University of Maryland

National Transportation Center

Project ID: NTC2014-R-SU-231

Vehicle Trajectory Tool: Application Pilot for AMS Test Bed

Final Report

By

Nagui M. Rouphail, PhD
Director ITRE and Professor of Civil Engineering
909 Capability Drive, Suite 3600
Research Building IV
Raleigh, North Carolina 27606
University Name(s)

Behzad Aghdashi, Research Associate, ITRE
Sangkey Kim, Post-Doctoral Scholar, ITRE
Ian Pike, Research Student, ITRE

For

April 2016

ACKNOWLEDGEMENTS

DISCLAIMER

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXCUTIVE SUMMARY

The objectives of this study are to test the feasibility of using in-vehicle data generated through a new on board vehicle unit (OBU) called i2D for potential applications related to safety and mobility. Off-line data applications of this research are focused on the generation of a region-specific database of 1Hz vehicle trajectories from a fleet of vehicles that can support fundamental traffic and safety research. These applications include the identification of locations with high crash rates, shock wave detection and propagation, arterial speed distribution near traffic signals, and free flow speed (FFS) estimation. A secondary objective was to upgrade and pilot-test the existing vehicle trajectory collection tool (VTT) for testing the feasibility of a Vehicle to Infrastructure to Vehicle (VIV) system, for the purpose of communicating specific speed or acceleration-triggered messages, which are also timely and spatially sensitive, to vehicles upstream of an actual bottleneck or queue. Data collected as part of this research effort is shown in the summary table below.

Data Collection Period	April 1 2014 to December 31 2015
Data Collection Duration	20 Months
Total number of Trips	38,875 Trips
Total Number of (different) Drivers	24 Drivers (mostly students and some staff)
Total Number of Seconds of HiRes Data	32,351,971 Seconds (~9,000 Hours)
Total Miles Driven	277,601 Miles
Average Trips Duration	13.9 Minutes
Average Trips Length	7.14 Miles

Trends in the data showed that the mean trip length observed was short, in the range of 1-5 miles, while mean trip duration was about 10-15 minutes. This is not surprising given the nature of the driver cohort, of students involved. In fact, about one third of all trips fell in the 1-5 miles and 1-20 minute duration category. Much more information was mined from the data, including the distribution of departure times (whose mode was the 5-6 pm period) and the number of one-way trips per day per driver, which averaged around 4.50. The latter number includes only those days were trips were made, or at least recorded by the system.

A summary of the utility of i2D data for various mobility and safety applications is as follows:
- The data was used to develop an improved model FFS based on individual driver and section attributes. This study, for first time, showed that "familiar" drivers exhibited a higher FFS than non-familiar ones, under certain conditions. Our data also developed a FFS model that highlighted the need to include the posted speed limit when estimating FFS. It had the best explanatory power of all models tested, including the one cited in the US Highway Capacity Manual.

- Our data showed a good association, but not causality, between rear-end crashes hotspot locations on a freeway facility and the frequency of high deceleration events (values 4 m/s/s or higher) observed from our i2D equipped fleet of 24 drivers as they traveled the facility. We also found that the incidence of such decelerations were not confined to one or two drivers, but were experienced by the majority of drivers crossing those segments.

- High resolution data, focused on the approach and departure links at a signalized intersection, were used to verify the fundamental difference between time and space mean speed definitions. The speed data produced an excellent estimate of space mean speed when weighted by density (in the high resolution data this means the number of GPS observations on a segment over a fixed interval). On the other hand, when the speed observations were weighed strictly by the number of trips, the resulting speed averaged produced the time mean speed. The data shows that near the intersection proper, those differences in speed estimation can be very significant by orders of magnitudes. The emerging speed distributions measured across segments also provides a good indication of the quality of the main street signal timing and progression.

- The shock wave detection and speed estimation experiments were found to be slightly difficult to control, both because of the timing (congestion did not occur precisely at the time vehicles were deployed), or because vehicles could not maintain adequate spacing to track the shockwave position. A successful experiment led to identifying four shockwave formations with speeds estimated from 20 to 40 mph. Promisingly, we found that only a minimal number of equipped vehicles are needed to track the shockwave.

Regarding the secondary objective related to the utility of the VIV experiments, which aims at emulating the effect of connected vehicle messaging, the findings were less promising. The research team found evidence of both false as well as missed alerts in addition to several latency issues between the transmitted and received messages. For example, the received message was delay on the freeway while the transmitted message was of a strong deceleration. Nevertheless, the team is still confident about the value of utilizing a real time event warning system using the VIV system. This new in-vehicle technology holds good promise for both mobility and safety applications. In the long run, and with the proliferation of equipped vehicles, the benefit of such data may eliminate the need for expensive, fixed monitoring infrastructure. Finally, the suggested VIV technology has the potential to mitigate the effect of secondary crashes, which generally occur because of a primary culprit, such as an exhisting crash, work zone, or poor weather. A timely alert can help drivers dodge these types of crashes, which are often unavoidable.

1.0 INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

In recent years, the issue of connected and autonomous vehicle research has received much attention, with the FHWA leading the way in testing and documenting how those technologies can benefit traveler safety, reduce congestion and improve the deleterious environmental impacts. Specifically, this study focuses on how high resolution probe vehicle data, alone or in conjunction with vehicle-to-vehicle (V2V) and vehicle to infrastructure (V2I) communications protocols, can be explored to improve both the safety and operations of traffic flow on congested urban facilities under both recurring (bottleneck) and non-recurring (weather, incident) congestion. At the USDOT Joint Program Office, the Dynamic Mobility Applications for Intelligent Network Flow Optimization (DMA-INFLO) program encompasses a number of high priority applications identified by the USDOT that show promise for safety and operational improvements. Initial applications of V2X technology in the DMA environment have included the field-testing of speed harmonization and queue warning (FHWA 2012).

It is evident that in order to implement either of these two protocols in a connected vehicle environment, a detailed knowledge of vehicle position, speed and acceleration are needed to detect a queue, and communicate targeted speed and queuing information will be necessary. In short, it is vital that *detailed geo-located vehicle trajectory data* is acquired in near-real time on the facility where the algorithms will be implemented (KAI 2013). This research provides support both for testing the ability to extract micro-scale vehicle data (at 1Hz resolution) from a real time database of equipped vehicles and for disseminating key information to other (equipped) drivers on the same facility depending on their location. In other words, two general applications using archival and real-time data constitute the core of this study. This research involved the testing of multiple traffic mobility and safety applications with the archival data explained in Section 1.3, including the identification of hazardous locations on the road network, estimating shock wave speed using platoons of, identifying speed profiles at a signal using probe report density and exploring the estimation of free flow speed (FFS) from a driver perspective. In regards to the real-time data, the focus was on defining and tracking events that would motivate the instantaneous generation of safety messages from the vehicle to the "cloud" and the resending of appropriate warning messages to other vehicles in the neighborhood of the impacted (or sender) vehicle. This process is termed VIV for Vehicle-Infrastructure-Vehicle. Just like the field of automated incident detection algorithms on freeways (Parkany 2005), a proper definition of what constitutes an event is critical. On the one hand, generating and sending too many events will needlessly overwhelm the communications bandwidth. Conversely, missing important events may lead to safety or mobility problems. This balance should always be considered when designing VIV systems.

3

1.2   RESEARCH OBJECTIVES

This research is exploratory in nature, and comprises two principal objectives:

1. Test the feasibility of a new on board vehicle unit (OBU) for potential applications related to safety and mobility. Off-line applications have focused on the generation of a region-specific database of second by second vehicle trajectories that can support fundamental traffic and safety research in the areas of high crash location identification, shock wave development and propagation, arterial speed distribution near traffic signals and free flow speed estimation.
2. A secondary objective was to upgrade and pilot-test the existing vehicle trajectory collection tool (VTT) for testing the feasibility of a VIV system in communicating timely and spatially sensitive messages to vehicles upstream of a recurring or non-recurring bottleneck queue. If successful, this work could support the DMA-INFLO work funded by USDOT and contribute to the ongoing research on connected vehicles sponsored by FHWA (Connected Vehicle 2016; CV Pilot 2016).

1.3   RESEARCH APPROACH

1.3.1 Description of the i2D System

Our approach relies on an existing, functional on-board diagnostics (OBD)-connected, GPS enabled, in-vehicle trajectory data collection tool developed in partnership with the Technical University of Lisbon and Live Drive (Live Drive 2016). The device, labeled i2D (for **i**ntelligence to **D**rive) does more than collecting trajectory data. It is equipped with an accelerometer that reports three-dimensional accelerations, and when the vehicle is registered it can provide good estimates of instantaneous fuel consumption (i2D 2013). More importantly, all data are gathered and archived while the vehicle is in motion at a resolution of 1Hz. Currently, i2D vehicles are equipped to gather trajectory data in four continents including Europe, Asia, North America, and South America. Figure 1 depicts the various components of i2D and their placement in a vehicle. These include an OBD-II connector, the OBU (circled), and the GPS antenna.

Figure 1 - Components and Installation of i2D System

Figure 2 depicts how the various components within the i2D system operate. The i2D device includes additional sensors, a memory card, a high-end processor, and a cellular network interface. All OBU raw data are collected at the 1Hz resolution and are transmitted over the mobile network every 23 seconds to a secure server and archived in a standard format for further analysis. Then, processed data is displayed on a secure website for the driver to access trip based information immediately after completing the trip.
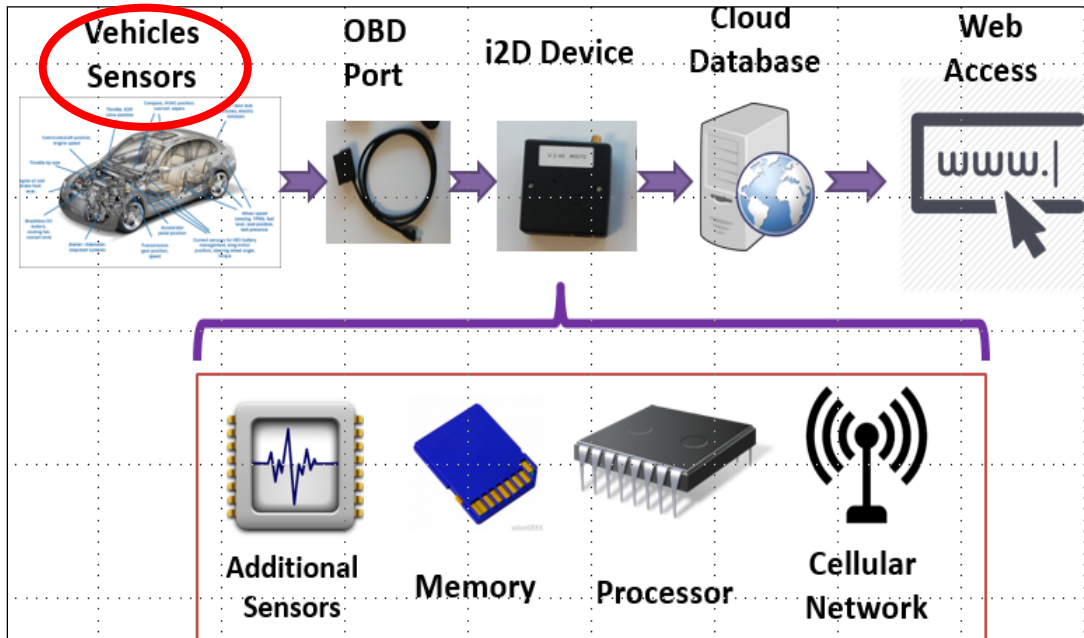
Figure 2 - In Vehicle System Data Path and Routing

Figure 3 shows the current capabilities of the i2D system in generating and visually displaying the micro-scale vehicle data. It shows a screenshot of what the driver accessing his/her private website

would observe when clicking on a specific trip and requesting the trip details. In the original screen (not shown), summary trip information such as start and end time, duration, distance traveled, fuel used, etc., is provided. Exhibit 3 on the right depicts timelines of, from top to bottom: (a) second by second vehicle speed (blue) and engine rpm (green); (b) 3D accelerations in the longitudinal, lateral and vertical dimensions indicated by the various line colors; (c) instantaneous and cumulative fuel consumption for a single trip whose origin, destination and path highlighted on the left side of the graph; and (d) the path elevations as measured by a barometric altimeter.

Fuel consumption is estimated for the specific vehicle making a given trip. When registering a vehicle for the i2D system, detailed information such as age, model year, model type, and engine size are entered in the secure website. Good estimates of fuel use can be made at the instantaneous and trip average level based on this information and the measured vehicle dynamics. The trip shown in Exhibit 3 was made on city streets and an urban freeway in Raleigh, NC. Those data and others not shown on the graph are all archived by trip for the driver (note that a trip begins as the engine is started, and ends when the ignition key is off). This data are used to develop driver-specific reports of fuel use and acceleration profiles. In addition, several icons on the top bar (not displayed here) are used to track specific events on the trip. For example, a displayed X speed icon provides a time stamp and location whenever the vehicle speeds is at or exceeds X mph. The STOP icon will indicate stopped vehicle positions with the engine is still running. Finally, there is a trip simulator feature (circled) that runs the actual trip and depicts the various events as travel is unfolding.
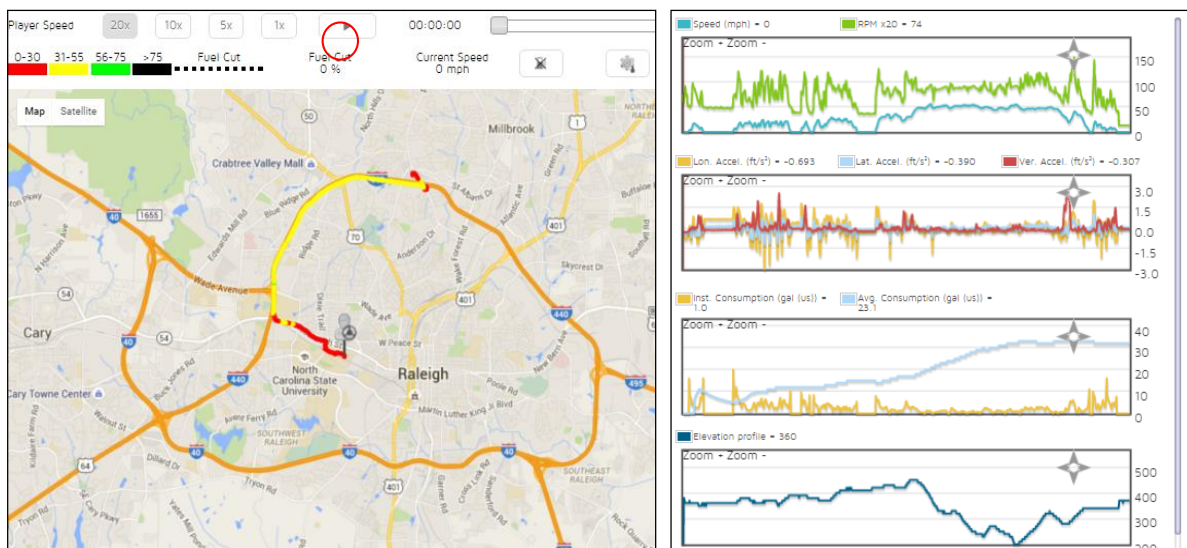
Figure 3 - i2D Trip Data Displays and Travel Dynamics Profiles

1.3.2 Data Description

This section describes the actual data that generated by the system, and the potential applications of the data for various mobility and other purposes. Figure 4 is an efficient way to show the data, which moves from raw to processed to "value" data. Raw data includes:
- Location
- Speed
- Engine temperature
- 3D accelerations
- Altitude
- Barometric air pressure

In turn, those data are processed to yield:
- Fuel consumption
- Road grade
- Trip based statistics (average, max speed, distance traveled, travel time)
- Event-based statistics (heavy accelerations and decelerations, excessive speed, stopping, danger events, etc.)

Finally, the processed data can be agglomerated to generate the following applications:
- Driving profiles
- Personalized driver aids
- Fleet management
- Congestion monitoring on the network
- Identification of high potential crash locations
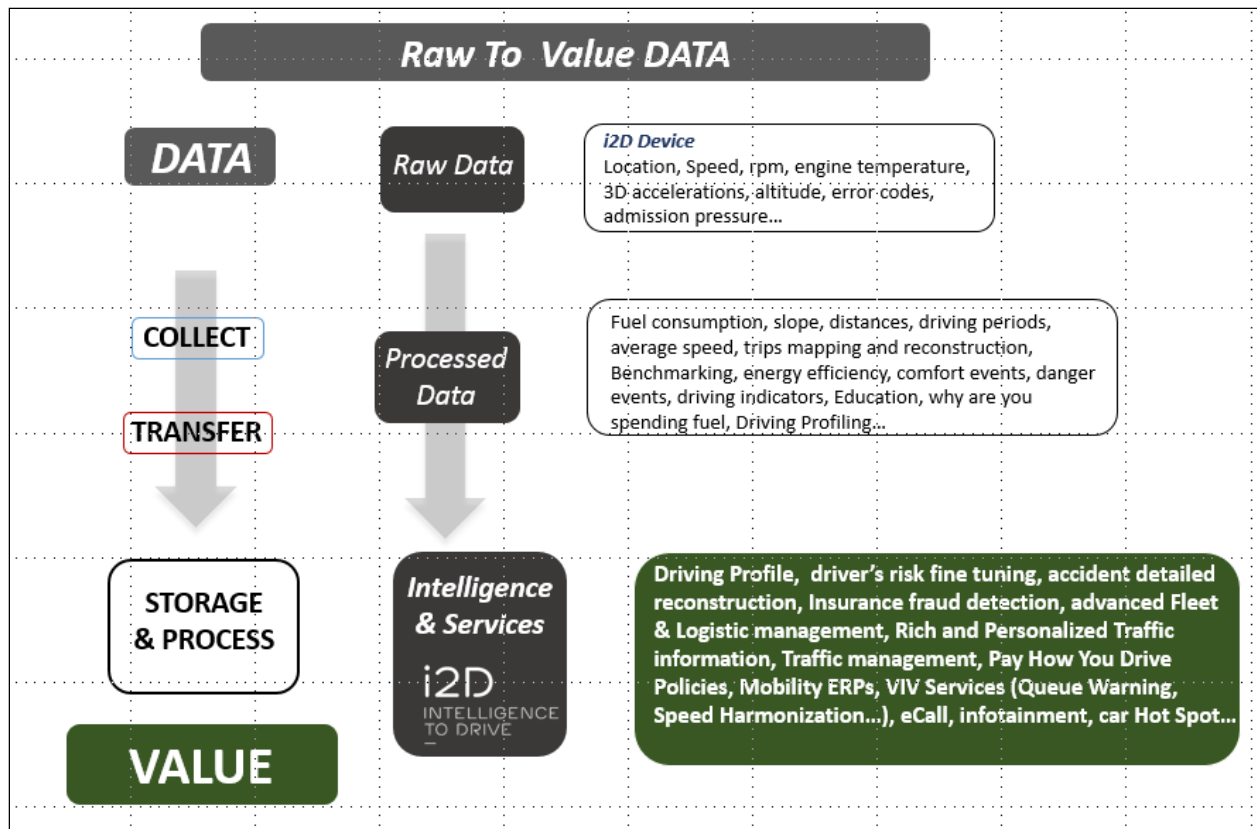
Figure 4 - Types and Utilization of Various Data Elements and Classes

1.3.3 Research Tasks

As indicated in the project proposal, this research involved the following series of tasks.

Task 1. Development of functional system requirements: This task lays out the functional requirements of the VTT database, including reporting frequencies, parameters to be collected, methods for data cleaning and imputations, and database access.

Task 2. Experimental design plan: In this task, the details of the experiments with archival data are laid out, including site selection, number of vehicles to be equipped, driver recruitment (if needed), data collection times and durations, etc. Most experiments were conducted in a naturalistic driving environment where data are collected while drivers are using their vehicles undisturbed. However, the shock wave experiments needed to be controlled to enable the research team to track its path by observing multiple i2D-equipped vehicles in the same vicinity.

Task 3. Conduct and analyze experiments with the archival data: The experimental plan in Task 2 is implemented in this task

Task 4. Develop and Test VIV capabilities: This task develop and test emerging capabilities of i2d to deliver targeted messages to vehicles that are equipped with the i2d system. The focus here was on event definition, and event tracking using the archival data.

8

Task 5. VIV Experiments: These controlled experiments included near-real time message generation based on vehicle-extracted data, execution of an algorithm to formulate a targeted message (exclusively queue warning) and delivery of messages to other targeted vehicles.

Task 6. Draft Final Report: Principal conclusions and findings of the study are summarized in this final report.

1.4   REPORT ORGANIZATION

This report is organized as follows. Following this introductory chapter, Chapter 2 discusses the functional description of the system starting from the Research Platform to generating the database that enables all following analyses with archival data. Chapter 3 describes the various tools developed in this study to facilitate the conduct of the application studies. These include the object-oriented SQL database, the interface with a GIS layer of the network to select specific trips and enable geo-fencing and a trajectory analysis tool to test the VIV events. Chapter 4 focuses on the application side of the archival data, describing the experiments and results from four distinct mobility and safety applications. Chapter 5 describes the additional in-vehicle features involving the VIV system including the i2D App and setting of message, in addition to some preliminary results. Chapter 6 gives an overview of the study, some of the key findings and conclusions and recommendations for future work.

2.0   FUNCTIONAL DESCRIPTION OF THE I2D SYSTEM

This section of the report describes the different functional areas of i2D technology. This chapter focuses on data collection methods and data management and storage, including the flow of data into different platforms and procedures used to generate meaningful information.

This chapter consists of four sections. At the beginning the web-based data storage methods, including the database housing i2D second by second raw data is discussed. In the second section, we present the process developed by the research team to download the raw data from the cloud and upload them into a local SQL database for faster processing and analysis. At the end of this section, we discuss high-level statistics that we have observed based high resolution data collected, followed by some discussion of Vehicle to Infrastructure to Vehicle (VIV) capabilities.

2.1   I2D UNITS AND WEB-BASED PLATFORMS

As mentioned in Chapter 1, i2D is an integrated Car Telematics solution for collecting, transferring, and processing data from Light Duty Vehicles (LDV), primarily designed to support services and policies associated with road mobility (virtual tolling, Pay As You Drive), road safety (Usage Based Insurance - Pay How You Drive), and energy efficiency by monitoring and quantifying driving behavior. The development of this solution was integrated into the framework of the project "I2D – Intelligence to Drive," promoted by ITds, an IT company and IDMEC - IST Mechanical Engineering Institute of Instituto Superior Técnico, Technical University of Lisbon, and co-funded by a Portuguese Renewable Energies Innovation Fund (FAI - Innovation Aid Fund). The main components of an i2D monitoring devices consist of:

- *An On-Board Unit (OBU) plugged to and powered from the vehicle OBD Port. This OBU integrates an OBD interface to collect standard OBDII/EOBD PIDs (Parameter ID), a GPS receiver, a tri-axial accelerometer, a barometric altimeter, a 4GB internal memory, a General Packet Radio Service (GPRS) module, and a Bluetooth module for short range communications (Figure 2);*
- *Machine to Machine (M2M) communications (integrated into the OBU) with automatic national and international roaming;*
- *A central server for processing and storage; and*
- *A set of web-based user interfaces.*

The OBU is available with both internal and external GPS antennas. The internal antenna simplifies the installation process at the cost of a slightly reduced GPS performance. Once connected, the OBU immediately starts recording data and stores it in the internal memory. As soon as a data connection is available, the data is sent through the GPRS connection. Although close to real time, (< 2 seconds delay), communications are possible and in practice, data is sent every 23 seconds in order to reduce communication costs. A second communication channel is

available for real-time communications, such as for Vehicle-Infrastructure-Vehicle (VIV) or Vehicle to Vehicle (V2V) applications. The data is processed and stored at four main levels:

1. In the OBU, data is collected from different sensors at different frequencies. The signals are locally filtered and stored at 1 Hz.
   a. *GPS: ~1Hz*
   b. *OBD: From 1 to 20 Hz per PID (Parameter ID) depending on OBD protocol*
   c. *Altimeter and accelerometer at 100Hz*
2. Raw data at 1 Hz is sent to the server and stored, grouped in trips.
3. From the raw data a second level of data is obtained, using proprietary algorithms, that includes: fuel consumption, CO_2 production, estimates of gaseous pollutant emissions, slope (grade), distances travelled, driving periods, driving events (stops, predefined alerts over speed, rpm, accelerations, etc.), and average speed.
4. The final level of data applies to trips taken, including energy efficiency for each trip, trip mapping and reconstruction, benchmarking and driver profiling, including why and where the driver is wasting fuel.

Every 23 seconds the data is sent to the cloud to be stored in a secure database. The team developed three separate web interfaces (websites) via which users and researchers can manage and download the high resolution trajectory data. Each web interfaces interacts with the cloud-based database based on its given purpose and design, including querying the database and providing the users with requested information. A user is either a driver who has an i2D device in his/her vehicle, or a fleet manager that oversees a group of drivers.

The first web interface (http://app.i2d.co) is designed to provide users with relatively high-level summaries of the data that has been created. Throughout this report, this specific web interface will be referred to as the *Fleet Manager and Drivers Website*. In this web-based interface, a number of computations are performed on high resolution trajectory data to generate meaningful information such as overall safety and comfort scores. As mentioned earlier in this report, trip summaries along with map visualization are provided to assess trips that have been made. When registering a device, there is an option to make it autonomous or part of a fleet, and in this case, the fleet manager can have access to data from all the vehicles. Fleet drivers also have access to this web interface and they can see their own statistics.

The processed data includes comfort scores, an eco-driving index and etc. Additionally, the users are provided with some details about their trips. The details of each trip includes second by second presentation of their trip characteristics such as speed, GPS coordinate, etc. In the case of registering i2D devices as a fleet of vehicles for data collection purposes, this web interface allows the administrator (fleet manager) to manage drivers, vehicles, and i2D devices. Figure 5 shows a screenshot of fleet manager and drivers' website.
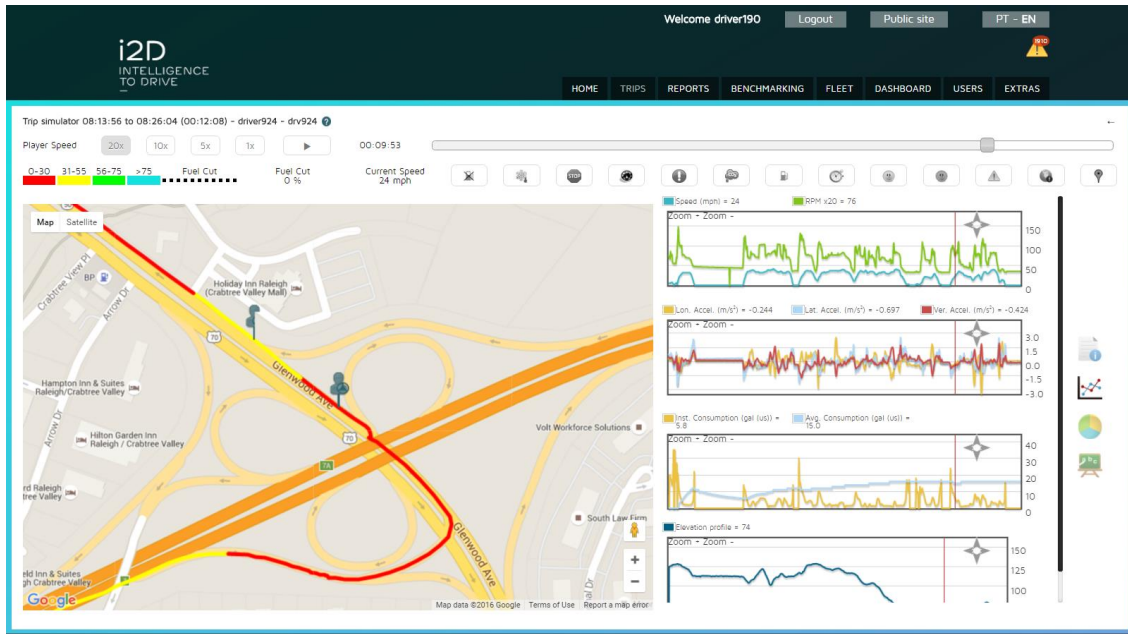
Figure 5 – Fleet Manager and Drivers Web Interface Showing Drivers Trips on the Map with Several Statistics

The second web interface (http://research.i2d.co), called *Research Platform* is designed to allow the users to download high resolution trajectory data. This web interface directly queries the main database in the cloud to provide users with requested information. The only information that users can download is high resolution trajectory data on a second by second basis. In this platform, none of the high level statistics can be viewed by users. The data are transmitted in a CSV comma delaminated file format. The research platform also provides the users with the ability to download data and other options such as transferring the raw data to an email account. In this case, users do not need to wait on the webpage for their download to start. Figure 6 shows a screenshot of the research platform web interface. This platform filters the trips made in a fleet and allows the user to select trips based on a number of criteria such as date, time, day of the week, etc.
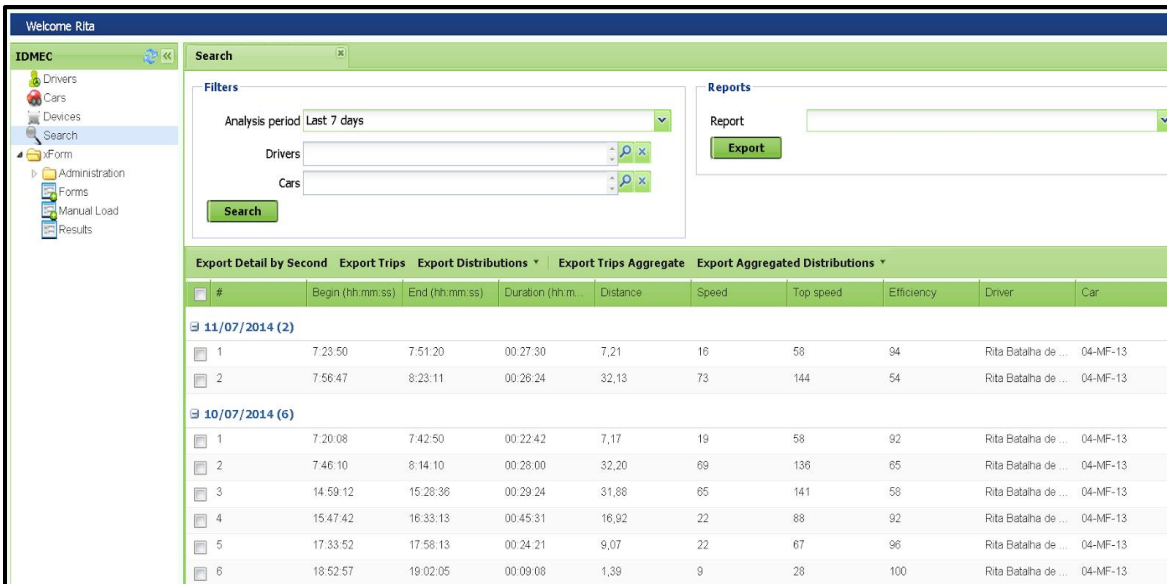
Figure 6 – Research Platform Web Interface

In the research-focused web interface, the following information is available for export:
- *Raw data at 1 Hz (GPS coordinates, OBD data, Accelerometer XYZ, Altitude, vehicle ID, driver ID)*
- *Trip aggregated data*
- *Distributions per trip or aggregated number of speed bands, vehicle specific power, and fuel consumption*

A third web interface (https://www.i2d.co/i2dpubportal/home_private.xvw) is designed to allow the users and researcher to observe and evaluate Vehicle to Infrastructure to Vehicle (VIV) communications. This web-based platform is called *VIV Control Panel*. More details regarding this web platform is provided later in this chapter in VIV functionality section. Figure 7 shows a screenshot of this web interface.
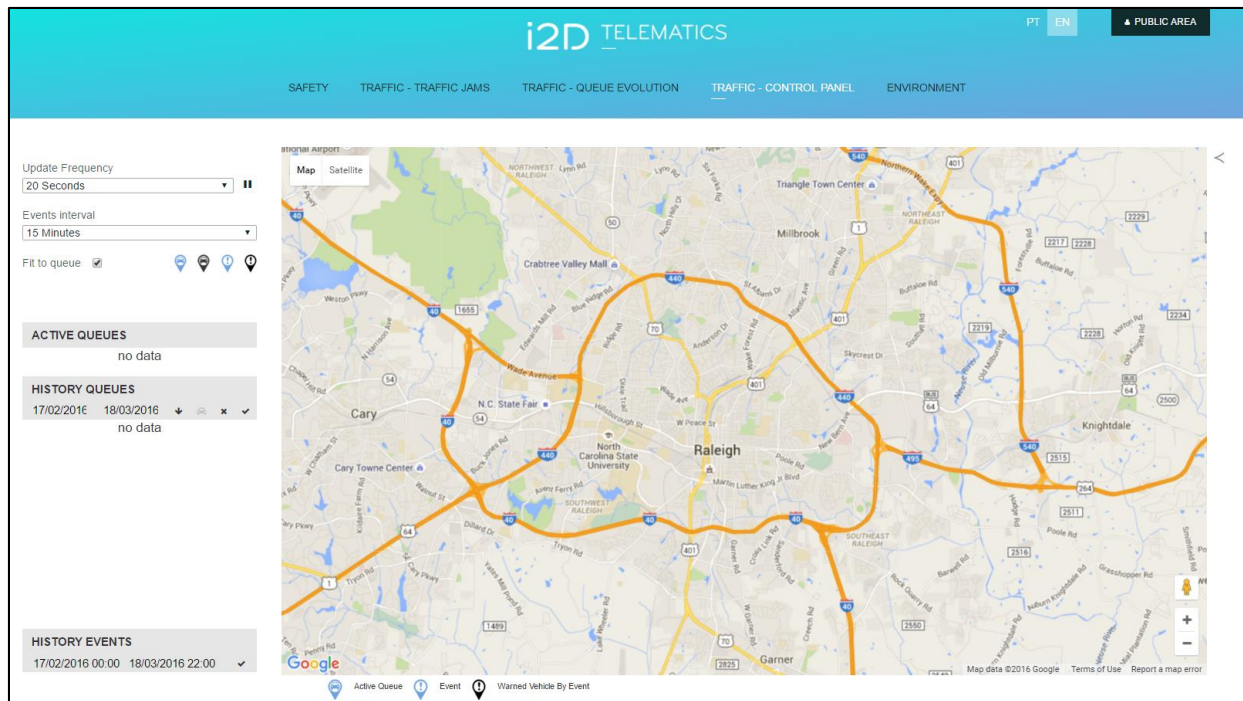
Figure 7 – Vehicle to Infrastructure to Vehicle Control Panel Web Interface

2.2   ATA MANAGEMENT AND STORAGE

The i2D units record a set of parameters that are coming from both onboard vehicles sensors as well as i2D the built-in sensors of units in a second by second format. As mentioned, the sensor readings along with a set of identifier parameters, such as vehicles ID, driver ID, date, and time of the trip, are stored in an online database, which is hosted in a secure server in Portugal. As mentioned, the i2D support infrastructure consists of a research-focused web-based platform via which the research teams are able to download these second by second data. The full list of 64-recorded parameters in the online database is presented in Table 1.

