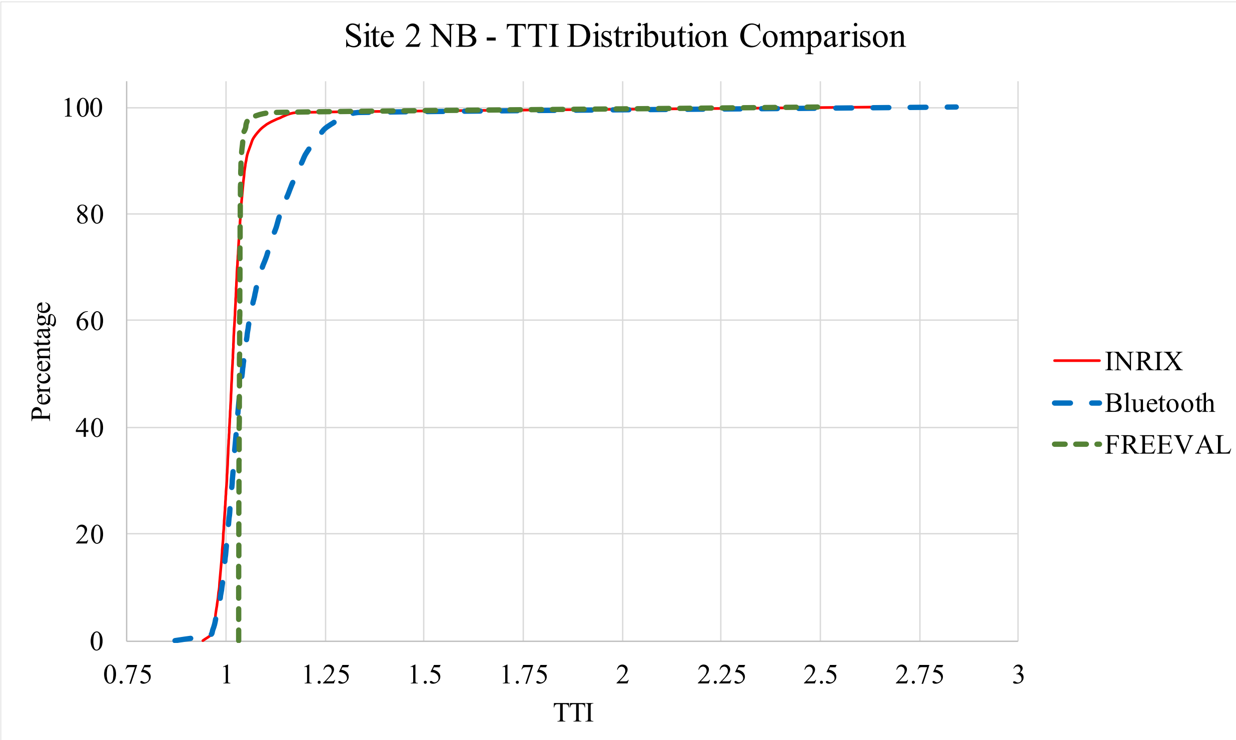

Figure 29: Travel Time Index Distribution Comparison for Site 2 – Northbound

The figure above shows that Bluetooth and INRIX are very similar on the left tail of the distribution, where FREEVAL is drastically different from both of them. However, on the right tail of the distribution its FREEVAL and INRIX that are very similar to each other and Bluetooth is completely different from the two. Bluetooth distribution peels off from the other two distributions somewhere around 40 percent and gap widens at the percentage increases. The main reason for this difference in Bluetooth and INRIX distributions is presence of a weigh station on this route, which a significant number of vehicles (20%) are trucks and need to go the weigh station. Following table shows the reliability performance measures for the northbound direction of the I-95. Visual observations of the table confirm similarity of INRIX to FREEVAL distributions and slight difference of Bluetooth from them. The PTI, a measure located on the right tail of the distributions, of Bluetooth has the highest value among the distributions.

