2017

Congestion Quantification Using The National Performance Management Research Data Set

Shaghayegh Rostami Hosuri
University of Alabama at Birmingham

CONGESTION QUANTIFICATION USING THE NATIONAL PERFORMANCE MANAGEMENT RESEARCH DATA SET

by

SHAGHAYEGH ROSTAMI HOSURI

VIRGINIA P. SISIOPIKU, CHAIR
ANDREW SULLIVAN
IAN HOSCH

A THESIS

Submitted to the graduate faculty of The University of Alabama at Birmingham,
in partial fulfillment of the requirements for the degree of
Master of Science

BIRMINGHAM, ALABAMA

2017

CONGESTION QUANTIFICATION USING THE NATIONAL PERFORMANCE MANAGEMENT RESEARCH DATA SET

SHAGHAYEGH ROSTAMI HOSURI

CIVIL ENGINEERING

ABSTRACT

Traffic congestion is a serious urban transportation problem, compounded by the fact that demand for travel grows all the time. Monitoring of transportation system performance is a key part of any transportation operation and planning strategy and depends on estimation of performance measures generated from analysis of large amounts of traffic data. This thesis focused on analyses of travel time data using the Regional Planning Commission of Greater Birmingham travel time database. The goal of this research was to develop an automated process to utilize the National Performance Management Research Data Set (NPMRDS) for generating performance measures for congestion monitoring applications. The Birmingham region was used as a test bed in a case study to demonstrate how NPMRDS data can be used to determine the extent of congestion along four main corridors (I20, I65, I59, I20/59) over space and time. The capabilities of the relational database management system (RDBMS) are employed in order to manage the large amounts of data. Powerful visual maps were developed by GIS software that helped to identify areas of congestion and illustrate congestion extent and severity.

The thesis describes the NPMRDS characteristics and the factors that affect data measurements and discusses ways to address issues related to the data such as data gaps and anomalies. The case study demonstrates the feasibility of the proposed methodology. Overall, this research showcases the use of NPMRDS database to understand congestion patterns and provides insights into travel time uncertainty in urban areas such as the Birmingham metropolitan region.

Keywords: Transportation Data Analytics; Performance Measurement; Congestion Management, NPMRDS, Bottlenecks.

ACKNOWLEDGMENTS

I would like to extend my sincere thanks and gratitude Dr. Virginia P. Sisiopiku for serving as my supervisor and mentor during my master's studies. The completion of this work would have not been possible without her vision and guidance. In addition, I express my sincere appreciation to Dr. Ian Hosch and Mr. Andrew Sullivan for serving as committee members. Appreciation goes as well to Dr. Ossama E Ramadan and Mr. Amin Pahlevannejad for their valuable time, suggestion, and advice. I also wish to thank the Department of Civil, Construction, and Environmental Engineering of the University of Alabama at Birmingham (UAB) and the Regional Planning Commission of Greater Birmingham (RPCGB) for sponsoring my master's studies.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure *Page*

LIST OF ABBREVIATIONS

AADT	Annual Average Daily Traffic
ALDOT	Alabama Department of Transportation
ATRI	American Transportation Research Institute
BTI	Buffer Time Index
CMS	Congestion Management System
DOT	Department of Transportation
FFS	Free-Flow Speed
FFTT	Free-Flow Travel Time
FHWA	Federal Highway Administration
FPM	Freight Performance Measures
GIS	Geographic Information System
IF	Impact Factor
LOS	Level of Service
MAP-21	Moving Ahead for Progress in the 21st Century
MPO	Metropolitan Planning Organizations
NHS	National Highway System
NPMRDS	National Performance Management Research Data Set
PTI	Planning Time Index
RDBMS	Relational Database Management System
RPCGB	Regional Planning Commission of Greater Birmingham
SHRP 2	The second Strategic Highway Research Program

SQL	Structured Query Language
TMC	Traffic Message Channel
TTI	Travel Time Index
TTRI	Travel Time Reliability Indices
UCR	Urban Congestion Report
UMS	Urban Mobility Scorecard
VPP	Vehicle Probe Project

CHAPTER 1: INTRODUCTION

1.1. Background

Growing traffic congestion on America's roadways has negative impacts on mobility, the environment, and the economy. According to a Texas A&M Transportation Institute report, the total congestion cost for 471 U.S. urban areas in 2014 was $160 billion and congestion caused travelers to waste 6.9 billion hours and a number of billion gallons of fuel [1]. Congestion can result from many factors such as the presence of physical bottlenecks (capacity), traffic incidents, work zones, weather, traffic control devices, and special events[2]. Traditional approaches to address the congestion involve widening roads by adding lanes or building new facilities. However, such approaches offer only temporary congestion relief, are costly, and often impractical, especially under the current economic uncertainty. In the absence of supply expansion, it is necessary to implement an effective and efficient congestion management process that makes the existing transportation system operate more efficiently.

Congestion management process is a systematic approach for managing congestion that greatly depends on accurate information on transportation system performance. Performance measures are used to measure congestion on both the local and regional level. Travel Time Reliability is an example of a performance measure utilized by many transportation agencies, and regional planning organizations to assess variability in travel time [3].

Recent improvements in data-driven technologies create new opportunities for gathering detailed data and using them to quantify transportation system performance. This, in turn, will enable development of more accurate methods for prioritizing and evaluating congested roadways. To facilitate the process, the federal government has acquired a national data set of average travel times called National Performance Management Research Data Set (NPMRDS) and made it available to States and Metropolitan Planning Organizations (MPOs) to utilize for their performance management activities.

1.2. Problem Statement and Objectives

Despite the many advantages resulting from systematic collection and sharing of detailed traffic related date, challenges also exist with respect to data management, ensuring data accuracy, storing data, and using them to develop meaningful and useful performance measures. The objective of this research is to showcase the development of an automated process to facilitate the management, storage, and processing of big transportation datasets such as NPMRDS for congestion monitoring applications. The study uses traffic data analytics and statistical analysis to extract Travel Time Reliability performance measures. Such measures can be used to determine the congestion extend and severity and guide optimization of traffic operations.

1.3. Thesis Structure

This thesis is organized in six chapters as follows:

- Chapter 1 discusses the scope and objectives of the research
- Chapter 2 presents information about congestion performance measures based on an extensive literature search and details on previous studies in this field.
- Chapter 3 presents the framework and tools utilized for NPMRDS data acquisition and preprocessing and discusses the characteristics of the case study and data analysis techniques
- Chapter 4 presents summary findings and interpretation of the related results
- Chapter 5 summarizes the study conclusions and provides recommendations for future research.

CHAPTER 2: LITERATURE REVIEW

2.1. The Importance of Travel Time Reliability

A Congestion Management System (CMS) is a systematic planning process for managing congestion that offers information on transportation system performance and on substitute strategies for alleviating congestion and enhancing mobility. In 1999, E. Lindquist summarized the CMS into three main areas, namely the identification of effectiveness measures, the collection of data and system monitoring, and the development of improvement plans and strategies [4]. Therefore, considering effective performance measures is an important part of having a successful Congestion Management System.