Table 1: The full list of raw stored parameters in the online i2D database

No.	Parameter	Sensor Location	Description
1	TRIP_ID	NA	A unique ID is assigned to each trip made
2	START_DATE	NA	Start date of the trip
3	START_TIME	NA	Start time of the trip
4	END_DATE	NA	End date of the trip
5	END_TIME	NA	End time of the trip
6	Seconds	NA	This is an integer number which shows the second from start of the trip that the data is associated with
7	Time	NA	GPS UTC time
8	Latitude	i2D	GPS Latitude Coordinate
9	Longitude	i2D	GPS Longitude Coordinate
10	Course	Server*	Course is the compass computed based on coordinates of two consequent seconds of data

15

No.	Parameter	Sensor Location	Description
11	GPS Speed	i2D	The speed computed based on GPS coordinates of two consequent seconds of data
12	Altitude GPS	i2D	The altitude that is reported by GPS
13	Barometer Pressure	i2D	The pressure reported by barometer
14	Barometer Altitude	i2D	The altitude that has recorded by barometer
15	Speed (km/h)	OBD	The speed reported by OBD in km/h
16	Engine Load	OBD	The percent turtle that is exposed to the engine.
17	RPM	OBD	The amount of revolution per minute that the engine is operating
18	Battery State (SOC) (%)	OBD	If available, as reported from OBD (only for hybrid vehicles)
19	Manifold air pressure	OBD	Air pressure in the engine intake manifold
20	In-take air temperature	OBD	Air temperature in the engine intake manifold
21	Air mass flow (g/s)	OBD	Intake air mass flow.
22	Throttle position (%)	OBD	Throttle positions as reported by OBD
23	Coolant temperature (?C)		Engine coolant temperature as reported by OBD
24	Acceleration X (m/s2)	i2D	Longitudinal Acceleration reported by i2D accelerometer
25	Acceleration Y (m/s2)	i2D	Lateral Acceleration reported by i2D accelerometer
26	Acceleration Z (m/s2)	i2D	Vertical Acceleration reported by i2D accelerometer
27	Slope (m/m)	Computed Value	Road grade
28	Injection Cut (s/n)	OBD	Engine running under fuel cutoff conditions
29	Instant Consumption (L/100km)	OBD	Instantaneous consumption in l/100km
30	Accumulated Consumption (L)	Computed Value	Accumulated consumption in liters
31	Average Consumption (L/100km)	Computed Value	Average consumption in l/100km
32	Gear shifting	N/A	Not available
33	Idle (s/n)	OBD	Engine running under and car stopped
34	Cold Running (s/n)	OBD	Engine running under cold conditions
35	FUEL_TYPE	Computed Value	Fuel in use
36	DISPLACEMENT	Computed Value	Engine displacement
37	CAR_DATE	Computed Value	Vehicle registration date
38	VSP (W/kg)	Computed Value	Vehicle specific power
39	Road Type	Computed Value	Type of road if reported in OpenStreetMaps
40	EXTREME_BRAKE_EVENTS	Computed Value	Driving events
41	EXTREME_ACCELERATION_EVENTS	Computed Value	Driving events
42	NEUTRAL_GEAR_EVENTS	Computed Value	Driving events
43	POSITIVE_VSP_INTEGRAL (J/kg)	Computed Value	Not applicable
44	EXCESS_RPM_EVENTS	Computed Value	Driving events
45	EXCESS_CUSTOM_RPM_EVENT	Computed Value	Driving events
46	EXCESS_SPEED_50_EVENTS	Computed Value	Driving events
47	EXCESS_SPEED_90_EVENTS	Computed Value	Driving events
48	EXCESS_SPEED_120_EVENTS	Computed Value	Driving events
49	COMFORT_Y	Computed Value	Driving events
50	SAFETY_Y	Computed Value	Driving events
51	COMFORT_Z	Computed Value	Driving events
52	HIGH_CO2_EMISSION_EVENT	Computed Value	High CO2
53	HIGH_FUEL_EVENT	Computed Value	High Fuel consumption

No.	Parameter	Sensor Location	Description
54	HIGH_VSP_EVENT	Computed Value	High Power
55	ZERO_SPEED_EVENT	Computed Value	Vehicle stopped
56	DRIVER_ID	Computed Value	Unique driver identifier
57	GENRE	Computed Value	Driver Gender (if available)
58	AGE_GROUP	Computed Value	Driver age group (if available)
59	EXPERIENCE	Computed Value	Years of driving license (if available)
60	CAR_ID	Computed Value	Unique vehicle identifier
61	Instant Consumption (g/s)	Computed Value	Instantaneous fuel consumption in grams per second
62	IN_AREA	Computed Value	Location identifier
63	AREA_IDS	Computed Value	Location identifier
64	REFERENCE_POINTS	Computed Value	Location identifier

* The data field is computed based on other data on the server

In order to speed up the data analysis, the data are ultimately stored in a local MySQL database. For this purpose, the research team have downloaded all the high resolution raw data from the online platform. The local database contains 8 tables: *raw_data*, *driver_data*, *calculated_data*, *trip_data*, *filtered_data*, *excluded_days*, *in_corridor_output*, and *thru_corridor_output*. It contains all the data for more than 32 million seconds of trip data for 38,875 unique trips. The *raw_data* table is the initial table that all of the data that has been downloaded from the research platform are uploaded to. It has 63 columns for each of the recorded parameters listed above, except for REFERENCE_POINTS, and contains a unique ID column, *uid*. The other tables contain the following data:

- *driver_data* contains information about each driver in the fleet
- *calculated_data* contains information that has been calculated within MySQL for each second of data, such as distance traveled and acceleration
- *trip_data* contains information about each unique trip, such as distance traveled and total trip time
- *excluded_days* contains days that are holidays or extreme weather days
- *filtered_data* contains all the second-by-second data for trips that lasted longer than five minutes, traveled farther than one mile, and do not occur on one of the days in the *excluded_days* table
- *in_corridor_output* and *thru_corridor_output* contain the output from the ArcGIS tool developed by the research team discussed in Chapter 3

Figure 8 shows all the tables in the local database. It also highlights the connections between each table in this database that the queries are using. The connection between tables in this database are vital because several tables are involved in the majority of information and statistics that have been extracted by SQL queries. Appendix A presents the detailed specification of each table in the local SQL database as well as queries that have been used to populate the tables with raw data.
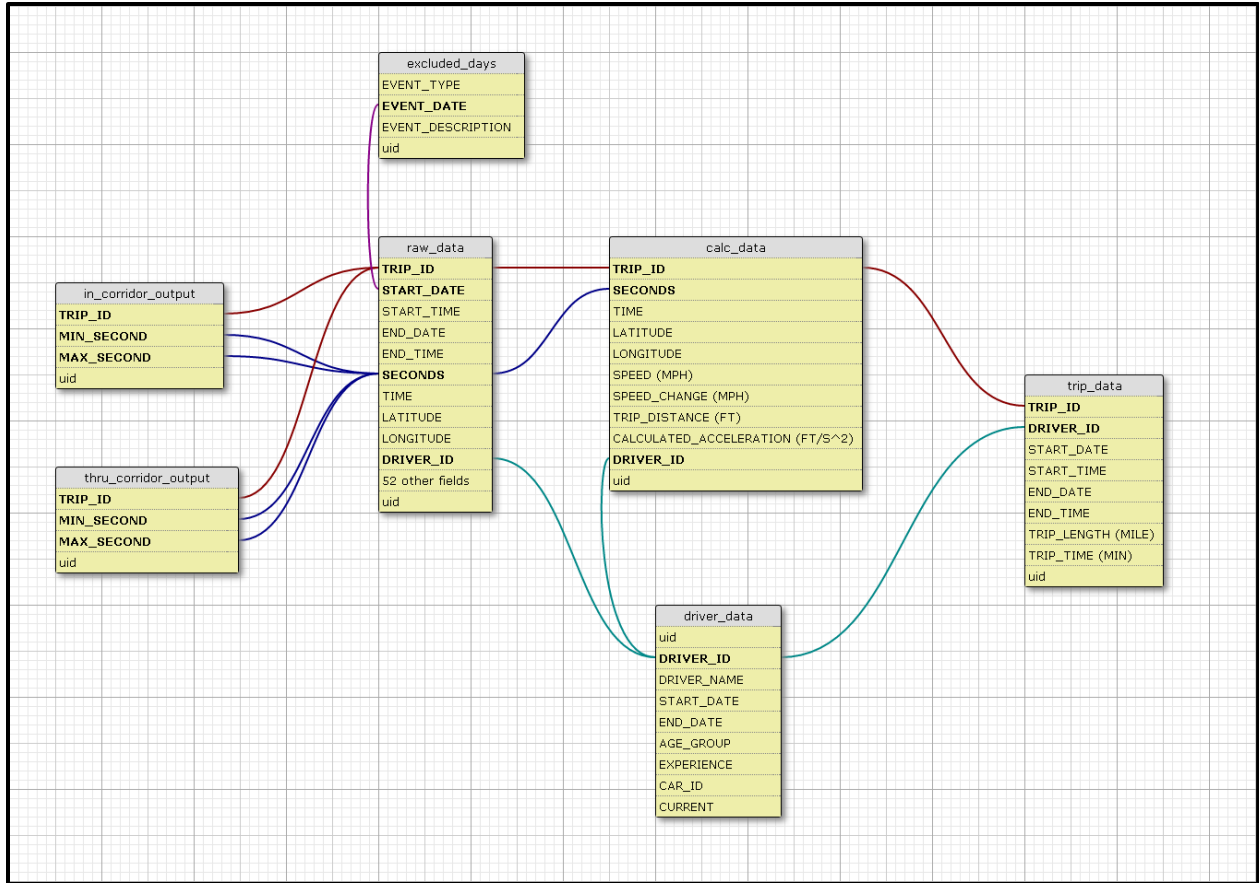
Figure 8 – Local SQL Database Tables Relationship

The research team used an open source web-based package called PHPMyAdmin to provide users with a user-friendly interface that queries the SQL database. Figure 9 shows a screenshot of the PHPMyAdmin interface.

18
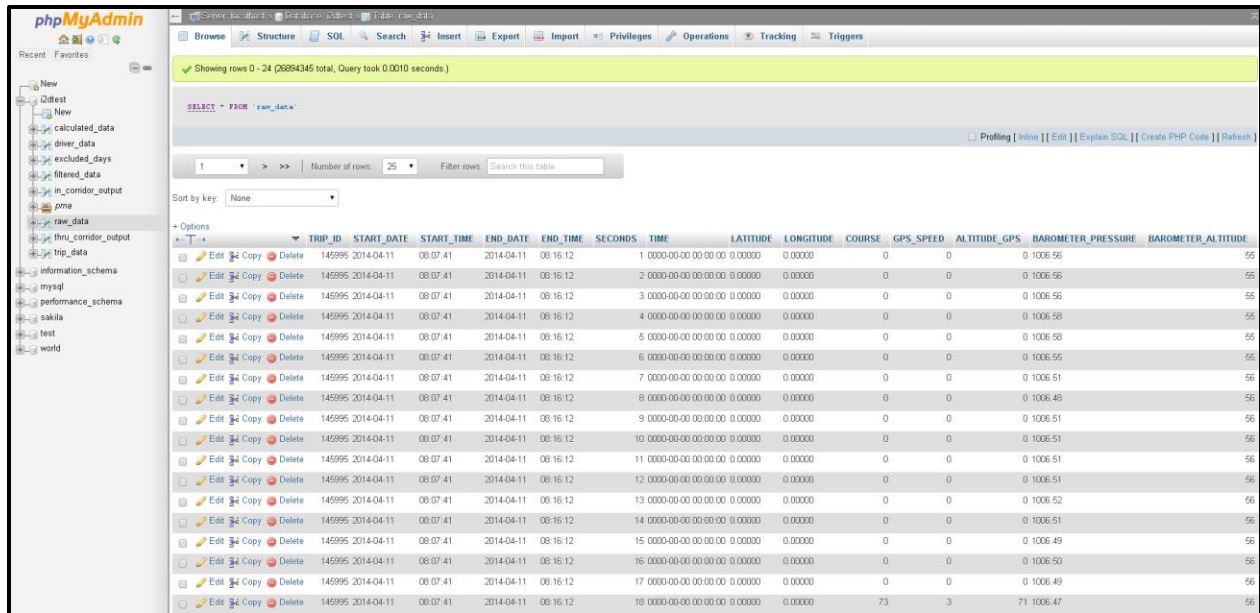
Figure 9 – PHPMyAdmin Interface

2.3 DATA COLLECTION AND CLEANING PROCESS

In order to perform different types of analyses within high resolution data collected, a four step process was designed by the research team to transfer data from the i2D cloud-based database to the local and more manageable SQL database. The local database enables the research team to conduct queries much faster. A set of tools is designed to interact with the local database to generate desirable outputs, which is discussed in Chapter 3.

2.3.1 Downloading the Raw Data

The research platform web interface provides the option to download second by second data for each trip. It also provides a batch mode download capability that allows the analyst can download data for multiple trips. There are two ways to download the data: 1) directly from the research platform and 2) via email. The latter option allows users to enter an email address and then the research platform sends the data as an attachment into that email address. Emailed and downloaded files are in a semicolon delimited CSV format. In this project, the research team has used the research platform to download all data that has been collected for the fleet of vehicles.

2.3.2 Cleaning the Downloaded Data

The research team has written a VBA script in Microsoft Excel to automate the necessary formatting steps to prepare the data for uploading into the MySQL database. The script runs on each CSV file in a local folder. It first converts the text to columns based on the semicolon delimiter and deletes the first row, which contains the name of each column. Next, the script changes the format of the four columns that contain date and time data to match the required format

19

for the MySQL database. Finally, it fills in all of the blank cells with "NULL" because MySQL cannot read blank cells. It is important to not open the CSV files once they have been cleaned with the VBA script because Microsoft Excel reverts the date and time columns back to Microsoft's format.

2.3.3 Uploading Data into the Local SQL Database

The research team has also written a batch file that uploads each CSV file from a designated folder into the MySQL database, then moves them to another folder with all of the other uploaded files. It is significantly faster to upload the CSV files through the MySQL command line than to use the PHPMyAdmin graphical interface. It should be noted that the research team has only used a command line SQL interface to upload the high resolution trajectory data to a local SQL database.

2.3.4 Updating the SQL Database

Once all of the data has been uploaded, there are several other tables that must be updated to reflect all of the new data. A query has been written for each table that needs updating in the following order – '*driver_data*', '*calculated_data*', '*trip_data*', and '*filtered_data*'. For example, the *trip_data* table, whcich contains high level summary information for each trip, should be populated based on a query that digs into the main table *raw_data*.

2.4 HIGH LEVEL STATISTICS GENERATED BY I2D SYSTEM

The research team organized a team of 25 volunteers who agreed to install i2D devices in their vehicles. As discussed, all the data recorded by the fleet were stored in the research platform. Nearly 1.6 million seconds of raw data were generated by the fleet each month. During the 20 month duration of the data collection period, the research team collected about 32 million second by second high resolution trajectory data points. The research team also provided support to the drivers in terms of installation, troubleshooting, and monitoring the correct performance of the i2D units. Since the i2D units are not commercial level products, some units stopped working due to bugs in the firmware software or incompatibility with the CAN bus system of some vehicles. The research team monitored the status of the i2D devices on a weekly basis to assure healthy working status of the i2D devices in the fleet. The fleet of 25 volunteer drivers was modified during the course of this project. This was due to graduation and leave of some students.

In this section a high level summary of the collected and processed data is presented. The summaries discussed in this section are generated by designed queries written in SQL language. The high level statistics that are provided in this section reveals insights into several trip characteristics such as trips distance, trips durations, trips distribution for each weekday and month of the year, etc. Table 2 shows a summary of the data that has been collected.

Table 2 – High Level Statistics of Collected High Resolution Trajectory Data

Data Collection Period	April 1 2014 to December 31 2015
Data Collection Duration	20 Months
Total number of Trips	38,875 Trips
Total Number of (different) Drivers	24 Drivers
Total Number of Seconds of High Resolution Data Recorded	32,351,971 Seconds (~9,000 Hours)
Total Miles Driven	277,601 Miles
Average Trips Duration	13.9 Minutes
Average Trips Length	7.14 Miles

The data collection started in April 2014 and continued to December 2015. In this effort 24 drivers volunteered to participate in the data collection. The data collection effort recorded about 9,000 hours of driving information that includes 277,601 miles. It is expected that the longer trips in terms of length tend to be associated with longer trips in terms of duration. Table 3 presents the duration and length distributions of the trips that have been made during the data collection effort. Table 3 is color coded to be presented as a contour plot. As shown only 0.4% of the trips lasted more than two hours and more than 100 miles in length.

Table 3: Distribution of Trips by Travel length and Duration

DURATION	TRAVEL LENGTH						
	<=1 MILE	1-5 MILES	5-10 MILES	10-25 MILES	25-50 MILES	50-100 MILES	>100 MILES
<=1 MIN	928	0	0	0	0	0	0
1-5 MIN	3,939	1,616	0	0	0	0	0
5-10 MIN	453	6,454	297	2	0	0	0
10-20 MIN	99	3,333	3,035	481	0	0	0
20-30 MIN	13	164	801	1,749	67	0	0
30-60 MIN	9	28	113	776	638	49	0
60-120 MIN	3	5	5	22	41	151	86
>120 MIN	4	2	0	0	0	7	145

2.4.1 Impact of Day of Week on Observations

During the data collection timeframe, from April 1, 2014 to December 31, 2015, 38,875 trips are recorded. Table 4 shows the average and standard deviation for durations and lengths of the trips by day of the week. The average of trip durations and lengths are 13.87 minutes and 7.14 miles respectively. As seen, the average trips time and distance are consistent between weekdays. However, the variation of trip durations and lengths is higher on the weekends compared to weekdays, as indicated by standard deviation in Table 4.

Table 4 – Average and Standard Deviation of Duration and Length of the Trips by Day of Week

Day of the Week	Average Of Trip Duration (Min)	Stddev Of Trip Duration (Min)	Average Of Trip Length (Mile)	Stddev Of Trip Length (Mile)
SUNDAY	13.84	21.03	8.29	21.08
MONDAY	14.00	16.80	7.04	15.87
TUESDAY	13.88	15.77	6.74	14.57
WEDNESDAY	13.53	15.75	6.52	14.48
THURSDAY	13.87	17.09	7.11	17.07
FRIDAY	14.21	17.11	7.11	16.00
SATURDAY	13.71	18.41	7.48	17.00
Grand Total	**13.87**	**17.36**	**7.14**	**16.55**

Table 5 presents the distribution of trip lengths as function of day of the week that the trips are generated. It is observed that the duration of the trips does not significantly vary due to the day of the week that they were made. Table 5 shows that there are more trips generated during the weekdays as opposed to the weekends. Table 5 is color coded to be presented as a contour plot.

Table 5: Distribution of Trips by Trip Length and Day of the Week

Day Of Week	Trip Length							Sum
	<=1 MILE	1-5 MILES	5-10 MILES	10-25 MILES	25-50 MILES	50-100 MILES	>100 MILES	
Monday	1071	2739	896	728	157	31	56	5,678
Tuesday	1015	2847	924	727	155	35	46	5,749
Wednesday	1178	2794	904	787	169	33	33	5,898
Thursday	1074	2904	905	771	173	42	49	5,918
Friday	1234	2813	991	874	174	44	48	6,178
Saturday	1134	2013	876	588	126	75	44	4,856
Sunday	964	2020	832	533	114	68	67	4,598
Sum	7,670	18,130	6,328	5,008	1,068	328	343	38,875

Similar to Table 5, Table 6 shows the distribution of trips durations based on the day of the week that the trip is made. There is a slight increase in trips duration on Fridays compared to the other weekdays.

Table 6: Distribution of Trips by Trip Duration and Day of the Week

DAY OF WEEK	TRIP DURATION								Sum
	<1 MIN	1-5 MIN	5-10 MIN	10-20 MIN	20-30 MIN	30-60 MIN	60-120 MIN	>120 MIN	
MONDAY	192	1,123	1,636	1,576	670	384	57	40	5,486
TUESDAY	174	1,107	1,649	1,655	694	380	63	27	5,575
WEDNESDAY	287	1,126	1,640	1,692	684	387	57	25	5,611
THURSDAY	183	1,186	1,769	1,631	639	411	57	42	5,735
FRIDAY	230	1,282	1,632	1,726	768	431	76	33	5,948
SATURDAY	181	1,175	1,264	1,390	474	249	95	28	4,675
SUNDAY	145	1,115	1,397	1,150	467	194	83	47	4,453
Sum	1,392	8,114	10,987	10,820	4,396	2,436	488	242	38,875

2.4.2 Impact of Hour of the Day on Observations

In this section, the focus is on the impact of the hour of day that the trips are generated on the high resolution trajectory data observations. Figure 10 shows the hourly distribution of the trips that have been created. The distribution shows two distinct peak at 8am and 5pm. As mentioned, majority of the drivers in the fleet are students and staff for Institute for Transportation Research and Education (ITRE). As such, the observed peaks were predictable.
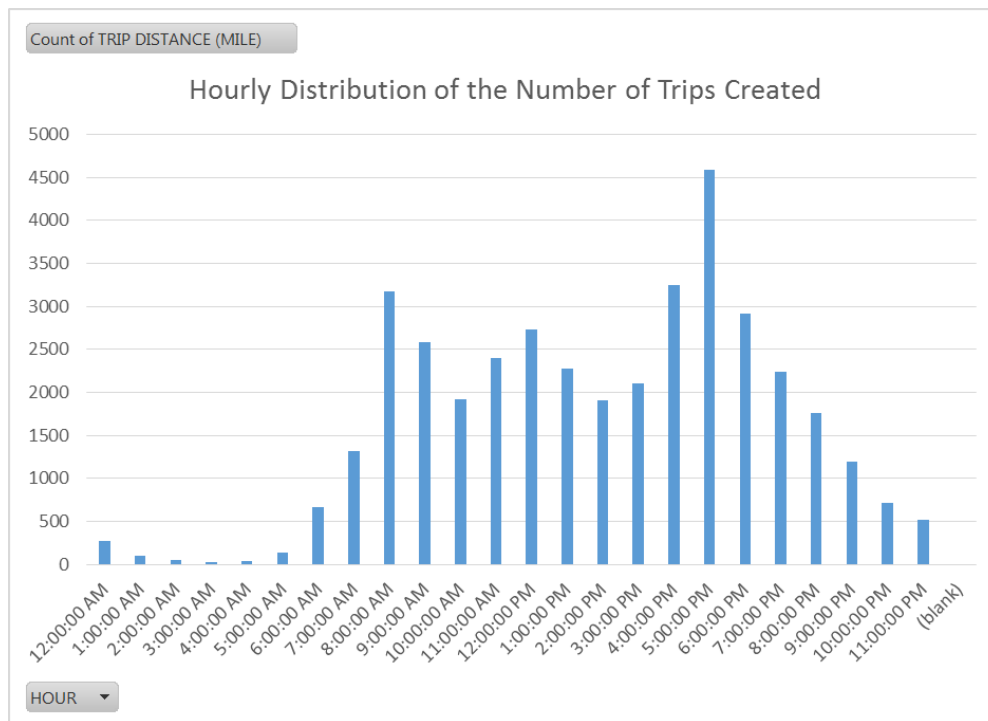
Figure 10 – Hourly Distribution of Number of Trips Generated

Table 7 and Table 8 breaks the hourly number of trips that have been generated based on trip lengths and durations. These observations will help to identify the impact of traffic level on the trip durations. Table 7 shows that the trip lengths are consistent with the trend presented in Figure 10. However, Table 8 shows slightly higher trips durations that have been made in PM peak (5pm).

23

It is assumed that this is caused by traffic congestion in PM peak. The data that have been analyzed in this report contains all the trips that have been made in urban and rural areas. Thus, not every trip has suffered longer duration due to congestion. However, the trend on Table 8 shows slightly higher delays for trips that have been made in the PM peak.

Table 7 : Trip Distribution by Trip Distance and Departure Time

| DEPARTURE TIME | TRAVEL DISTANCE | | | | | | | |
	<=1 MILE	1-5 MILES	5-10 MILES	10-25 MILES	25-50 MILES	50-100 MILES	>100 MILES	Sum
12:00 AM	52	116	53	37	15	1	3	277
1:00 AM	27	34	17	13	2	2	1	96
2:00 AM	16	21	5	2	1	1	0	46
3:00 AM	6	14	0	1	1	1	0	23
4:00 AM	2	8	9	13	1	0	2	35
5:00 AM	32	27	16	34	8	2	16	135
6:00 AM	103	83	93	214	145	10	16	664
7:00 AM	218	419	257	346	43	10	25	1,318
8:00 AM	425	1,659	445	529	81	25	16	3,180
9:00 AM	369	1,337	447	337	40	26	26	2,582
10:00 AM	411	910	308	214	38	23	16	1,920
11:00 AM	590	1,151	385	208	36	18	16	2,404
12:00 PM	668	1,365	386	241	28	22	24	2,734
1:00 PM	508	1,148	363	176	38	20	20	2,273
2:00 PM	412	875	320	206	44	26	20	1,903
3:00 PM	448	913	406	198	78	33	35	2,111
4:00 PM	568	1,549	539	380	137	34	40	3,247
5:00 PM	845	2,158	737	744	67	14	18	4,583
6:00 PM	603	1,428	480	301	72	19	12	2,915
7:00 PM	500	1,086	374	211	47	12	16	2,246
8:00 PM	385	814	273	231	37	10	7	1,757
9:00 PM	246	522	192	193	24	6	7	1,190
10:00 PM	125	307	125	112	40	7	3	719
11:00 PM	111	186	98	67	45	6	4	517
Sum	7,670	18,130	6,328	5,008	1,068	328	343	38,875

The group of participating drivers for this study consists of students and employees of NC State University. As such, two peaks of trip start time are 8am and 5pm. However, due to the expected behavior of the student drivers, there are is considerable amount of trips that have been made during the day between these two peaks. The main reason to observe these trips is that the students are not regular employees working from 8am to 5pm. The interesting observation on trips length comes from the trips that have been made early in the morning at 6am that have relatively longer trips lengths. Overall the peak departure time was 5-6 pm, constituting over 11.3% of all trip departures times. This is akin to the K hourly factor used in transportation planning analyses. Also, the peak trip duration was in the 5-20 minutes in duration, which when combined accounted for over 56% of all trips.

Table 8: Distribution of Trips by Trip Duration and Departure Time

DEPARTURE TIME	TRIP DURAITON								
	<1 MIN	1-5 MIN	5-10 MIN	10-20 MIN	20-30 MIN	30-60 MIN	60-120 MIN	>120 MIN	Sum
12:00 AM	9	76	80	67	22	19	3	1	268
1:00 AM	4	29	29	20	6	7	1	0	92
2:00 AM	1	22	11	7	1	3	1	0	45
3:00 AM	0	13	5	2	1	1	1	0	23
4:00 AM	0	3	6	12	3	9	1	1	35
5:00 AM	5	25	19	26	20	23	3	14	130
6:00 AM	17	61	78	106	128	246	21	7	647
7:00 AM	43	197	193	386	236	220	29	14	1,275
8:00 AM	123	396	900	1,038	507	174	28	14	3,057
9:00 AM	109	460	886	679	311	78	43	16	2,473
10:00 AM	83	448	557	537	175	70	37	13	1,837
11:00 AM	91	603	715	646	216	97	26	10	2,313
12:00 PM	87	704	859	722	222	93	27	20	2,647
1:00 PM	69	515	729	650	178	89	29	14	2,204
2:00 PM	84	399	578	523	175	101	26	17	1,819
3:00 PM	100	449	583	588	194	118	52	27	2,011
4:00 PM	97	619	955	859	369	269	48	31	3,150
5:00 PM	171	791	1,150	1,351	645	431	31	13	4,412
6:00 PM	90	613	796	956	304	121	27	8	2,825
7:00 PM	84	584	648	635	191	79	15	10	2,162
8:00 PM	53	481	504	444	195	65	10	5	1,704
9:00 PM	32	299	370	280	142	52	10	5	1,158
10:00 PM	18	192	196	172	92	35	13	1	701
11:00 PM	22	135	140	114	63	36	6	1	495
	1,392	8,114	10,987	10,820	4,396	2,436	488	242	38,875

2.4.3 Impact of Month of Year on Observations

In this section, the effect of month of the year on the observed data has been investigated. Since the data collection period started from April 1 2014 to December 31 2015, the period during which data has been collected has only one month worth of data for the months of January, February, and March. Considering this fact, the research team felt confident to compare all the months of the year in this section. Figure 11 shows the average number of trips for each month of the year. The average has been made for the month April to December that the research team had access to two months of the year worth of data. As expected, the data showed some seasonal trends, with slightly higher incidence of trips in March and in August.
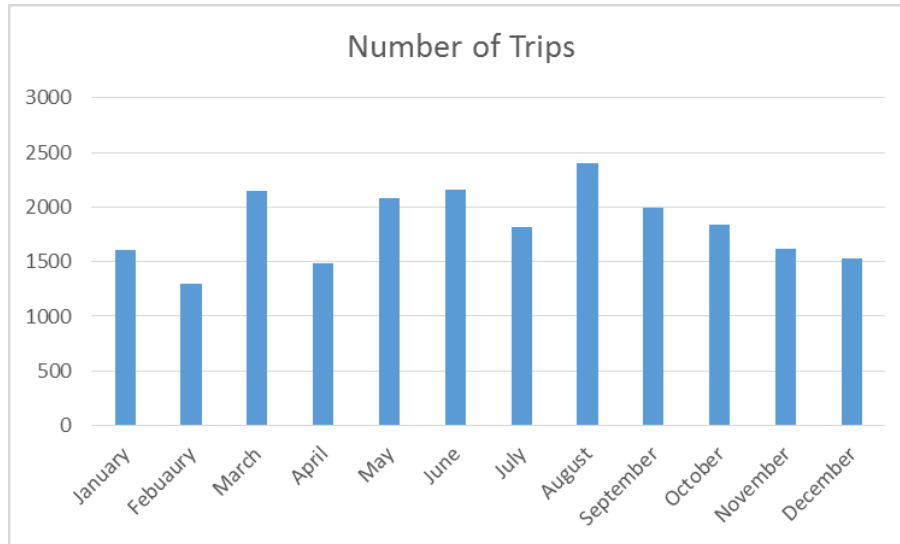
Figure 11 - Average Number of Trips by Month of Year.

Table 9 shows the average and standard deviation of trip length and duration for each month of data collection period. The standard deviation for trips lengths and durations for the months of December, January, and June is higher than other months, which could be due to longer trips related to holiday or special travel. Also, the average trip length and duration is higher for the months of December and June.

Table 9 - Average and Standard Deviation of Trip Length and Duration by Month

Month	Average of Trip Duration (MIN)	StdDev of Trip Duration (MIN)	Average of Trip Length (MILE)	StdDev of Trip Length (MILE)
January	14.1	14.7	7.0	12.4
February	13.3	11.8	6.1	10.1
March	13.8	16.8	6.9	16.3
April	13.0	17.2	6.8	16.9
May	14.0	17.6	7.4	16.2
June	15.2	21.8	8.2	20.4
July	13.5	16.4	7.1	15.6
August	13.6	18.1	7.2	17.1
September	14.1	17.5	6.8	16.5
October	13.6	15.4	6.3	13.4
November	13.3	15.6	6.5	14.4
December	14.3	17.2	8.4	20.2
Grand Total	**13.9**	**17.4**	**7.1**	**16.5**

Table 10 and Table 11 present the breakdown of monthly trips based on their duration and length. Both tables are color coded as contour plot.

Table 10 : Trip Distribution by Trip Duration and Month of Year

Month/Year	Trip Length							
	<=1 MILE	1-5 MILES	5-10 MILES	10-25 MILES	25-50 MILES	50-100 MILES	>100 MILES	
Apr-14	241	401	95	42	22	1	2	804
May-14	497	796	291	201	50	16	6	1,857
Jun-14	473	1,110	446	295	83	29	17	2,453
Jul-14	375	681	317	215	53	12	22	1,675
Aug-14	495	1,242	503	290	76	16	31	2,653
Sep-14	377	789	297	268	37	7	15	1,790
Oct-14	396	679	300	231	21	9	13	1,649
Nov-14	349	818	240	218	36	11	19	1,691
Dec-14	321	775	293	199	64	27	19	1,698
Jan-15	275	757	255	239	67	9	7	1,609
Feb-15	242	604	217	184	45	3	2	1,297
Mar-15	542	939	297	279	56	15	22	2,150
Apr-15	436	1,014	321	281	65	19	26	2,162
May-15	429	997	379	360	71	33	30	2,299
Jun-15	296	926	288	242	58	21	38	1,869
Jul-15	340	952	320	288	31	16	11	1,958
Aug-15	428	1,036	319	275	39	30	19	2,146
Sep-15	404	1,159	304	232	44	21	24	2,188
Oct-15	325	1,044	332	277	40	11	10	2,039
Nov-15	218	844	245	157	59	10	3	1,536
Dec-15	211	567	269	235	51	12	7	1,352
	7,670	18,130	6,328	5,008	1,068	328	343	38,875

Table 11 – Trip Distribution by Trip Distance and Month of Year

Month/Year	Trip Duration							
	<1 MIN	1-5 MIN	5-10 MIN	10-20 MIN	20-30 MIN	30-60 MIN	60-120 MIN	>120 MIN	
Apr-14	49	231	231	212	38	36	3	4	804
May-14	82	465	509	503	145	130	21	2	1,857
Jun-14	71	493	663	696	272	209	36	13	2,453
Jul-14	71	367	405	462	191	144	23	12	1,675
Aug-14	68	515	980	636	211	191	32	20	2,653
Sep-14	39	387	466	487	257	125	19	10	1,790
Oct-14	61	392	420	438	220	92	14	12	1,649
Nov-14	50	396	493	448	177	99	14	14	1,691
Dec-14	40	349	491	494	146	131	29	18	1,698
Jan-15	41	315	439	467	199	129	13	6	1,609
Feb-15	28	270	353	364	189	87	5	1	1,297
Mar-15	149	454	554	555	255	134	37	12	2,150
Apr-15	92	455	596	596	265	116	26	16	2,162
May-15	87	466	606	590	336	149	47	18	2,299
Jun-15	60	402	544	475	220	103	33	32	1,869
Jul-15	91	390	575	577	221	72	24	8	1,958
Aug-15	95	427	603	630	247	90	41	13	2,146
Sep-15	72	413	696	624	237	98	29	19	2,188
Oct-15	72	355	578	639	259	108	21	7	2,039
Nov-15	36	284	456	500	142	100	16	2	1,536
Dec-15	38	288	329	427	169	93	5	3	1,352
	1,392	8,114	10,987	10,820	4,396	2,436	488	242	38,875

2.4.4 Impact of Driving Style on Observations

Table 12 shows number of trips that each driver has generate along with the number of days that they were participating in the data collection. It also provides the average number of trips that each individual driver has generated during the course of participation in data collection. The average and standard deviation of the number of trips by different drivers are 5.2 and 1.2 respectively.

Table 12 – Number of Trips per Driver

Driver	TRIP_COUNT	DAYS_WITH_TRIPS	TRIP PER DAYS
Driver #1	1,252	257	4.9
Driver #2	1,066	227	4.7
Driver #3	3,656	478	7.6
Driver #4	1,462	328	4.5
Driver #5	1,915	378	5.1
Driver #6	2,540	452	5.6
Driver #7	2,567	435	5.9
Driver #8	2,211	465	4.8
Driver #9	1,258	164	7.7
Driver #10	559	130	4.3
Driver #11	594	97	6.1
Driver #12	1,772	445	4.0
Driver #13	670	99	6.8
Driver #14	3,048	519	5.9
Driver #15	65	14	4.6
Driver #16	766	210	3.6
Driver #17	1,205	223	5.4
Driver #18	2,069	473	4.4
Driver #19	1,722	368	4.7
Driver #20	449	97	4.6
Driver #21	835	110	7.6
Driver #22	1,356	298	4.6
Driver #23	445	149	3.0
Driver #24	1,573	379	4.2
Driver #25	1,622	333	4.9

Since the population of the drivers is quite diverse (students and employees), the number of trips that have been made by participating drivers is variable as well. However, the majority of the drivers average about 5 trips per day, as presented in Figure 12.
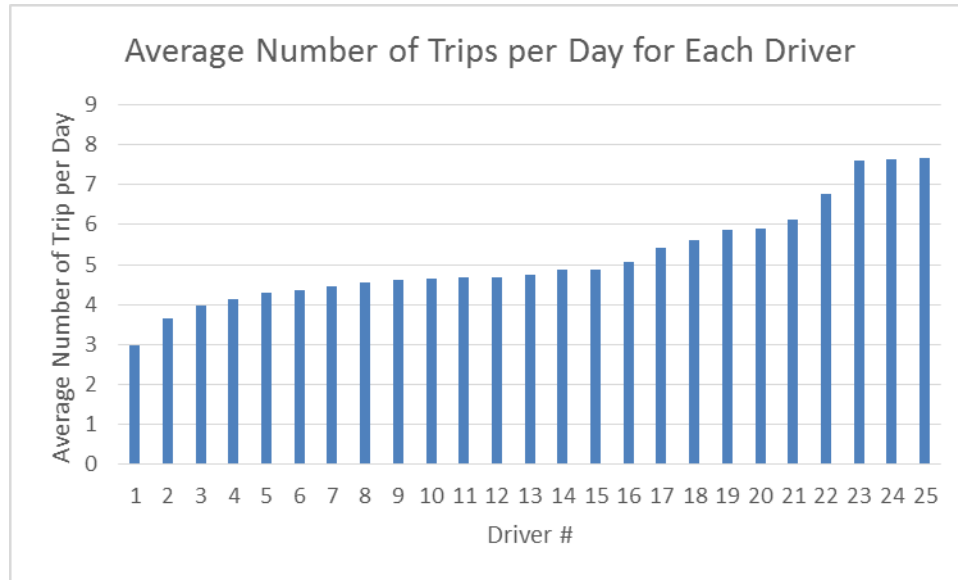
Figure 12 – Average Number of Trips per Day Distribution

The research team also investigated the trips that each driver made and did not find any correlation based on day of week or month of year. However, by breaking down the number of trips each driver made as a function of hour of travel, some interesting observations emerged. As Table 13 shows, some drivers (such as driver #14) tended to perform trips in a narrow time slice of the day, as opposed to other drivers who had a more dispersed travel pattern throughout the day (for example driver #3.)

Table 13 – Trips breakdown by each individual driver based on the hour of the day

Driver #	12_AM	01_AM	02_AM	03_AM	04_AM	05_AM	06_AM	07_AM	08_AM	09_AM	10_AM	11_AM	12_PM	01_PM	02_PM	03_PM	04_PM	05_PM	06_PM	07_PM	08_PM	09_PM	10_PM	11_PM
1	5	1	0	0	0	2	6	26	31	58	58	120	138	69	48	77	116	137	112	105	49	51	17	15
2	0	0	0	0	6	24	26	141	82	64	71	65	60	49	53	46	77	99	65	49	41	17	7	2
3	78	28	8	1	2	4	7	39	139	228	195	226	231	175	134	119	232	367	289	299	309	202	114	98
4	4	0	0	0	0	0	40	37	107	150	69	80	104	76	75	107	83	184	117	73	62	39	9	9
5	3	0	0	0	2	0	7	206	162	70	62	84	101	90	65	67	134	381	133	84	119	49	23	8
6	33	8	0	2	1	4	10	47	229	52	44	113	197	172	118	88	127	258	230	206	118	133	140	130
7	15	4	6	3	3	3	11	27	310	256	108	93	98	118	125	138	221	329	200	114	101	113	62	25
8	3	2	1	0	0	4	122	147	166	135	162	174	130	150	97	139	336	145	85	50	47	19	8	3
9	4	1	1	0	0	45	36	81	39	48	86	88	98	94	92	111	117	94	36	37	55	34	39	4
10	0	1	0	0	0	0	0	3	16	35	53	104	79	53	51	31	37	22	27	17	10	10	6	3
11	3	0	0	0	0	1	1	28	65	48	33	24	35	27	31	23	51	64	48	44	33	20	9	4
12	4	3	0	0	3	2	2	74	159	195	79	106	120	98	86	88	173	262	129	68	53	22	5	2
13	14	3	1	0	1	1	0	1	3	29	41	46	64	50	47	58	34	33	61	54	46	38	23	17
14	20	5	6	8	3	0	13	53	428	201	149	145	130	140	152	86	243	514	151	173	127	70	34	19
15	0	0	0	0	0	1	1	4	3	1	4	9	5	9	7	7	4	3	0	3	2	2	0	0
16	10	6	0	0	0	1	2	4	13	57	54	39	51	46	46	59	57	70	81	66	44	11	28	10
17	6	4	3	1	2	9	193	22	13	25	45	131	112	64	59	140	182	42	33	41	9	15	15	9
18	2	0	0	0	1	6	15	124	381	139	116	143	131	103	102	83	83	157	288	90	34	14	5	4
19	17	9	5	1	0	4	4	17	106	304	92	87	140	127	55	62	88	162	55	67	84	52	42	27
20	1	0	0	0	0	1	12	9	58	29	13	16	34	21	24	30	61	77	30	11	4	1	1	4
21	10	5	4	3	1	1	2	10	14	25	108	44	44	35	32	41	94	97	55	51	63	37	25	28
22	10	4	2	2	2	1	7	35	50	52	53	71	78	63	65	69	120	137	127	85	52	33	30	
23	0	0	0	0	1	1	0	26	97	15	7	20	24	10	6	8	50	117	26	13	11	7	0	1
24	5	1	1	0	1	6	26	50	284	131	55	95	116	67	58	70	117	232	74	46	33	38	22	11
25	10	2	3	1	0	1	2	7	6	26	23	86	173	153	109	111	106	191	219	184	93	61	17	15

2.5 i2D VEHICLE-INFRASTRUCTURE-VEHICLE (VIV) SYSTEM

Each i2D unit is equipped with two communication channels. The first communication channel transmits second by second raw trajectory data every 23 seconds to the server. These data are used for all offline analyses, as will be discussed comprehensively in Chapter 4. Earlier sections of this chapter discussed the flow of data from the web-based cloud to the local SQL database. All these communications were performed under Channel 1.