Table 25: Comparison of Reliability Performance Measures for Northbound of I-95

Reliability Measures	Bluetooth	INRIX	FREEVAL	Abs. Dif. (BT - FVL)	Abs. Dif. (INRIX - FVL)
Mean TTI	1.070	1.021	1.037	0.034	0.015
50th % TTI	1.039	1.015	1.034	0.005	0.020
80th % TTI	1.136	1.036	1.036	0.099	0.000
Reliability Rating (%)	99	100	100	1	0
95th % TTI (PTI)	1.237	1.077	1.044	0.193	0.033

Following figure compares the travel time index distributions for the southbound of site 2.

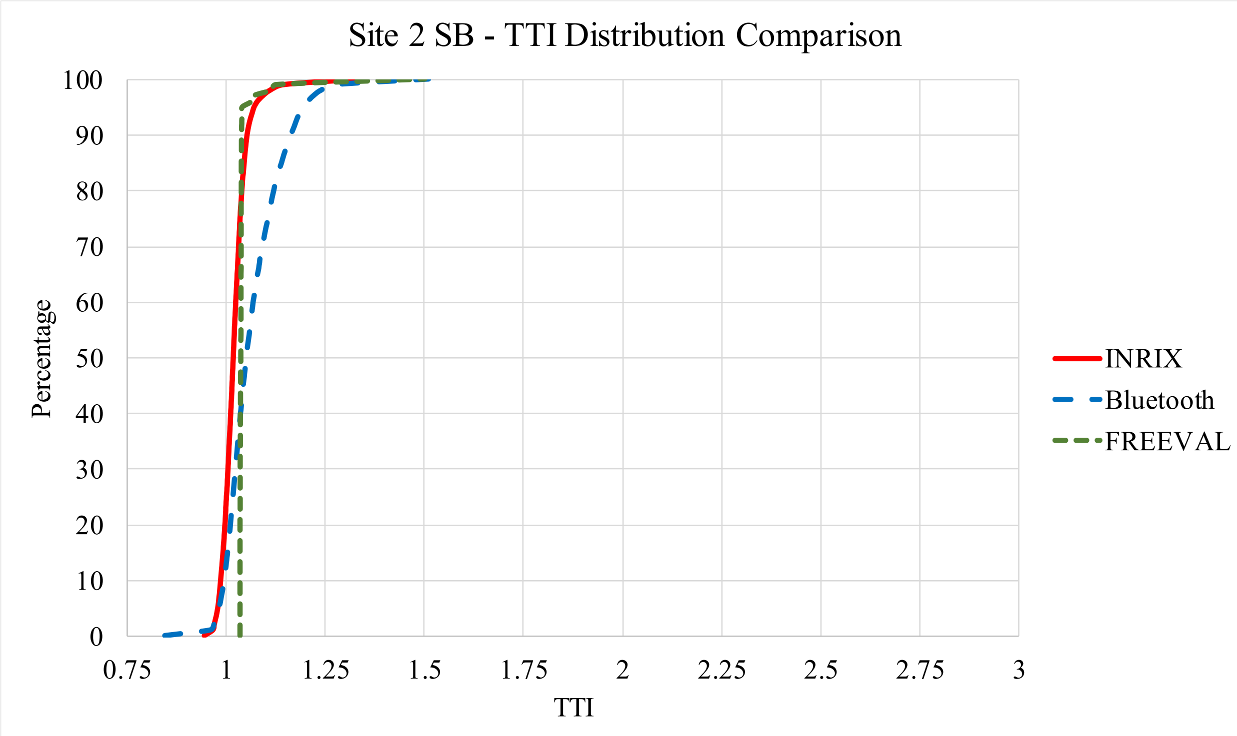

Figure 30: Travel Time Index Distribution Comparison for Site 2 – Southbound

The figure above shows similar patterns that were highlighted for the northbound of site 2. Following table shows the reliability performance measures for the southbound direction of I-95. The table shows that the reliability performance measures are not significantly different between the distributions. The measure with the biggest difference is the PTI, located on the right tail of the distribution. This measure is relatively larger for Bluetooth due to the fact that a significant number of vehicles (trucks) need to report to the weight station. These vehicles are believed to be filtered out by the INRIX algorithms.