Traditionally, assessing system performance was based on the average travel times. However, travel time alone is not capable of representing adequately the quality of service that commuters experienced every day and may lead to underestimation of the level of congestion. To measure the effect of unexpected congestion on user perceived quality of service, Travel Time Reliability measures should be also considered. Travel Time Reliability assists transportation planners and decision-makers in quantifying traffic congestion better and implementing successfully congestion monitoring applications.

In 1999, Lida defined Travel Time Reliability as the probability of on-time arrival [5]. In addition, Lodex et al in 2003 described Travel Time Reliability as a measure that accounts for the variability of travel time experienced by commuters and as an indicator of the consistency of a certain mode during a time period [6]. Within that context, the impact of congestion that causes the variability of travel time across different times of day, and from day to day can be measured through Travel Time Reliability indices.

Figure 1 illustrates an actual travel time distribution for a particular segment of road and implies how reliability metrics can be defined. The distribution shape depends on the events skewed toward higher travel times. In fact, the skew reflects the impacts of interruptions, such as incidents, work zones, and bad weather. Therefore, the Travel Time Reliability metrics are placed mostly on the right half of the distribution. It should be noted that the free-flow travel time is the benchmark for any Travel Time Reliability analysis.

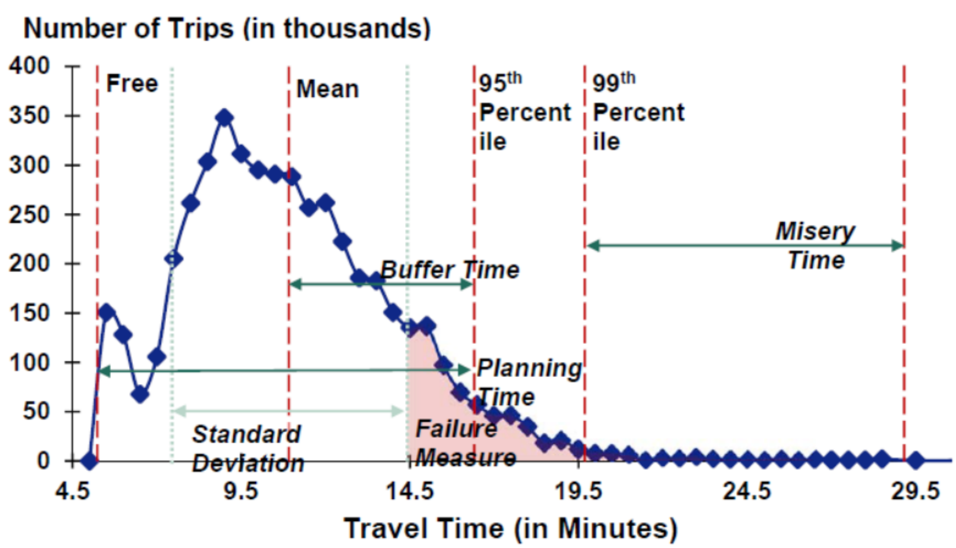

Figure 1. Travel Time Distribution [7]

2.2. Travel Time Reliability and Level of Service

The Level of Service (LOS) analysis is defined in the Highway Capacity Manual (HCM) as a process that assigns the quality of service levels to the facility based on performance measures such as density, speed, etc. [8]. This is done using a rating system from A to F, where A corresponds to free flow conditions and F denotes failure of the facility to accommodate the demand. In 2002, Chen et al proposed a formula that creates a link between travel time and Level of Service (LOS). By plotting travel time mean and standard deviation during each Level of Service rate, this relationship was proved. The study highlighted that Travel Time Reliability is a useful measure of freeway service quality that is superior to LOS since it can capture travel time variability experienced by commuters [9]. In 2012, the SHRP 2 L08 Project team conducted research to integrate Travel Time Reliability to HCM and develop Travel Time Reliability LOS measures [10]. Table 1 captures the recommended range of Travel Time Index (TTI) for equivalent speed and corresponding LOS based on the research findings reported in [10]. Travel Time Index refers to the ratio of the travel time during the peak period to the time required to make the same trip at free-flow speeds.

Table 1. SHRP 2 L08 Project Team's Freeway Reliability LOS defined by Travel Speed and TTI Ranges [10]

		Percent of Trips in Each LOS Range, Weekdays, 4:30 – 6:00 PM			
	Travel Speed* and Equivalent TTIs	Seattle, I-405		Atlanta, I-75, Northside	
LOS		NB	SB	NB	SB
A	>=60 (TTI <= 1.083)	27.7%	1.6%	0.4%	15.4%
B	50—59 (1.083 < TTI <= 1.300)	71.9%	48.3%	6.6%	80.5%
C	45—49 (1.300 < TTI <= 1.444)	0.3%	12.0%	3.2%	1.4%
D	40 – 44 (1.444 < TTI <= 1.625)	0.0%	9.3%	8.5%	0.3%
E	35 – 39 (1.625 < TTI <= 1.857)	0.1%	11.0%	14.4%	0.8%
F	< 35 (TTI > 1.857)	0.0%	17.8%	66.8%	1.6%
	Mean TTI	1.016	1.352	1.984	1.050

* Travel speed is the space mean speed over the facility

To better summarize the user perceptions of congestion based on Table 1 three different thresholds were chosen in this study that categorizes congestion into four different level (Table 2):

Table 2. TTI Threshold Values

Level of Congestion	**TTI Threshold value**
Little/None	TTI <1.1
Moderate	1.1 <= TTI <1.5
Significant	1.5 <= TTI < 2
Severe	TTI >= 2

2.3. Studies on the National Performance Measurement Research Dataset

Federal Highway Administration (FHWA) has acquired a national data set of average travel times called National Performance Management Research Data Set (NPMRDS) and made it available to States and Metropolitan Planning Organizations (MPOs) to utilize for their performance management activities in November 2013 [11]. Since then, FHWA holds webinars quarterly to introduce NPMRDS to users and share findings of ongoing research in an effort to assist agencies to meet the MAP-21 requirements for performance-based reports [5]. While the benefits of gaining access to a comprehensive database such as NPMRDS are tremendous, some challenges and difficulties have been reported by DOTs, MPOs and researchers in their efforts to utilize the NPMRDS dataset to develop performance measures and generate the reliability reports.

Wisconsin Traffic Operation and Safety Laboratory was one of the first institutes that applied probe data for developing performance measures. In 2014, they offered a performance measure process that describes the steps that should be taken for data processing and developing mobility measures such as Travel Time Reliability and Vehicle Delay by integrating hourly volume into NPMRDS (Figure 2) [12]. Regarding data management, they declared that the dataset required the usage of database and scripting skills for this purpose.