Channel 2 is designed to communicate with the cloud every 6 seconds. Channel 2 also enables two-way communication between the cloud and each i2D unit. This enables the i2D units to be used for VIV communication purposes. A designated server on the cloud monitors real time data (with 6 second resolution), used GPS coordinates of the neighboring i2D vehicles and transmits a message to i2D devices based on some predefined rules. Also, i2D devices are equipped with Bluetooth networking. A cell phone application (app) is designed for Android based operating systems that can communicate with i2D devices. The message that has been sent by the server under channel 2 can be ultimately seen on the driver's cell phone. A more detailed discussion on the VIV functional system is provided in Chapter 5.

The third web interface called VIV platform is also designed to enable the research team to monitor VIV systems functionality. The VIV platform can show the event generation as well as communications between the infrastructure and vehicles. Figure 7 shows a screenshot of the VIV web interface.

2.6 CONCLUSIONS

This chapter explains the flow of data from i2D units to different storage and management platforms. The majority of these processes have been designed and implemented during the course of this project. High-level summaries of the findings in this chapter are:

- Three separate web interfaces were designed to help the research team, manage the fleet, download raw data, and conduct VIV tests.
- Setting up the local SQL database enabled the research team to generate summary statistics regarding the raw data to explore the high level driving behavior of the drivers.
- Channel 2 of i2D technology enables 6 second resolution communication between cloud and i2D devices. This enables the research team to conduct VIV experiments.
- With the assistance of 25 drivers who vehicles were equipped with i2D devices, and traveled over a period of 20 months, a total of 32,351,971 seconds (~9,000 hours) and 277,601 miles of trip data have been recorded. In total, those involved 38,875 driving trips by all participants.

- The average trip duration and length observed in the data was 13.9 minutes and 7.1 miles. The standard deviation for weekend trips tended to be higher than that of weekday trips.
- Only 0.4% of the trips had a duration longer than 120 minutes or length of over 100 miles.
- Overall, two peak periods for trip, 8AM and 5PM, were observed. The PM peak is larger than AM peak.
- Trips conducted during early morning (6AM to 8AM) have a longer duration and distance compared to trips between 8AM and 12PM.
- The standard deviation of trips lengths and durations for the months of December, January, and June is higher than that of other months, which could be an indication of longer trips taken for vacation or other special traveling events. Also, the average trip length and duration is higher for the months of December and June.
- The average and standard deviations for the number of trips by different drivers are 5.2 and 1.2 respectively.

3.0   GEOFENCING AND TRAJECTORY GENERATION TOOL

This chapter presents the tools that have been developed during the course of this project to assist data filtration and analysis. Since the collected high resolution data spans beyond 32 million seconds, the use of automated tools are essential. Two tools are developed in this project.

The first tool is developed within ArcGIS software. It connects to the local SQL database and can filter the trips that have used a specific corridor. The information regarding the desired corridor can be entered in this tool and by executing the filteration, it process all the trips with in local SQL database. The trips that have used the secified corridor will be identified and will be recored in a separate table in the SQL database. Each time that the tool is executed, it clears the last run's output and fills with new results. Therefore, the analyst should copy identified trips from the SQL database after completion of each run.

The second tool is called i2D Trajectory Generator. The i2D Trajectory Generator Tool is an analysis tool designed to identify and trace the different kind of instantaneous events that can occur on freeways or surface streets, as derived from the high resolution vehicle sensing data. In addition, this tool can be used for V-I-V event verification

3.1   ARCGIS TOOL

3.1.1 Introduction

Between April 2014 and December 2015, i2D devices captured more than 32 million seconds of travel data across the fleet. As a next step, the research team decided to utilize ArcGIS to filter the data by location so that statistics by corridor. The analysis process was streamlined using code written in Python and the ArcGIS Model Builder. Two different versions of the resulting analysis tool were developed: 1) one that creates a corridor based on intersections and 2) one that creates a corridor based on latitude and longitude points. The Intersection Corridor Tool is applicable to arterial roads while the Latitude/Longitude Corridor Tool is most useful for the analysis of freeways.

3.1.2 High-Level Description

Below is a description of the two tools and the accompanying processes.

1. The user inputs the necessary road information, depending on the version of the tool, as shown in Figure 13.
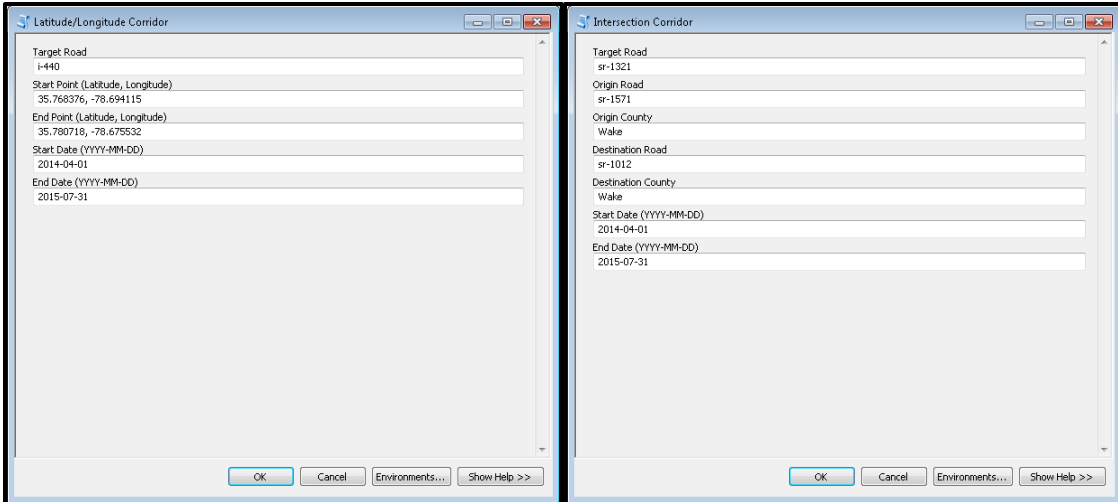
Figure 13 – (Left) Latitude/Longitude Corridor Tool; (Right) Intersection Corridor Tool

2. Based on the given inputs, ArcGIS creates a geofence rectangle that includes the entire corridor from the start point to the end point. The coordinates of the rectangle are used to query the MySQL database for all the data within that rectangle.
3. The MySQL data within the geofence are sent back to ArcGIS for additional filtering and data analysis.
4. A buffer along the desired corridor is used to filter data further. Then, several ArcGIS tools are used to determine which trips in the corridor are travelling in the desired direction. These trips are then sent back to the MySQL database.
5. Trips are placed into two tables – *in_corridor_output* and *thru_corridor_output*. All trips that contain at least one second of data located within the corridor buffer are fed into the first table. Those trips that completely traversed the corridor from the start point to the end point populate the second table. Both tables include data on the trip's unique identification code (TRIP_ID), the first second that the trip enters the corridor (MIN_SECOND), and the last second that the trip enters the corridor (MAX_SECOND). This information stays in the MySQL database until the tool is run again or the tables are manually cleared.

3.1.3 Detailed-level Description

1. Creating the corridor: This is the only step that differs between the Intersection Corridor Tool and the Latitude/Longitude Corridor Tool, although this variance is minor and is primarily due to the difference in the required inputs.
   a. *Intersection tool*
      i. A layer is created for each input – one for the target road(s), one for the road marking the start of the corridor, and one for the road marking the end of the corridor.

33

ii. A point is created at each intersection – one where the target road meets the start road and one where the target road meets the end road.

iii. The target road layer is split into segments at each intersection and then the segment between the start road and end road is copied to a new layer.

b. *Latitude/longitude tool*

i. A layer is created for each input – one for the target road(s), one for the start point, and one for the end point.

ii. A 50-foot buffer is created around the start point and the end point to ensure that it intersects with the target road.

iii. A point is created at each intersection – one where the target road touches the start point buffer and one where the target road touches the end point buffer.

iv. The target road layer is split into segments at each intersection and then the segment between the start point and end point is copied to a new layer.

2. Calculating direction and length of the target road: From here on out the two tools operate identically.

a. The direction of a trip is calculated using Python's *atan2* function.

b. The length is identified using ArcGIS's *native length* field.

3. Creating the geofence:

a. A 100-foot buffer is created along the target road segment. The extents of this buffer are used as the geofence when pulling data from the MySQL database.

b. The geofence is used as the first filter when selecting data so that the tool is not selecting more than 20 million records each time.

4. Collecting the data:

a. A connection is made to the MySQL database.

b. Two queries are immediately run to clear out any data in the output tables from the last time a script was run.

c. Another query is run to pull in the data. It selects the TRIP_ID, SECONDS, LATITUDE, and LONGITUDE for each point that is within the buffer geofence and within the timeframe given by the user.

d. For each record selected, a new point is added to a layer in ArcGIS.

5. Cleaning the data

a. Once all the points are created, a clip is used to select all the points that are inside of the actual buffer, not just the geofence.

b. Entire trips are then removed based on whether they enter and exit the buffer multiple times. This is done by checking whether all the seconds for a given trip are consecutive or if there are any gaps.

6. Finding the correct outputs

a. An ArcGIS tool converts all the points to lines based on their TRIP_ID so that a single line represents each trip.
b. Another ArcGIS tool calculates the average direction of each trip line, along with other statistics.
c. Based on these trip directions, those trips that are within ±30° of the target segment direction are added to a list. The first second and last second of each of these trips are also added to a list.
d. A query is created to insert each trip ID with its first and last second into a table in the MySQL database for trips that have at least one second within the target corridor. Subsequently, new layer is created within ArcGIS for these trips.
e. A new list is created based on the length of each trip line. If it is between the original length of the target corridor and the length of the target corridor plus 250 feet, it is considered to be a trip that has completely traversed the corridor. Due to use of a buffer in selecting data, there will not be any trip that its length is smaller than the length of corridor.
f. One last query is then created to insert each of these trips with their first and last seconds into a table in the MySQL database for trips that have travelled the length of the target corridor. A new layer is also created within ArcGIS for these trips.

7. The connection to the database is closed and the tool's analysis is complete.

By performing the described process, the research team was able to extract corridor level data and statistics.

3.1.4 Limitations

Due to the characteristics of the road layer used, there are some limitations to these two tools. The user identified the desired road by its name; however, within ArcGIS each road also has a 10-digit code that ties it to the county where it is located. Therefore, when trying to collect data along a corridor that spans multiple counties, the tool breaks down. The best workaround for this issue when working with the current version of the code is to split the road into multiple parts, run the tool multiple times, and then find the common trips amongst each of the results. Figure 14 shows an example facility that expands into two counties. As shown, the facility is divided into two pieces in a way that each piece is located in only one county. The analysis should be carried on two times for two stretches of the facility. Then, by comaring the trips that have been identified in MySQL database, the common trips can be identified.
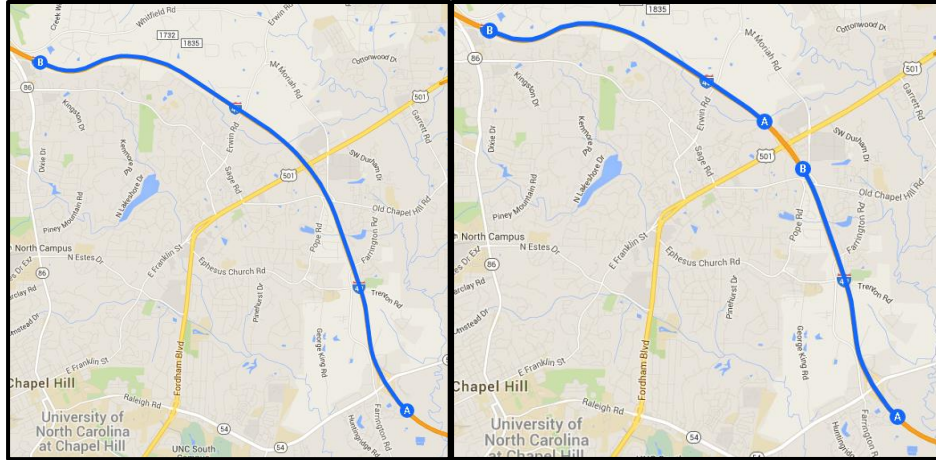
Figure 14 An example of the workaround for the limitation of the tool.

(Left) A single corridor that spans two counties in North Carolina
(Right) Two corridors that are split along the county line

3.2 I2D TRAJECTORY GENERATOR TOOL

The i2D Trajectory Generator Tool is an analysis tool designed to identify and trace the different kind of instantaneous events that can occur on freeways or surface streets, as derived from the high resolution vehicle sensing data. In addition, this tool can be used for V-I-V event verification. VIV events are generated using vehicle actuation data (high resolution vehicle trajectory), and these events need to be verified. However, verifying the data manually is not realistic due to the significant number of data points. Consequently, the i2D Trajectory Generator Tool is valuable because it enables researchers to test i2D V-I-V application accuracy.

This tool reads i2D high resolution data (vehicle OBD speed data) and detects the various events, which occurred during the trip. The i2D Trajectory Generator Tool allows the user to configure the various input parameters and analyze the trips and the various alerts generated. The tool generates two different graphs: "Speed vs Distance" and "Speed vs Time". The various event alerts are shown on the graphs (shown in Section 3.4).

3.2.1 Events

The i2D Trajectory Generator Tool identifies the three events as defined by the user. The pre-defined events are detected by user input (user defined event thresholds). The pre-defined events are following:

- Delay Event: Vehicle experiences a speed less than a pre-defined speed threshold for more than a certain number of seconds on a highway as well as vehicle has stopped on the local road more than a prespecified number of seconds, the delay event will be triggered.

- Awkward Deceleration Event: Vehicle has a hard deceleration on the road

- Crash Event: Vehicle possibly has crashed or incident on road

36

Depending on the roadway type, the events are marked Delay (DL), Awkward Deceleration (AL), Crash (CL) for Local or surface streets. Similarly, for Highways, the events are marked Delay (DH), Awkward Deceleration (AH), Crash (CH).

3.2.2 User Input (Event Threshold)

The software allows user to upload the trip details through a file (i2D high resolution vehicle trajectory data). The file contains the trip ID, start date and time, end date and time as well as the position of the vehicle in latitude and longitude coordinants, the OBD speed of the vehicle, the GPS speed, etc. There are two roadways types: 1) Highway and 2) Local. Default values can be customized for various inputs depending on the roadway type as well as event type selected.

Freeways and Highways (uninterrupted flow)

There are eight Highway Inputs, each of which can be customized by the user. The user interface in

Figure 15 shows the various inputs that can be customized by the user.

Figure 15 Highway Input

1. **Delay Highway (DH)**

A "Delay Highway event" is designed to help detect a delay event and provide information to following vehicles' drivers. When the i2D device-equipped vehicles experiences a certain pre-designed condition, the device will trigger an event alert and send specific event information to the server.

The Delay threshold and Time Window threshold are coded in the tool and Delay Highway event will be displayed when the two conditions below are satisfied:

a. Delay threshold (Moving average speed): The threshold for the average speed to identify the delay alert. The default average speed threshold is 40 mile/hour.

b. Time Window: The duration of the window for which the average speed needs to be computed. The default time window is 60 seconds.

2. Awkward Deceleration on Highway (AH):

An Awkward Deceleration event on a highway involves three user input thresholds, including "Deceleration," "Average Velocity," and "Duration." During the last "x" minutes (Duration) a vehicle traveled above a certain speed (Average Velocity), and then the vehicle had hard deceleration (Deceleration), the AH event will be detected and displayed via i2D trajectory generation tool.

a. Deceleration: The threshold for the deceleration to identify the AH event. The user has to enter a negative number to indicate that it is a deceleration. If non-negative number is entered then the application will notify the user. The default value is -4m/sec^2.

b. Average Velocity: The threshold for the average velocity to identify the AH event. The default value is 50 mile/hour.

c. Duration: The duration of the window for which the average speeds needs to be computed. The default duration is 60 seconds.

3. Crash Highway (CH):

A crash event on highway also includes three user input fields: "Deceleration," "Speed," and "Average Velocity." When a vehicle has a hard deceleration (Declaration threshold) under a certain speed (Average Velocity) then the vehicle speed (Speed threshold) became 0, and the Crash event will be detected and displayed.

a. Deceleration: The threshold for the deceleration to identify the CH event. The user has to enter a negative number to indicate that it is a deceleration. If non-negative number is entered then the application will notify the user. Default value is -4m/sec^2.

b. Speed Threshold: The exact speed threshold at which the crash event occurred. Default value is 0.

c. Average velocity: The threshold for the average velocity to identify the AH event. Default value is 50 mile/hour.

Local Roadway

There are 10 Local Road Inputs that can be customized by the user. Figure 16 shows all the inputs available to the user.

1. **Delay Local (DL):**

   a. Delay Threshold (Moving average speed): the moving average speed via user defined time window to identify the delay evnet. The default average speed threshold is 40 mile/hour.

   b. Time Window: The duration of the window for which the average speed needs to be computed. The default time window is 60 seconds.

   c. Delay: The number of seconds at which the instantaneous velocity should be zero (a vehicle stopped somewhere more than given value and then the DL event will be detected and displayed). The default value is 30 seconds.

2. **Awkward Local (AL):**

   a. Deceleration: User defined hard deceleration value. The user has to enter a negative number to indicate that it is a deceleration. If non-negative number is entered then the application will notify the user. Default value is -4m/sec^2.

   b. Average velocity: Moving average velocity before a vehicle makes hard deceleration. The default value is 40 mile/hour.

   c. Duration: The duration of the moving window for which the average speeds needs to be computed. The default duration is 60 seconds.

3. **Crash Local (CL):**

   a. Deceleration: User defined hard deceleration value. The user has to enter a negative number to indicate that it is a deceleration. If non-negative number is entered then the application will notify the user. The default value is -4m/sec^2.

   b. Speed Threshold: The exact speed threshold at which the crash event occurred. The default value is 0.

   c. Average velocity: Moving average velocity before a vehicle makes hard deceleration. The default value is 40 mile/hour.

   d. Duration: The duration of the window for which the average speeds needs to be computed. The default duration is 60 seconds.

Figure 16 Local Road Input

3.2.3 i2D trajectory Generator Instruction

The user can start the application by clicking on the JAR file. The user needs to have Java 8 installed on their system to run the application.

- Step 1: The user must first select the trips file (i2D high resolution vehicle trajectory data) which contains the detailed information about the trips.

- Step 2: After selecting the file, the Roadway type (Highway or Local) needs to be selected. The user can customize the input parameters for each of the roadway type, as described in section 3.2.

- Step 3: The user needs to click "Get Trip Info" to view the summaries for all unique trips. The trip summary will be display on the bottom of i2D trajectory generator toll and it shows trip unique ID, trip start date and time, as well as travel distance with total travel time. Figure 17 shows an example, which the user can view in the application.

- Step 4: The user must then select one of the trips from trip summary list. This can be done by clicking on the trip from the list. Once a trip is selected, the user can generate the I2D graph by clicking on "Plot I2D Graph." The user can view the graphs "Speed vs Distance" and "Speed vs Time." The user can choose to save the graphs in PNG Image format. The detailed explanation of the output is explained in the following section.
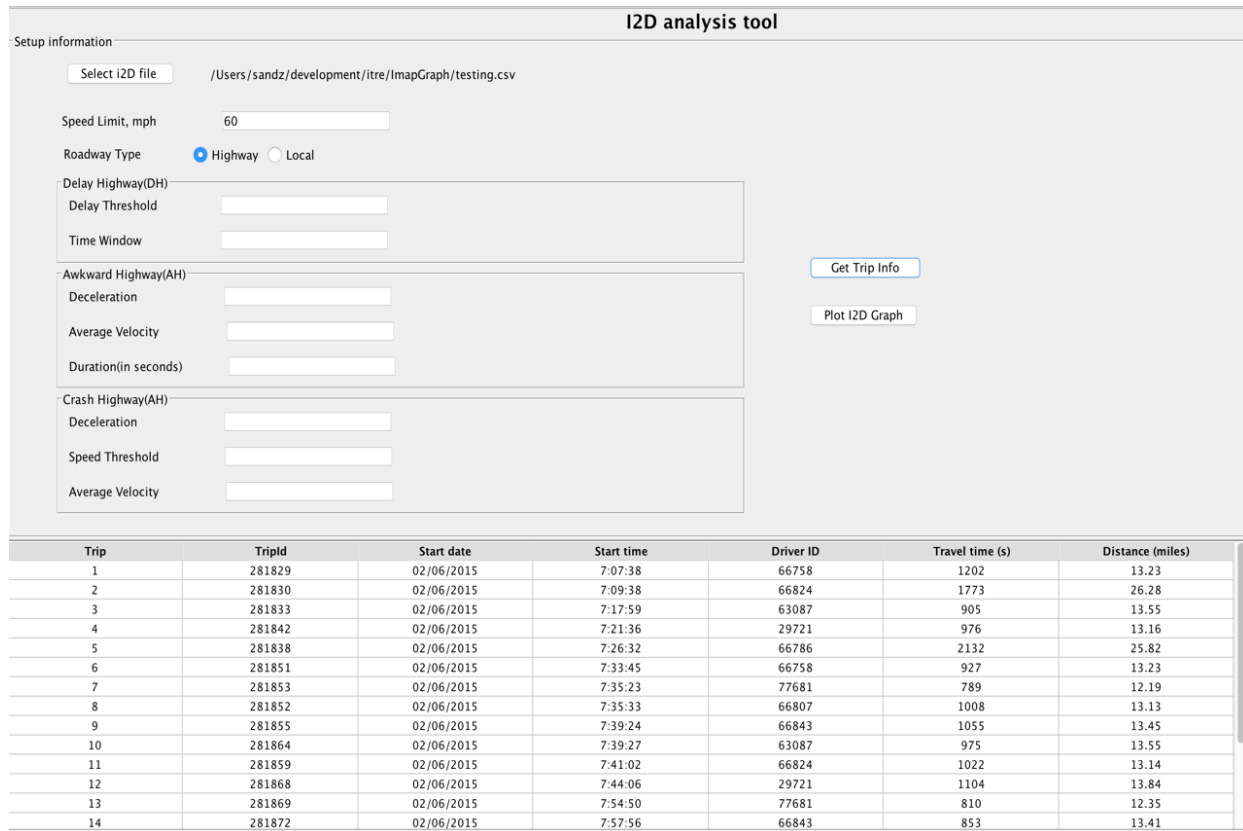
Figure 17 Unique Trip Summary List in the Application

3.2.4 Application Logic

The user should upload the trips data as a CSV file. Once the user clicks "Get Trip Info." the application collects all the unique trip information and its corresponding instantaneous speed. The i2D Trajectory Generator plots two graphs: 1) Speed vs Distance and 2) Speed vs Time for selected unique trip.

Speed vs Distance Graph: The speed vs distance graph plots the instantaneous speed and the distance travelled by the vehicle. It also plots the moving average speed of the vehicle on the graph along with the speed limit, as shown in Figure 18.
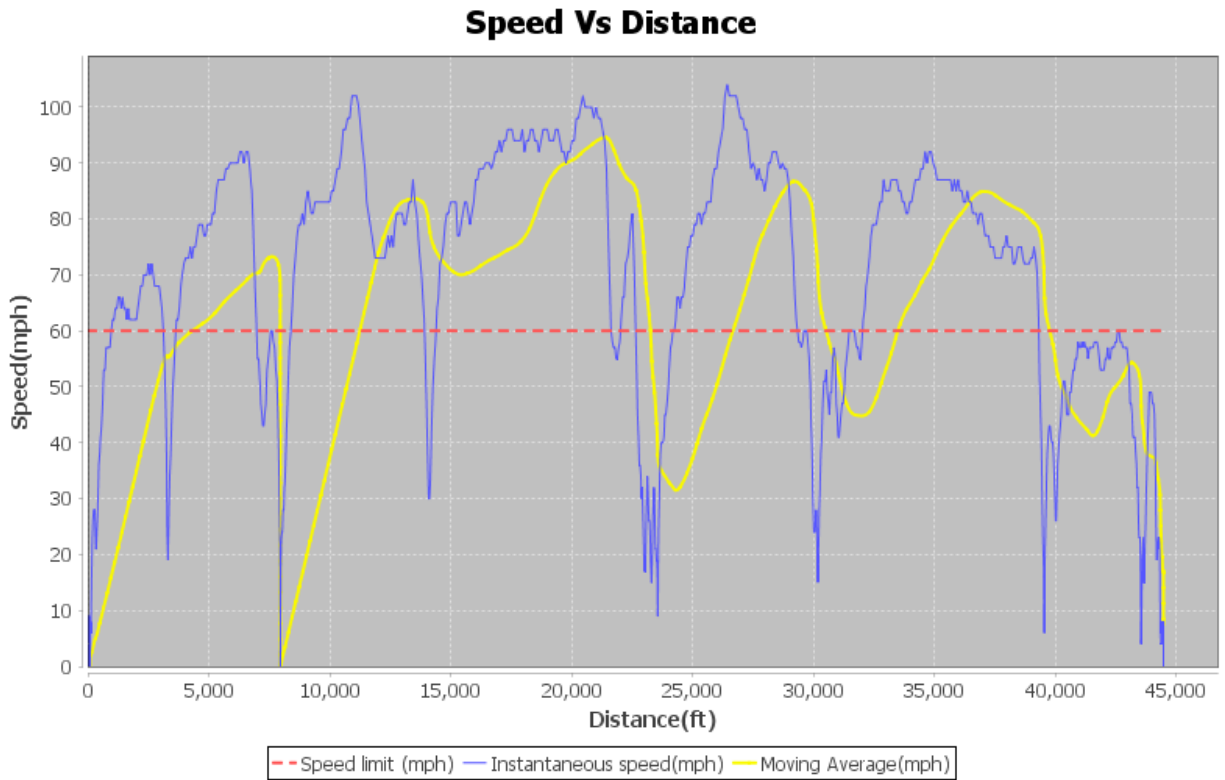
Figure 18 Speed Vs Distance for Highway

Depending on the type of roadway selected, the i2D trajectory generator tool detects the corresponding "Delay," "Awkward Deceleration," and "Crash" alerts. This is the logic for detecting an event for Highway Roadway type:

- Delay (DH): At any point of time if the moving average speed of the vehicle is less than the user input value (value from "Delay Threshold" field) for more than "X seconds" (value from "Time Window") seconds, then a Delay (DH) event is generated.

- Awkward Deceleration (AH): If at any point of time the vehicle deceleration is less than "E" (value from "Deceleration" field) and the average velocity is greater than "F" (value from "Average Velocity" field) during the last "G" (value from "Duration") seconds, then an Awkward Deceleration (AH) event is generated.

- Crash (CH): When the vehicle is decelerating less than "L" (value from "Deceleration" field), the average velocity in the last minute is greater than "N" (value from "Average Velocity" field), and the speed is equal to "M" (value from "Speed Threshold"), then a Crash (CH) event is generated.

42
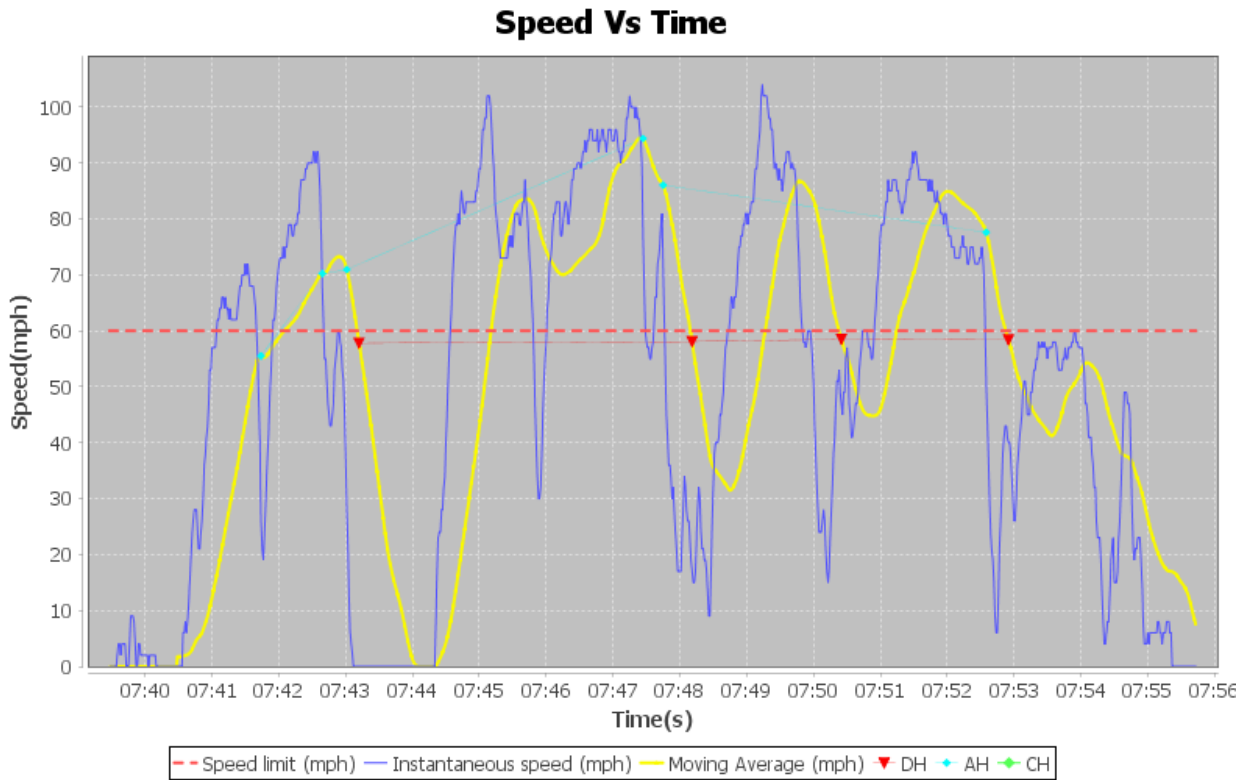
Figure 19 Speed Vs Time Graph for Highway

The logic for detecting an event for Local Roadway type is as follows:

- Delay (DL): If at any point in time if the average speed of the vehicle is less than the "Y" (value from "Delay Threshold" field) for more than "Z" (value from "Time Window") seconds, and the vehicle has a zero velocity for more than "X" (value from "Delay"), then a Delay (DL) event is generated.

- Awkward Deceleration (AL): If at any point in time the vehicle deceleration is less than "B" (value from "Deceleration" field), and the moving average velocity is greater than "C" (value from "Average Velocity" field) during the last "D" (value from "Duration") seconds, then an Awkward Deceleration (AL) event is generated.

- Crash (CL): When the vehicle is decelerating less than "H" (value from "Deceleration" field), the average moving velocity in the last "K" seconds (value from "Duration") is greater than "J" (value from "Average Velocity" field) and the speed is equal to "I" (value from "Speed Threshold"), then a Crash (CL) event is generated.

43
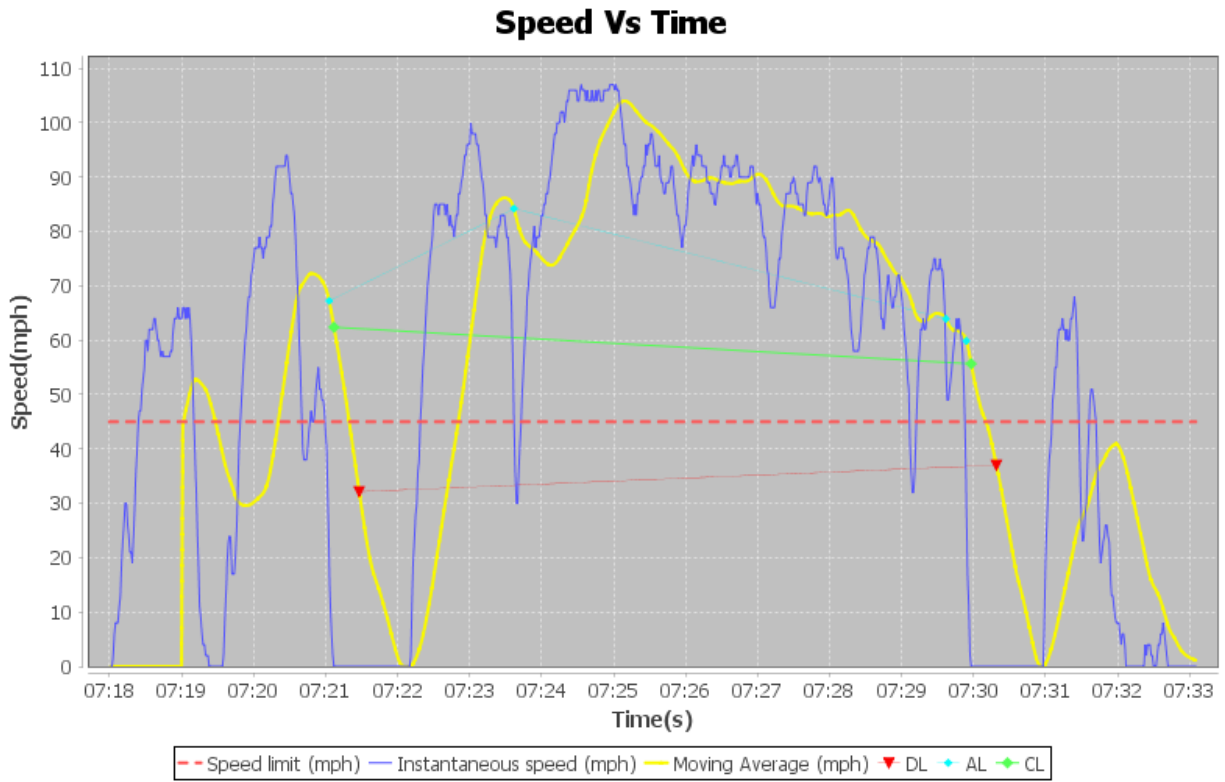
Figure 20 Speed Vs Time Graph for Local

4.0   DATA APPLICATIONS

This chapter presents four specific and highly diverse applications of the high resolution data, using the various tools described in Chapter 3, along with the key findings. The first three applications utilize the naturalistic, archival data while the last one is based on a set experiment using vehicles deployed in a pre-specified manner. These applications are as follows:

- Theoretical analysis of observed time and space mean speed spatial distributions along a signalized arterial street,
- Analysis of the association between rear-end collision frequency and the incidence of severe decelerations on freeways,
- Freeway free flow speed analysis from the driver perspective, and
- Shock wave analysis using controlled experiments of instrumented vehicle dispatch near freeway bottlenecks.

4.1   SPEED ANALYSIS ON SIGNALIZED ARTERIALS

The relationship between time-mean and space mean speed identified by Wardrop (1952) has been widely recognized. Traffic stream speeds are traditionally measured or collected in the field using a variety of spot speed measurement technologies such as inductive loop detectors and side-fire radar detectors. Since most studies tend to use spot speed observations, link-based space mean speed and time mean speed, explained futher in the next section, are not well understood.

In reacent years, several vendors (INRIX, HERE and TomTom) collected nationwide link-based traffic management center (TMC) speed (travel time) data from probe vehicles. Probe vehicles can provide instantaneous speed data and GPS location at every designated time interval, producing data that enables analysts to calculate link-based space mean speed and time mean speed for a certain time-space domain directly. Data vendors, however, do not share their high resolution data. Instead, only aggregated link speed or travel time data are reported for each TMC at resolutions of 1 minute or less. The high resolution data collected in this project overcomes this constraint by enabling the calculation of both types of speeds from fundamental principles.