Table 26: Comparison of Reliability Performance Measures for Southbound of I-95

Reliability Measures	Bluetooth	INRIX	FREEVAL	Abs. Dif. (BT - FVL)	Abs. Dif. (INRIX - FVL)
Mean TTI	1.066	1.021	1.064	0.003	0.043
50th % TTI	1.050	1.018	1.037	0.014	0.019
80th % TTI	1.120	1.039	1.038	0.082	0.001
Reliability Rating (%)	100	100	99	0	1
95th % TTI (PTI)	1.196	1.070	1.041	0.155	0.030

Bluetooth sensors were installed on site 3 from October 21, 2015 and collected data until March 8, 2015. Only data collected in 2015 was used for generation of the TTI distribution and travel time reliability performance measures. Following figure compares the travel time index distribution for eastbound direction of site 3 (I-40 at Asheville). The figure is followed by the table that compares the reliability performance measures obtained from these distributions.

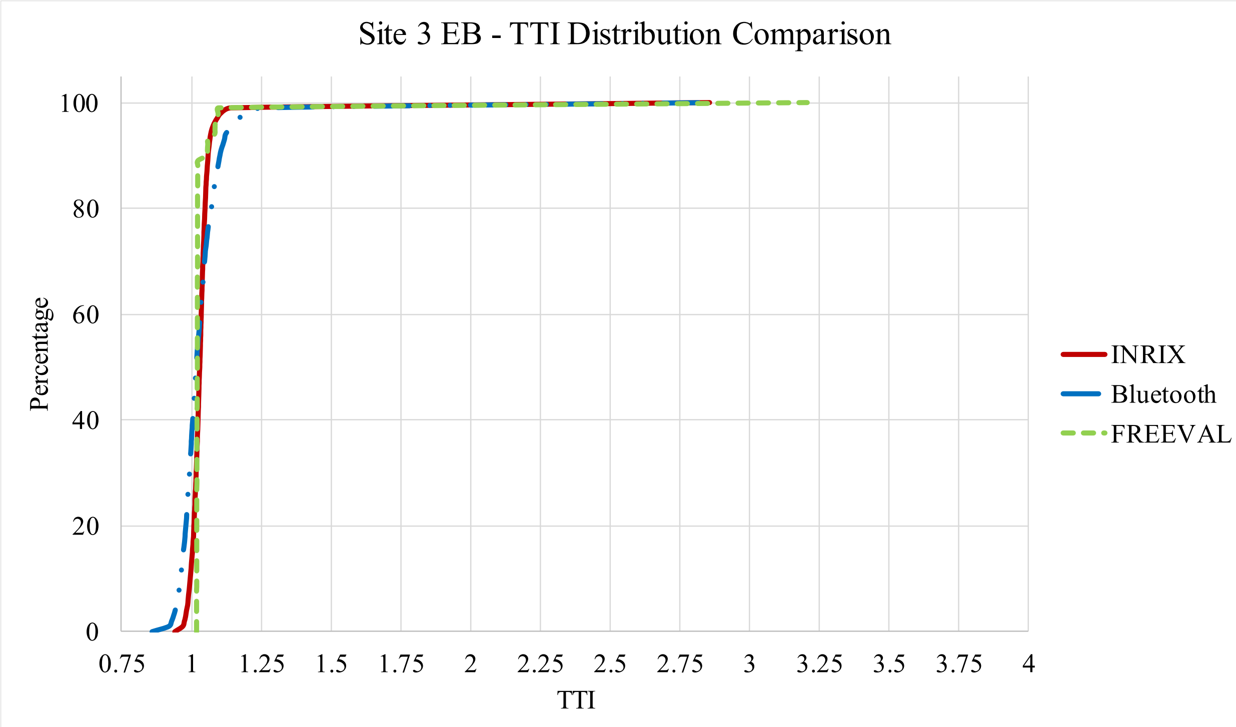

Figure 31: Travel Time Index Distribution Comparison for Site 3 – Eastbound

Table 27: Comparison of Reliability Performance Measures for Eastbound of I-40 at Asheville