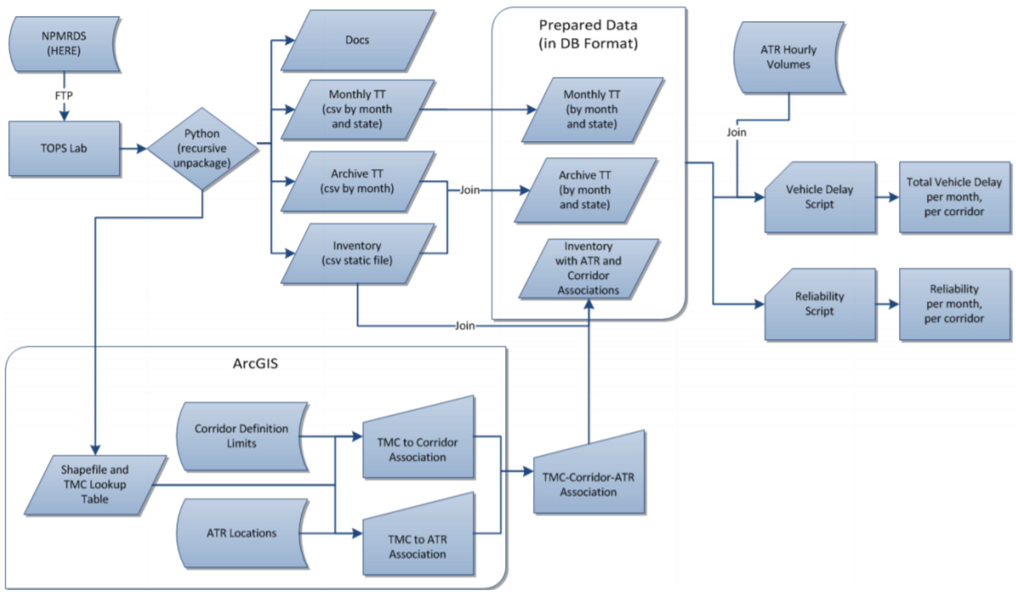

Figure 2. Data Processing Flowchart by Wisconsin TOPS [12]

They also studied travel time data distributions and confirmed the presence of outliers and data gaps in the dataset. In addition, they investigated ways to integrate NPMRDS dataset with ARCGIS in order to enable visualization of the findings on a map, and suggested to employ a tool called "Shape2SQL" that allows importing shapefile into the database and assigning the result to the shapefile.

The University of Minnesota and Minnesota DOT provided another valuable report on freight performance analysis, A total of 38 freight corridors were studied, and SQL scripts were used for data processing [13]. This work demonstrates the feasibility of travel time data records obtained from freight trucks as a source for calculation of speed variation and truck delay during peak hours.

In another study, the American Transportation Research Institute (ATRI) reported their analysis results regarding the cost of delay and congestion experienced by the freight industry[14]. The University of Maryland conducted a validation analysis between NPMRDS and I-95 Corridor Coalition's Vehicle Probe Project (VPP) data. The researchers pointed out that the comparison between different data sources is complicated as it requires careful consideration of the differences in segments as every data collection source uses different segmentations for collecting traffic data [15].

Another research institute that directed a validation analysis was the Upper Midwest Reliability Resource. They reported that the travel time data records in the NMPRDS dataset display a higher variation and a lower mean of travel time compare to data records INRIX dataset. This finding raises a question of whether there is a need for filtering outliers and imputing missing data. To address this issue, PostSQL and Psycopg were utilized to store the dataset and perform the analysis by writing codes in Python [16].

In May 2014, Iteris Inc. offered a module called "MAP-21 Module" to help agencies meet the requirement for reliability and congestion mitigation reports established by MAP-21. This module stores NPMRDS into the series of databases and users can easily query the data through a web interface, to develop performance measures and maps for visualization purposes [17]. They also developed a tool to impute data in data gaps. This work is considered a useful tool to overcome the issue of handling big data for the agencies with the lack of knowledge and capabilities to process big data and apply big data analytics [17].

To date, all the published researches and work have relied on the usage of complex programming languages and databases and have been done by experts in such fields. However, mid-size city MPOs employees and staff of transportation agencies that have encountered difficulties in utilizing the NMPRDS dataset for congestion monitoring purposes due to the lack of experience in big data analytics and related resources. This thesis develops tools that assist those agencies that need economic and less complex algorithms for managing big data and data processing to quantify the level and extent of congestion, generate performance-based reports and make effective strategic decisions for congestion mitigation in their jurisdictions.

CHAPTER 3: CASE STUDY DESCRIPTION AND METHODOLOGY

3.1. Introduction

This chapter first offers a description of the Birmingham case study used as the study testbed. Details on the National Performance Management Research Data Set (NPMRDS) are provided next, followed by information about the required steps to develop an automated process to utilize NPMRDS dataset for generating performance measures for congestion monitoring applications.

3.2. Site Location

The Birmingham region has been used as a test bed in a case study. Four major freeways were selected for data analysis namely I-65, I-20, I-59, and I-20/I-59. Figure 1 shows a map of the four study corridors on the region's transportation system. As can be seen in Figure 3, the study corridors extend over two counties from Jefferson/Blount County line on the North to the Shelby/Chilton County line on the South and from Tuscaloosa/Jefferson County line on the East to the Jefferson/St. Clair County line on the West. Originally, the study corridors were divided into the total number of 182 Traffic Message Channels (TMCs)[1] but new segmentations were defined that combine them to 14 major segments in each direction based on the location of the TMC, and their average annual daily traffic counts. The segment attributes are illustrated in Table 3.

---

[1] Traffic Massage Channel: The TMC code is an industry convention that defines a particular directional segment of the road.