4.1.1 Defining Space Mean and Time Mean Speeds

Space mean speed is defined as the *harmonic mean of speeds passing over a point* during a period of time. Therefore, it corresponds to the average speeds over a length of roadway. *Time mean speed* is defined as the *arithmetic mean of speeds of vehicles passing a point*. Accordingly, for point-based measurements:

$$\text{Space Mean Speed: } u_S = \frac{N}{\sum_{n=1}^{N} \frac{1}{v_N}} \tag{1}$$

$$\text{Time Mean Speed: } u_T = \frac{1}{N} \sum_{n=1}^{N} v_N \tag{2}$$

As mentioned previously, Wardrop (1952) established the relationship between the two speeds:

$$u_T = u_S + \frac{\sigma_S^2}{u_S}$$

(3)

Where,

u_T: time mean speed
u_T: space mean speed
v_N: spot speed measured at the point
N: total number of vehicles sampled

On the other hand, because probe vehicles can provide instantaneous speed information every few of seconds, a single probe vehicle space mean speed can be derived from the relationship between velocity and travel distance:

$$x(t) = x_0 + \int_{t_0}^{t} v(t)dt$$

(4)

Where, $x(t)$: travel distance
v_t: instantaneous speed
x_0: initial travel distance

Therefore, the space mean speed for a *single probe vehicle* traveling through a specific segment, $\bar{u}_{SMS}^{Probe}$, can be calculated as the total distance traveled divided by total travel time, as follows:

$$\bar{u}_{SMS}^{Probe} = \frac{Total\ distance\ traveled\ of\ a\ vehicle}{Total\ travel\ time\ of\ a\ vehicle} = \frac{\lim\limits_{dt \to 0} \int_{n=1}^{N} u_n dt}{\sum\limits_{n=1}^{N} t_n} = \frac{\sum\limits_{n=1}^{N} u_n \times \frac{T}{N}}{T} = \frac{\sum\limits_{n=1}^{N} u_n}{N}$$

(5)

Where,

T = total segment travel time (sec),
n = data reporting interval (sec),
N = number of data reports received during the segment travel time
dt = time interval duration (sec) or T/n,
t_n = travel time during a time interval n (sec),
u_n = instatenous during a time interval, n (sec),

This essentially becomes the arithmetic average of the probe speed reports. Individual probe vehicles provide information at a specified reporting interval, Δt. To precisely calculate the space mean speed for a probe vehicle, Δt should converge to 0, which would yield the exact travel distance for the TMC. Figure 21 depicts the time-space diagram and instantaneous speeds as a

function of Δt; here Δt is 0.1 sec or 1 sec. As seen in the graphs, a larger sampling interval Δt would cause more errors in calculating the space mean speed.

| (a) Distance traveled | (b) Instantaneous speeds |
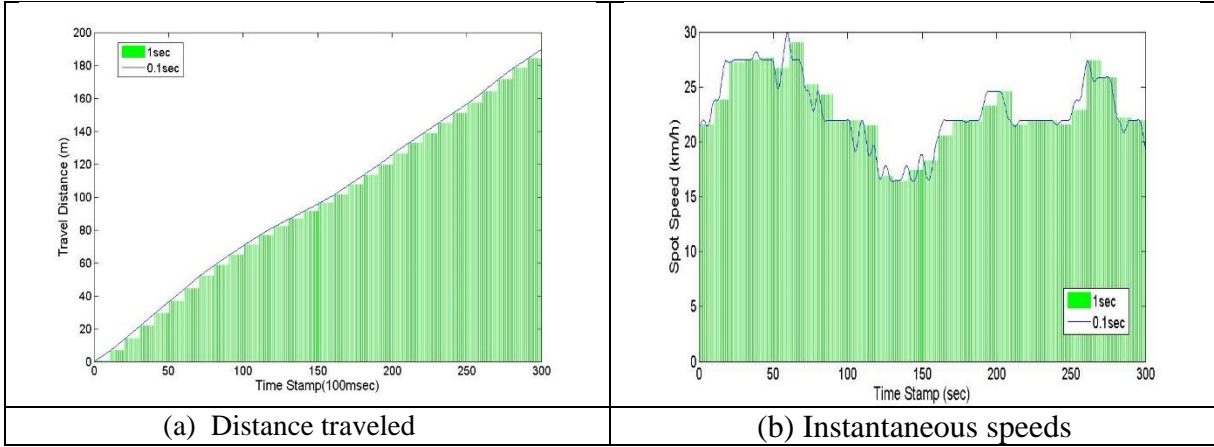
Figure 21 Total Predicted Travel Distance and Speed by Reporting Interval Δt

On the other hand, total travel time, T can be easily calculated by summing times over a segment. Δt can be converted to T/N. Thus, the space mean speed can be derived as the probe vehicle instantaneous speeds averaged over the segment as seen at the right side of Equation 1. This derivation can be directly applied to calculate the average space mean speed for all probe vehicles on a segment. The average space mean speed of all probe vehicles occupying a given segment is calculated as the total distance traveled by all probe vehicles divided by their total travel time. The total distance can be calculated directly by summing the probe vehicles' instantaneous speeds, while total travel time is a summation of time interval of probe vehicles. The average space mean speed, $\bar{u}_{SMS}^{avg}$, can be calculated as:

$$\bar{u}_{SMS}^{ave} = \frac{\sum_{v=1}^{V} \sum_{n=1}^{N} u_{nv}}{\sum_{v=1}^{V} N_v} \qquad (6)$$

Where,

n = the number of data reporting time

v = vehicle number,

V = number of vehicles occupying the segment (vol), and

N_v = total reporting frequency for vehicle v.

According to Equation 2, the unbiased link space mean speed estimation is the arithmetic mean of the entire reported probe vehicle's instantaneous speed within certain time and space domains. This definition is important since the link-based space mean speed estimation does not require tracking the probe vehicles, which helps ensure individual driver privacy.

4.1.2 Speed Distribution

The i2D database reports each vehicle's instantaneous speed and GPS position information in one-second resolution. As discussed in section 4.1.1, it is possible to derive link-based space mean speed and time mean speed, as well as each link speed distribution. This study shows the relationship between link (section) average speed using reported densities of GPS data points wthin the segment and the number of trips made on the segment to compare the values of the space mean speed with time mean speed, recalling the space mean speed is actually a density-weighted speed.

4.1.2.1 Defining Average Speed Using Density and Number of Trips

Figure 22 depicts an i2D GPS data plot in a spatial representation. Each color represents different vehicles. For example, in Figure 22 there are two vehicles in a given time space domain. Vehicle 1 was tracked 6 times in the section while Vehicle 2 was tracked 14 times.
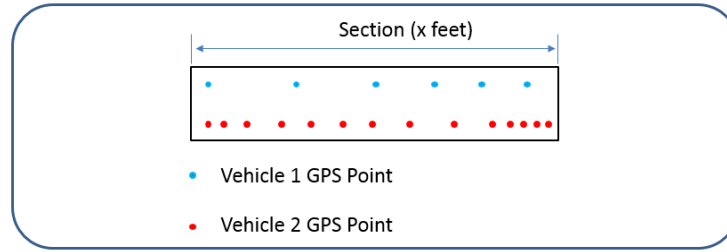
Figure 22 Example of GPS Data

Each point genereates an OBD based speed and GPS position. The label $v1_0$ denotes vehicle 1's OBD speed and $v2_0$ denotes Vehicle 2's OBD speed. Therefore, $v1_{oi}$ denots Vehicle 1's ith OBD speed and $v2_{oj}$ denots Vehicle 2's jth OBD speed, where $i \in I$ and $j \in J$ so that I and J represent the total number of observations of the vehicle position within the segment. In the figure then I = 6 and J= 14. Using strictly trips, one calculates the average speed for each vehicle (for two vehicles in Fig 23) and divide it by the number of trips (also two).

$$\text{Average link speed by density} = \frac{Sum\ of\ all\ instantaneous\ OBD\ speeds}{Total\ number\ of\ GPS\ points}$$

$$= \frac{\sum_{\forall i \in I} v1_{oi} + \sum_{\forall j \in J} v2_{oj}}{n_{v1i} + n_{v2j}} \tag{7}$$

$$\text{Average link speed by trip} = \frac{Sum\ of\ each\ trip\ average\ speed}{Total\ number\ of\ Trips}$$

$$= \frac{\frac{\sum_{\forall i \in I} v1_{oi}}{n_{v1i}} + \frac{\sum_{\forall j \in J} v2_{oj}}{n_{v2j}}}{\sum_{\forall k \in K} v_k} \tag{8}$$

In the Equation 4, k denotes k^{th} vehicle and $k \in K$. In fact, Equation 2 and Equation 3 provide identical speed results, which means that computing the average link speed using reported density is an unbiased estimation of the link based space mean speed. Equation 4 is by definition an unbiased estimation of link-based time mean speed.

4.1.2.2 Speed Distribution on Arterial Streets

High resolution data observed at the entrance and departure approaches to a single traffic signal was analyzed to compare the speed estimates based on reported density versus trip average. The selected site is in Raleigh, North Carolina in the eastbound direction at Avent Ferry Road between Lake Dam Road and Trailwood Drive intersections. The selected corridor length is 1,800 feet and includes one signalized intersection at Avent Ferry Road and Gorman Street. The upstream intersection is located 3,340 ft from the intersection. The site was subdivided into 18 analysis sections, each 100 ft in length. Figure 23 depicts the study arterial.

Figure 23 Selected Site on Avent Ferry Rd

Two months of i2D data, starting from April 1 to May 31, 2015 were used to derive each section average speed using both density and trip based methods. The selected trips consisted strictly of through traffic on Avent Ferry. Seven different drivers were found to have made 38 trips in that two-month analysis period between the 8am and 9am morning peak. Figure 24 illustrates each section average speed by density (or occupancy) and per trip. As mentioned earlier, the link average speed by occupancy is the same as link space mean speed. The blue bar indicates link space mean speed and it is always less than or equal to the section trip-based average speed.

Figure 25 shows the space mean and time mean speed data when aggregating sections into zones. The observed section space mean and time mean speeds were nearly 40 mph on sections between 1,200ft – 700 ft. upstream of the intersection stopline, or zone A in Figure 16. The occupancy based SMS starts to decrease in section B, reaching its lowest value at the stop-line. Subsequently, it begins recovering in Zone C on the departure end of the signal. Beyond zone C, the SMS is stable at the same initial approach speed of 40 mph (Zone D in Figure 26). Notice the very large differences between the two mean speed estimates near the stopline (5mph for SMS vs. 25 mph for TMS). This outcome is the result of the heavier weight given to slow (or stopped) vehicles that the SMS accounts for in computing its speed.

49
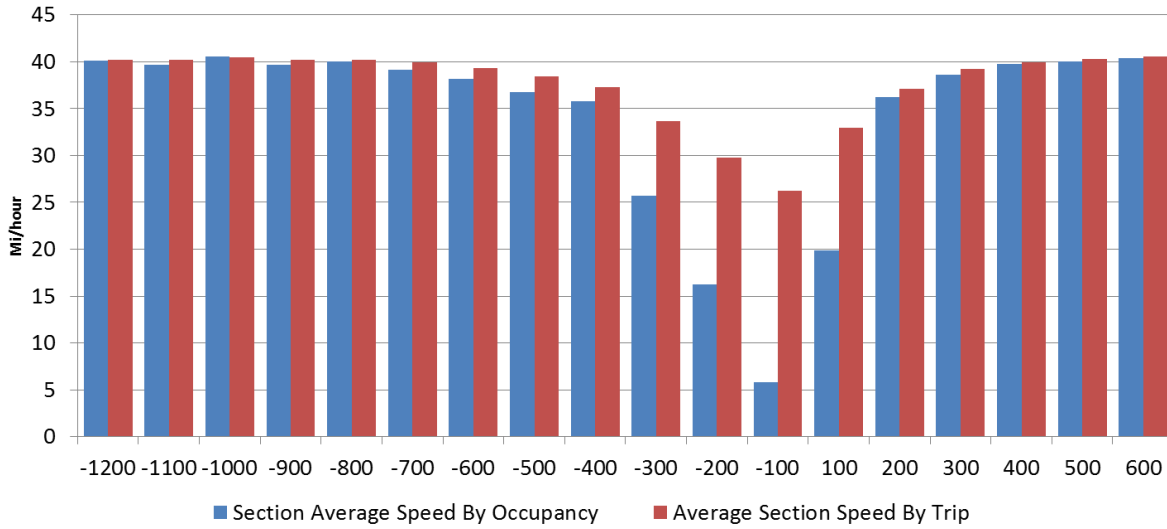
Figure 24 Section Average Speed by Occupancy and Trip
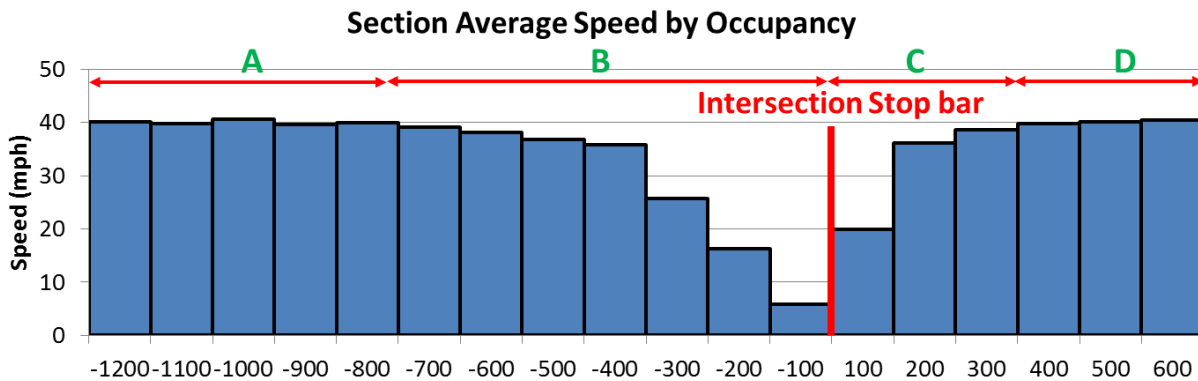
Figure 25 Section Average Speed by Occupancy

This phenomenon is better explained in Figure 26, which depicts the total number of GPS observations measured in each section. When the vehicle approaches the intersection influence area, there is an increased probability of stopping or slowing down. Therefore, there are more observations in the intersection influence area. In fact, the density of observations is highly correlated with traffic signal delay. A flat, non-peaked, histogram would actually be indicative of good signal progression, which is not the case in this particular study.
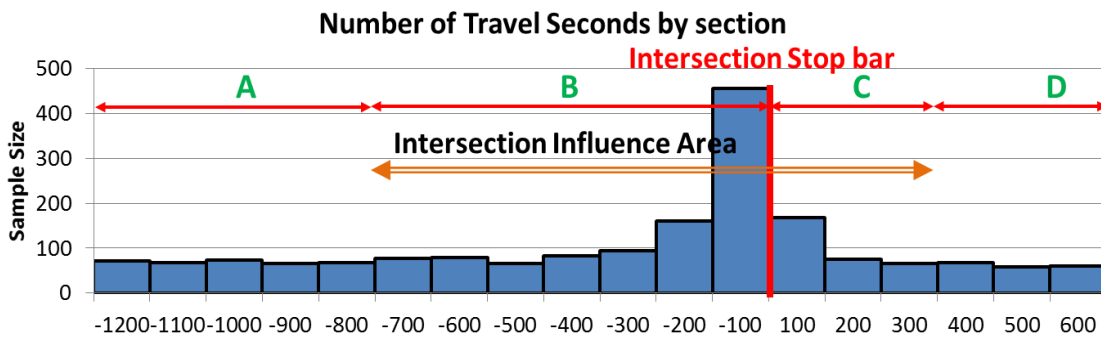
Figure 26 GPS Data Points Reported by Section (8-9 am)

The next set of analyses focuses on speed estimation based on the number of trips (ot time mean speed) in each section, as depicted in Figure 27. Speed variability across sections is much lower than that observed in calculating SMS. Figure 28 presents this result. As shown, fewer than 38 trips were recorded for some sections because the travel speed was high enough to pass the section in less than a second--- while other sections may show more than 38 trips. Thus, the calculated speed corresponding to the trip average speed is similar to the time mean speed (TMS).
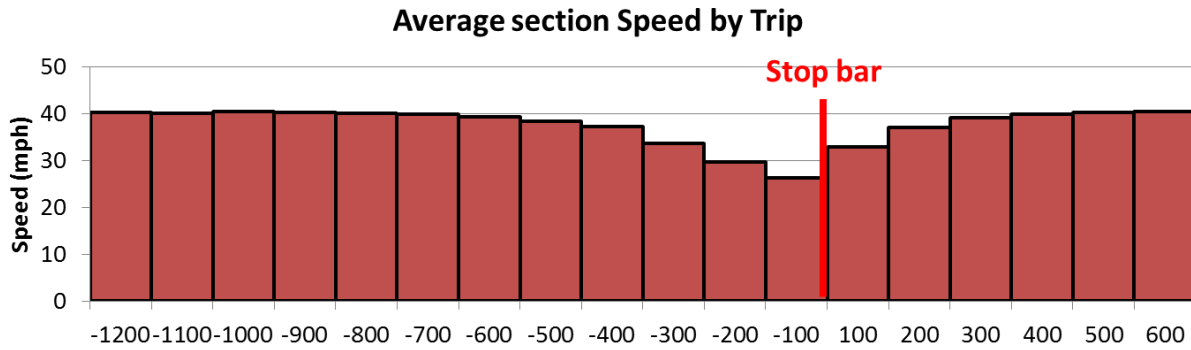
Figure 27 Section Average Speed by Trip
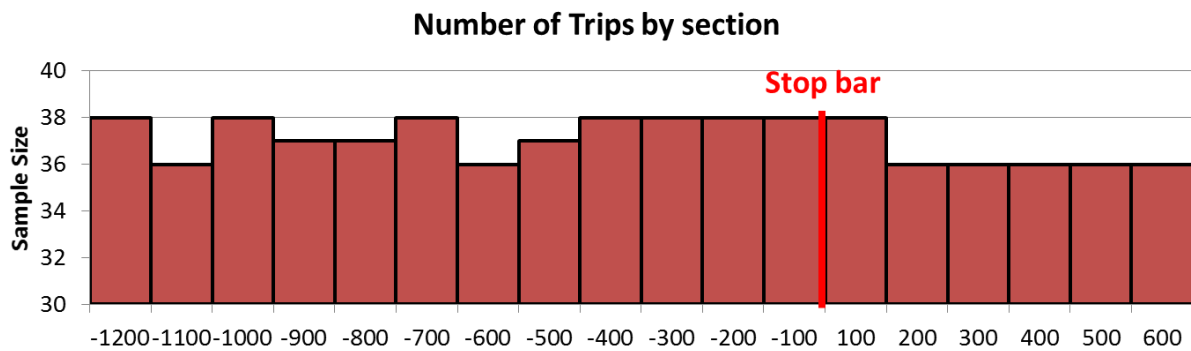
Figure 28 Number of Trips by Section

Figure 29 shows the theoretically derived TMS using Wadrop's equation and the section average speed averaged by trip.
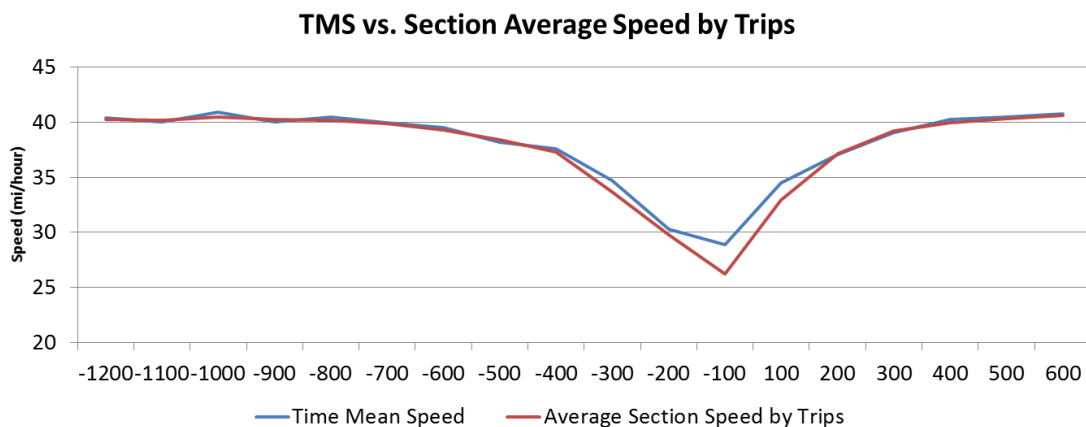
Figure 29 Comparison between TMS vs. Section Average Speed by Trip

51

Figure 24 to 31 have shown how to estimate the two types of section average speed. However, even though in some cases section average speeds are similar to each other, the speed distribution based on the instantaneous speeds measured by the i2D device across different drivers might be different. Figure 30 depicts the observed speed distribution in each section. Each point from the distribution represents a speed observation from a given i2d unit on thes section during the 8-9 am period over two months. The results illustrate the impact of the signalized intersection as well as defining its spatial impact area. For example, in sections far from the intersection, the distribution is both left skewed and very tight around a high speed value of 40 mph. As the section gets closer to the signal more and more low speed observations become apparent. Within 100 ft of the signal, (Figure 31 (l)) the speed distribution has a mode of zero, indicating that most vehicles actually have come to a stop at the signal.

Finally as vehicles reach sections far downstream of the intersection their speed distributions begin to mirror those observed on the upstream end (compate for example the speed distribution in Fig 31(b) and 31 (r)). Those results are also summarized at key points on the arterial in Figure 32. Those types of results can enhance the understanding and management of both the mobility and safety aspects of signalized intersections.

(a) 1200 to 1100   (b) 1100 to 1000   (c) 1000 to 900   (d) 900 to 800

(e) 800 to 700   (f) 700 to 600   (g) 600 to 500   (h) 500 to 400
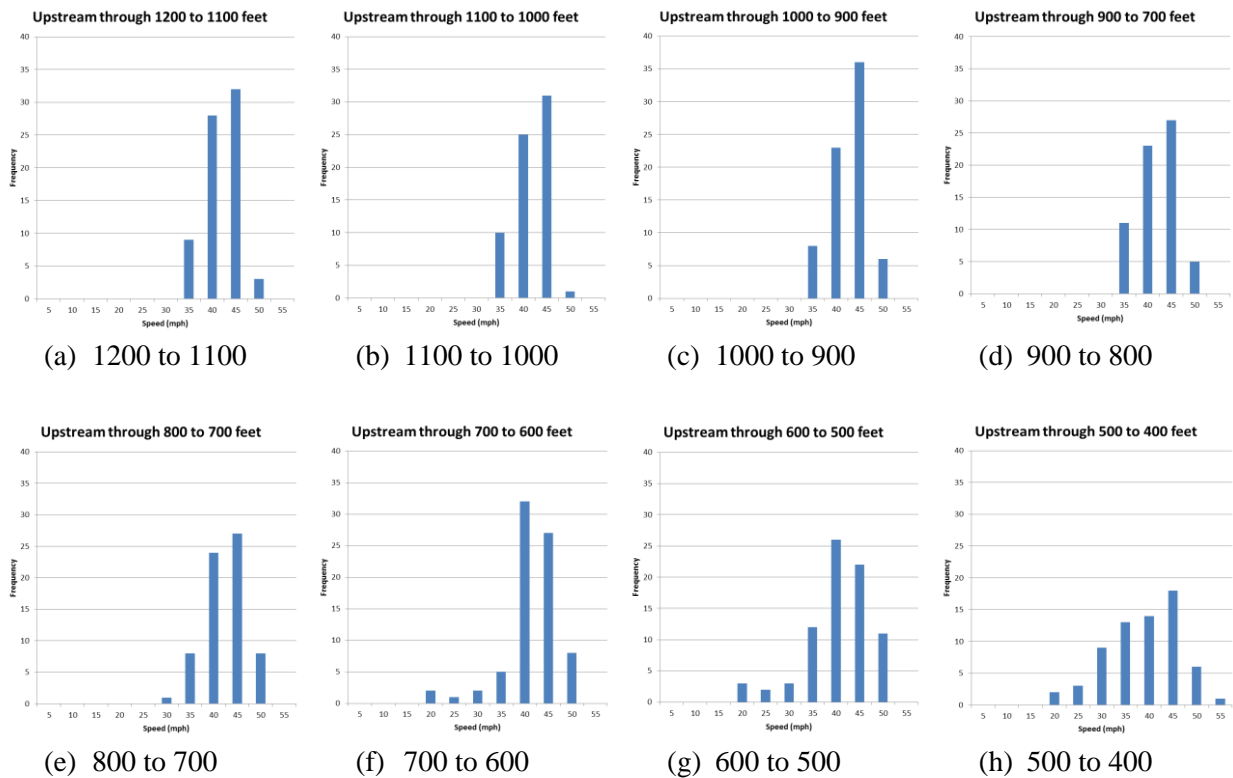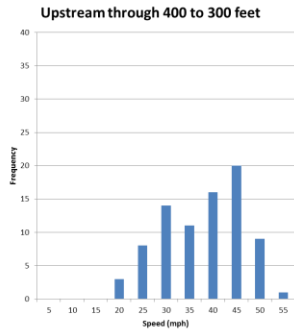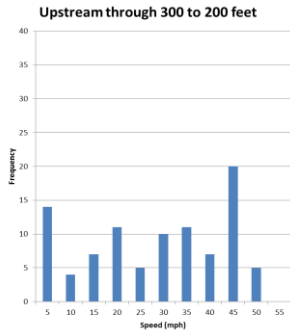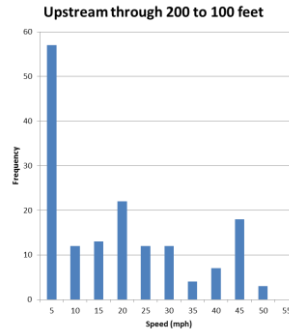
(i)　400 to 300　　　　　(j)　300 to 200　　　　　(k)　200 to 100　　　　　(l)　100 to 0
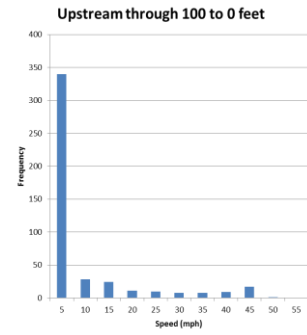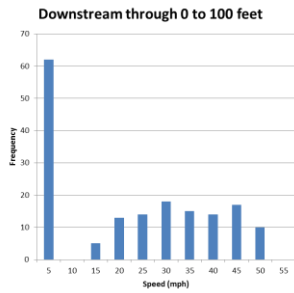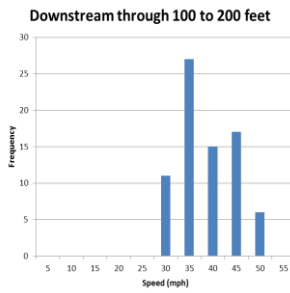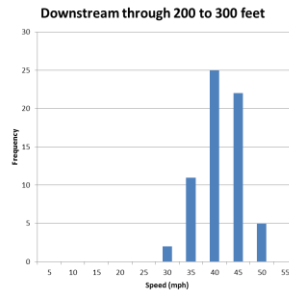
(m) 0 to 100　　　　　(n)　100 to 200　　　　　(o)　200 to 300　　　　　(p)　300 to 400
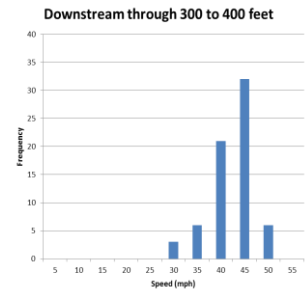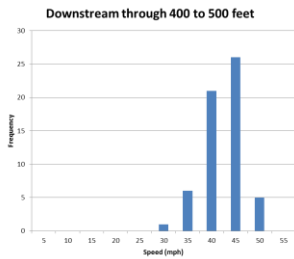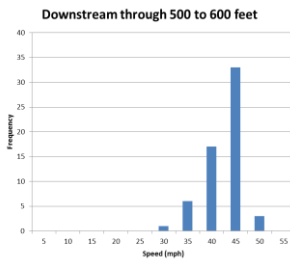
(q)　400 to 500　　　　　(r)　500 to 600

Figure 30 Speed Distribution by Section

53
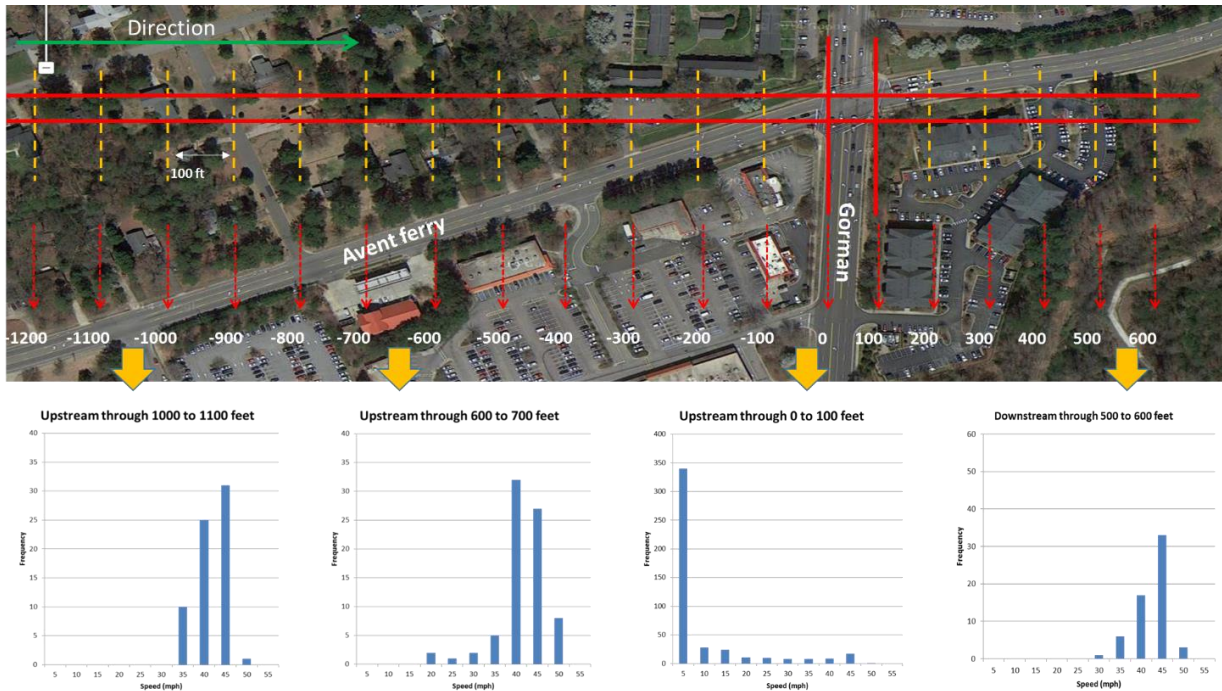
Figure 31 Speed Distribution by Section Position on an Arterial Segment

4.1.3 Summary

This section introduced the concept of link-based average speed using both time-mean and space mean speed definitions. It was first explained that an unbiased space mean speed for a roadway section can be calculated from high resolution vehicle trajectory data (i2D) without necessarily continuously tracking a vehicle. Secondl the researchers found that the average section speed by occupancy resulted in identical values to the true space mean speed formulation by Wardrop. In addition, the average section speed averaged by trip is nearly identical to the time mean speed. Finally, each section on a signalized intersection corridor yielded various speed distributions whose parameters vary as measurements are taken closer to the intersection proper. By exploring the space mean speed with speed distribution, the intersection impact area can be more accurately defined.

54

4.2   FREEWAY REAR-END CRASHES HOTSPOT DETECTION

The second application of the high resolution data derived from i2D technology explores the relationship between driver behavior at the short term tactical level and long term crash experience. Researchers have long sought to better identify the association between these two variable. Generally speaking, crash models have investigated such associations based on demographics, generalized driver behavior, driving conditions, roadway geometry, and traffic exposure. The availability of an in-vehicle sensing devices capable of capturing and documenting micro-scale dynamic driver behavior offers the opportunity to expand this field of research.

One of the primary objectives of this application is to explore the relationship between rear-end crash occurrences on freeway segments against microscopic, high resolution vehicle dynamic data for a sample of drivers using those segments, along with the link attributes data. Vehicle dynamic data was collected using i2D technology, as explained earlier. The research team developed and analyzed a link-focused integrated database which combines crash data, high resolution vehicle trajectory data via i2D, and GIS highway data.

4.2.1 Methodology

The empirically based methodology adopted in this study utilizes observed crash data and exposure information to generate long-term crash rates for each link on the freeway facility that is the focus of this study. Other link attributes, both static (e.g. length) and dynamic (average speed in time period of interest), are also incorporated. Furthermore, high resolution (second by second) vehicle data on the facility of interest are mapped onto each link, to coincide with the other link-based information. Finally, all data designed to be available for visual display in an ArcGIS environment.

Figure 32 describes the framework of the integrated database. While the focus of this study is on safety indicators, the integrated database can be used for a variety of transportation applications including reliability, environmental impacts, and planning. Each component is explained below in detail.
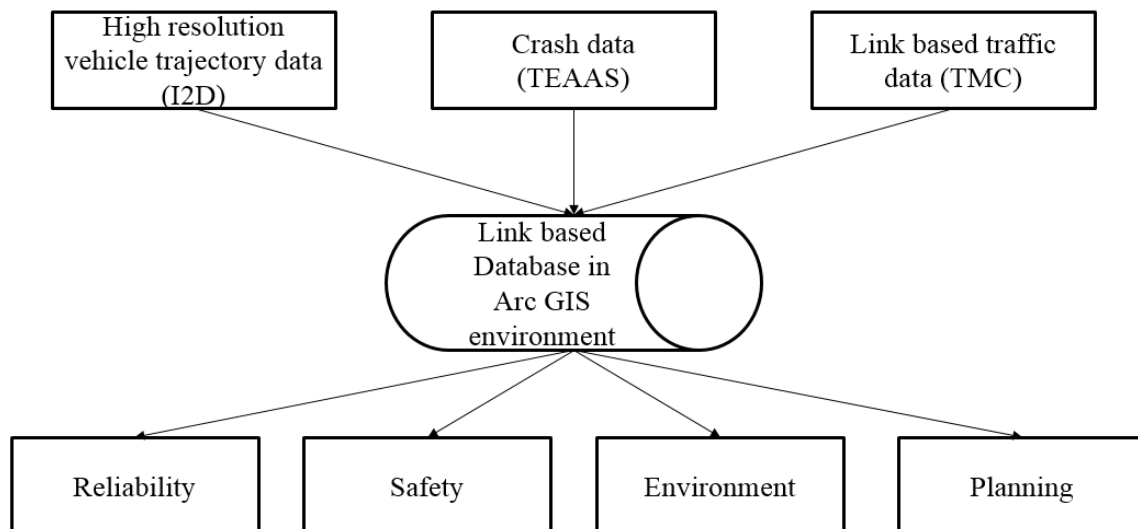
Figure 32 Integrated Database with High Resolution Vehicle Data

4.2.1.1 High Resolution Vehicle Trajectory Data – i2D

As explained earlier, i2D is an integrated Car Telematics solution to collect, transfer and process data from Light Duty Vehicles (LDV). I2D is primarily designed to support services and policies associated to road mobility, road safety, and energy efficiency by monitoring and quantifying driving behavior. The resulting data are processed and stored at four main levels as explained in detail earlier in Section 2.1.

4.2.1.2 Crash Data (TEAAS)

The North Carolina Department of Transportation (NCDOT) manages all reportable traffic crashes that are archived in the Traffic Engineering Accident Analysis System (TEAAS), a crash analysis software system that has been updated since 1990 and allows for data downloads. For each incident, TEAAS provides information on the crash type, injury severity, roadway characteristics, crash position (latitude and longitude), and environmental characteristics including weather and lighting conditions.

4.2.1.3 Link-Based Traffic Data

NCDOT has an ongoing contract with INRIX to provide link based travel time data for all freeways and most major arterials in the state. Data can be downloaded at a one minute resolution. INRIX uses GPS-enabled probe vehicles to collect speed information on over one million miles of roads across the United States. All freeway and arterial Traffic Message Channel (TMC) codes defined by Tel Atlas and Navteq, represents a directional roadway segment with a geo-located beginning and ending point. INRIX distinguishes between *internal* and *external* TMCs, as depicted in Figure 33. An external TMC is defined as starting from the merge area to the diverge area. An internal TMC is defined as starting from the diverge area to the merge area. INRIX also labels a TMC as a recurring bottleneck using the Average Historic Congestion Index (AHCI) as a guide. Simply stated, AHCI denotes the percentage of time congestion occurs on a given TMC in a specific time period based on long-term observations, typically 2 years.
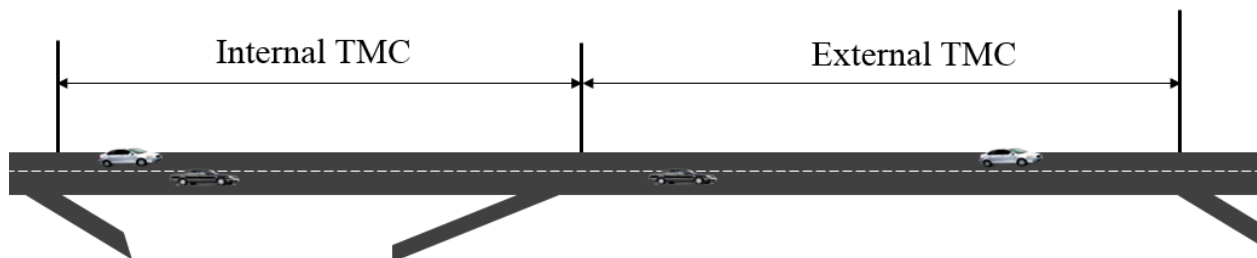
Figure 33 Definition of TMC Segment in INRIX

4.2.1.4 Integrated Database Format

TMC link information includes its unique identifier, length, TMC definition, and traffic direction. Link TMC traffic data include an estimate of aggregated traffic volumes (AADT) and the average speed for a selected time stamp. Both TEAAS and i2D generate a wealth of information on a TMC. For each crash, TEAAS reports crash time, location, type and severity, in addition to many other details. With respect to i2D data, second by second driver maneuvers are geo-located on a given TMC, include instantaneous speed and three-dimensional acceleration (longitudinal, lateral, and vertical), and are all tagged by a unique trip ID and driver ID. It should be noted that both TMC and i2D data are time-synchronized, while the TEAAS data by its nature extends over multiple years in order to produce a meaningful statistical sample.