Reliability Measures	Bluetooth	INRIX	FREEVAL	Abs. Dif. (BT - FVL)	Abs. Dif. (INRIX - FVL)
Mean TTI	1.066	1.021	1.064	0.003	0.043
50th % TTI	1.050	1.018	1.037	0.014	0.019
80th % TTI	1.120	1.039	1.038	0.082	0.001
Reliability Rating (%)	100	100	99	0	1
95th % TTI (PTI)	1.196	1.070	1.041	0.155	0.030

The above figure shows that apart from the left tail of the distributions, FREEVAL and INRIX are very close to each other and don't seem to deviate a lot. However, Bluetooth distribution is slightly different at both tails of the distribution when compared to INRIX and FREEVAL distributions. Similar results can be obtained from the table that compares the reliability performance measures. Overall, FREEVAL does a good job of predicting the travel time reliability performance measures for the eastbound direction of the route. Following figure compares TTI distributions for the westbound of site 3. The figure is followed by the table that compares the reliability performance measures obtained from the distributions.

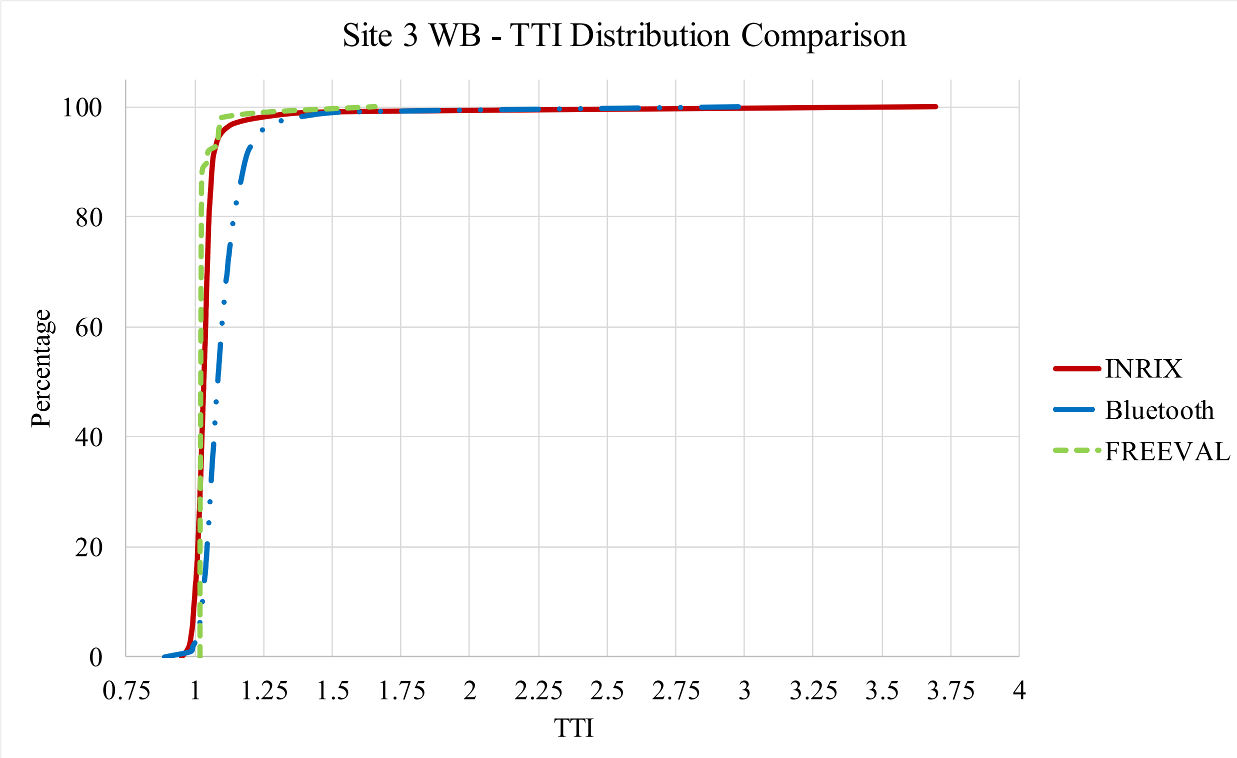

Figure 32: Travel Time Index Distribution Comparison for Site 3 – Westbound

Table 28: Comparison of Reliability Performance Measures for Westbound of I-40 at Asheville