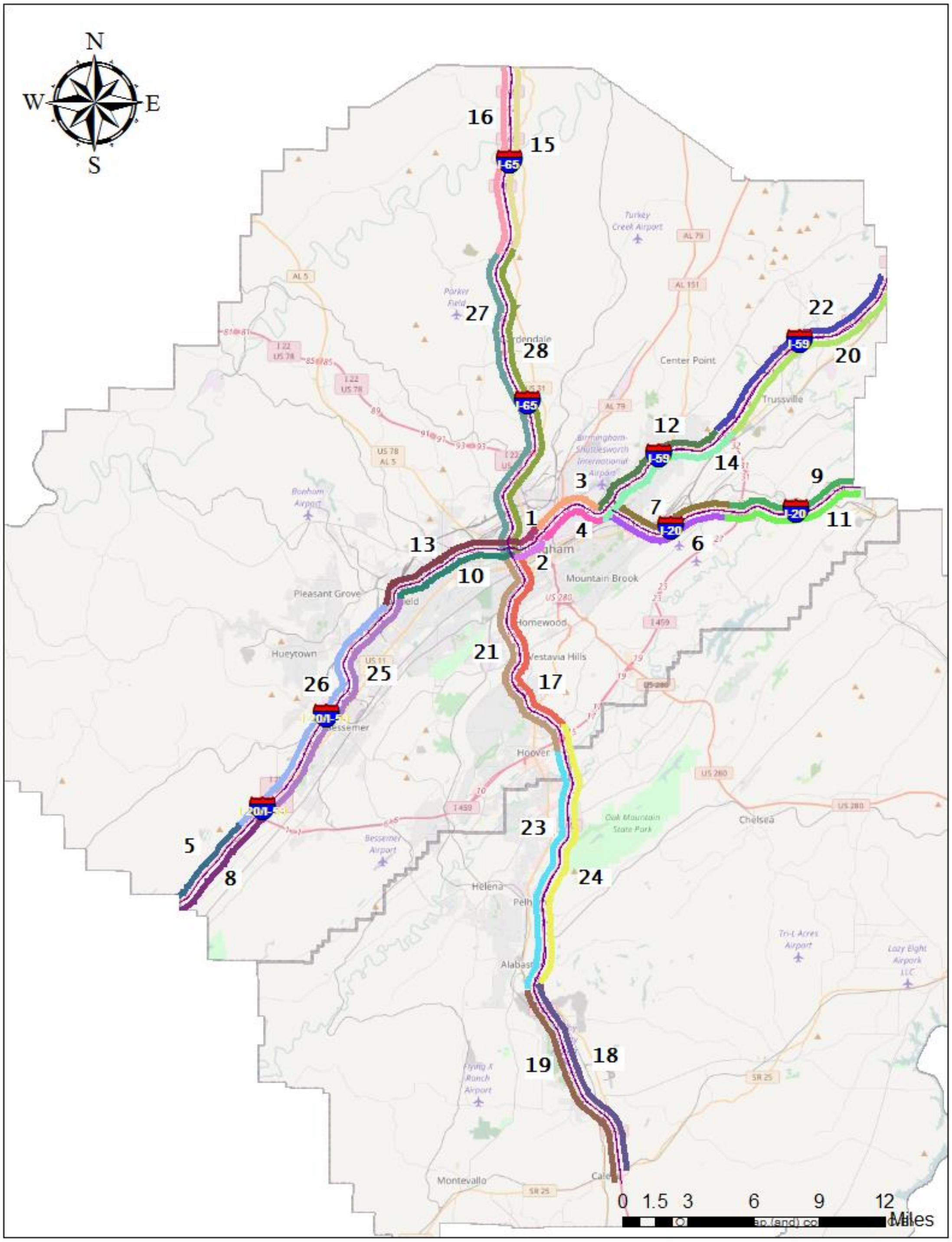

Figure 3. Case Study Freeway Network

Table 3. Study Segment Attributes

Road Number	Segment Name	Segment Code	TMC Count	Length (mile)	Travel Direction
I-20	I20/59 to I459	6	6	5.87	Eastbound
		7	6	5.96	Westbound
	I459 to St. Clair County	11	3	6.65	Eastbound
		9	3	6.43	Westbound
I-20/I-59	I459 to Valley Road	25	6	12.09	Eastbound
		26	6	12.74	Westbound
	I65 to RME	2	3	1.39	Eastbound
		1	3	1.27	Westbound
	RME to I20/59 Split	4	4	3.44	Eastbound
		3	4	3.34	Westbound
	Tuscaloosa Co. Line to I459	8	2	5.98	Eastbound
		5	3	6.54	Westbound
	Valley Road to I65	10	8	7.04	Eastbound
		13	9	7.76	Westbound
I-59	I20/59 to I459	14	7	7.76	Northbound
		12	6	7.47	Southbound
	I459 to St. Clair County	20	4	10.45	Northbound
		22	5	10.85	Southbound
I-65	Chilton County Line to US31 in Alabas	18	3	9.99	Northbound
		19	3	10.04	Southbound
	I20/59 to US31/Mary Buckelew	28	10	15.1	Northbound
		27	9	14.42	Southbound
	I459 to I20/59	17	10	9.44	Northbound
		21	11	10.69	Southbound
	US31 (Exit 275) to Cullman County Li	15	8	13.96	Northbound
		16	9	16.65	Southbound
	US31 in Alabaster to I459	24	6	12.53	Northbound
		23	6	11.83	Southbound

In this study, analyses have been performed on all of the TMCs located along the study corridors over an one-year period (from January 2015 to December 2015) in order to measure congestion along the study segments.

3.3. Data Sources

3.3.1. Dataset Acquisition

This study utilized NPMRDS datasets that were obtained from the Federal Highway Administration (FHWA) with the help of the Regional Planning Commission of Greater Birmingham (RPCGB). Travel time estimates derived from NPMRDS can provide valuable insights on travel times, speeds, and reliability.

3.3.2. NPMRDS Dataset Overview

The NPMRDS database was acquired by the FHWA Office of Operations back in 2013. Primarily, the objective of FHWA with the purchasing of NPMRDS data was to use the database in its performance measurement reports (FPM[1] and UCR[2]). However, with the passage of the MAP-21[3], FHWA made the dataset available for use by State DOTs and MPO partners to generate metrics and analyses[18].

NPMRDS is a form of vehicle probe-based travel time data set meaning data records are collected from a variety of sources that reports vehicle speed, position, and heading. The sources of these probe points can be connected cars, mobile devices, portable navigation, commercial fleet, and sensors. More than hundreds of billions probe points have been processed in 2016, and they continue to growth exponentially due to the increase in the number of connected cars, mobile devices sensors, and commercial fleet every day.

---

[1] Freight Performance Measures
[2] Urban Congestion Report
[3] Moving Ahead for Progress in the 21st Century

Multiple observations on a TMC segment during any 5-minute intervals (EPOCH) are aggregated to compute average travel speeds; then travel times are computed by dividing the segment length by average travel speeds. The data set provides average travel time in seconds for every 5 minutes, 24 hours per day, and seven days per week and covers the entire National Highway System (NHS) containing all interstates and US highways. It also offers three different categories for travel time estimate include freight trucks, passenger cars and all vehicles. NPMRDS data for each month is available in the following month with no estimates, historical data substitution, and imputation[19].

Passenger vehicles probe speed data is collected by HERE, and the ATRI[1] collects freight trucks probe speed observations. All vehicles travel times are a weighted average determined by combining passenger cars and freight trucks average travel speed based on a respective number of observations. Travel time data is referenced to TMC[2] codes which represent locations of collecting data. TMC codes are a unique reference that breaks down NHS roads into unequal segments for each direction. Moreover, the NHS shapefile is supplied to the dataset that enables mapping and spatial analysis in ArcGIS. The NHS shapefile contains precise road geometry and attributes of each road section. Tables 2 and 3 represent a sample of travel time data and static file for TMC from the study site.

3.3.2.1. NPMRDS File Layout

The NPMRDS dataset is composed of large files as detailed below:

[1] American Transportation Research Institute
[2] Traffic Message Channel

1. Monthly Average Travel Time Data File: The data file format is "FHWA_TASK_AL_YYYY_MM _TT.csv", where yyyy and mm represent the year and month of data collection, respectively. Table 4 summarizes the types of travel time data available, and Table 5 offers a sample of travel time data from the study site.