4.2.2 Database Analysis

A key objective of this study is to verify a number of hypotheses as well as to test some unproven concepts regarding the association between micro-scale driver maneuvers and the occurrence of rear-end crashes on freeway TMCs. Three hypotheses are posed as follows:

1. Crash rate can be shown to be associated with microscopic driver maneuvers
2. Rear-end crashes occur more frequently in internal TMCs
3. Rear-end crashes are more likely to occur at recurring bottleneck areas

The first step in the analysis is to calculate and normalize crash rates across TMCs, in order to understand the impacting factors generated from both i2D and link characteristics. Crashes per million vehicle miles traveled (VMT) is considered to be the dependent variable, calculated as:

$$CVMT = \frac{C \times 1,000,000}{AADT \times 365 \times N \times L} \tag{9}$$

Where,

$CVMT$: TMC Crash per million vehicle-miles of travel
C : Total number of crashes in the study period for the TMC
$AADT$: Annual Average Daily Traffic Volume on TMC
N : Number of years considered
L : Length of the TMC in miles

In addition to the dependent variable above, we considered and generated a variety of independent variables which can be used to characterize microscopic driving behavior on each TMC, all of which are generated from i2D data, as listed in Table 14. Focus was placed on variables that are known to have some association with rear-end crashes, most notably those related to speeds and accelerations. Note that all the independent variables are also TMC-focused.

Table 14 Independent Variables Tested in the Analysis

Independent Variables		Description
TMC –Based	Avg.Spd	Average OBD speed (mi/h)
	SDSpd	Standard deviation OBD speed (mi/h)
	Avg.Acel	Average OBD aceleration or deceleration (m/s^2)
	SDAcel	Standard deviation of OBD aceleration or deceleration (m/s^2)
	D0	Number of seconds drivers decelerate on TMC
	D3	Number of seconds drivers decelerate below -3 m/s^2 on TMC
	D4	Number of seconds drivers decelerate below -4 m/s^2 on TMC
	D5	Number of seconds drivers decelerate below -5 m/s^2 on TMC
	D6	Number of seconds drivers decelerate below -6 m/s^2 on TMC
	nD0	Fraction of time drivers deccelerate on TMC (%)
	nD3	Fraction of time drivers deccelerate below -3 m/s^2 on TMC(%)
	nD4	Fraction of time drivers deccelerate below -4 m/s^2 on TMC(%)
	nD5	Fraction of time drivers deccelerate below -5 m/s^2 on TMC(%)
	nD6	Fraction of time drivers deccelerate below -6 m/s^2 on TMC(%)
	Min.Spd	Lowest OBD speed observed on TMC (mi/h)
	Max.Spd	Highest OBD speed observed on TMC (mi/h)
	Min.Acel	Lowest OBD deceleration (m/s^2) on TMC
	Max.Acel	Highest OBD acceleration (m/s^2) on TMC
	Avg.AcelY	Absolute average lateral acceleration on TMC (m/s^2)
	SDAcelY	Absolute standard deviation lateral acceleration on TMC (m/s^2)
	Avg.AcelZ	Absolute average vertical acceleration on TMC (m/s^2)
	SDAcelZ	Absolute standard deviation vertical acceleration on TMC (m/s^2)
Across TMCs	Avg.Spd Diff	Downstream TMC – current TMC average OBD speed (mi/h)
	SD Spd Diff	Downsttream TMC – current TMC standard deviation OBD speed (mi/h)
	InTMC	1 if internal TMC; 0 = otherwise
	Rbottle	1 if TMC tagged as recurrent bottleneck; 0 = otherwise

A correlation analysis was carried out using the entire set of independent variables to identify those independent variables that were highly correlated with the TMC rear-end crash rate. In addition, a spatial analysis of the location of rear-end crashes along the TMCs was used to identify the location of crash hotspots on a given TMC segment. This finding prompted further investigation of the microscopic driving behavior along those sections of TMCs identified as hotspots.

Finally, for those hotspots exhibiting the highest crash rates, it was important to distinguish between the frequency of those driving events deemed unsafe, but are predominated by a small number of drivers traversing the TMC, and those events that were frequented by most drivers traveling that same TMC. This distinction can be made by comparing the trip or driver based coefficient of variation (CV) of a selected performance measure at those hotspot locations. A high CV tends to indicate the likelihood that a hotspot is governed by a few drivers applying extreme levels of speeds or acceleration, and therefore, may be unlikely to be remedied by some type of treatment. Lower CV's on the other hand may be indicative of a design or operational flaw related to that specific TMC. The intent then was to test this hypothesis to see if low CV's and high crash rates were correlated. CV is mathematically defined as the ratio of the standard deviation σ to the mean μ of a specified measure of performance:

$$c_v = \frac{\sigma}{\mu} \qquad (10)$$

Where,

c_v : coefficient of variation (CV)

σ : standard deviation

μ : mean value

4.2.3 Database Analysis Results

4.2.3.1 Study Site Description

A 63-mile section of interstate I-40 between Exits 266 and 328 in North Carolina (NC) was selected for this case study. This freeway contains 60 TMCs, with an average TMC length of 1.04 miles. The posted speed limit of the section varies from 65 mph to 70 mph. Vehicle trajectory data was collected using a fleet of 20 volunteer vehicles from April 1 to June 30, 2014. A total of 150,322 seconds (~43 hours) of i2D data was extracted, averaging about 2,521 seconds per TMC. Figure 34 (a) denotes the trip travel time distribution comprising a total of 224 trips traversing portions of or the full network. It shows that most trips last about 30 minutes or less on the selected study area. On the other hand, Figure 34 (b) shows number of trips on any given TMC varied from 8 to 293seconds. The crash data consisted of a total of 1,311 rear-end crash records for the studied section covering the years 2010 to 2013 (4 years). The average number of rear-end crashes per TMC is 5 crashes per year.

With the focus of the study being on rear-end collisions and it association with driver behavior, it is important to get a sense of the range of acceleration and deceleration rates recorded via i2D on the study site. In Figure 34, (c) shows the speed acceleration contours as computed from the OBD instantaneous speeds (km/h) for all trips across the study area. As one would suspect, the general trend is a decrease in the range of accelerations and decelerations as speeds increase, with some notable exceptions. As mentioned earlier, it is the set of extreme deceleration events (in the range of 3-4 m/sec/sec or 10-14 ft/sec/sec) that is of interest in this study.

(a) Trip travel time distribution (i2D)   (b) Number of trips by TMC segment

(c) Speed acceleration contours from i2D in the study area
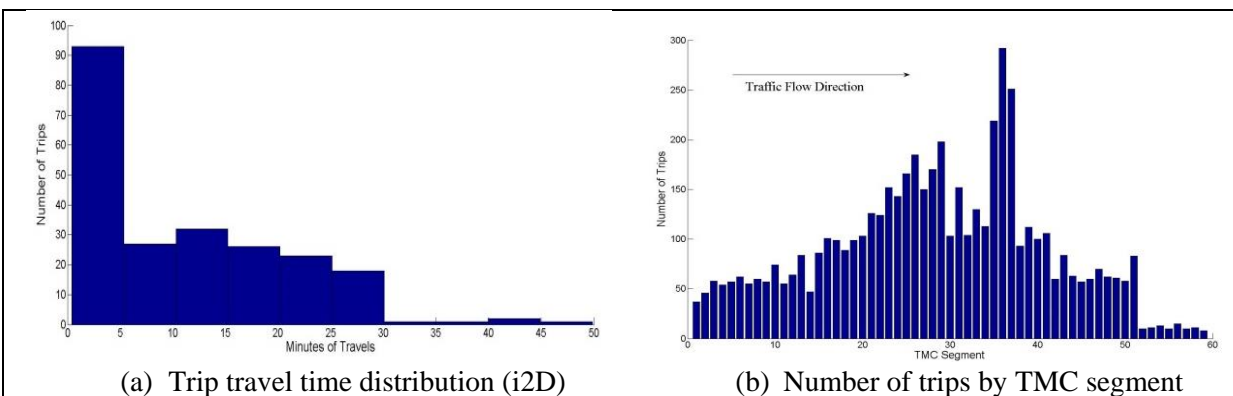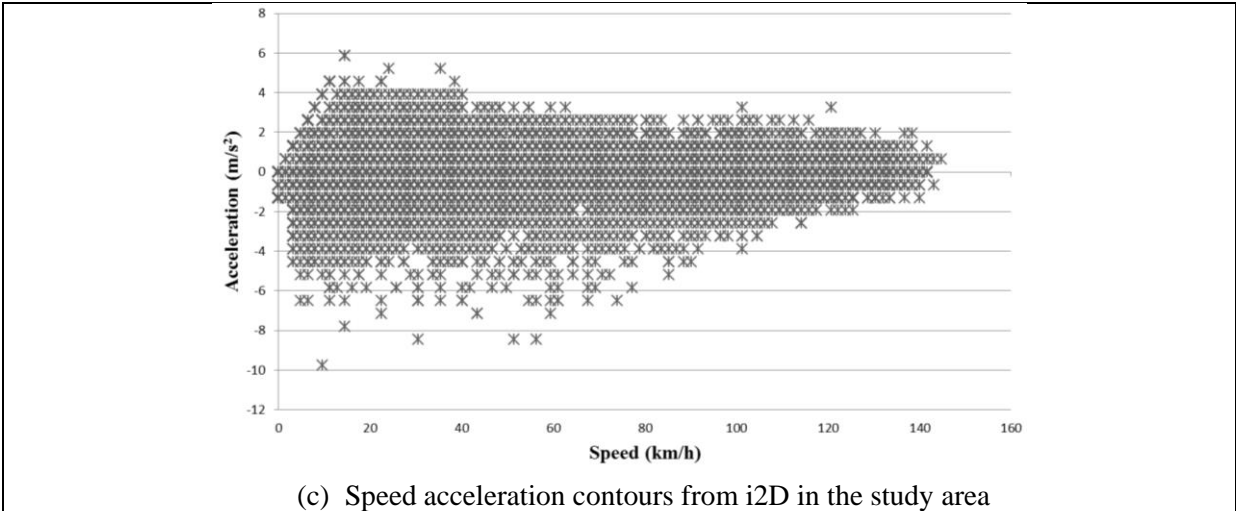
Figure 34 i2D Raw and Processed Data Plots

4.2.3.2 Vehicle Maneuver Characteristics and Crashes

Figure 35 shows the average and standard deviation of OBD speeds and crash rates on the study site by TMC segment (segmented were numbered 1 to 60 going from west to east). The left axis shows the rear-end crash rate and the right axis depicts average TMC speed (mi/h). The blue line represents average speed and the green line the speed standard deviation. The black bar indicates the crash rates for internal TMC, while the gray bars shows crash rates for external TMCs, as defined in Figure 5. In this figure, it is very evident that internal TMC have comparatively higher crash rates than those reported for external TMCs. The data also shows a significant speed drop in the western portion of I-40, around the TMC labeled 15 to 25.
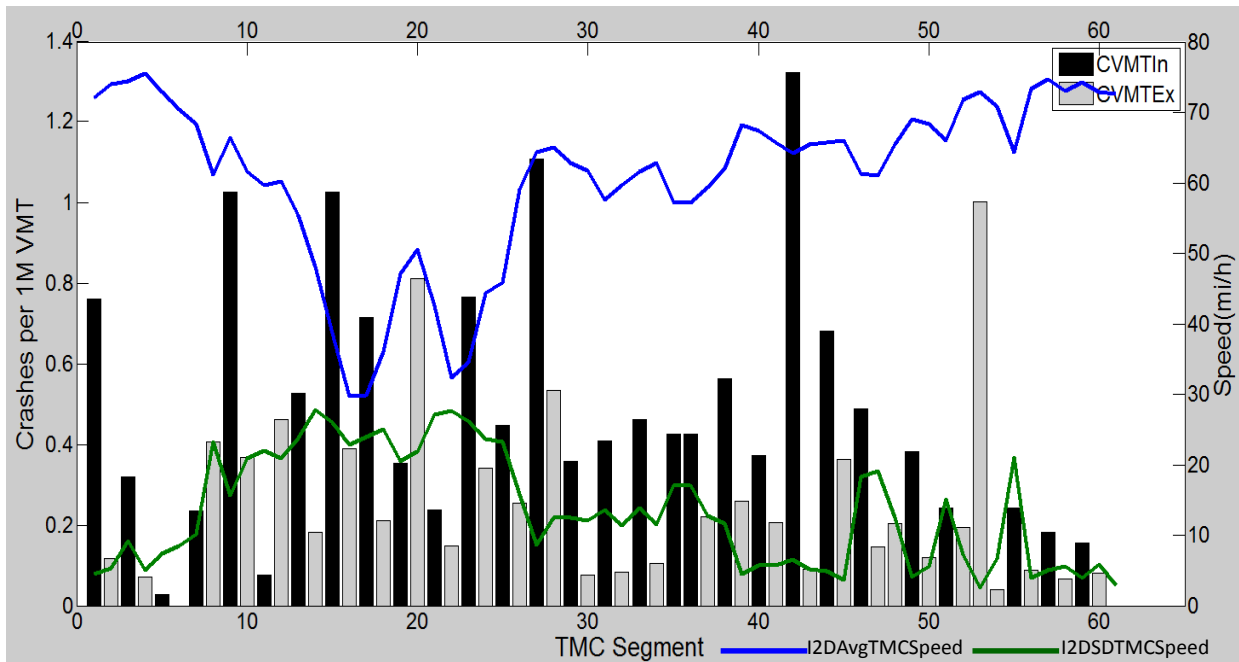
Figure 35 Rear-end crash rate and speed by TMC

Correlation analysis results between the dependent variable (CVMT) and the set of independent variables cited in Table 14 is shown in Table 15. Those independent variables exhibiting strong correlation with the crash rate are SDAcelZ, nD3, D4, nD4, InTMC. The nD4 variable is the only one selected as an explanatory variable due to its high correlation among all deceleration-based variables. Strong associations with crash rate were detected on the TMCs in which drivers frequently decelerated at rates below -4m/s2. This seems to be consistent with the finding by Tarko et al. There was also a positive significant correlation between rear-end crash rate and InTMC. Interestingly, the absolute standard deviation of the vertical acceleration Z was also highly associated with crash rate.

Table 15 Correlation between rear-end crash rate and selected independent variables

	CVMT	SDAcelZ	SDAcel	D0	nD0	D3	nD3	D4	nD4	D5	nD5	D6	nD6	InTMC	Rbottle
CVMT	1.00														
SDAcelZ	**0.33**	1.00													
SDAcel	0.24	0.36	1.00												
D0	0.17	0.38	0.25	1.00											
nD0	0.08	-0.08	0.33	0.10	1.00										
D3	0.28	0.28	0.69	0.62	0.24	1.00									
nD3	0.37	0.25	0.87	0.32	0.38	0.85	1.00								
D4	0.30	0.33	0.69	0.61	0.23	1.00	0.85	1.00							
nD4	**0.39**	0.30	0.84	0.37	0.36	0.89	0.98	0.89	1.00						
D5	0.19	0.26	0.72	0.58	0.25	0.99	0.84	0.99	0.87	1.00					
nD5	0.26	0.31	0.92	0.28	0.35	0.81	0.95	0.82	0.95	0.84	1.00				
D6	0.02	0.21	0.74	0.52	0.25	0.88	0.76	0.88	0.75	0.94	0.81	1.00			
nD6	0.21	0.31	0.90	0.27	0.35	0.78	0.90	0.79	0.90	0.82	0.98	0.83	1.00		
InTMC	**0.39**	0.21	0.10	-0.11	-0.22	0.16	0.10	0.13	0.11	0.17	0.08	0.14	0.10	1.00	
Rbottle	0.05	0.18	0.59	-0.03	0.11	0.28	0.52	0.31	0.48	0.28	0.55	0.31	0.52	0.08	1.00

4.2.3.3 Spatial Distribution of Rear-end Crashes

The correlation analysis indicated some association between crash rates and internal TMCs. Further graphical analysis confirmed an interesting and unsurprising trend in crash locations. As highlighted in Figure 36 below, most crashes which occurred on internal TMCs tended to be concentrated just upstream of the on-ramp merge, and for a distance further upstream. External TMCs, on the other hand, had fewer and were more spatially dispersed crash locations.

61

(a) Stack interchange at internal TMC

(b) Folded diamond at internal TMC

(c) Cloverleaf interchange at internal TMC
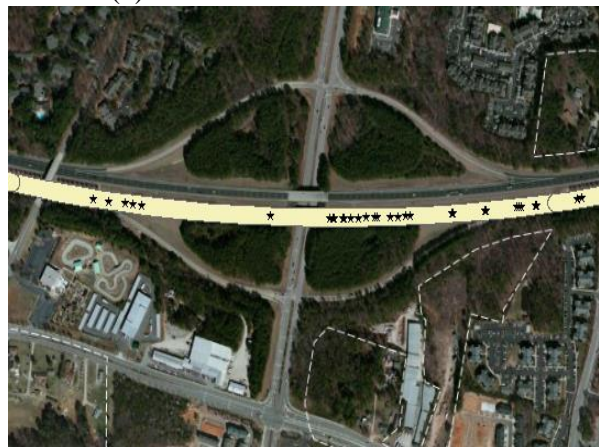
(d) Diamond interchange at internal TMC
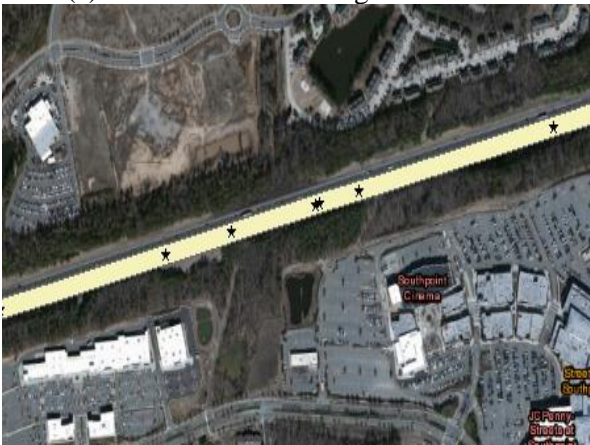
(e) Example 1 of external TMC

(f) Example 2 of external TMC

Figure 36 Spatial pattern of Crashes on Internal and External TMCs

A spatial analysis of the crash occurrences for the 30 on-ramps for the internal TMCs at the site showed that 85% of all rear-end crashes occurred within 2,000 feet of the on ramp gore (see Figure 37). This portion of the TMC is hereafter labeled a "hotspot" to indicate the high density of crashes in this region.
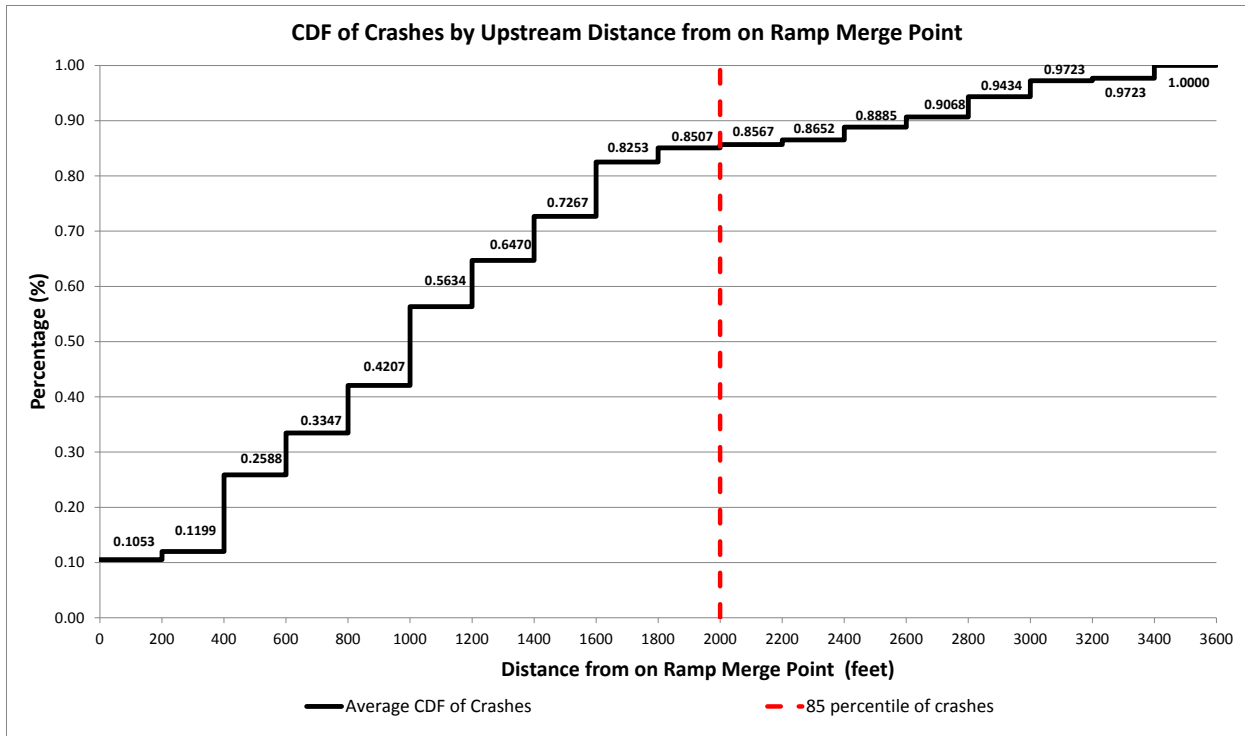
Figure 37 Average Crash Rate per Specified TMC Segment or Sub-Segment

In terms of crash propensity, the average internal TMC for the site had a crash rate of 0.53, while the average external TMC had a crash rate of 0.1, a 3.21 fold increase. Furthermore the average crash rate within 2,000 feet upstream of on-ramp, which the research team labeled as the hotspot, had a rear-end crash rate 4.45 times higher compared to that of the average external TMC. Figure 38 shows overall average rear-end crash rates for various TMC segment definitions.
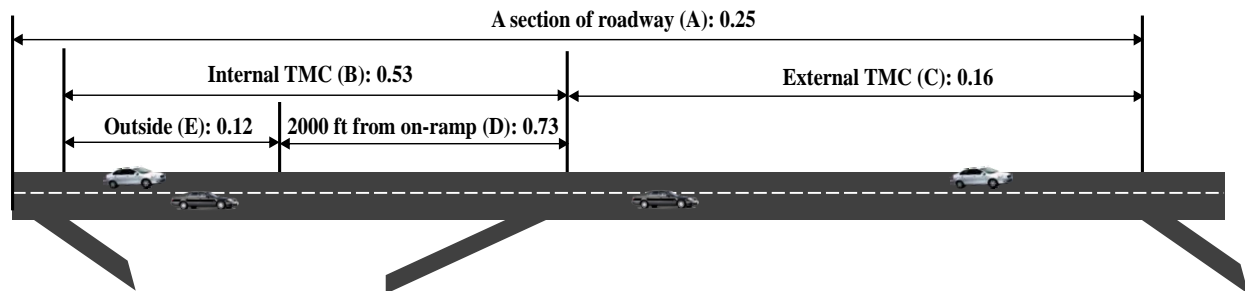
Figure 38 Average Crash Rate per Specified TMC Segment or Sub-Segment

4.2.3.4 Drivers' Deceleration Patterns on Internal TMCs

This section focuses on investigating the micro-scale driver behavior within the four internal TMCs that exhibited the highest rear-end crash rates in the study area. For comparison purposes the average TMC had a crash rate of 1.10 CVMT. An internal TMC (B) is subdivided into the hotspot (D) and outside segment (E), as indicated in Figure 38 above. The intent is to highlight differences, assuming such differences exist in driver behavior, on sub-segments D and E.

63

Table 16 Observed Speeds and Accelerations at Hotspots and Outside Hotspots for Internal TMCs with Highest Rear-End Crash Rates

	I-40 Exit 266		I-40 Exit 276		I-40 Exit 291		I-40Exit 298	
	Outside (E)	**Hotspot (D)**	Outside (E)	**Hotspot (D)**	Outside (E)	**Hotspot (D)**	Outside (E)	**Hotspot (D)**
Rear-End Crash Rate(CVMT)	0	**1.40**	0	**1.41**	0	**1.78**	0.25	**1.74**
Seconds of i2D observations	252	487	597	442	1406	1344	687	1003
Avg.Spd (mi/h)	72.69	71.84	68.07	64.96	62.60	60.60	63.75	64.49
SDSpd (mi/h)	3.97	4.41	13.63	16.40	12.45	15.35	5.77	7.11
Avg.Acel (m/s^2)	0.10	-0.08	-0.01	-0.07	0.04	-0.05	-0.12	-0.04
SDAcel	0.65	0.70	0.59	0.78	0.93	0.90	0.67	1.00
Avg.Acel Z	0.29	0.32	0.01	0.00	0.19	0.17	0.31	0.31
SDAcelZ	0.16	0.16	0.07	0.07	0.15	0.17	0.20	0.28
nD0 (%)	11.00	19.00	13.73	17.40	13.51	16.96	18.77	13.26
nD3 (%)	0	0	0	1.58	1.28	1.49	0.72	2.39
nD4 (%)	**0**	**0**	**0**	**0.45**	**0.64**	**0.67**	**0**	**1.69**

Table 16 shows comparison of hazardous driving maneuvers for the top four hotspot versus the area outside of the hotspot, all on internal TMCs. Predictably, the average speed on three of the four hotspots was lower. In addition, the standard deviation of the vertical acceleration was higher at hotspots. With respect to the normalized deceleration rates, higher values were reported for all hotspots. For comparison purposes, the normalized deceleration rate (nD4) across the entire study area is 0.10%. Thus, the non-zero fraction of the time drivers were decelerating at rates 4 m/sec2 or more appeared almost exclusively on hotspots.

It is evident that drivers often decelerate at high rates at hotspot locations, as shown in Table 4. To ascertain whether this propensity to decelerate is a common across drivers or is governed by a few select drivers, Figure 39 depicts the CV's for nD3, nD4 reported on a trip basis. This was analysis was conducted to avoid a potential bias in the data stemming from outlier drivers with more erratic travel behavior. At the hotspots at both Exit 291 and 298 the CVs are lower compared to the outside segment. This finding suggests there is evidence that many drivers often decelerate below -3 or -4m/s2 to avoid potential rear-end crashes in the location of hotspot.
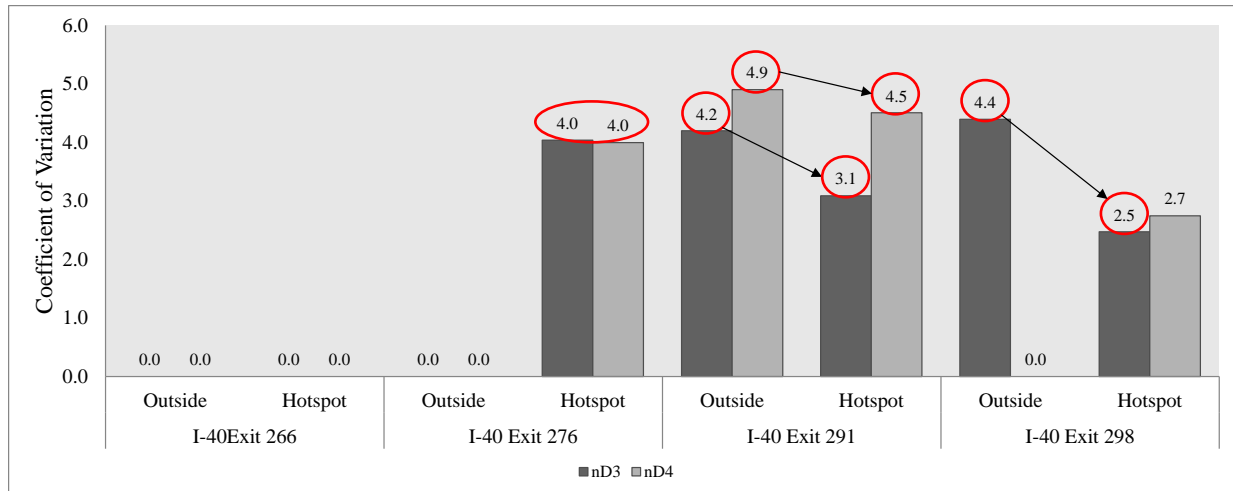
Figure 39 Spatial Analysis of the Coefficient of Variation for Excessive Deceleration Rates

4.2.4 Conclusions

This study discusses exploratory work in the area of micro-driver behavior and its association with real world crash data. The association is indirect in that it relies on the spatial distribution of crashes and driver events over specific links, or TMCs. This type of analysis was possible because of a high resolution, in-vehicle data sensing system (i2D) that allows the collection and instantaneous transmittal of driver-based data to a server, enabling a full characterization of events across multiple drivers, trips and TMCs. When integrated with geo-located crash and operational data, the analysis power increases significantly. The results presented here should be considered as preliminary due to the short-term nature of the high resolution data, as contrasted with the 4 years of crash data. The benefit of this approach is that additional high resolution data are still being collected and will continue to augment the initial database developed through this study.

One important finding gleaned from this work is initial evidence of an association between crash rates and driver behavior. This evidence was clearer at specific locations on the freeway facilities, referred to as "hotspots" and defined to occur within 2,000 feet upstream of an on-ramp merge (there were 30 such segments on the facility we studied). In fact, 85% of rear-end crashes occurred within hotspots on internal TMCs. Another finding is related to the highest rear-end crash rate locations, which also indicated a preponderance of high incidents of extreme decelerations (above 4/m/s/s). The performance measure indicating this preponderance is the fraction of all time spent on a given TMC where drivers exercised those extreme decelerations (labeled nD4). Finally, and in order to ascertain that the performance measure was not tainted by a handful of drivers applying extreme accelerations, this paper tested for the nD4 coefficient of variation and found it to be lower on a trip-basis at hotspots. This indicated to us that the extreme deceleration problem is actually the norm rather than the exception at high-crash locations.

Clearly, additional research needs to be done to extend the temporal scale of the high resolution data in order to reaffirm the findings of this study. The system is, however, is now in place and more data is currently being collected for the purpose of enhancing this study. Additional analyses will be forthcoming.

4.3 FREEWAY FREE FLOW SPEED ESTIMATION

The availability of high resolution vehicle data also opens up opportunities to take a fresh look at how and why free flow speed (FFS) estimations have been conducted in the past. Related literature reveals that the FFS is affected by lane width, weather, horizontal alignment, sight distance, lateral clearance, design speed, driver behavior, and adjacent environment (Gaber et. al., 1989). A FFS model is provided in the 1997 HCM (Transportation Research Board, 1997), in which a value of 75 mph is recommended for ideal rural freeways and 70 mph for ideal urban and suburban freeways. Adjustments to the ideal for estimating FFSs are also provided.

The 2010 US HCM provides a couple of approaches to estimate FFS (Transportation Research Board, 2010). The HCM recommends field data collection as the preferred approach. This is done by measuring the average speed of vehicles at very low flow rates (less than 1,000 pcph/lane) and use that value as the estimate of FFS. The other option is to use a predictive model as shown below (Roess 2009). This model has been developed and calibrated based on extensive field data.

$$FFS = BFFS - f_{LW} - f_{RLC} - 3.22 \times TRD^{0.84} \tag{11}$$

Where FFS represents the estimated FFS and $BFFS$ is the base FFS with a fixed value of 75.4 mph. Parameters f_{LW} and f_{RLC} are adjustments for lane width and right side lateral clearance respectively. TRD Is the total ramp density, defined as the number of directional ramps located within 3 mi upstream and 3 mi downstream of the midpoint of the segment under study, divided by 6 mi. The total ramp density has been found to be a measure of the impact of merging and diverging vehicles on FFS (Transportation Research Board, 2010).

The limited number of parameters in the FFS estimation model has generated some accuracy concerns. One notable criticism is the absence of the segment speed limit in the prediction in Equation 1. Matthew et al. for example used radar guns and manual tally sheets to collect 1,668 individual vehicle speeds at ten different sites with speed limits ranging from 20 mi/h to 75 mi/h (Mathew et. al, 2011). The authors found a very strong correlation between the resulting FFS and posted speed limit. In this study, we have designed a set of hypotheses to assess the adequacy of the FFS model and provides alternative FFS estimation models.

4.3.1 Data Collection and Preparation

Based on one full year of high resolution i2D data collected from a local fleet of about 20 vehicles driven by volunteer participants, more than 20 million seconds of trip records have been accumulated. These include instantaneous speed (1Hz), three dimensional acceleration, route choice, stopped delays, and link and route travel times. Local trips occurred on both freeway and surface street segments in the Research Triangle Region in North Carolina. Several trips outside the region were also tracked and archived. The collected data were downloaded and archived in a local SQL database. This eases the process of data filtration based on time of day and day of week, so as to remove any abnormal driving conditions, e.g. under severe weather condition. An ArcGIS tool was used to map the coordinates of any second of travel data with facilities geometry to extract spatially defined data for the analysis. Algorithms have been developed to distinguish the direction

of travel. Based on the described process, the team collected and filtered an appropriate set of high resolution trajectory data for ten specific freeway segments to test the key hypotheses.

4.3.2 Hypotheses Regarding FFS and Supporting Data

This application is focused on evaluating the following three hypotheses:

1. H1: The empirically measured FFS from i2D is impacted by the posted speed limit;
2. H2: driver familiarity plays a role in determining FFS, regardless of segment. This hypothesis can only be tested using our i2d dataset;
3. H3: The density of ramps inside a section affects the variability in FSS, more so than the density estimated over a 6 mile area centerd on the segment.

In order to test these three hypotheses, 10 freeway sections in North Carolina were selected, with varying geometric and traffic conditions. Table 17 summarizes the section attributes (all had adequate shoulders and ideal lane widths). Four of the sites have a posted speed limit of 55 mph and the remaining six 65 mph. Sections were selected in a similar manner to the INRIX definition of external (no ramps; the first 4 sections) and internal (within interchange areas; the last six sections) TMCs. Table 18 shows the extent of the (unfiltered) travel data collected in terms of trips, drivers and travel seconds on each selected section.

Table 17 Ten selected freeway sections for testing proposed hypotheses

Section ID	Section length (mi)	# of lanes	Posted speed limit (mph)	Total number of ramps within 6 miles	Ramp density per mile*	Total number of ramps in section	Section ramp density per mile**
S1E	0.6	3	65	12	2.00	0	0
S1W	0.6	3	65	12	2.00	0	0.00
S2E	0.5	4	65	9	1.50	0	0.00
S2W	0.5	4	65	7	1.17	0	0.00
S3N	0.3	2	55	17	2.83	2	6.67
S3S	0.3	2	55	17	2.83	2	6.67
S4N	0.4	2	55	21	3.50	1	2.50
S4S	0.4	2	55	17	2.83	1	2.50
S5N	0.4	2	55	20	3.33	1	2.50
S5S	0.4	2	55	16	2.67	2	5.00

*TRD variable in Equation (1)   **Ramp density within the section length proper only*

Table 18 Summary of the collected data by site

Section	# of Trips	# of Drivers	Total section travel seconds
S1E	201	21	6,999
S1W	165	18	5,649
S2E	470	22	15,233
S2W	438	23	13,781
S3N	434	22	10,377
S3S	462	21	11,148
S4N	362	20	9,842
S4S	265	20	7,193
S5N	307	20	8,619
S5S	185	20	4,569

Trip travel times were used to estimate the space mean speed on each section as was shown in the first application in this chapter Figure 31depicts the range of trip-based space mean speeds by section, showing considerable variability within and across sites. Note that those data are not filtered by level of congestion and thus do not necessarily represent strictly FFSs.