Reliability Measures	Bluetooth	INRIX	FREEVAL	Abs. Dif. (BT - FVL)	Abs. Dif. (INRIX - FVL)
Mean TTI	1.105	1.043	1.031	0.074	0.012
50th % TTI	1.084	1.031	1.022	0.062	0.009
80th % TTI	1.142	1.050	1.024	0.118	0.026
Reliability Rating (%)	98	99	99	1	0
95th % TTI (PTI)	1.230	1.092	1.086	0.143	0.006

Both the figure and the table above show that INRIX and FREEVAL is very close to each other and that Bluetooth starts same as INRIX, but soon distances itself from both distributions. This divergence of Bluetooth can be associated to the weigh station that is located on the westbound of the site. The traffic on this interstate is composed of a relatively higher percentage of heavy weight trucks that need to be weighted by the static scale at the weigh station. Their stop at the weigh station causes Bluetooth's average speeds to drop significantly and impact the overall travel time reliability on the facility. The weigh station was modeled in FREEVAL as a combination of off and on ramps with percentage of trucks as the off-ramp volume that exits the freeway and returns back to the freeway via the on ramp. However, the speed of vehicles inside the weigh station environment can't be adjusted to mimic the operation of a weigh station. As a result, FREEVAL does not perfectly fit the distribution obtained by Bluetooth data, but does a fairly good job of matching INRIX's distribution. the main reason INRIX's distribution looks very similar to that of FREEVAL is the fact that INRIX smooths its data and filters out vehicles that stop at the weigh station for weight enforcement purposes. In summary, FREEVAL does a fairly good job of predicting the travel time reliability at both directions of site 3 – at least for passenger vehicles.

5.0 CONCLUSIONS AND RECOMMENDATIONS

SHRP 2 Project L08 L08 – *Incorporation of Travel Time Reliability into the Highway Capacity Manual* – delivered implementable methodologies and tools for prediction of reliability performance on freeways and arterial streets. Although data sets employed in the development of those methodologies were chosen with robustness and generalizability, validation of the methodologies by applying them to multiple locations and conducting detailed assessment of the accuracy and usefulness of the reliability performance predictions were deemed necessary. This project sat out to tackle the validation of the developed methodologies. The central objective of this research was to validate the freeway reliability performance prediction methods developed under SHRP 2 Project L08 for incorporation into the Highway Capacity Manual. The inherent secondary objective was to identify and clearly define any modifications to the freeway reliability performance prediction methods that may be necessary.

The validation efforts included validation of 1) facility travel time estimates under base-model, 2) facility performance modeling under non-recurring congestion, 3) scenario generation approach for travel time distribution estimation and 4) resulting travel time distributions.

Validation of the facility travel time estimates under base-model conducted both for uncongested and recurring congestion states, revealed that the developed tools are capable of estimating the travel times with reasonable accuracy. Although the reliability performance measures estimates provided by FREEVAL for the left tail of the distributions were slightly different than those obtained through INRIX and Bluetooth, other measures located in the center and right tail of the distribution are very similar for the two data sources and FREEVAL.

Three conditions were considered for validation of non-recurring congestion. Presence of an incident, presence of an adverse weather event and combination of the two. For the first condition, presence of only an incident on the facility, FREEVAL does not do a very good job of predicting travel times for the facility. The distribution of travel time index generated using FREEVAL's travel time estimates is not matching those of Bluetooth and INRIX at both tails. In addition, the reliability performance measures given by FREEVAL is significantly different than those given by Bluetooth and INRIX. However, FREEVAL does a better job predicting travel time estimates under adverse weather events. Its distribution is slightly different from INRIX and Bluetooth starting 50th percentile to about 85th percentile but gets closer to them on both tails. The reliability performance measures show similar pattern, with highest difference for 80th percentile TTI, and misery index. Under the third condition, presence of both adverse weather and accident on the facility, FREEVAL's travel time predictions are considered to be fairly representative of true travel times on the facility as obtained from INRIX data sets. Interestingly, right tail of Bluetooth's distribution departs from both INRIX and FREEVAL. In summary, FREEVAL can reasonably predict travel times for a situation where both incident and adverse weather event are present.