Table 4. Travel Time Data

Field Name	Data Type	Example	Description
TMC	Short Text	101N04496	Traffic Message Channel (TMC) is bi-directional codes that defines a particular directional segment of the road. It is coded in the map as Positive and Negative. These directions relate to other TMC codes in the location table and do not relate to road travel Traffic Location code in the format of: CLLDTTTTT C is the Country Code (1 digit). LL is the Country Code (2 digit). D ('P' Positive or 'N' Negative direction of the TMC). • TTTTT is the Country Code (5 digit).
DATE	Long Integer	1012016	The Date that travel time data is collected in the format of m/dd/yyyy
EPOCH	Long Integer	30	A value from 0 through 287 that defines the 5-minute period the average speed applies (local time), where: 0 = 00:00:00 to 00:04:59 1 = 00:05:00 to 00:09:59 2 = 00:10:00 to 00:14:59 … 287=23:55:00 to 23:59:59
Travel_TIME_ALL_VEHICLES	Long Integer	180	Travel times are calculated in seconds as the ratio between the segment length and the average speed on the segment. Average segment speed is determined from a combination of the passenger and freight trucks individual GPS probe speed observations.
Travel_TIME_PASSENGER_VEHICLES	Long Integer	154	Travel times are calculated in seconds as the ratio between the segment length and the average speed on the segment. Average segment speed is determined from only passenger individual GPS probe speed
Travel_TIME_FREIGHT_TRUCKS	Long Integer	185	Travel times are calculated in seconds as the ratio between the segment length and the average speed on the segment. Average segment speed is determined from only freight trucks individual GPS probe speed

Table 5. Travel Time Data Sample

TMC	Date	EPOCH	Travel_TIME_ALL_VEHICLES	Travel_TIME_PASSENGER_VEHICLES	Travel_TIME_FREIGHT_TRUCKS
101N04362	12012015	0	157	157	157
101N04363	12012015	1	186	186	186
101N04364	12012015	2	155	158	151
101N04365	12012015	3	157	157	157
101N04366	12012015	4	160	147	162
101N04367	12012015	5	162	158	164
101N04368	12012015	6	161	160	161
101N04369	12012015	7	148	147	149
101N04370	12012015	8	157	157	157

2. TMC Static Data File: The data file format is "FHWA_Monthly_Static_File_yyyyQn.csv", where yyyy and n represent the year and the quarter number of data collection. TMC Static File contains descriptive information about the road segment (TMC code, names, admin info, segment lengths, lat/long) as shown in Table 6. For illustration purposes, Table7 provides a sample of the Static File Data.

Table 6. Static File

Field Name	Data Type	Example	Description
TMC	Short Text	101N04496	Traffic Message Channel (TMC) is bi-directional codes that defines a particular directional segment of the road. It is coded in the map as Positive and Negative. These directions relate to other TMC codes in the location table and do not relate to road travel direction. Traffic Location code in the format of: CLLDTTTTT C is the Country Code (1 digit). LL is the Country Code (2 digit). D ('P' Positive or 'N' Negative direction of the TMC). • TTTTT is the Country Code (5 digit).
ADMIN_LEVEL_1	Short Text	USA	The Country where the listed Traffic Location Code is located.
ADMIN_LEVEL_2	Short Text	Alabama	The State / Province where the listed Traffic Location Code is located.
ADMIN_LEVEL_3	Short Text	Jefferson	The County where the listed Traffic Location Code is located.
DISTANCE	Double	2.73544	The length of the TMC, measured in Miles to five decimal places.
ROAD_NAME	Long Text	Richard Arrington Jr Blvd N	The local name of the road.
ROAD_NUMBER	Short Text	I-20	The route number of the road.
LATITUDE	Double	33.54717	WGS84 coordinate to five decimal places represents the latitude of the beginning of the TMC
LONGITUDE	Double	-86.77939	WGS84 coordinate to five decimal places represents the longitude of the beginning of the TMC
ROAD_DIRECTION	Short Text	Southbound	The direction of travel based on the road sign

Table 7. Static File Data Sample

TMC	ADMIN_LEVEL_1	ADMIN_LEVEL_2	ADMIN_LEVEL_3	DISTANCE	ROAD_NUMBER	ROAD_NAME	LATITUDE	LONGITUDE	ROAD_DIRECTION
101N04362	USA	Alabama	Jefferson	2.89	I-20		33.54636	-86.5859	Eastbound
101N04363	USA	Alabama	Jefferson	1.13576	I-20		33.54784	-86.63269	Eastbound
101N04364	USA	Alabama	Jefferson	2.4618	I-20		33.5458	-86.65443	Eastbound
101N04365	USA	Alabama	Jefferson	1.19288	I-20		33.53871	-86.69048	Eastbound
101N04366	USA	Alabama	Jefferson	0.51208	I-20		33.53147	-86.7124	Eastbound
101N04367	USA	Alabama	Jefferson	1.33533	I-20		33.53365	-86.71868	Eastbound
101N04368	USA	Alabama	Jefferson	0.29241	I-20		33.54408	-86.73904	Eastbound
101N04369	USA	Alabama	Jefferson	0.07651	I-20		33.5466	-86.74474	Eastbound
101N04370	USA	Alabama	Jefferson	1.04314	I-65		33.37246	-86.77859	Southbound

3. NHS shapefiles: The polyline shapefile provided is customized to include detailed road geometry of the NHS and attributes about the road segment. The shapefile helps to create the network dataset which, in turn, enables modeling of the transportation network. It provides the user the ability to view the roads and data contained in the files spatially.

3.4. Data Management

3.4.1. Relational Database Management System

The NPMRDS contains a significant amount of information. For instance, it covers 4,727 TMC segments in Alabama, each of which is generating 288 epochs per day. These figures scale to approximately 1,361,376 records per day, and 495,540,864 records annually. Despite the fact that the NPMRDS is a well-structured data set, the amount of travel time data records filed inhibits the ability of using typical desktop software, like excel for processing the data. Handling travel time files require a more advanced database and scripting expertise.

To address this challenge, a Relational Database Management System (RDBMS) implemented in this study to structure data into relational tables and allow for data to be encrypted and analyzed through Structured Query Language (SQL). RDBMS is a well-proven database management system invented by computer scientist, Edgar F. Codd [20]. The technology used in a relational database management system stores data in a systematic way that makes retrieving, manipulating, and producing information more efficient and secure. Relational database management system supports a relational model which is the most popular data model in the open source and commercial tools and databases for data storage and processing. Data models demonstrate the relationships between data together with the procedure of processing and storing data inside the system. The most important relational model features include an ability to a. store the data in tables (or “relations”) of rows and columns which represent the records and the attributes, and b. normalize relations in order to eliminate unnecessary data, and ensuring that data is logically stored. The NPMRDS dataset structure meets the requirements of using a relational database management system, and thus RDBMS was employed in this study.