4.3.3 Hypothesis 1

Figure 40 shows the average and distribution of observed trip FFS from high resolution trajectory data for the ten grouped by posted speed limit. Clearly, the lower speed limit (55 mph) Group 2 is associated with a lower observed FSS. Considering the number of trips (1,890) and different drivers (20) traveling these sections, it is readily apparent that the posted speed limit has considerable impact on the prevailing average FSS on these ten freeway sections.
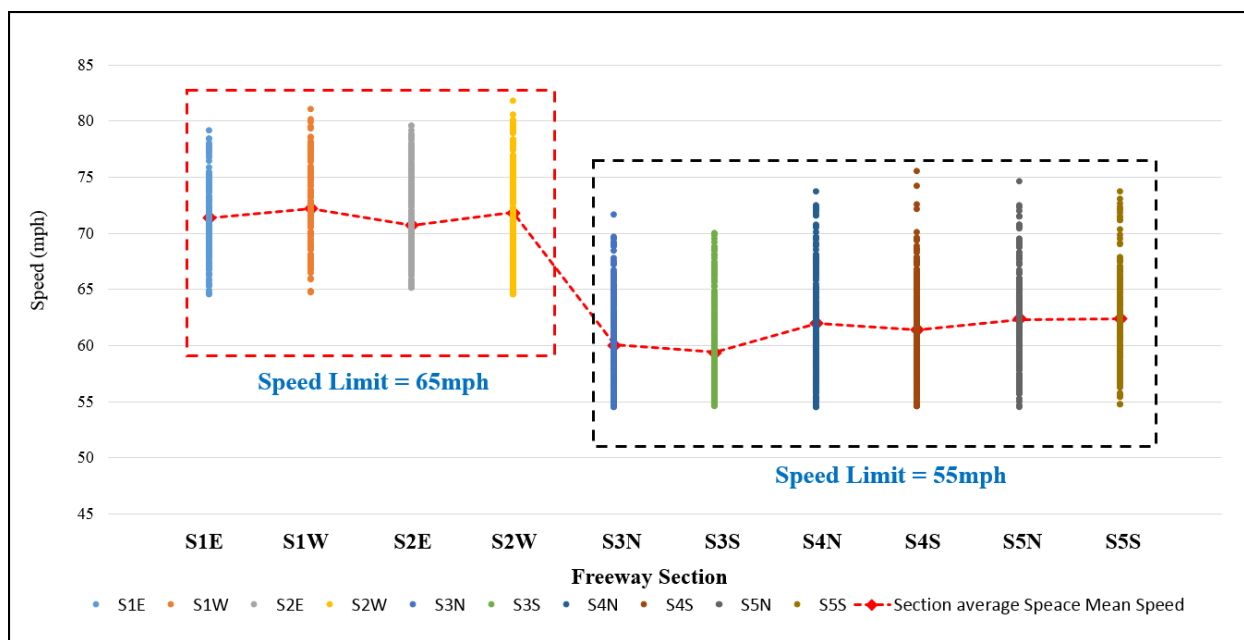
Figure 40 Distribution of filtered Trip FFS observations by site and posted speed limit
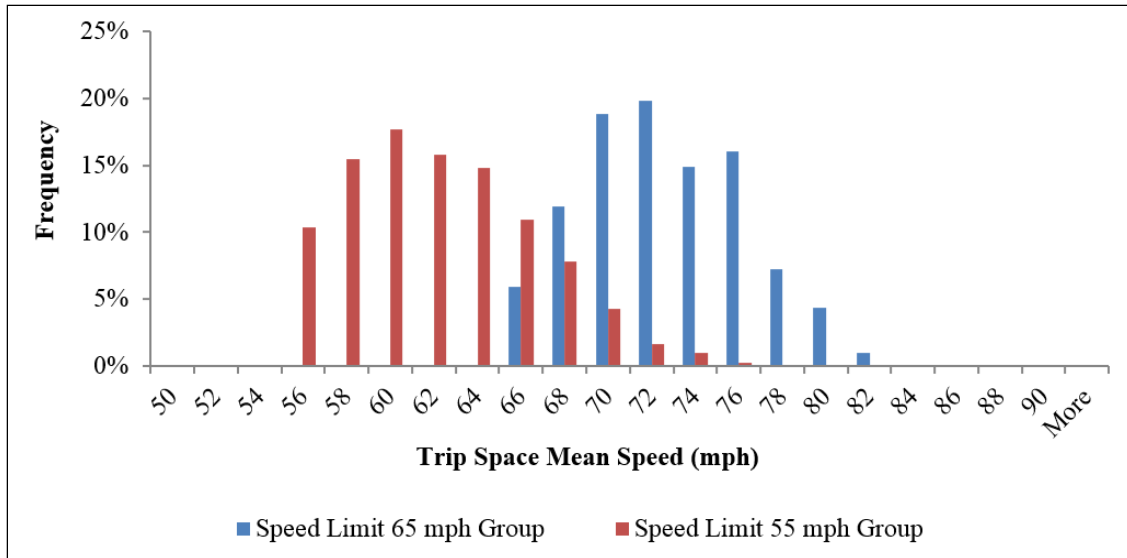
Figure 41 PDF comparison of FFS by Speed Limit Group

Figure 41 shows the probability density function (PDF) of the observed trip speeds for the two groups of sections with different posted speed limits. From this data, which was confirmed by a two-sample t statistical test with unequal variance, it is evident that the hypothesis that speed limit has an impact on drivers' FFS cannot be rejected.

4.3.4 Hypothesis 2

The second hypothesis investigated in this paper, is whether there is (positive) correlation between the observed FFS and *the level of driver familiarity* with the freeway section. For this purpose the data was sorted into two nearly equally sized groups, with those drivers who traversed a section less than once every 7 days assumed to be unfamiliar with the route. The emerging speed CDF distributions for those two groups shown in Figure 42 show an average speed difference of 5 mph in favor of the familiar drivers at the 65mph posted speed limit sites. However, the speed distributions were nearly identical in the 55 mph freeway sections. A K-S test between the two CDF groups revealed statistically significant differences between the two sets of distributions, thus the hypothesis that driver familiarity impacts the level of FFS could not be rejected either, although the data are inconclusive in terms of direction.
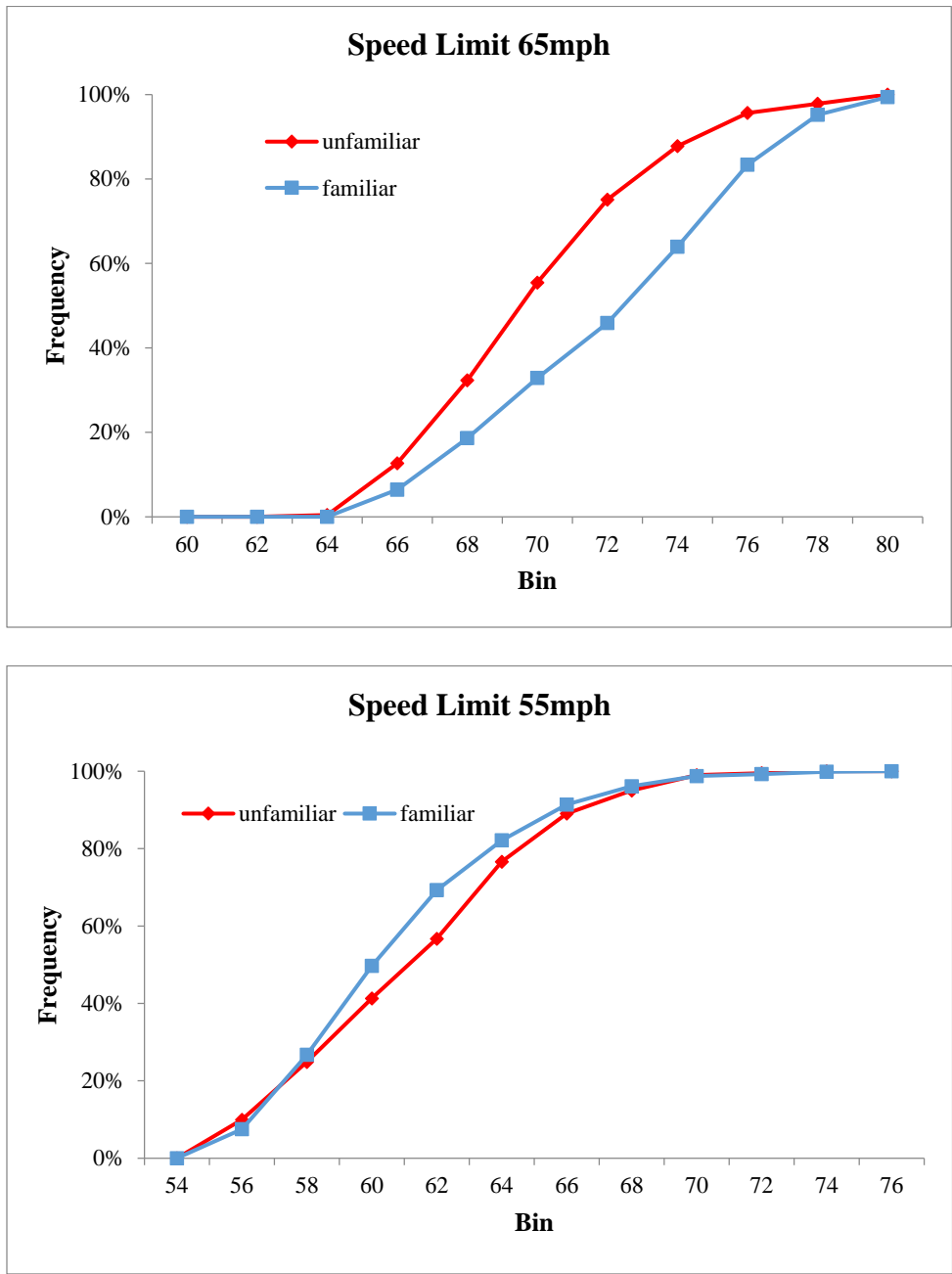
Figure 42 Distribution of FFS by Driver Familiarity with the Section and by Posted Speed Limit

4.3.5 Hypothesis 3

The variable section ramp density for each of the ten sites is listed in Table 1. This test focused on two sections (S3N and S5N) that had varying section densities but operate at the same speed limit. S3N has a section density of 6.67 compared to 2.50 for section S4N. This hypothesis stipulates that where ramp density is high, even under free flow conditions, through drivers on the section will tend to experience a wider range of accelerations and decelerations in anticipation of merging and diverging traffic, which inevitably occurs at lower speeds. Figure 43 shows the probability density functions of accelerations and decelerations recorded for those two sections. As shown, it

70

is inferred that the low ramp density section exhibited less variation in the range of accelerations (-2 to +3) compared to the high ramp density section (-2 to +5)
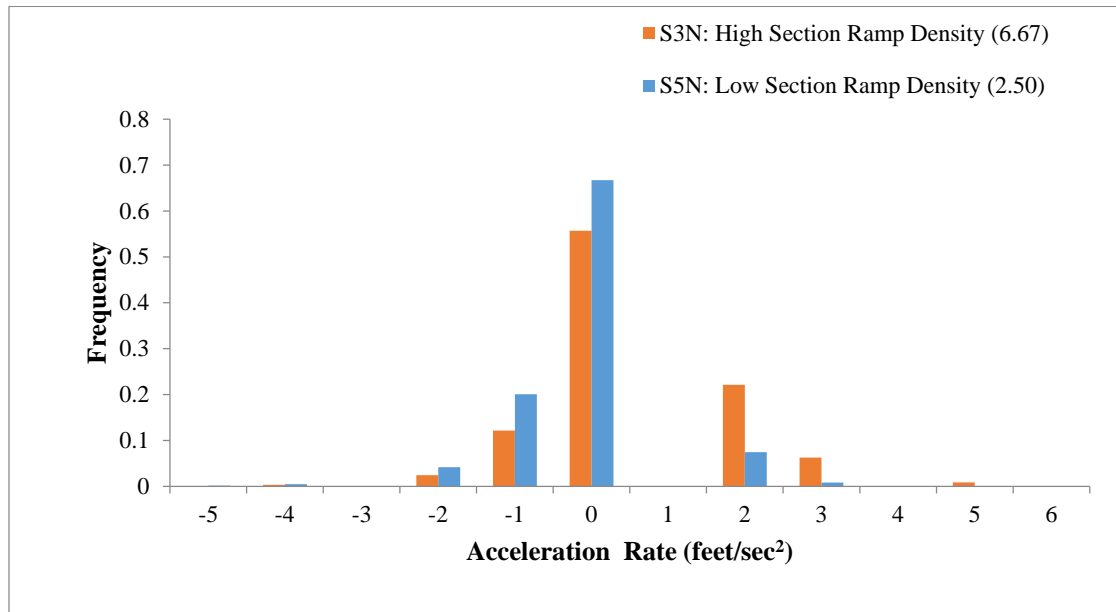
Figure 43 Acceleration and deceleration distributions for high and low ramp density freeway section groups

4.3.6 Estimating FFS from High Resolution Data

Based on the findings from our hypothesis tests, three empirical models were constructed for estimating FFS. In the first model, the form of HCM 2010 FFS estimation model was used to calibrate the parameters based on observed data. In the second, section ramp density was substituted for overall ramp density in the HCM 2010 model, which produced some improvement in the model fit. The third model was constructed to incorporate posted speed limit along with section ramp density as explanatory variables based on the findings of the hypothesis testing. The results shown in Table 19 indicate that the third model produced the better estimates of FFS based on observed high resolution data.

Table 19 Summary of the empirical FFS estimation models

Model #	FFS Model (mph)	R^2
1 (HCM 2010 Form)	$FFS = 81.185 - 7.651 \times (Total\ Ramp\ Density)^{0.84}$	0.415
2 (Section Density)	$FFS = 69.786 - 2.198 \times (Section\ Ramp\ Density)^{0.84}$	0.475
3 (+ Speed Limit)	$FFS = 19.763 + 0.798 \times (Speed\ Limit) - 0.771 \times (Section\ Ramp\ Density)$	0.576

Table 20 compares FSS estimates for the ten sites using predictive models from the literature and those shown in Table 4. The former group includes FFS models in the 1997, 2000, 2010 HCMs along with the model by Matthew et al (Matthew et. al., 2011). All estimates are compared to field observations, shown in the second column in the table, and the percent differences from the field

shown in parentheses. The bottom row shows the mean absolute percent error (MAPE) by model. They clearly show that the proposed model 3 yielded the lowest MAPE among all models considered at 1.3%. On the other hand, the HCM 2010 did not fare as well, with an MAPE of 6.6%. In fact, as a group Models 1 to 3 were the top performers in terms of explaining the observed FFSs.

Table 20 Summary of Selected FFS Model Predictions and Errors

Site	Field FFS (mph)	HCM 1997, 2000	HCM 2010	Matthew et al (12)	Model #1 (Table4)	Model #2 (Table 4)	Model #3 (Table 4)
S1E	71.3	65.0 (-8.8%)	69.6 (-2.3%)	65.7(-7.9%)	67.5(-5.3%)	69.8(-2.1%)	71.6 (0.5%)
S1W	72.2	65.0 (-10.0%)	69.6 (-3.5%)	65.7 (-9.1%	67.5 (-6.5%)	69.8 (-3.3%)	71.6 (-0.8%)
S2E	70.8	67.0 (-5.4%)	70.9 (0.1%)	65.7 (-7.3%)	70.4 (-0.5%)	69.8 (-1.4%)	71.6 (1.2%)
S2W	71.9	67.3 (-6.4%)	71.7 (-0.2%)	65.7 (-8.7%)	72.5 (0.8%)	69.8 (-2.9%)	71.6 (-0.4%)
S3N	60.1	62.7 (4.3%)	67.7 (12.6%)	55.6 (-7.6%)	62.9 (4.6%)	59.0 (-1.9%)	58.5 (-2.6%)
S3S	59.4	62.7 (5.5%)	67.7 (13.9%)	55.6 (-6.5%)	62.9 (5.8%)	59.0 (-0.7%)	58.5 (-1.5%)
S4N	62.0	62.0 (0.0%)	66.2 (6.7%)	55.6 (-10.4%)	59.3 (-4.4%)	65.0 (4.9%)	61.7 (-0.4%)
S4S	61.4	62.7 (2.1%)	67.7 (10.2%)	55.6 (-9.5%)	62.9 (2.4%)	65.0 (5.9%)	61.7 (0.5%)
S5N	62.3	62.2 (-0.2%)	66.6 (6.8%)	55.6 (-10.8%)	60.2 (-3.4%)	65.0 (4.4%)	61.7 (-0.9%)
S5S	62.4	62.8 (0.7%)	68.1 (9.1%)	55.6 (-11.0%)	63.7 (2.1%)	61.3 (-1.8%)	59.8 (-4.2%)
MAPE		4.3%	6.6%	8.9%	3.6%	2.9%	1.3%

4.3.7 Conclusion

This research investigated the factors affecting free flow speed (FFS) in the HCM 2010 using high resolution i2D data. Ten freeway sections in North Carolina, representing a diverse set of parameters that potentially can affect FFS, were selected for this study. The resulting data was filtered and organized and key parameters, such as average travel speed and acceleration rates, were estimated. Three hypothesis were constructed and tested. The first hypothesis confirmed that speed limit contributes to the observed FFS. The second hypothesis regarding driver familiarity effects was inconclusive and depended on the prevailing speed limit. The research team also confirmed that variability in the observed FFSs were impacted by section ramp density, which counts ramps within the section proper only.

Finally, a comparison of several FFS predictive models have shown that a model that includes both speed limit and section ramp density as explanatory variables produced the best fit to the observed FFS data, with an average mean absolute error across all ten sites of 1.3%. This compares with an error rate of 6.6% with the current HCM 2010 FFS prediction model.

4.4 SHOCK WAVE DETECTION AND ANALYSIS

Shock waves are byproducts of traffic congestion and queuing. They are transition zones between two traffic states that move through a traffic environment in a manner similar to a propagating wave. Most drivers on urban freeways can identify shoke waves as a transition from a flowing, speedy state to a congested, standstill state. However, shockwaves are also present in opposite cases during which drivers idling or slowly moving in traffic are suddenly able to accelerate. Shock waves are a major safety concern for transportation agencies because the sudden change of conditions drivers experience as they pass through a shockwave can lead to rear-end crashes, including a chain reaction of multiple crashes.

However, shock waves are difficult to observe and measure strictly using point detectors such as inductive loops and side fire radars sensors. The research team has attempted to observe shock wave formation and propagation through the use of i2D high resolution vehicle trajectory data. After conducting several field-controlled (as opposed to naturalistic driving) experiments, the project team successfully observed shock wave propagation on a freeway section through a series of i2D instrumented vehicles traveling in close proximity. In this section, we provide results gathered from three consecutive data collection trips that were attempted to characterize the shock wave formation. Some of the experiments were not successful in yielding useful data.

4.4.1 Data Collection Effort 1

The purpose of this experiment is to collect i2D high resolution vehicle trajectory data for testing the accuracy of device and shock wave propagation. In addition, the collected data will also be used to understand microscopic vehicle maneuver characteristics under various traffic stream conditions.

Site Selection

For the data collection and offline data analysis, the research team developed site selection criteria. The selected route should satisfy the following criteria:

- The selected route should include various road hierarchies (freeway, ramp, primary arterial, minor arterial, and collector);
- The selected route should have at least one recurrent congested segment section on the freeway portion of the route;
- The average travel time of the selected route should be less than 20 minutes to enable the conduct of multiple runs in a single peak period;
- The route should have supplemental traffic data sources such point based detector data and link based travel time data;
- The selected route should be easily accessible from the launching point.

Route Description

Figure 44 describes the selected data collection route. The route includes two freeway facilities (I-40 and I-440), a major arterial (Gorman St. and Western Blvd.), a minor arterial (Avent Ferry Rd.),

and a collector (Varsity Dr.). The total route length is 9.1 miles and the approximate travel time in an uncongested condition is about 15 minutes. The I-40 freeway section has three lanes with posted speed limit of 60 mile/hour. The I-440 section has two lanes with a speed limit of 55 miles/hour. The selected route had eleven intersections; ten of those intersections are signalized and only one intersection (at Capability Dr. & Varsity Dr.) is unsignalized as a two-way stop control.

Link-based speed information is available for the entire route. The I-440 section has two point-based side fire radar detectors. The highest speed resolution of both databases (Traffic.com and INRIX.com) is one minute. The link database provides one minute resolution of link travel speed and time information based on the pre-defined segment. The point-based database provides one minute resolution of lane-by-lane speed, occupancy, and volume information.
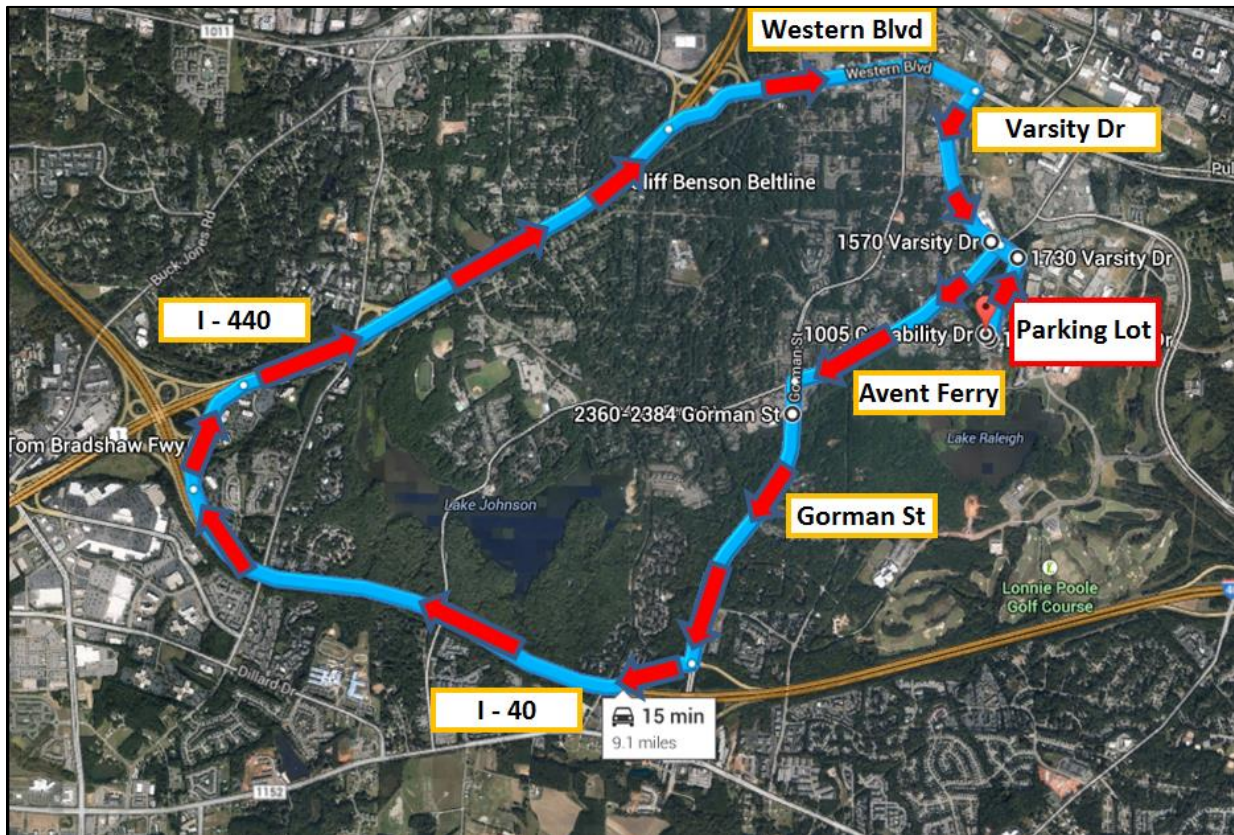
Figure 44 Data Collection Route

Figure 45 depicts the location of the two point-based sensors on I-440 (Sensor 72704), which is located near the Jones Franklin Rd. interchange and Sensor 72703 which is located upstream of the NE bound Western Blvd. off-ramp.
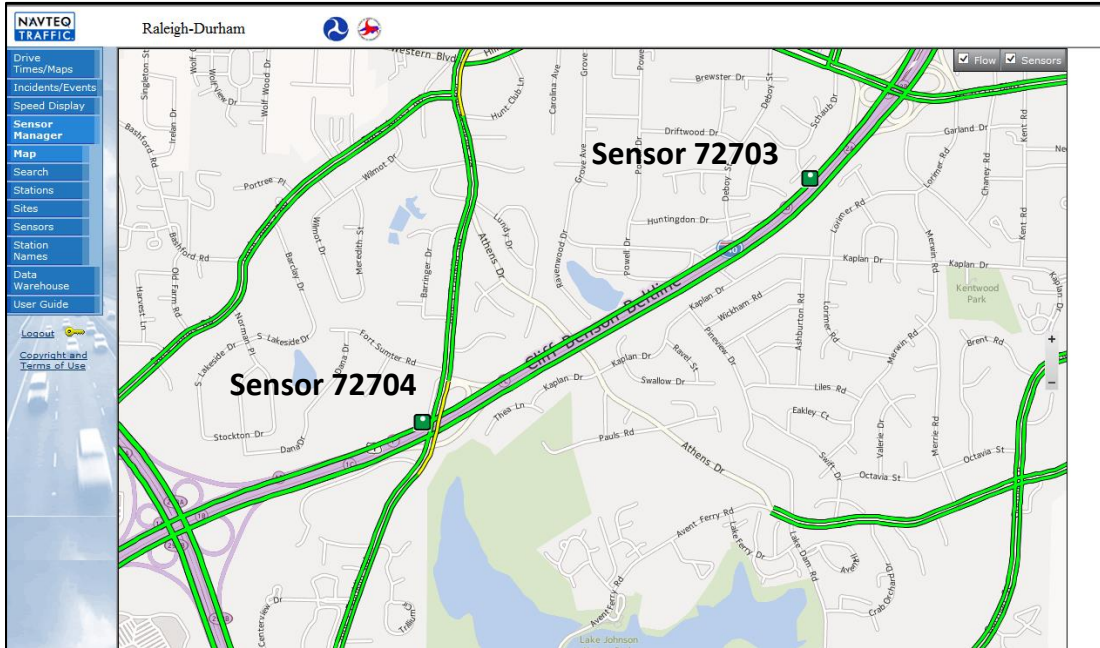
Figure 45 Traffic.com Sensor Location on I-440 selected section

To ascertain the timing and duration of congestion based on the sensor data, diurnal speed data for four individual Fridays during the Fall of 2014 (Oct. 17, 24, 31, and Nov. 7) were downloaded to analyze traffic patterns. Figure 46 and Figure 47 illustrate the travel speed for each lane. Lane 1 is to the left side lane and Lane 2 is to the right side. According to Sensor 72704 data, congestion usually started around 7:35am and lasted for about 30 minutes except on Nov. 7 when congestion lasted much longer, possibly due to an incident. Congestion was also detected for the Sensor 72703 location but lasted less than 10 minutes.
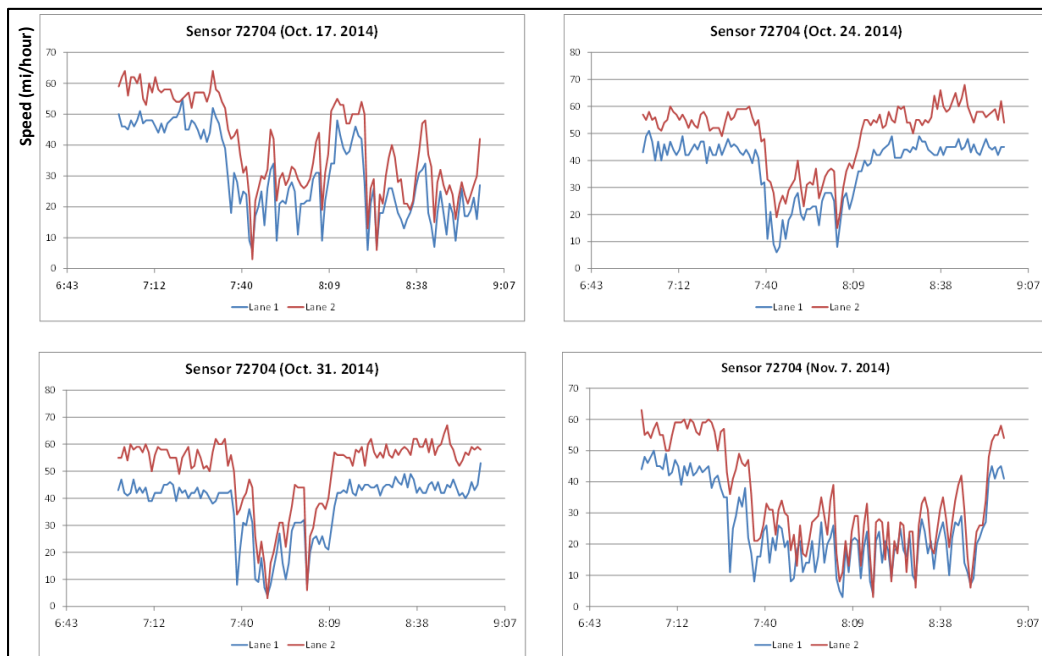
Figure 46 Speed Profiles on Sensor 72704 on four Fridays

75
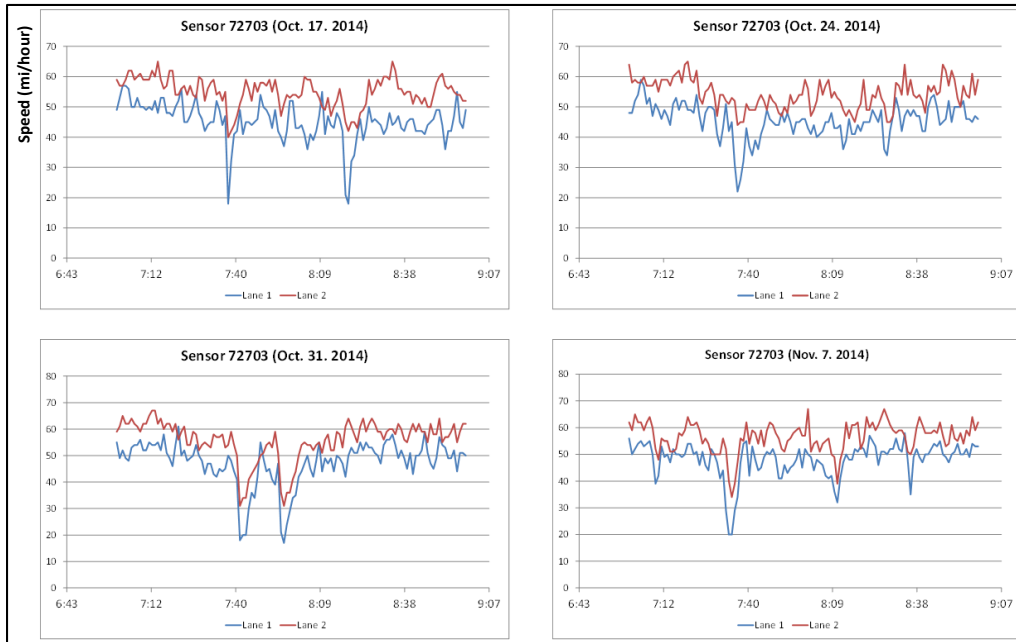
Figure 47 Speed Profile on Sensor 72703 on four Fridays

Field Data Collection

Based on previous historical traffic information, the research team conducted field data collection on December 5, 2014. Eleven participants in their own i2D-equipped vehicles carried out floating car runs on the selected route following predetermined guidance. Drivers were instructed to follow the route shown in Figure 39 starting at the parking lot location and then traveling clockwise through the route. The experiment began at 7:10am to enable the team to capture the formation of the 7:35 recurring congestion observed at sensor 72704. Vehicles were let go at 5 minute intervals to enable the coverage of a long stretch of the freeway simultaneously to capture the shape of the shock wave.

Results

A total 30 of high resolution vehicle trajectory data were collected during the morning runs on December 5, 2014. Despite every effort to schedule the experiments during the recurring congestion period, it was observed that most of the trips did not experience any congestion on the I-40 freeway section, as shown in Figure 48. The x -axis represents time of day and the y-axis cumulative distance of travel since the starts (in feet). A few trips revealed minor speed disturbances, but those disturbacnces were not sufficiently strong enough to create an observable shock wave, as see in Figure 49.

76
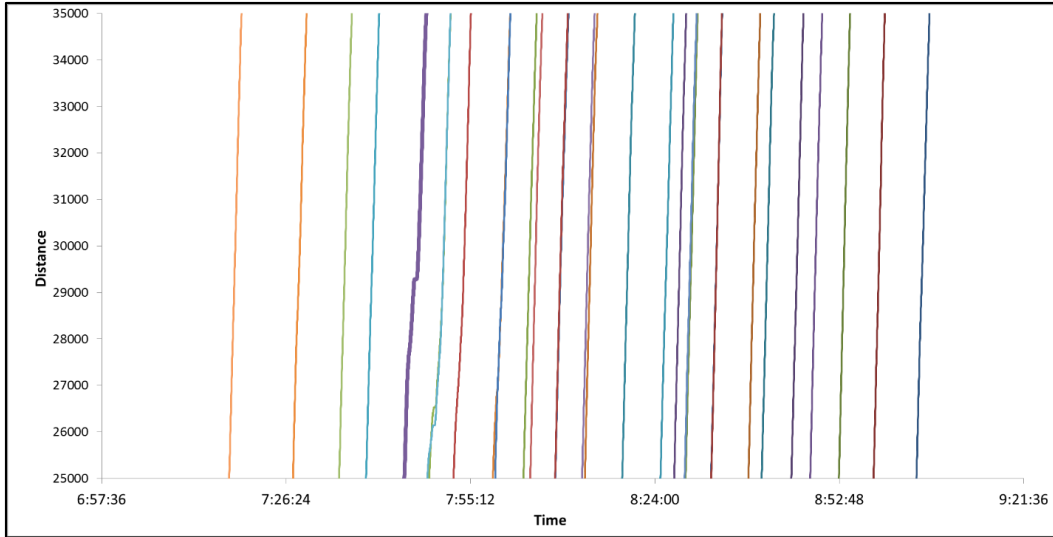
Figure 48 High Resolution Trajectories on the Freeway Section on December 5th Experiment
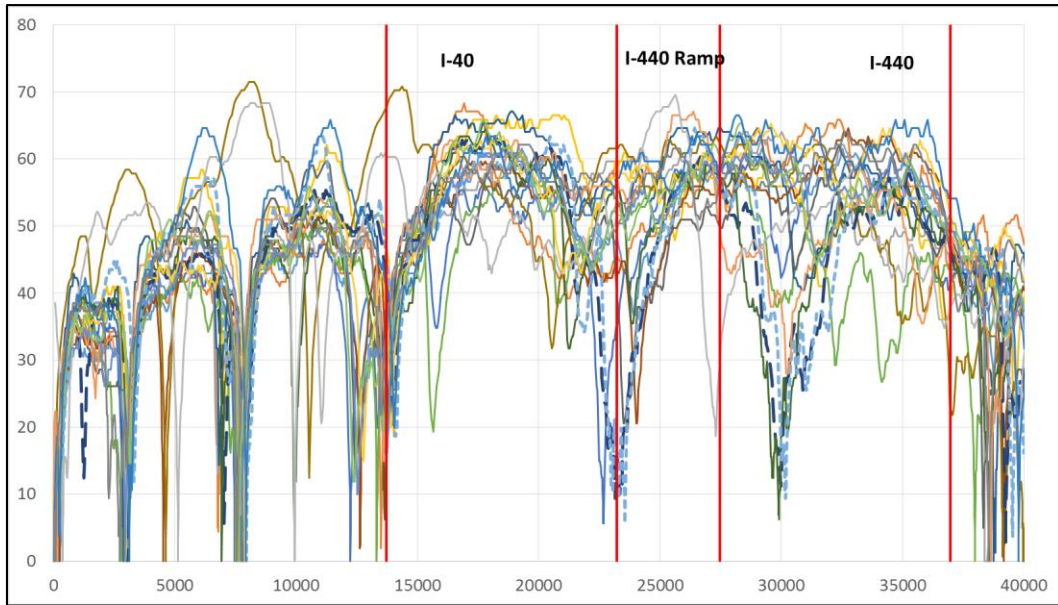
Figure 49 High Resolution Speed vs. Distance Data on December 5th Experiment

4.4.2 Data Collection Effort 2

The research team learned from the previous field experiment and attempted to solve the problem by changing the site location and the experimental method. First, the site was selected based on historical *link-based* speed data from the RITIS database, and not from point detector data.. Figure 53 shows average historical link speed contour map over 13 weekdays for nine TMC's that were considered for the second set of experiments. The data clearly indicates recurring congestion on the three middle TMC's for a one hour period between 7:55 and 8:55 am.
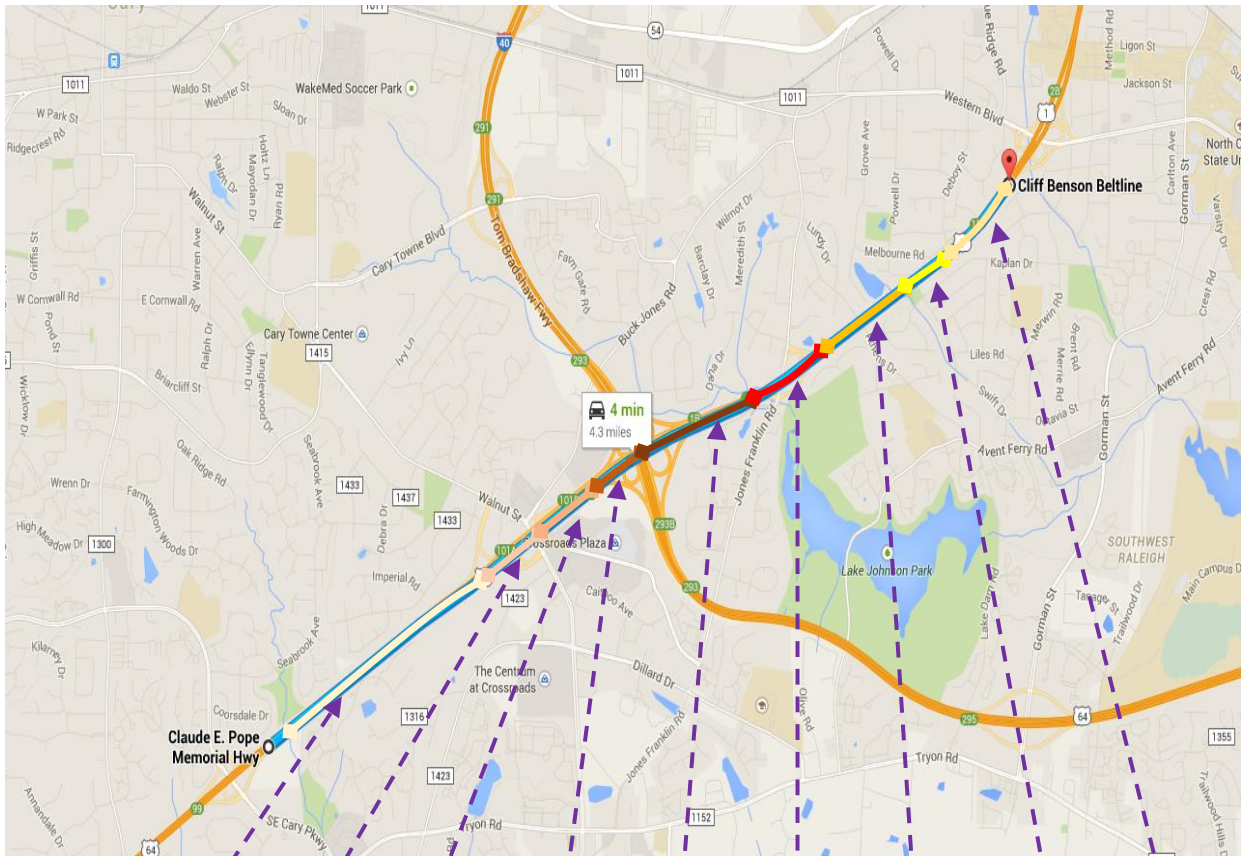
	Upstream to Downstream (SE Cary Bvld to Western Blvd)								
	13 weekdays Average								
	125+04994	125P04994	125+04995	125P04995	125-04984	125N04984	125-04983	125N04983	125-04982
07:00:00	66.43	63.67	61.25	62.32	64.20	64.83	63.85	63.67	63.38
07:05:00	65.64	63.32	61.17	62.29	62.79	63.57	63.66	64.29	63.65
07:10:00	63.98	60.05	59.02	60.99	62.86	63.97	64.17	64.39	63.72
07:15:00	63.54	58.61	57.57	58.19	62.91	63.87	63.06	63.17	62.11
07:20:00	62.74	57.72	56.53	57.88	61.58	61.74	62.09	63.14	62.33
07:25:00	60.80	53.34	52.72	54.60	59.46	59.67	61.08	62.15	61.84
07:30:00	57.69	52.36	53.15	54.46	57.88	57.97	57.71	60.00	57.83
07:35:00	52.21	48.24	50.60	51.75	54.03	54.51	54.46	55.72	53.55
07:40:00	49.39	44.94	47.37	48.68	49.41	47.16	48.50	52.52	51.45
07:45:00	45.92	41.02	46.53	45.52	43.74	41.96	44.12	49.98	48.62
07:50:00	41.60	40.69	43.90	40.25	36.02	33.57	41.34	49.36	49.77
07:55:00	40.09	38.53	40.80	32.37	29.03	29.57	41.27	49.13	50.06
08:00:00	43.19	39.49	40.10	30.14	28.43	29.47	40.71	48.72	49.09
08:05:00	44.66	40.61	41.39	30.26	28.59	29.82	41.07	51.83	51.67
08:10:00	46.79	43.66	45.31	34.57	28.55	30.34	42.11	52.19	51.17
08:15:00	49.16	44.89	45.01	36.22	27.77	30.11	42.28	50.58	51.70
08:20:00	50.77	44.95	44.29	34.86	29.00	31.13	43.24	53.56	53.70
08:25:00	50.74	45.09	43.38	34.97	30.49	32.90	42.86	51.53	52.20
08:30:00	51.31	44.64	42.14	35.13	31.60	33.22	39.52	48.46	48.63
08:35:00	52.76	46.41	43.14	35.07	33.74	33.95	38.86	47.70	49.09
08:40:00	51.84	45.43	42.95	37.70	32.75	35.67	41.83	48.22	49.75
08:45:00	53.00	45.89	43.53	38.93	34.79	37.38	43.85	50.99	50.91
08:50:00	57.01	47.48	44.07	38.39	38.60	38.42	42.69	49.60	49.84
08:55:00	60.25	50.98	47.21	40.81	37.56	37.03	42.13	49.33	48.96