Scenario generation for estimation of travel time distribution was conducted considering three data environments: data-rich (facility specific values), data-poor (national default values) and data-moderate (statewide default values). It was found that FREEVAL does a very poor job generating the required number of events that need to be modeled in all the study periods in a data-poor environment. However, the gap between FREEVAL generated scenarios and real number of events narrowed as national default values were substituted with statewide default values. With facility specific data, the gap between FREEVAL generated scenarios and ground truth number of events dropped significantly (negligible).

Travel time index distributions generated by FREEVAL matched both INRIX and Bluetooth distributions very well for both directions of the first site (I-540 at Raleigh). The distributions generated for both directions of the second route by FREEVAL, matched quite well with INRIX's distributions. However, there was a stark difference between FREEVAL's generated TTI distribution and that of Bluetooth's for both directions of site two (I-95 at Lumberton) and westbound of site three (I-40 at Asheville). Not only did Bluetooth's TTI distribution not match well with FREEVAL's distributions, but also it did not match with INRIX's distributions. The reason for this difference was associated to the presence of weigh stations on both directions of these facilities. It was found that 20% of traffic on these facilities consisted of trucks that needed to report to the weigh station for weighing purposes. These stops at the weigh stations caused Bluetooth average speeds to drop significantly and impacted both the overall shape of travel time index distribution and travel time reliability on the facility. Another repeating difference that was found between FREEVAL's TTI distribution and other distributions was located on the left tails of the distributions. While INRIX and Bluetooth always had some data points with TTI less than one (translating to speeds greater than free flow speed), FREEVAL never predicted any travel time estimates that would translate to speeds greater than the free flow speed specified in the model. In summary, FREEVAL fairly generated the TTI distribution for site one, but did a poor job for site two and one direction of site three where weigh stations were located. However, if INRIX were to be used as the ground truth TTI distribution for the comparison, it can be concluded that FREEVAL generated the TTI distribution extremely accurate.

In summary, the findings of this project showed that the methodologies and tools developed for prediction of travel time reliability performance by SHRP 2 Project L08, could sufficiently predict travel times and its associated reliability performance measures for freeways.

6.0 REFERENCES

Cambridge Systematics Inc., Texas Transportation Institute, University of Washington, Dowling Associates, 2003. F-SHRP Web Document 3: Providing a Highway System with Reliable Travel Times: Study 3-Reliability. Transportation Research Board of the National Academies, Washington, D.C.

National Research Council, 2013. Analytical procedures for determining the impacts of reliability mitigation strategies. Transportation Research Board.

Manual, H.C., 2016. A Guide for Multimodal Mobility Analysis. Transportation Research Board, Washington, DC.

INRIX, 2016. INRIX [WWW Document]. INRIX - INRIX. URL http://inrix.com/products/parking/ (accessed 1.4.18).

Parker, S.P., 2003. McGraw-Hill Dictionary of Scientific and Technical Terms. McGraw-Hill Companies, Inc.

Sisiopiku, V.P. and Islam, M., 2012. A freeway travel time reliability study. International Journal of Engineering Research and Development, 3(10), pp.83-101.

Toppen, A. and Wunderlich, K., 2003. Travel time data collection for measurement of advanced traveler information systems accuracy (Vol. 23). Mitretek Systems.

Traffic Survey Group, NCDOT [WWW Document], 2018. URL https://connect.ncdot.gov/resources/State-Mapping/Pages/Traffic-Survey-Group.aspx (accessed 1.5.18).

TRASK, L., AGHDASHI, B., SCHROEDER, B. and ROUPHAIL, N., FREEVAL 2015e Users Guide.

Weather Underground, 2018. Weather Undergr. URL https://www.wunderground.com/ (accessed 1.14.18).

Zegeer, J., Bonneson, J., Dowling, R., Ryus, P., Vandehey, M., Kittelson, W., Rouphail, N., Schroeder, B., Hajbabaie, A., Aghdashi, B. and Chase, T., 2014. Incorporating Travel Time Reliability into the Highway Capacity Manual (No. SHRP 2 Report S2-L08-RW-1).