One of the most popular desktop applications for RDBMS is Microsoft Access database. In this project, Microsoft Access has been used since it is affordable, easy to learn, significantly easier to implement and maintain compared to larger database systems such as Oracle or SQL Server, integrates well with Excel, Word, Outlook. Most RDBMS use SQL as a powerful language for querying and maintaining the database. This was also the case in the Birmingham case study.

3.4.2. Database Architecture

In order to downsize the Access database and avoid exceeding the 2GB limit, the database has been split into two files, a front-end and a back-end. The back-end database contains only tables and relationships, and the front-end provides queries, forms, reports, and modules. The remarkable advantages which have been gained by splitting the Access database as described were performance improvements, the reduction in corruption, and the ability to create a multi-user database (Figure 4). In addition, deploying updates to the design of queries, forms, reports and modules was made reasonably convenient by replacing the front-end database.

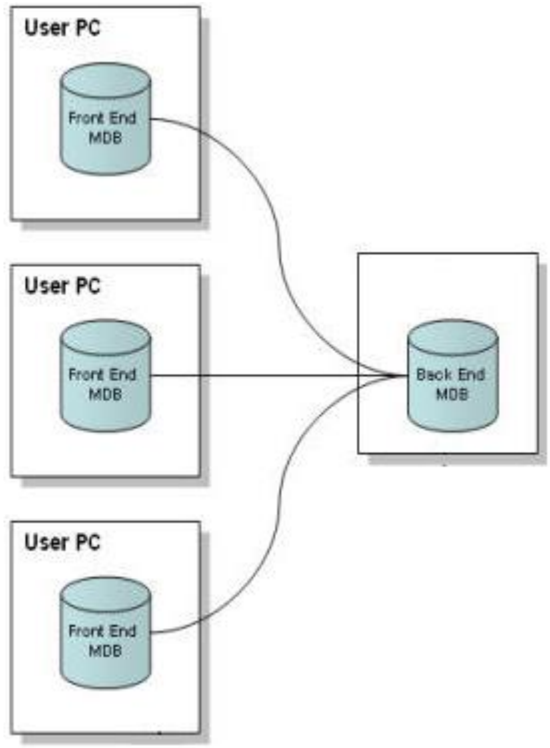

Figure 4. Multi-user Database

3.4.2.1. Back-end file

The back-end file contains a series of tables uploaded and then stored as an accessible, queryable file that has relationships with a primary (foreign) key, which allows manipulation and processing of data in any order. In this research, in addition to the Travel Time table and Static file that came with NPMRD, a series of tables have been created to leverage usage of data. For example, Table 8 is a calendar table which provides more information for each data collection date. Table 9 is a sample of an epoch table which relates each epoch to different time periods and assigns a unique code to each 15-mintue period of 24 hours. Table 10 shows segment data that define new segmentations by combining TMCs.

The attribute data stored in these tables by relating them to the unique entities known as relation keys in order to create a good database structure. Each attribute holds a different data type. A relation key can identify each record in the database, uniquely.

Table 8. Calendar

Field Name	Data Type	Example	Description
Date	Long Integer	1012016	The Date that travel time data is collected in the format of m/dd/yyyy
Day	Short Text	Monday	The day of the week for any particular date
WorkDay	Integer	0 or 1	Binary variables that differentiate workdays from Non-workdays, where: 1 = Workdays 0 = Weekends
Comment	Short Text	New Year holidays	It enables data analysist to provide a reason for holidays

Table 9. EPOCH Data

Field Name	Data Type	Example	Description
EPOCH	Long Integer	30	A value from 0 through 287 that defines the 5-minute period the average speed applies (local time), 0 = 00:00:00 to 00:04:59 1 = 00:05:00 to 00:09:59 2 = 00:10:00 to 00:14:59 … 287=23:55:00 to 23:59:59
StartTime	Date/Time	12:00:00 AM	The beginning time of each 5-mintute period in the format of hh:mm:ss AM/PM
EndTime	Date/Time	12:04:59 AM	The Ending time of each 5-mintute period in the format of hh:mm:ss AM/PM
15minInterval	Long Integer	15	A value from 0 through 95 that defines the 15-minute period
TrafficPeriod	Short Text	PM_Peak	It defines the traffic period, where: OFF_Peak = 19:00:00 to 5:59:59 AM_Peak = 6:00:00 to 9:59:59 Mid_Day = 10:00:00 to 14:59:59 PM_Peak = 15:00:00 to 18:59:59
StartHour	Date/Time	10	A value from 0 through 23 that defines when the 1 hour period 0 = 00:00:00 … 23 = 23:00 :00

Table 10. Segment Data

Field Name	Data Type	Example	Description
TMC	Short Text	101N04496	
Segment	Short Text	I20/59 to I459	A name that describe the beginning and the end of segment
Ffs	Integer	60	A value for Free flow speed
Seg_length	Double	5.87347	The length of segment, measured in Miles
Seg_Code	Integer	1	A value from 1 to 28

Figure 5 illustrates the relationships that have been defined between the multiple tables introduced in the previous paragraphs. Ensuring that the data is logically stored and the same data has not been stored in more than one tables, is a worthy goal as it reduces the amount of required database space.

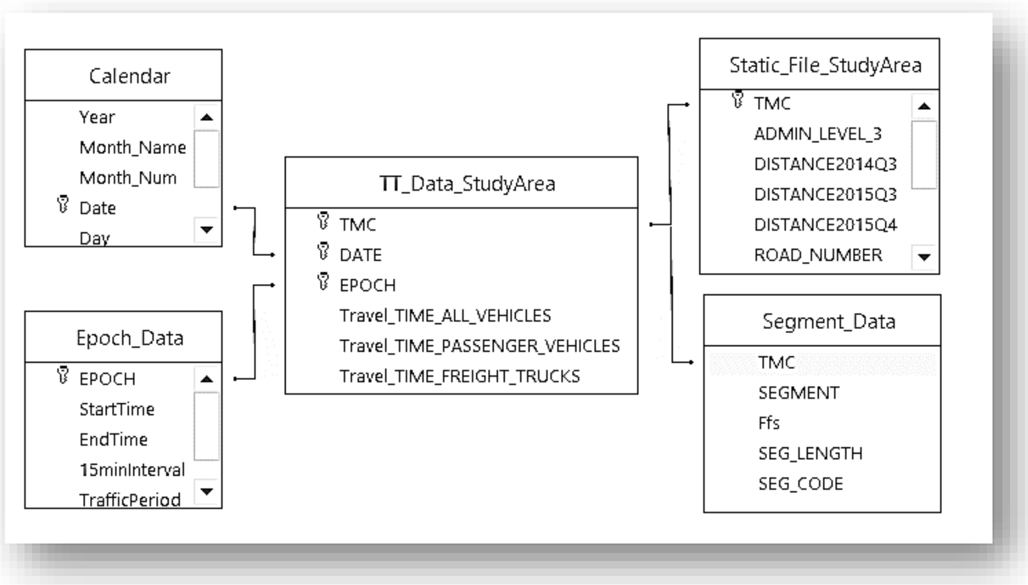

Figure 5. Relationships and Primary Keys

3.4.2.2. Front-end file

The front-end database enables users to access the raw data stored in back-end dataset and display data. The Structured Query Language (SQL) was used for providing data summaries, queries, and analyses. Table 11 illustrates examples of query commands utilized in this study.