Figure 50 Average Historical Link Speed Contour Map

Field Data Collection Plan

Based on the average historical link speed contour map, the research team conducted the second field data collection on May 5, 2015. Eight participants in i2D-equipped vehicles carried out floating car runs on the selected route under the following predetermined guidance. Participants were instructed to follow the route shown in Figure 51, starting at the Marla Dorrel Park parking lot, and traveling I-440 through US-1. The total route length is approximately 5.4 miles long and generally requires 7 minutes to travel in uncongested conditions. The experiment began at 7:45am, to enable the team to capture the formation of the 7:50 recurring congestion observed via the average historical link speed contour map. Vehicles departed at regular intervals and drivers tried to stay in the second lane on US-1 and on the left lane on I-440 NB just before exiting at Exit 2A onto Western Blvd.
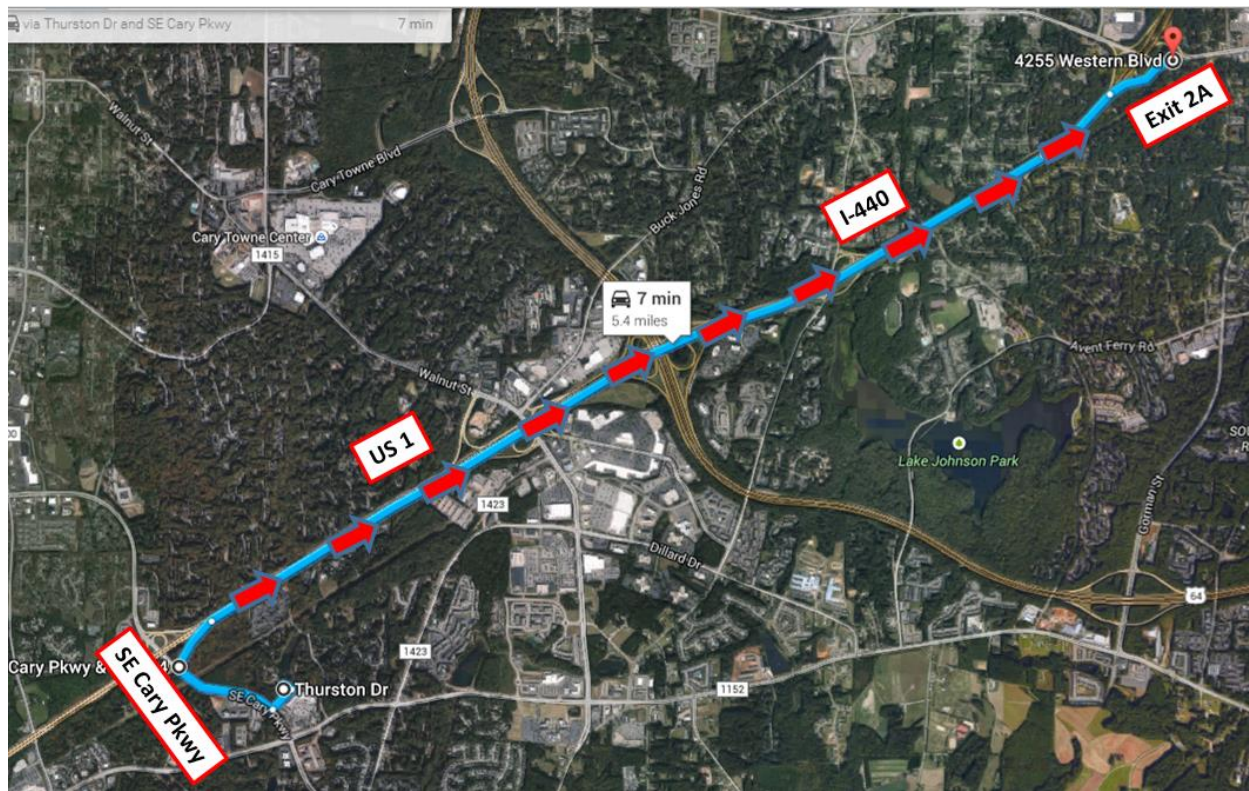
Figure 51 Data Collection Route, Attempt 2

Results

After downloading the i2D data from the second experiment, the research team successfully observed a shock wave formation using the i2D-equipped fleet. Figure 52 and 56 depict the time-space diagram and time vs. speed trajectories of five floating cars that were in the vicinity of the shockwave.
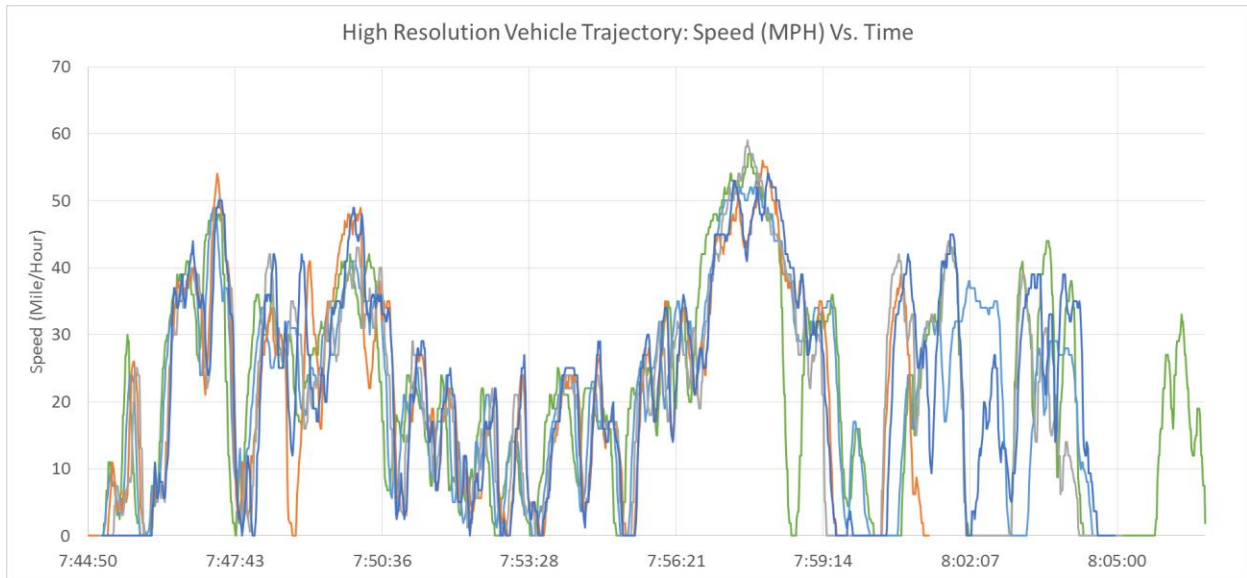
Figure 52 Five Floating Cars Trajectories on May 5ᵗʰ, 2015
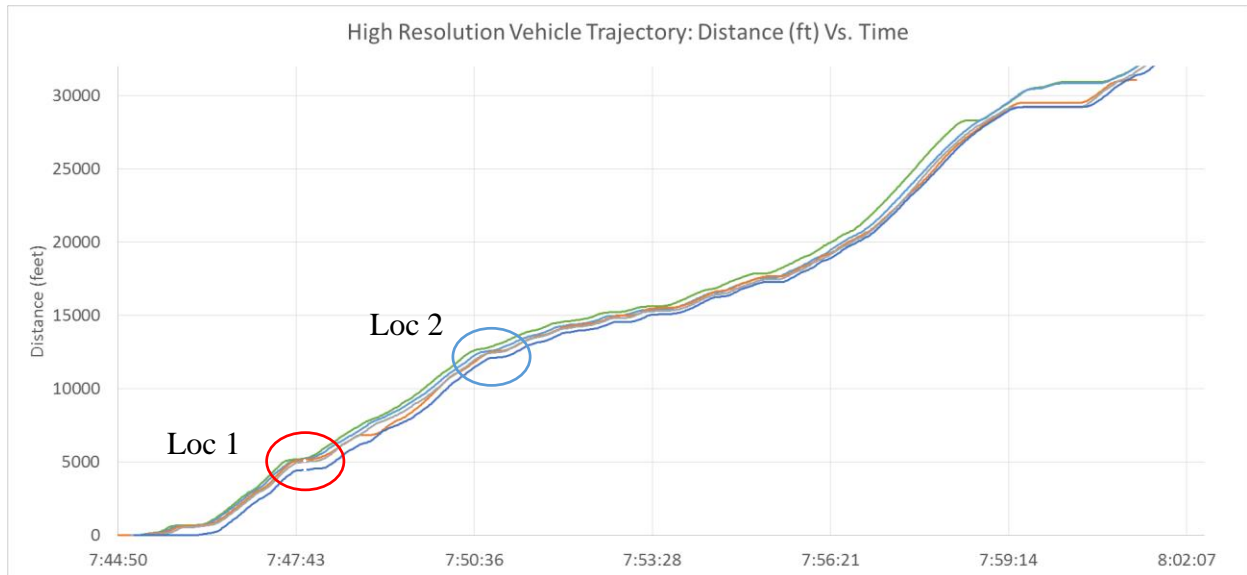
Figure 53 Five Vehicle Time-Space Diagram on May 5ᵗʰ, 2015

The first observed shock wave included five vehicles at "Loc 1". Figure 54 shows the time-space diagram for these five vehicles. The deceleration time clock of the first vehicle was at 7:47:32 and the fifth vehicle's deceleration time clock was at 7:47:52. The corresponding GPS vehicle locations were at 5041.52ft and 4455.58ft for the first and fifth car, respectively. Based on the vehicle locations and deceleration time clocks, the estimated shock wave speed was found to be 19.98 miles/hour (or 29.297 ft/sec).
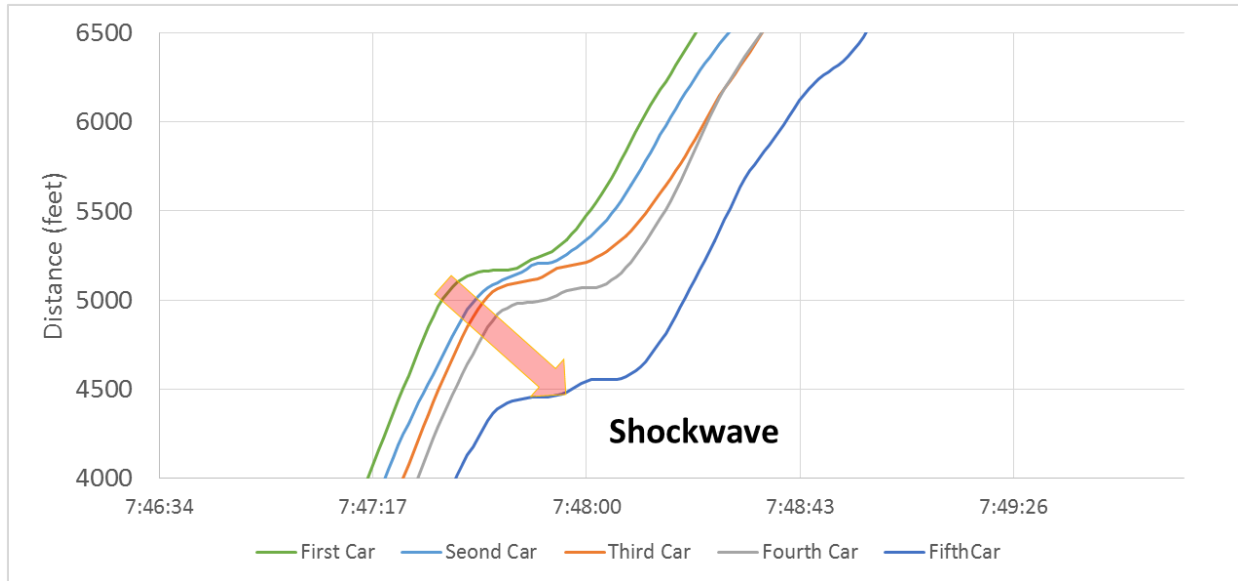
Figure 54 Loc 1 Shock Wave Speed Estimation

The Loc 2 shock wave included four floating vehicles and there appears to be at least four shock waves at that location. Drivers repeatedly accelerated and decelerated, resulting in several shock waves being created. From the first to the last shock wave, the observed sequence of shock wave speeds were clocked at 21.20 mi/hour, 37.61 mi/hour, 31.97 mi/hour, and 43.87 mi/hour.
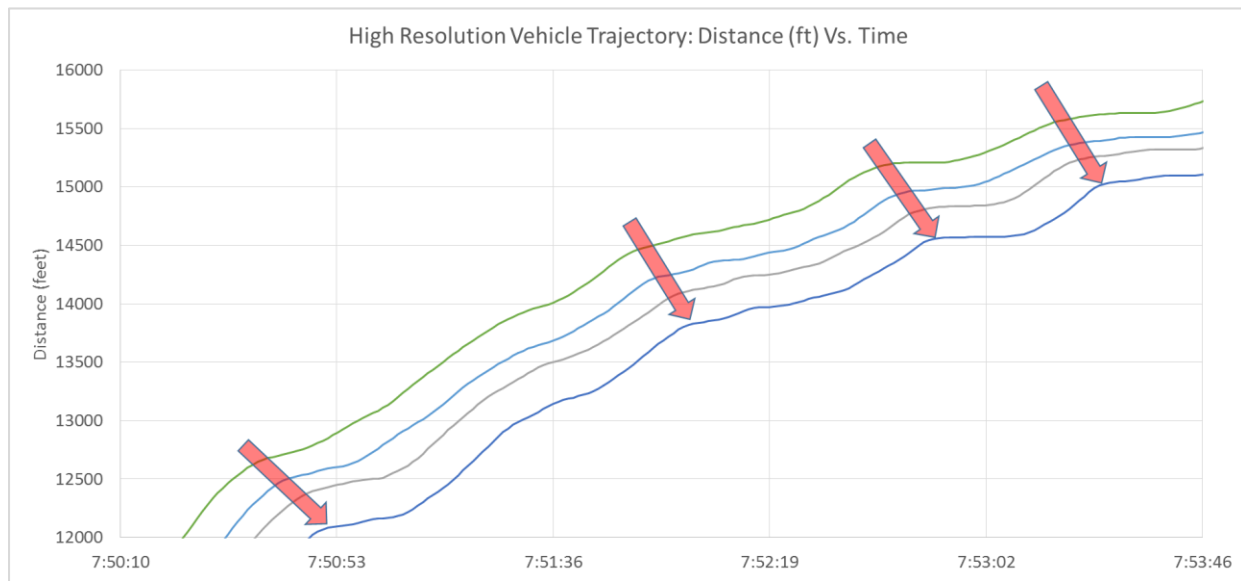
Figure 55 Loc 2 Shock Wave Speed Estimation

4.4.3 Summary and Findings

In this study, the research team generated experiments designed to to observe shock waves using the high resolution i2D data in a controlled sequence of vehicle trajectories. The first challenge was to select the proper site for observing shock waves, which on the first attempt did not produce any results due to the random nature of congestion which did not materialize on that particular day. The second challenge was to distinguish whether vehicles were on the same lane or not to

ascertain that the sequence of decelerations that were associated with the same identified shock wave. Even though the second controlled experiment instructed the participants to drive on a pre-assigned lane, drivers may have changed lanes to maintain safe headways. Finally, the resolution quality of the GPS data used in this study is not high enough to confirm the position of a vehicle on a certain lane. Therefore, the research team relied on drivers to briefly explain details about their experiences such as lane location, queue, etc., during floating car data collection.

Nevertheless, the research team believes that this series of experiments represents the first step to understand real world shock wave formation and dissipation involving equipped vehicles. Once connected vehicle capabilities, especially as they relate to vehicle to infrastructure communication technology, becomes available for a reasonable sample of vehicles the system may be capable of monitoring and even predicting real time shock wave propagation. In this case, rear-end collisions and secondary crashes may be minimized if the system can provide real time, timely warning to approaching upstream drivers, all of which will depend on the monitored shock wave speed.

5.0   TESTING VEHICLE TO INFRASTRUCTURE TO VEHICLE (VIV) CAPABILITIES

5.1   VIV FUCTIONAL SYSTEM

The VIV system described in this chapter is implemented as a new capability onto the i2D vehicle telematics infrastructure. It can be simply viewed as an implementation of connected vehicle systems within the i2D framework. The connection between vehicles is not direct, but is done through the server on the infrastructure side (or in the connected vehicle parlance a virtual Road Side Unit or RSE); and through a mobile device app at the driver interface end. The VIV system described here is not intended to be commercially available at this time, but is evaluated with the intent to identify all the functional requirements that enable such as a system to perform at a high level of reliability. Functionally the VIV system should meet certain criteria regarding its levels of reliability (i.e. low incidence of communication loss); latency (low latency between an event and an alert); and efficiency (i.e. message sent only to relevant users).

As explained in an earlier chapter, this second Channel is designed to communicate with the cloud every 6 seconds. It also enables two-way communication between the cloud and the i2D unit. A designated server on the cloud monitors real-time (with 6 second resolution) GPS coordination of the i2D vehicles and based on the defined rules transmits a message to i2D devices where can be seen by drivers cell phone app.

5.1.1   System Architecture

To reiterate, the main components of the i2D system are:
- An OBU (On-Board Unit) plugged to and powered from the vehicle OBD Port. This OBU integrates an OBD interface to collect standard OBDII/EOBD PIDs (Parameter ID), a GPS receiver, a triaxial accelerometer, a barometric altimeter, a 4GB internal memory, a GPRS module and a Bluetooth module for short range communications.;
- M2M Communications (integrated into the OBU), with automatic national and international roaming;
- A central server for Processing and Storage
- A web-based user interface
- A smartphone based application

The i2D OBU is powered from the vehicle battery and detects when the vehicle is turned on or off, enabling the use of sleep mode, under which total current used is ~10mA (the "wake-up" of the OBU can also be triggered by other sensors, even with the engine off, providing a warning in case of theft or towing).

For the purpose of implementing VIV, besides the development of a specific mobile phone app a number of changes have been made to the firmware of the on-board device and the server.

On-board device

The i2D device installed on the vehicle has limited processing capabilities and is used essentially for three functions: providing the vehicle location, triggering events based on pre-programmed rules (e.g hard braking) and serving as a relay between the server and the mobile app that interfaces with the user. For its function as data collection node in a vehicle telematics platform the communication costs are a relevant variable, and for this reason the information sent to the server is compressed in packets of 23 seconds of data (that includes the GPS data plus all other data collected from the vehicle). When operating in VIV mode the information on position is sent every second to the server.

The on-board device also relays the messages from (warnings) and to (configuration parameters) the server to the smartphone app through the Bluetooth interface.

Server

The server needs to identify the vehicles with active VIV function, their locations, headings and speed. The server also determines when the devices switch from standard and every second communication modes, the geographical metadata of each vehicle location and the rules that trigger the warnings sent to each vehicle.

Smartphone app

A dedicated smartphone app was developed for the VIV system. The app is not intended as a final, commercially viable user app but as a development tool for the VIV system. Two screenshots of the app can be seen in Figure xx and yy
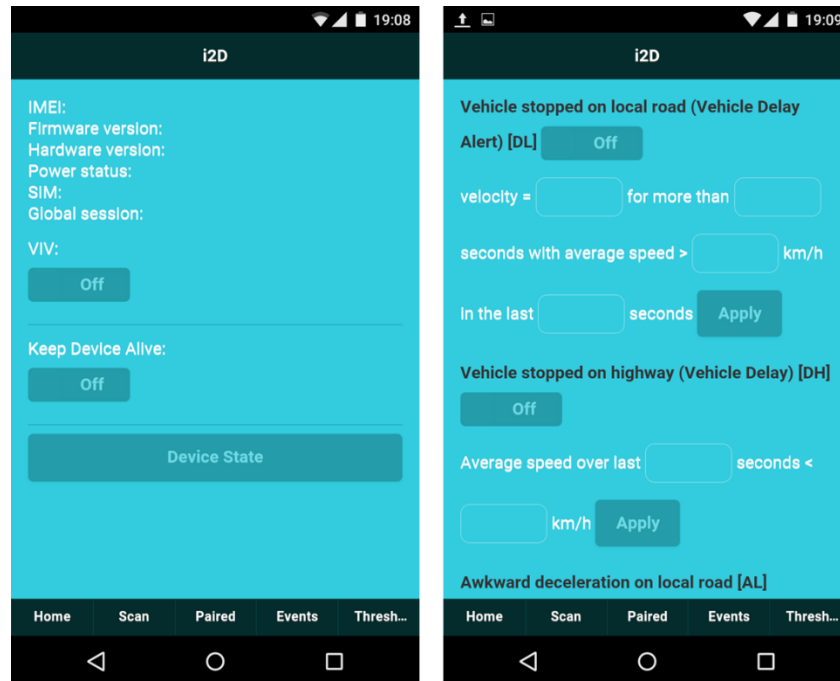
Figure 56 Two screenshots of the smartphone app

On the left is the home screen, with information on the on-board device connected (identified by the IMEI - International Mobile Station Equipment Identity), the firmware and hardware version, the power status, SIM number and global session (each session established between the on-board device and the server has a unique identifier). The next two switches are to enable or disable the VIV function of the device (this information is relayed to the server and also activates the triggers defined in the thresholds tab), and the Keep device alive switch to prevent it entering sleep mode.

On the bottom are 5 tabs:

- Home screen
- Scan: here the user can select the i2D device to connect through the Bluetooth interface
- Pair: Status of the paired device
- Events: All the messages exchanged between the on-board device and the server are displayed here and logged for further analysis
- Thresholds: here the user can define the events that will be triggered by the on-board device (the decision to send warnings to other devices is made at server level) and the thresholds for each event as described in previous chapters. Multiple triggers and events can be active simultaneously, e.g. high deceleration and a vehicle stop can be triggered at the same time.

5.1.2 Decision process

When the devices have the VIV functionality activated the server starts calculating the linear distances between the vehicles from the GPS position. When that distance falls below a certain value (e.g. 2 km) the server sends a message to the device to start sending the position every second. The GPS information is now used by the server to obtain two parameters:

- The name of the road where the vehicles are circulating and of the adjacent roads. This is extracted from the metadata of a map of the area where the vehicle is located (in this case "Openstreetmaps" was used). This is an important step in order to make sure that the vehicles are circulating in the same road, particularly in urban areas.
- The heading of the vehicles to check if they are on converging or diverging courses. Of course, only relevant vehicles that are converging on the alerting vehicle course should be receiving the message.

Once the server determines the relative position of the vehicles it is now in position to start checking every second if any of the VIV conditions are verified or not and sends messages to each on-board device in accordance to the rules that have been defined (e.g., deliver a warning when the distance to a stopped car on the same road and heading in the same direction falls below 300 meters)

5.1.3 Communications

Device and Server

The communications between the device and the server are all carried through the mobile network using a M2M connection. There are two bi-directional communication channels established: one that sends the complete information set from the vehicle to the server at low frequency and a priority channel where information is sent back and forth with no delays. This second channel is used for the VIV implementation.

Device and App

The communications between the device and the smartphone app are done through a Bluetooth connection, the app does not communicate directly with the server. The smartphone must be paired with the on-board device for the connection to be established, and once that is done the connection will be active until disconnected from the smartphone or the on-board device powers down.

5.1.4 Limitations and areas of improvement

The i2D platform on which the VIV system was implemented was designed as a vehicle telematics platform, and as such there are some limitations regarding its use as a VIV platform, mainly regarding network availability and latency.

Mobile network coverage

The i2D VIV platform does not rely on local infrastructure, as all communications between the on-board device and the server are performed through the mobile network. As such, whenever there are shadows in the mobile network coverage the VIV system is rendered inoperative. The on-board device will still trigger events but they are not sent to the server until the connection is re-established.

Latency

Latency is a critical restriction for a VIV implementation. On systems that rely on local communication infrastructure latency is very low. In the present implementation, with the devices running in the US and the server in Portugal, latency is usually under 2-3 seconds. This value is increased in situations of degraded mobile network connection.

Multiple requests

Due to the way the system was implemented, there is a considerable burden on the server, at each second it needs to match all the VIV active devices, including extracting the geographical metadata from the map and calculating all the relative positions. The tests were performed with a handful of vehicles and as such no limitations where observed, with more vehicles active in the same region this could not be the case and may require refining the algorithm used.

Map quality

As mentioned before the server uses the street name as part of the process. If this information is not present on the metadata, then only heading and speed will be available, which may result in false positives or negatives, especially in the case of roads with multiple changes in direction or in dense urban areas.

5.2 INITIAL FIELD TEST

The purpose of these trials is to test, under near real-time conditions, the VIV capabilities of the i2D platform, namely:

- Identification of events based on micro vehicle dynamics
- Selection of message recipients based on distance, road and direction
- Selection of type of message to be sent to each targeted vehicle
- Low latency from event to warning of other users
- Persistence of events (configurable)

In order to safely test the VIV possibilities, a minimum of two vehicles / conditions are required (optimally three) and a suitable road with no curves, good network coverage and a length of approximately 1 mile, where one of the vehicles can safely stop.

The vehicles are equipped with specific i2D devices for this test, using the i2D Development Environment. The firmware installed in this device allows them to communicate with a smartphone (Android only at the moment), where a specific installed app display messages to the vehicle occupants (sound, images and text).

5.2.1 Test Procedure

First the testers must assure that all devices are connected (i2D-vehicle, i2D-network, i2D-smartphone). It is advisable that both drivers be able to communicate via mobile phone if needed. When using the i2D VIV platform the first parameter is warning distance, which at the moment is set at half a mile (~800 meters). Whenever a vehicle enters in the "Queued state" (speed < 5 mph) any vehicle within the defined distance that is travelling in the same road and *heading towards* the queued vehicle, on the same direction, receives a warning. The queued event has a lifetime (configurable) of two minutes and during that period any vehicle that enters within the specified range, on the same road and on the same direction, will receive a warning. While the warning is valid, it is displayed on the smartphone connected to the i2D device. The suggested procedure is to select three locations on the road that are easily recognizable:
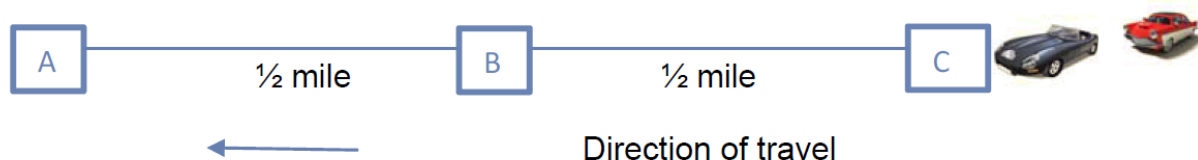
Figure 57 Concept of VIV Initial Test

87

- **Location A:** location where the lead vehicle 1 stops or slows down to less than 5 mph.
- **Location B:** a location that is at the warning threshold from location A (at the moment it is set at half a mile but this can be changed in order to better suit the road selected)
- **Location C:** a location from where both vehicles 1 and 2 depart. This should be at least a mile away from location A

Assuming a constant speed of 35 mph, vehicle 1 departs from location C towards A. Vehicle 2 can depart with a delay between 1 and 2 minutes. When vehicle 1 stops at location A, two events can happen:

- If vehicle 2 had already passed location B, it will receive a warning within 2-6 seconds (see limitations for factors that affect delay) after vehicle 1 had stopped
- If vehicle 2 had not passed location B when vehicle 1 stops, it should receive a warning 2-6 seconds immediately after it passes location B

Using the same two vehicles, it is possible to confirm that a third vehicle travelling in the *opposite direction* will not receive a warning assuming any of the two initial vehicles at any point stops, and having the third vehicle past that location in the opposite direction within 2 minutes. With a third vehicle available it will also be possible to test the progression of the back of the queue by having vehicle 1 stop at say location A, vehicle 2 stop halfway between A and B and then vehicle 3 will receive a warning a similar distance (800 m) before arriving to location B.

However, the results obtained may be influenced by several factors:

- GPS coverage and signal strength may limit the accuracy of the location
- Communication delays resulting from poor mobile network coverage add latency to the system and delay the warning messages
- While the system identifies the road where the vehicle is, at the moment the distances are not calculated according to the road profile and better results are obtained with straight roads
- The only event being transmitted at the moment is the queued state of the vehicle (queued or not queued) but other events are configurable based on vehicle dynamics

5.2.2 Selected Sites and Test Results

Initially, two sites (one arterial and one freeway) were selected to test the VIV basic communication function. Figure 58 shows the first test site on Avent Ferry Rd. between Varsity Dr. and Gorman St. The purpose of this test was to verify the vehicle stop functionality on the street warning. When the first vehicle stopped on the assigned complex parking lot near by street, the following second car did receive the stop warning message through the mobile App. Through the test, the research team confirmed the i2D unit could communicate with the server in near real time (2-6 seconds) and that the server could detect specific events and provide information to the other i2D users.

Figure 58 VIV Initial Test on Arterial

However, the travel speed in the first test site was less than 40 mi/h since the test site was on a signalized arterial so the team selected one freeway site to carry out the test under high speed travel. The selected freeway section is in the West bound direction of Interstate 40 in Raleigh, NC between mile post 297 and mile post 295 (between Lake Wheeler on ramp and Gorman off ramp).

The experiment was conducted successfully but the team also detected that a VIV equipped vehicle traveling in the opposite direction (Eastbound) also received a vehicle stop warning alert, indicating that the system could not properly distinguish the movement direction yet. It thus appears that the server broadcasted the vehicle stop event to all i2D user who were within 0.5 mile of the detected event, regardless of direction. This finding also implies that any vehicle, no matter which road they are, will receive a warning from the server if the vehicles are in 0.5 mile physical distance from event triggered vehicle. The research team used map matching technology with "openstreetmap" to solve stated issues. Openstreetmap has road layer which contains the road name. Map matching technology allow identifying events triggered by the vehicle's location on the specific road and it is possible to not broadcast the event alert to all vehicle inside of boundary but send event alert to only targeted vehicles.

Figure 59 VIV Initial Test on I-40 Freeway

5.3 THRESHOLD DESIGN FOR VIV TEST

The specified VIV criteria of reliability, latency and efficiency are very much dependent on the level of criteria threshold that are set to trigger an event. Just like the situation encountered in automated incident detection, there are two inherent risks in setting those thresholds. First if the threshold values are too restrictive, in other words alerts are reserved for the most dangerous situation, the investment in VIV would be very inefficient as the system would have very little impact or use. On the other hand, a system that generates too many alerts will also be undesirable as it (a) will overload the communication system and thus increase latency; and (b) it will render the system less reliable, and (c) it may result in many alerts being ignored as it increases the driver information load.

This section therefore is intended to test the impact of selecting specific variables and specific threshold values of those variables on the intensity of the messaging system. For obvious reasons we separate alerts on high speed travel and those on surface streets. We also limit the number of variables that will be used so that latency does not become an issue.

Table 21 shows six events code with designated event triggers. Three of those events are related to local roads and the other three are freeway related events. When any i2D user with VIV enabled experiences one or more of the pre-designed events--- for example if a vehicle is traveling on a arterial with the average speed that exceeds 50 km/h over the last one minute, but then suddenly stops on the arterial, it will trigger a DL (delay on Local Road) event and any *following* vehicle within the pre-specified 1,000ft threshold will receive DL event alert. The full set of variables and thresholds were coded in the i2D VIV App, which was shown in Figure 56.

90

The App communicates with the i2D unit through Bluetooth and detects specific events using the pre-defined threshold via the high resolution vehicle travel information. The detected event will be instantaneously sent to the server and the server will sent an event alert to other i2D units which are within the "Distance threshold". When i2D units receive event alert, it will be displayed on the i2D App on the mobile phone.

Table 21 Pre-Specified VIV Events and Thresholds

Event Code	Event	Threshold	Data sent to the server	Distance Threshold for Alert Dispatch
DL	Vehicle stopped on local road (Vehicle Delay Alert)	Velocity = 0 for more than 30 seconds (Delay > 30 seconds) along with an average speed > 50 km/h over the last 60 seconds	Alert	1,000ft (300 m)
DH	Vehicle stopped on highway (Vehicle Delay)	Average speed over last 60 seconds < 50km/h	Alert	2,500ft (750 m)
AL	Awkward deceleration on local road	Deceleration ≤ -0.4g, with average velocity > 50 km/h in the last 60 seconds	Alert + X+Y+Z Acceleration values in the previous 20 seconds and the next 10 seconds.	1,000ft (Note: alert given 10 seconds after the event ends)
AH	Awkward deceleration on highway	Deceleration ≤ -0.4g, with average velocity > 100 km/h in the last 60 seconds	Alert + X+Y+Z Acceleration values in the previous 20 seconds and the next 10 seconds.	2,500ft (Note: alert given 10 seconds after the event ends)
CL	Possible Crash on local road	Deceleration ≤ -0.4g and speed=0, with average velocity > 50 km/h in the last 60 seconds	Alert	1,000ft
CH	Possible Crash on highway	Deceleration ≤ -0.4g and speed=0, with average velocity > 80 km/h in the last 60 seconds	Alert	2,500ft

Whenever events are triggered, the mobile App automatically generates a series of log files to allow tracking and for verifying the quality of the VIV system performance. An example of a created log file was tested for quality using the i2D Trajectory Generator Tool which was explained in detail in section 3.2.

Figure 60 shows a sample of archived event logs on the mobile App. The panel to the left shows a "TRANSMITTED" event which means the corresponding i2D unit triggered a specific event (DL) at 19:40:04 at the GPS position (latitude: 35.767395, longitude: -78.676963). The panel to the right shows a "RECEIVED" event from the server. The received event was also the DL and while the event hour was at 19:53:57, the unit received this event at 19:56:41. The distance from event location to current vehicle was 12 ms, which is directly calculated by latitude and longitude.
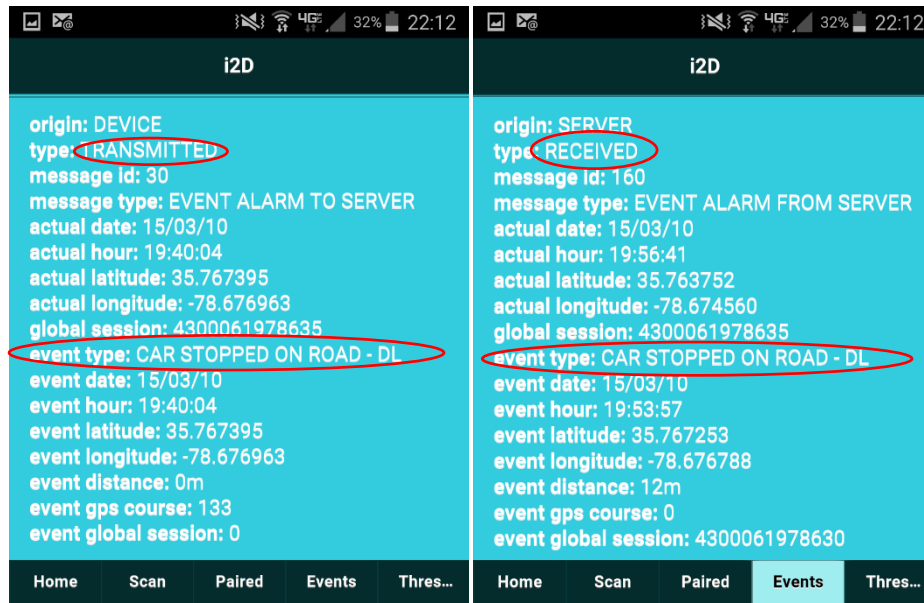
Figure 60 Event Logs in Mobile App

The i2D mobile app for VIV communication shows 17 rows of information. Table 22 shows the detail information of i2D VIV mobile App information.

Table 22 i2D VIV mobile App Data Description

Information	Description
Origin	Where the message comes from (i2D device or Sever)
Type	Message is "Transmitted" or "Received" (Transmitted: the i2D unit trigger a specific event, Received: someone triggered event and i2D unit received event alert from server)
Message ID	Unique ID for event message
Message Type	"Event Alarm to Server" or "Event Alarm from Server"
Actual Date	Date stamp for the event message
Actual Hour	Time stamp for the event message
Actual Latitude	Latitudinal location for i2D vehicle
Actual Longitude	Longitudinal location for i2D vehicle
Global session	Randomly assigned i2D device unique ID (It will be changed every time when the car engine is restarted)
Event Type	Explain event type (Currently six event types are codded)
Event Date	Date stamp for the event
Event Hour	Time stamp for the event
Event Latitude	Latitudinal location for event triggered i2D vehicle
Event Longitude	Longitudinal location for event triggered i2D vehicle
Event Distance	The distance between event triggered vehicle and received vehicle
Event GPS Course	Event triggered vehicle's GPS course
Event Global Session	Randomly assigned i2D device unique ID for event triggered vehicle

5.4 VIV TEST RESULTS

The final VIV test was conducted on the NC State University Centennial Campus. Three vehicles were used for testing and verifying the VIV warning system. The test site was selected under consideration of safety issue and in consideration of the wireless communication quality (mobile network) to minimize latency. The selected route is a 0.9 mile long section on Main Campus Dr. with very low volumes. Figure 61 shows the test site.
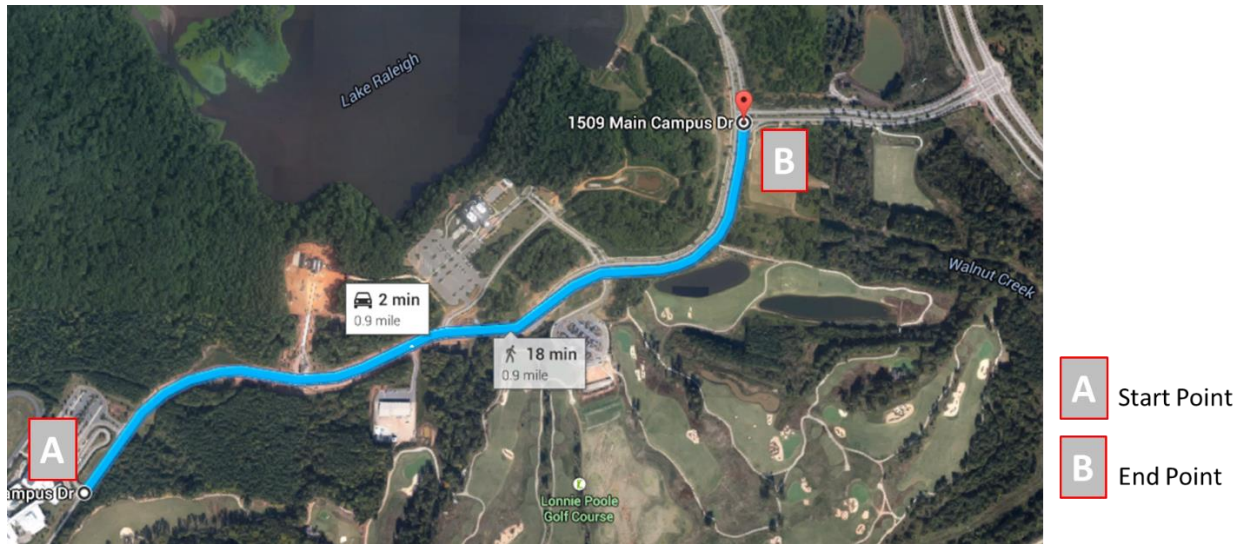
Figure 61 VIV Test Site on Centennial Campus

The local road related event thresholds were slightly modified for testing purpose since the site is only 0.9 mile in length. The idea here was to reduce the values of the thresholds so that we can observed various alerts that include not only DL (Delay on Local Road) events but also AL (Awkward Deceleration on Local Road) and CL (Potential Crash in Local Road) events. Table 23 shows the modified test thresholds for this Local Road Experiment.

Table 23 Modified Local Road Event

Event Code	Event	Threshold	Data sent to the server	Distance Threshold for Alert Dispatch
DL	Vehicle stopped on local road (Vehicle Delay Alert)	Velocity = 0 for more than 3 seconds (Delay > 3 seconds) along with an average speed > 25 km/h over the last 20 seconds	Alert	1 km
AL	Awkward deceleration on local road	Deceleration ≤ -0.2g, with average velocity > 25 km/h in the last 20 seconds	Alert + X+Y+Z Acceleration values in the previous 20 seconds and the next 10 seconds.	1 km
CL	Possible Crash on local road	Deceleration ≤ -0.2g and speed=0, with average velocity > 25 km/h in the last 20 seconds	Alert	1 km

The test procedure worked as follows:

- First car leaves point "A" and drive naturally and stop point "B" with high deceleration
- The Second car leave point "A" 20 second later and stops behind the first car
- The Third car leaves point "A" 40 second later and stops behind the second car

The *expected sequence* of alerts based on this experiment is as follows:

- First car trigger AL event first, followed by a CL and DL events.
- Second and third car will receive event alert from first car in the same order
- Second car also creates event in the same order as the first car
- Third car receives the event alerts from the second car

The three vehicle trajectories involved in this experiment were tracked. Figure 62 shows time vs. speed trajectory of three test cars on Main Campus Dr. The gray line depicts the trajectory for the first vehicle, and the blue and orange line trajectories for the second and third car vehicles respectively. The green dash line is the first vehicle's acceleration profile. When the first car arrived at point B the driver stopped with high deceleration value (-0.22). Therefore, the second and third car should have received an AL event first.
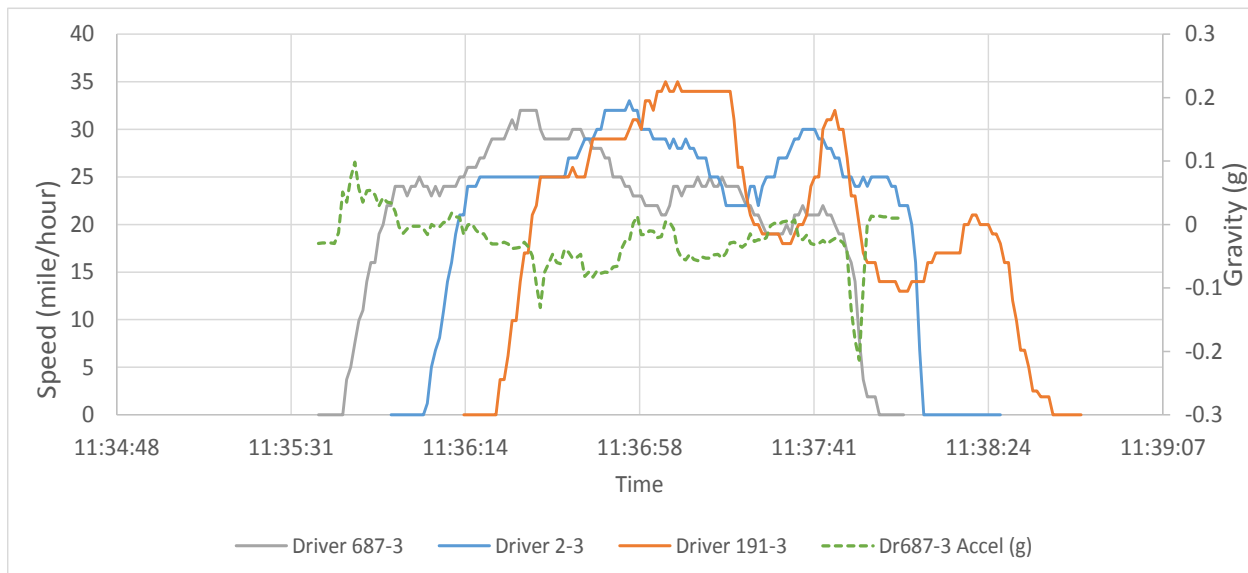
Figure 62 Time vs. Speed Trajectory for Three Test Cars

Figure 63 also shows clock time vs. travel distance trajectory of three test cars. The "0" travel distance point corresponds to start point "A" and point "B" is approximately 4390ft away from point "A" on Figure 61. However, it is hard to recognize where the specific events were triggered by time vs. travel distance plot.
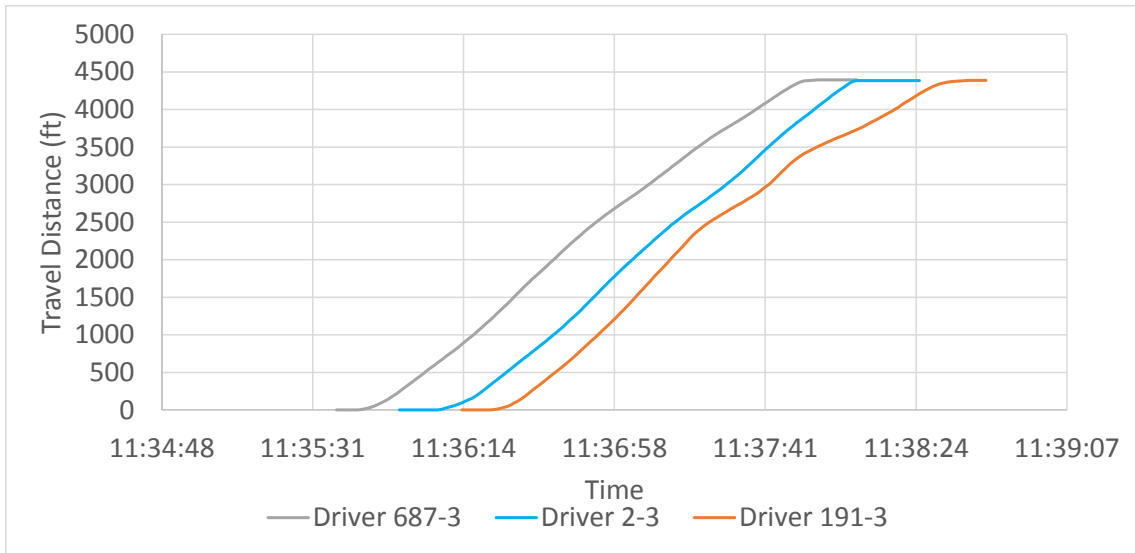
Figure 63 Time vs. Travel Distance Trajectory for Three Test Cars

The test results are shown in Figure 64 which is extracted from the 2D VIV website. The i2D VIV website provides queuing event information as well as specific events history over time. The left side shows the historical queue event lists and middle area on the web describes geographical location and specific message which i2D unit generates and receives. The right side displays detail information of events which are basic safety message (BSM). The message basically consists of three types: 1) delay, awkward deceleration (hard deceleration), crash for both highway and local road. Each BSM has timestamp, event triggered vehicle secure ID, event received vehicle secure ID and event type.
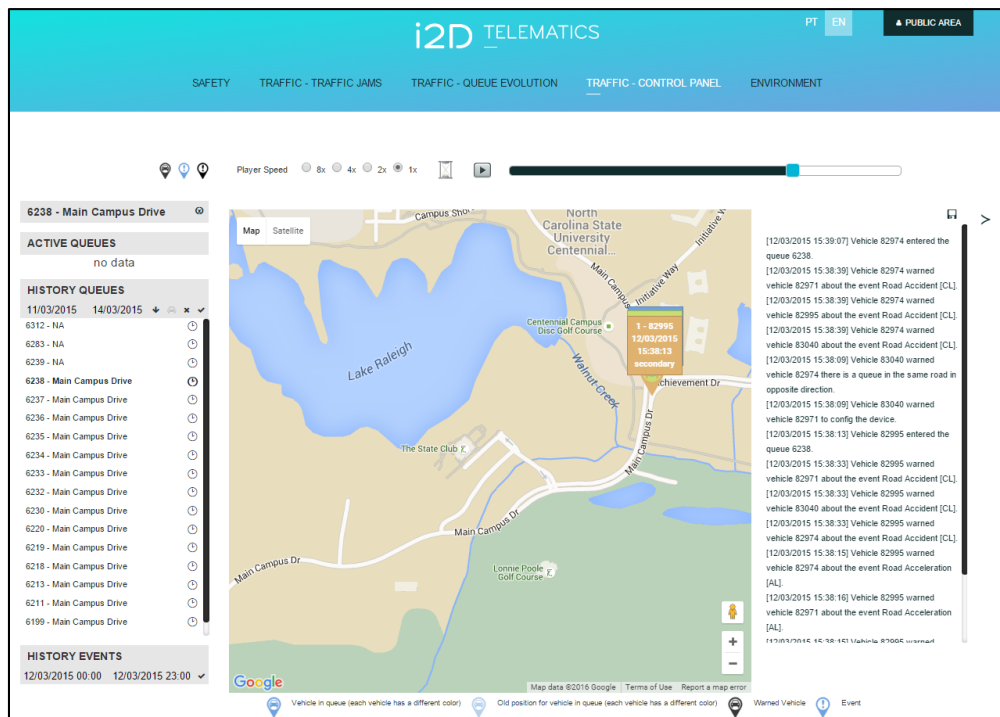
Figure 64 i2D VIV Log on Website (http://www.i2d.co/i2dpubportal/Login.xvw)

6238 - Main Campus Drive

SESSION	ORDER	LATITUDE	LONGITUDE	COURSE	TIME
83040	1	35.764133	-78.674294	6°	12/03/2015 15:38:00
82995	1	35.763939	-78.674316	7°	12/03/2015 15:38:13
83040	2	35.764149	-78.674294	6°	12/03/2015 15:38:31
82974	2	35.764065	-78.674309	7°	12/03/2015 15:39:07
82995	3	35.764153	-78.674294	4°	12/03/2015 15:39:07
83040	-	-	-	6°	12/03/2015 15:39:06
82995	-	-	-	4°	12/03/2015 15:39:07
82974	-	-	-	7°	12/03/2015 15:39:07

When there is no order information, the vehicle exited the queue

Figure 65 Queue Event on i2D VIV Website

Figure 65, also extracted from the VIV website, gives the session ID, vehicle order, location and time stamp for queue events. VIV website does not use the actual vehicle ID or driver ID but generates randomly re-assigned unique session ID for event related vehicles in order to protect the individual drive privacy. According to Figure 65 information, the re-assigned vehicle ID is 83040, 82995 and 83040 for first, second and third car, respectively. The first car creates a queue event at 15:38:00 and the second car joined the queue at 15:38:31. Finally, the third car joined at 15:39:07. Those result are close to three vehicles' trajectory shown in Figure 62. Vehicles are then departing from the queue at 15:39:06 and 15:39:07, respectively.

Figure 66 shows the generated VIV event logs. The first vehicle triggered a CL event and the following two vehicle received the CL event alert at 15:38:08. After 1 second, the first vehicle again sent a DL event to the other two vehicles. However, the very first expected AL event was not triggered and none of following vehicle received an AL event alert.

Interestingly, the second vehicle actually did trigger both AL and CL events. As expected, the third car received both event alerts; in addition the first car also received the second car's event warning alert.

In summary, the research team has found evidence of both false as missed alerts, in addition to latency issues. However, the team is still confident about the possibility of real time event warning system using VIV system. This new technology hold good promise for both mobility and safety application. Especially, the suggested technology have clear potential to avoid secondary crash which generally happens due to inclement weather conditions such as heavy rain, snow or fog.

[15:38:00 2015-03-12] Vehicle 83040 entered the queue 6238
[15:38:08 2015-03-12] Vehicle 83040 warned vehicle 82995 about the event Road Accident [CL]
[15:38:08 2015-03-12] Vehicle 83040 warned vehicle 82974 about the event Road Accident [CL]
[15:38:09 2015-03-12] Vehicle 83040 warned vehicle 82995 there is a queue in the same road in same direction
[15:38:09 2015-03-12] Vehicle 83040 warned vehicle 82974 there is a queue in the same road in opposite direction.
[15:38:13 2015-03-12] Vehicle 82995 entered the queue 6238
[15:38:15 2015-03-12] Vehicle 82995 warned vehicle 83040 about the event Road Acceleration [AI].
[15:38:15 2015-03-12] Vehicle 82995 warned vehicle 82974 about the event Road Acceleration [AI]
[15:38:33 2015-03-12] Vehicle 82995 warned vehicle 82974 about the event Road Accident [CL].
[15:38:33 2015-03-12] Vehicle 82995 warned vehicle 83040 about the event Road Accident [CL]

Figure 66 VIV Event Log

6.0 SUMMARY, FINDINGS AND RECOMMENDATIONS

Summary
This research investigated the feasibility of using high resolution data gathered by an in-vehicle on-board unit (OBU) for a variety of traffic flow and safety applications. In particular, this research examined data collected from the vehicle OBD (II) port that was subsequently disseminated to a web-based cloud for real-time analysis. A secondary objective of the study was to test the reliability of a vehicle to infrastructure to vehicle (VIV) system, which enables personalized messages to be delivered to vehicles within the influence area of an "alerting" vehicle to enhance safety and mobility. The underlying vehicle data for carrying out the objectives of the study was gathered from a fleet of 24 roaming vehicles in a naturalistic driving setting primarily within the Research Triangle region of North Carolina, U.S. Within the test period April 2014 to December 2015, the fleet generated over 38,800 one-way trips, covering nearly 9,000 hours of driving in which over 275,000 miles were driven. At the 1Hz level of resolution, the data yielded over 32 million seconds of driver activity, making it exceedingly useful for various types of travel analyses. Parameters recorded by the OBU and transmitted to the cloud via a mobile cellular network include instantaneous speed, GPS positioning, accelerations and decelerations in the three principal dimensions, road elevation, and several OBD internal parameters that can be used for estimating fuel use and emissions. All data was archived in a SQL database environment to enable quick retrieval and analysis. During the coruse of this study, several software tools were developed to enable researchers to configure the large amount of data for analysis. The VIV experiments were focused on testing the ability of the system to generate, transmit, and receive messages that were accurate and timely for both the alerting and alerted vehicles.

Findings
This section covers findings related to the research team's experience working with the voluminous dataset generated by the system, as well as those that are specific for the four safety and mobility applications investigated in this study. These include:

- The installation and calibration of the OBUs on a variety of vehicles was, in general, quite successful. There were only in a handful of times that certain vehicle models were not able to integrate with the OBU to generate the requisite OBD data.

- Attempts at visualizing the high resolution data in a dynamic fashion (time sequenced) in a GIS environment were not very successful and required significant processing time, highlighting a need to improve such environments to enable efficient analysis and visualization at a regional network level.

- Attempts at downloading the archived data from the central server website dedicated to our vehicle fleet were also challenging, especially as the size of the data escalated over time. At the conclusion of this study, alternate data exchange environments were explored, which significantly expedited data transfer.

- The first mobility application to use the OBU data yielded an improved free flow speed (FFS) model on freeways. For the first time, it also enabled the testing of level of familiarity on indicdually measured FFS. We found that driver familiarity was positively corrletaed with FFS under certain conditions. The research team also found that the model's explanatory power improved significantly when the value of the posted speed limit was added to the pool of variables.

- Our second mobility application explored speed distribution and definitions on signalized arterials. We confirmed empirically the theory that density-weighted speed yielded space mean speed while trip-weighted speed yielded time mean speed. In this data environment, density was substituted with *the number of GPS observations* made on a segment over a finite time interval. The team also found that the weighting scheme yielded very different values near the signalized intersection stopline, and that the density-based weighting portrayed a more accurate picture of the delay situation on segments nearest the signal.
- Our third mobility application explored the feasibility of measuring shock waves on freeway segments, using a strategically deployed platoon of vehicles equipped with OBUs. While it took a couple of trials to succeed (either because congestion never occurred on the day of the test, or because vehicle separation was not optimal), we were able to identify four different shock waves, with wave speeds varying from 20 mph to 40 mph. This was done using less than five vehicles in the experiment.
- Our lone safety application focused on identifying freeway hotspots (i.e. potential high crash locations) based on the relative incidence of high levels of longitudinal decelerations (above 4m/s/s), as measured from the OBU accelerometer. We then contrasted those locations with the actual occurences of rear-end collisions on the selected freeway facility, based on several years of observations. We found strong association – *but not necessarily causality* – between the two. This finding shows that with larger samples of high resolution data, our ability to predict hotspots before crashes actually materialize may be significantly enhanced.
- Finally our VIV experiments showed initial evidence of both false as well as missed alerts (e.g. the received message was "delay on the freeway" while the transmitted message was of a "strong deceleration"), in addition to observing several latency issues between the transmitted and received messages.

Recommendations

In conclusion, the in-vehicle technology tested in this study appears to hold good promise for generating several mobility and safety applications. In the future, the massive set of data collected from the vehicle sample will significantly reduce the need to build expensive, fixed monitoring infrastructure, as the network performance observability will be ubiquitous. Natural extensions of this study should explore research questions related to the the following research:

- The market penetration of OBU equipped vehicles needed to provide statistically meaningful conclusions for specific application needs (e.g queue detection; incident detection and transmission, crash prediction, etc.).
- The technological improvements are required to bring the reliability of the VIV system to the real-world testing realm.
- The extent to which individual driving high resolution data can be used to classify driving styles and its impact on fuel use and emissions.
- Testing at the macro scale actual route choice from origins to destinations and explain based on the experienced route attributes, which could address the perennial question of whether drivers actually use the shortest path.
- Integration of contextual with high resolution data, including GIS layers of network links, TMC-based link speeds, weather information and incident attributes.

- Use of the integrated database in the previous bullet, to enable the identification of regional network bottlenecks and other vulnerabilities.

REFERENCES

Aghdashi, S., Rouphail, N. M., Hajbabaie, A., & Schroeder, B. J. (2015). Generic Speed Flow Models for Basic Freeway Segments on General Purpose and Managed Lanes in Under-saturated Flow Conditions. In Transportation Research Board 94th Annual Meeting (No. 15-4563).

Blincoe, L., Miler, T.R., Zaloshnja, E., (2010). The Economic and Societal Impact of Motor Vehicle Crashes. National Highway Traffic Safety Administration.

Chin, H.C., Quek, S.T., (1997). Measurement of Traffic Conflicts. *Safety Science* 26, 169-185.

Chin, H.C., Quek, S.T., Chu, R.L., (1992). Quantitative Examination of Traffic Conflicts. *Transportation Research Record: Journal of the Transportation Research Board* 1376.

Dunlap, D.F., Fancher, P.S., Scott, R.E., McAdam, C.C., Segal, L., (1978). Influence of Combined Highway Grandeand Horizontal Aligment on Skidding. National Cooperative Highway Research Program (NCHRP).

Garber, N. J. and R. Gadiraju (1989), Factors Affecting Speed Variance and Its Influence on Accidents. In Transportation Research Record 1213, pp. 64-71. Washington, D.C.: Transportation Research Board, National Research Council.

Glennon, J.C., Neuman, T.R., Leisch, J.E., (1986). Safety and Operational Considerations for Design of Rural Highway Curves. United States Department of Transportation.

Google Inc. (2015), http://maps.google.com, Last Accessed Oct 10 2015.

Here Inc. (2015), http://here.com, Last Access Oct 20 2015.

INRIX Inc. (2015), http://www.inrix.com, Last Access Oct 15 2015.

Itds & Idmec (2015), http://www.i2d.co, Last Accessed Oct 10 2015.

Knipling, R.R., Wang, J.S., Yin, H.M., (1993b). Rear-End Crashes: Problem Size Assessment and Statistical Description. National Traffic Safety Administration.

Lee, J., Par, B., Malakorn, (2013). Sustainability Assessment of Cooperative Vehicle Intersection Control at an Urban Corridor. *Transportation Research Part C* 32, 193-206.

Matthew D. D., Brady N. W. and Joseph, F. (2011), Estimating Free-flow Speed from Posted Speed Limit Signs, *6th International Symposium on Highway Capacity and Quality of Service*, Vol. 16, pp. 306-316

Oh, C., Kim, T., (2010). Estimation of rear-end crash potential using vehicle trajectory data. *Accident Analysis and Prevention* 42, 1888-1893.

Ottensen, J.L. and R.A. Krammes (2000). Speed-Profile Model for A Design-Consistency Evaluation Procedure in the United States, *In Transportation Research Record* 1701, pp. 76-85. Washington, D.C.: Transportation Research Board, National Research Council.

Roess, R. (2009), Re-Calibration of the 75-mi/h Speed-Flow Curve and the FFS Prediction Algorithm for HCM 2010. Research Memorandum, NCHRP Project 3-92. Polytechnic Institute of New York University, Brooklyn, N.Y.

Rowan, N.J. (1959). A Study of the Effects of Posted Speed Limits on Traffic Speeds, Thesis. Texas A&M College.

TEAAS, (2015). Traffic Engineering Accident Analysis System (TEAAS).

TRB, (2015). Naturalistic Driving Study (NDS), Strategic Highway Research Program 2 (SHRP2) Safety Reserach Reports.

Transportation Research Board (1997), Highway Capacity Manual, Washington D.C.

Transportation Research Board (2000), Highway Capacity Manual, Washington D.C.

Transportation Research Board (2010), Highway Capacity Manual, Washington D.C.

Wadrop J.G., (1952). Some Theoretical Aspects of Road Traffic Research, *Proceedings of the Institute of Civil Engineers*, Vol. 1-2, pp. 325-378

Yang, H., Ozbay, K., (2011). Estimation of Traffic Conflict Risk for Merging Vehicles on Highway Merge Section. *Transportation Research Record: Journal of the Transportation Research Board* 2236, 58-65.

Zegeer, C.V., Deacon, J.A., (1987). Effects of Lane Width, and Shoulder Type on Highway Safety, Transportation Research Board State-of-the-Art Report 6: Relationship Between Safety and Key Highway Features - A Synthesis of Prior Research.

APPENDIX A

This appendix presents the table structure of the local SQL database as well as queries that have been used to populate different tables with in SQL database from raw data.

1.   Table Structures

Table 24 – "Raw_data" Table Properties In Local SQL Database

COLUMN	TYPE	SIZE	UNSIGNED	NULL
TRIP_ID	int	11	Y	N
START_DATE	date			N
START_TIME	time			N
END_DATE	date			N
END_TIME	time			N
Seconds	smallint	11	N	N
Time	datetime			Y
Latitude	decimal	8,5	Y	Y
Longitude	decimal	8,5	N	Y
Course	smallint	11	Y	N
GPS Speed	tinyint	11	Y	N
Altitude GPS	smallint	11	N	N
Barometer Pressure	decimal	6,2	Y	N
Barometer Altitude (m)	smallint	11	N	N
Speed (mph)	tinyint	11	Y	N
Engine Load	tinyint	11	Y	N
RPM	smallint	11	Y	N
Battery State (SOC) (%)	tinyint	11	Y	N
Manifold air pressure	tinyint	11	Y	Y
In-take air temperature	tinyint	11	N	N
Air mass flow (g/s)	decimal	5,2	Y	Y
Throttle position (%)	tinyint	11	Y	N
Coolant temperature (?C)	smallint	11	N	N
Aceleration X (m/s2)	decimal	5,3	N	N
Aceleration Y (m/s2)	decimal	5,3	N	N
Aceleration Z (m/s2)	decimal	5,3	N	N
Slope (m/m)	decimal	4,3	N	N
Injection Cut (s/n)	tinyint	11	Y	N
Instant Consumption (mpg (us))	decimal	6,2	Y	Y

Accumulated Consumption (gal (us))	decimal	8,4	Y	Y
Average Consumption (mpg (us))	decimal	6,2	Y	Y
Gear shifting	tinyint	11	Y	N
Idle (s/n)	tinyint	11	Y	N
Cold Running (s/n)	tinyint	11	Y	N
FUEL_TYPE	char	15		N
DISPLACEMENT (L)	decimal	2,1	Y	N
CAR_DATE	date			N
VSP (W/kg)	smallint	11	N	N
Road Type	char	15		N
EXTREME_BRAKE_EVENTS	tinyint	11	Y	N
EXTREME_ACCELERATION_EVENTS	tinyint	11	Y	N
NEUTRAL_GEAR_EVENTS	char	15		N
POSITIVE_VSP_INTEGRAL (J/kg)	char	15		Y
EXCESS_RPM_EVENTS	tinyint	11	Y	N
EXCESS_CUSTOM_RPM_EVENT	tinyint	11	Y	N
EXCESS_SPEED_50_EVENTS	tinyint	11	Y	N
EXCESS_SPEED_90_EVENTS	tinyint	11	Y	N
EXCESS_SPEED_120_EVENTS	tinyint	11	Y	N
COMFORT_Y	tinyint	11	Y	N
SAFETY_Y	tinyint	11	Y	N
COMFORT_Z	tinyint	11	Y	N
HIGH_CO2_EMISSION_EVENT	tinyint	11	Y	N
HIGH_FUEL_EVENT	tinyint	11	Y	N
HIGH_VSP_EVENT	tinyint	11	Y	N
ZERO_SPEED_EVENT	tinyint	11	Y	N
DRIVER_ID	int	11	Y	N
GENDER	char	15		N
AGE_GROUP	char	15		N
EXPERIENCE	char	15		N
CAR_ID	int	11	Y	N
Instant Consumption (g/s)	decimal	4,2	Y	Y
IN_AREA	tinyint	11	Y	Y
uid	int	11	Y	N

Table 25 – "Driver_data" Table Properties In Local SQL Database

COLUMN	TYPE	SIZE	NULL
uid	int	20	N
DRIVER_ID	int	11	N
DRIVER_NAME	Char	15	N
START_DATE	Date		N
END_DATE	Date		N
AGE_GROUP	Char	15	N
EXPERIENCE	Char	15	N
CAR_ID	int	11	N
CURRENT	tinyint	1	N

Table 26 – "Calc_data" Table Properties In Local SQL Database

COLUMN	TYPE	SIZE	UNSIGNED	NULL
TRIP_ID	int	11	Y	N
SECONDS	smallint	11	N	N
TIME	datetime			Y
LATITUDE	decimal	8,5	Y	Y
LONGITUDE	decimal	8,5	N	Y
SPEED (MPH)	tinyint	11	Y	N
SPEED_CHANGE(MPH)	tinyint	11	N	N
TRIP_DISTANCE (FT)	decimal	6,2	Y	N
CALCULATED_ACCELERATION (FT/S^2)	decimal	6,2	Y	N
DRIVER_ID	int	11	Y	N
uid	int	11	Y	N

Table 27 – "Trip_data" Table Properties In Local SQL Database

COLUMN	TYPE	SIZE	UNSIGNED	NULL
TRIP_ID	int	11	Y	Y
DRIVER_ID	int	11	Y	Y
START_DATE	date			
START_TIME	time			
END_DATE	date			
END_TIME	time			
TRIP_LENGTH (MILE)	decimal	6,2	Y	Y
TRIP_TIME (MIN)	decimal	6,2	Y	Y
uid	int	11	Y	N

Table 28 – "Filtered_data" Table Properties In Local SQL Database

COLUMN	TYPE	SIZE	UNSIGNED	NULL
TRIP_ID	int	11	Y	Y
DRIVER_ID	int	11	Y	Y
START_DATE	date			
START_TIME	time			
END_DATE	date			
END_TIME	time			
TRIP_LENGTH (MILE)	decimal	6,2	Y	Y
TRIP_TIME (MIN)	decimal	6,2	Y	Y
uid	int	11	Y	N

Table 29 – "Excluded_days" Table Properties In Local SQL Database

COLUMN	TYPE
EVENT_TYPE	VARCHAR
EVENT_DATE	DATE
EVENT_DESCRIPTION	VARCHAR
uid	INT

Table 30 – "In_corridor_output" Table Properties In Local SQL Database

COLUMN	TYPE	SIZE
TRIP_ID	int	11
MIN_SECOND	smallint	11
MAX_SECOND	smallint	11
uid	int	11

Table 31 – "Thru_corridor_output" Table Properties In Local SQL Database

COLUMN	TYPE	SIZE
TRIP_ID	int	11
MIN_SECOND	smallint	11
MAX_SECOND	smallint	11
uid	int	11

2. Queries

Update driver_data table

UPDATE driver_data d
LEFT JOIN (SELECT driver_data.uid AS UID,

```
        driver_data.DRIVER_ID AS DRIVER_ID,
        driver_data.DRIVER_NAME AS DRIVER_NAME,
        MAX(raw_data.START_DATE) AS LAST_DATE
        FROM `driver_data`
        JOIN raw_data
        ON driver_data.DRIVER_ID = raw_data.DRIVER_ID
        WHERE driver_data.uid IN (2,3,9,15,17,19,26,28,31,24,13,6,27,21,22,12,32)
        GROUP BY driver_data.uid) r
ON d.DRIVER_ID = r.DRIVER_ID AND d.DRIVER_NAME = r.DRIVER_NAME
SET END_DATE = r.LAST_DATE
WHERE d.uid IN (2,3,9,15,17,19,26,28,31,24,13,6,27,21,22,12,32);

UPDATE driver_data d
LEFT JOIN (SELECT driver_data.DRIVER_ID AS DRIVER_ID,
        driver_data.DRIVER_NAME AS DRIVER_NAME,
        COUNT(DISTINCT raw_data.TRIP_ID) AS TRIP_IDS
        FROM `driver_data`
        JOIN raw_data
        ON driver_data.DRIVER_ID = raw_data.DRIVER_ID
        WHERE raw_data.START_DATE BETWEEN driver_data.START_DATE AND
driver_data.END_DATE
        GROUP BY driver_data.DRIVER_ID, driver_data.DRIVER_NAME) r
ON d.DRIVER_ID = r.DRIVER_ID AND d.DRIVER_NAME = r.DRIVER_NAME
SET TRIP_COUNT = r.TRIP_IDS;

UPDATE driver_data d
LEFT JOIN (SELECT d1.DRIVER_ID,
        d1.DRIVER_NAME,
        COUNT(DISTINCT r1.START_DATE) AS DAY_COUNT
        FROM driver_data d1
        JOIN raw_data r1
        ON d1.DRIVER_ID = r1.DRIVER_ID
        WHERE r1.START_DATE BETWEEN d1.START_DATE AND d1.END_DATE
        GROUP BY d1.DRIVER_ID, d1.DRIVER_NAME) r
ON d.DRIVER_ID = r.DRIVER_ID AND d.DRIVER_NAME = r.DRIVER_NAME
SET DAYS_WITH_TRIPS = r.DAY_COUNT;
```

Update calculated_data table

```
TRUNCATE TABLE calculated_data;

INSERT INTO calculated_data(calculated_data.TRIP_ID,
                 calculated_data.SECONDS,
                 calculated_data.TIME,
                 calculated_data.LATITUDE,
                 calculated_data.LONGITUDE,
                 calculated_data.SPEED,
                 calculated_data.DRIVER_ID)
SELECT t1.TRIP_ID,
       t1.SECONDS,
       t1.TIME,
       t1.LATITUDE,
       t1.LONGITUDE,
       t1.SPEED,
       t1.DRIVER_ID
FROM raw_data t1;

UPDATE calculated_data
SET TRIP_DISTANCE = SPEED * 1.47;

UPDATE calculated_data c
JOIN (SELECT * FROM raw_data) p
ON p.TRIP_ID = c.TRIP_ID AND p.SECONDS = c.SECONDS - 1
SET ACCELERATION = ROUND((c.SPEED - COALESCE(p.SPEED, 0)) / 1.47, 2);
```

Update trip_data table

```
TRUNCATE TABLE trip_data;

INSERT INTO trip_data(trip_data.TRIP_ID,
                      trip_data.DRIVER_ID,
                 trip_data.START_DATE,
                 trip_data.START_TIME,
                 trip_data.END_DATE,
                 trip_data.END_TIME
                 )
SELECT DISTINCT t1.TRIP_ID,
       t1.DRIVER_ID,
       t1.START_DATE,
       t1.START_TIME,
       t1.END_DATE,
       t1.END_TIME
FROM raw_data t1;
```

```
UPDATE trip_data t
JOIN(SELECT DISTINCT TRIP_ID AS TRIP_ID,
        SUM(TRIP_DISTANCE) AS TRIP_LEN
        FROM `calculated_data`
        GROUP BY TRIP_ID) c
ON t.TRIP_ID = c.TRIP_ID
SET t.TRIP_LENGTH_MILES = ROUND(c.TRIP_LEN / 5280, 2);

UPDATE trip_data t
JOIN(SELECT DISTINCT TRIP_ID AS TRIP_ID,
        MAX(SECONDS) AS SECONDS
        FROM `calculated_data`
        GROUP BY TRIP_ID) c
ON t.TRIP_ID = c.TRIP_ID
SET t.TRIP_TIME_MINUTES = ROUND(c.SECONDS / 60, 2);
```

Update filtered_data table

```
TRUNCATE TABLE filtered_data;

INSERT INTO filtered_data(filtered_data.TRIP_ID,
                        filtered_data.DRIVER_ID,
                filtered_data.START_DATE,
                filtered_data.START_TIME,
                filtered_data.END_DATE,
                filtered_data.END_TIME,
                filtered_data.TRIP_LENGTH_MILES,
                filtered_data.TRIP_TIME_MINUTES
                )
SELECT DISTINCT t1.TRIP_ID,
        t1.DRIVER_ID,
        t1.START_DATE,
        t1.START_TIME,
        t1.END_DATE,
        t1.END_TIME,
        t1.TRIP_LENGTH_MILES,
        t1.TRIP_TIME_MINUTES
FROM trip_data t1
WHERE t1.TRIP_LENGTH_MILES > 1
AND t1.TRIP_TIME_MINUTES > 5
AND t1.START_DATE NOT IN (SELECT EVENT_DATE FROM excluded_days);
```