Table 11. Necessary SQL Command

SQL Command	Function
SELECT column_name FROM table_name	Used to fetch data from a database. Every query will begin with SELECT
JOIN table_a ON table_b.column_name = table_a.column_name	Is the primary key in each table is used to join two tables
GROUP BY column_name;	Used with aggregate functions to arrange identical data into groups. It is used in collaboration with the SELECT statement.
WHERE column_name operator value	Used to filter the result set to include only rows where the following condition is true
HAVING column_name operator value	Used for filtering data based on the group functions. Group functions cannot be used in WHERE Clause but can be used in HAVING clause
ORDER BY "column_name"	Used to list the output in a particular order

In particular, Table 12 shows an attempt to compute the monthly average travel time during peak hours for all TMC segments located in Birmingham region during peak hours in January 2015.

Table 12. Sample Query for Calculation of Monthly Average Travel Time

```
SELECT
      Segment_Data.TMC, Epoch_Data.[15minINTERVAL],   Epoch_Data.Traffic_Period,
      Round(Avg([TT_Data_MPO]![Travel_TIME_ALL_VEHICLES]),0) AS MAPTT,
      Round(Avg([Static_File_MPO]![DISTANCE]*3600/[Segment_Data]![Ffs]),0) AS FFTT
FROM
      Static_File_MPO INNER JOIN (Segment_Data INNER JOIN (Epoch_Data INNER JOIN (Calendar
      INNER JOIN TT_Data_MPO ON Calendar.Date = TT_Data_MPO.DATE) ON Epoch_Data.EPOCH
      = TT_Data_MPO.EPOCH) ON Segment_Data.TMC = TT_Data_MPO.TMC) ON
      Static_File_MPO.TMC = TT_Data_MPO.TMC
GROUP BY
     Segment_Data.TMC, Epoch_Data.[15minINTERVAL], Epoch_Data.Traffic_Period,
     Calendar.WorkDay
HAVING
     (((Epoch_Data.Traffic_Period) In ("AM_PEAK","PM_PEAK")) AND ((Calendar.WorkDay)=1));
```

3.4.3. Cloud-based storage

In order to facilitate file sharing and collaboration within users, all the databases and the data records obtained from RPCGB were restored in UABbox. UABbox is a free cloud-based storage environment that provides scalable, high performance and reliable unlimited storage service for personal and group accounts. UABbox includes a Box sync desktop application that allows conducting the analysis directly on the files that have been stored on its server.

3.5. Mobility Performance Measures

Quantifying the congestion along the study corridors was accomplished on the basis of some popular mobility performance measures. In this study, using the NPMRDS data for the study period and the Relational Database Management System Travel Time Reliability indices were developed and used to quantify the presence, severity and extend of the congestion along the study corridors. More specifically, four indices were generated in this study, namely, the Travel Time Index, Duration of Congestion, Buffer Time Index, and Planning Time Index. Definitions of those indices are provided next.

3.5.1. Travel Time Index (TTI)

The Travel Time Index (TTI) is a measure that indicates congestion and reliability of roadway segments. The TTI index is defined as a ratio of average travel time to free-flow travel time for a given roadway segment [21] as shown in Eq. (1) that follows:

$$\text{Travel Time Index} = \text{TTI} = \frac{\text{Average Travel Time}}{\text{Travel Time Based on Free Flow Speed}} \qquad \textbf{(1)}$$

The TTI is simply a comparison of the time it takes to travel a given segment during the peak period to the time it takes to travel that same segment under free flow conditions. According to literature review, threshold values were chosen to reflect whether congestion was moderate, significant, or severe as summarized below. These threshold values have been selected to reflect user perceptions of congestion and its impact on their travel times and are summarized as follows.

1.10 < TTI < 1.50 moderate congestion

1.50 < TTI < 2.00 significant congestion

TTI > 2.00 severe congestion

3.5.2. Duration of Congestion (DOC)

To study the frequency of congestion during peak periods, congestion duration was also computed for each segment. Congestion duration was captured by summing all of the 15-minute intervals during peak periods that contain the TTI values greater than 1.1. In this study, threshold values were chosen in a similar fashion as with the TTI but with new threshold values as follows:

0 < DOC < 30 min moderate congestion persistency

30 < DOC < 60 min significant congestion persistency

DOC > 60 min severe congestion persistency

3.5.3. Planning Time Index (PTI)

The planning time index (PTI) estimates the extent by which the free-flow travel time will be exceeded. It is represented by the ratio of the 95th Percentile travel time as compared to the free-flow travel time as shown in Eq. (2). It should be noted that the 95th Percentile indicates an excessively high travel rate, one that only five percent of all travel rates exceed for the time period under consideration.

$$\text{Planning Time Index (PTI)} = \frac{\text{95th Percentile Travel Time}}{\text{Travel Time Based on Free Flow Speed}} \qquad \textbf{(2)}$$

3.5.4. Buffer Time Index (BTI)

Buffer Time Index (BTI) is a measure of Travel Time Reliability that states the extra time required as a percentage of the average to be on time for 95 percent of the trips as shown in Eq. (3). This measure is considering the unexpected delay as the 95th Percentile of travel time to estimate the extra travel time that the trip may take compare to average time of travel.

$$\text{Buffer Time Index}(\%) = \frac{\text{95th Percentile Travel Time} - \text{Average Travel Time}}{\text{Average Travel Time}} \quad \textbf{(3)}$$

For instance, a BTI equal to 50% means that, in addition to the expected delay that the average time of travel is accounting for, there is an unexpected delay that requires that travelers increase their average travel time by 50 percent.

3.5.5. Procedure for Travel Time Reliability Indices calculation

The Travel Time Reliability analyses were applied to all TMCs on the selected study corridors during AM peak hours (6:00 AM to 10:00 AM) and PM peak hours (3:00 PM to 7:00 PM) on weekdays from January 2015 to December 2015. Public holidays were excluded from the analysis. The calculation of some of Travel Time Reliability indices required Free Flow Speed data that were provided by the RPCGB. The step-by-step process is described next and the following tables provide a sample of the analysis performed for generating the Travel Time Reliability indices in the case study for one segment (code 02) in January 2015.

1. APTT (Average Peak Travel Time): Travel time data were aggregated over 15-minute intervals during peak periods (Table 13).

2. MAPTT (Monthly Average Peak Travel Time): Travel time data were aggregated by month in 2015 (Table 14).
3. TTRI (Travel Time Reliability Indices): The indices for each TMC were calculated based on aforementioned formulas (Table 15).
4. STTRI (Segment Travel Time Reliability Indices): To calculate the TTR indices for the selected roadway segments, Eq. (4) was utilized. It provides the average of TTRI values across several TMCs along the selected roadway segments by using a weighting factor (Table 16).

$$\text{STTRI}_i = \frac{\sum_1^n (TTRI_n \times Length_n)}{Length_i} \quad \textbf{(4)}$$

Where:

i: Segment code from 1 to 28

n: TMC number along with segment i

5. Max STTI: the maximum of STTI over AM peak and PM peak separately was chosen to illustrate the reliability of travel time for a certain month (Table17).

Table 13. AAPT Calculation Sample

Year	Month Number	SEG Code	TMC	Date	15min Interval	Traffic Period	APTT	AFFTT
2015	1	2	101P04446	1122015	24	AM_Peak	24.50	22.88
2015	1	2	101P04447	1122015	24	AM_Peak	39.00	33.81
2015	1	2	101P04448	1122015	24	AM_Peak	39.67	43.16
2015	1	2	101P04446	1122015	25	AM_Peak	22.33	22.88
2015	1	2	101P04447	1122015	25	AM_Peak	32.50	33.81
2015	1	2	101P04448	1122015	25	AM_Peak	39.67	43.16
2015	1	2	101P04446	1122015	26	AM_Peak	19.00	22.88
2015	1	2	101P04447	1122015	26	AM_Peak	30.67	33.81
2015	1	2	101P04448	1122015	26	AM_Peak	40.00	43.16
2015	1	2	101P04446	1122015	27	AM_Peak	23.67	22.88
2015	1	2	101P04447	1122015	27	AM_Peak	32.67	33.81
2015	1	2	101P04448	1122015	27	AM_Peak	41.67	43.16
2015	1	2	101P04446	1122015	28	AM_Peak	22.67	22.88
2015	1	2	101P04447	1122015	28	AM_Peak	29.67	33.81
2015	1	2	101P04448	1122015	28	AM_Peak	42.33	43.16

Table 14. MAPTT Calculation Sample

Year	Month Number	SEG Code	TMC	15min Interval	Traffic Period	MAPTT	TT 95Percentile	FFTT
2015	1	2	101P04446	24	AM_Peak	21.6786	23.71	22.88
2015	1	2	101P04447	24	AM_Peak	31.1667	35.63	33.81
2015	1	2	101P04448	24	AM_Peak	38.9762	42.82	43.16
2015	1	2	101P04446	25	AM_Peak	21.369	23.95	22.88
2015	1	2	101P04447	25	AM_Peak	31.869	35.31	33.81
2015	1	2	101P04448	25	AM_Peak	121.107	626.56	43.16
2015	1	2	101P04446	26	AM_Peak	23.5833	29.53	22.88
2015	1	2	101P04447	26	AM_Peak	34.3095	42.29	33.81
2015	1	2	101P04448	26	AM_Peak	39.5476	43.03	43.16
2015	1	2	101P04446	27	AM_Peak	25.8452	30.26	22.88
2015	1	2	101P04447	27	AM_Peak	38.8095	49.54	33.81
2015	1	2	101P04448	27	AM_Peak	39.8571	43.49	43.16
2015	1	2	101P04446	28	AM_Peak	28.9405	47.48	22.88
2015	1	2	101P04447	28	AM_Peak	36.3333	45.84	33.81
2015	1	2	101P04448	28	AM_Peak	39.3929	42.69	43.16

Table 15. TTRI Calculation Sample

Year	Month Number	SEG Code	TMC	15min Interval	Traffic Period	TTI	PTI	BTI
2015	1	2	101P04446	24	AM_Peak	0.95	1.04	9.38
2015	1	2	101P04447	24	AM_Peak	0.92	1.05	14.32
2015	1	2	101P04448	24	AM_Peak	0.90	0.99	9.87
2015	1	2	101P04446	25	AM_Peak	0.93	1.05	12.09
2015	1	2	101P04447	25	AM_Peak	0.94	1.04	10.79
2015	1	2	101P04448	25	AM_Peak	2.81	14.52	417.36
2015	1	2	101P04446	26	AM_Peak	1.03	1.29	25.23
2015	1	2	101P04447	26	AM_Peak	1.01	1.25	23.26
2015	1	2	101P04448	26	AM_Peak	0.92	1.00	8.80
2015	1	2	101P04446	27	AM_Peak	1.13	1.32	17.10
2015	1	2	101P04447	27	AM_Peak	1.15	1.47	27.64
2015	1	2	101P04448	27	AM_Peak	0.92	1.01	9.11
2015	1	2	101P04446	28	AM_Peak	1.26	2.08	64.08
2015	1	2	101P04447	28	AM_Peak	1.07	1.36	26.17
2015	1	2	101P04448	28	AM_Peak	0.91	0.99	8.37

Table 16. STTRI Calculation Sample

Year	Month Number	SEG Code	15min Interval	Traffic Period	STTI	SPTI	SBTI
2015	1	2	24	AM_Peak	0.31	0.34	3.75
2015	1	2	25	AM_Peak	0.58	2.29	62.3
2015	1	2	26	AM_Peak	0.33	0.38	5.82
2015	1	2	27	AM_Peak	0.35	0.41	5.74
2015	1	2	28	AM_Peak	0.35	0.45	9.05

Table 17. Max STTRI Calculation Sample

Month Number	SEG Code	MaxSTTI_AM	MaxSPTI_AM	MaxSBTI_AM	DOC_AM
2	1	1	2.29	62.28	0

The purpose of developing Travel Time Reliability (TTR) indices is to measure the variability of congestion on monthly bases and provide a predictable travel time for travelers to adjust their travel plans accordingly. To better rank and prioritized congested segments, and provide a more comprehensive understanding of the extent and severity of congestion over space and time throughout the year, further quantitative analyses were performed. The 85th Percentile Intensity and Speed Drop measures were developed for this purpose. These two measures are effective performance metrics that benefit policy makers to better assign resources for improving network function to the area needs the most. Definitions of these measures and examples demonstrating their use in the Birmingham case study are presented next.

3.5.6. Congestion Intensity

Congestion Intensity is a two-dimensional measure which accounts for the percentage of congested area in the time-space map. Any time-space map includes two dimensions, i.e., the temporal dimension which is the study period (6:00 AM to 10:00 AM and 3:00 PM to 7:00 PM), and the spatial dimension which is the length of TMCs along with selected segment.

For illustration purposes, Figure 6 shows a sample of time-space map developed for study segment 1. Each cell depicted on the map represents the TTI value in July

2015. The associated range of color that reflects the level of congestion is set by defining different threshold values for TTI as shown on the left-hand side of Figure 6.

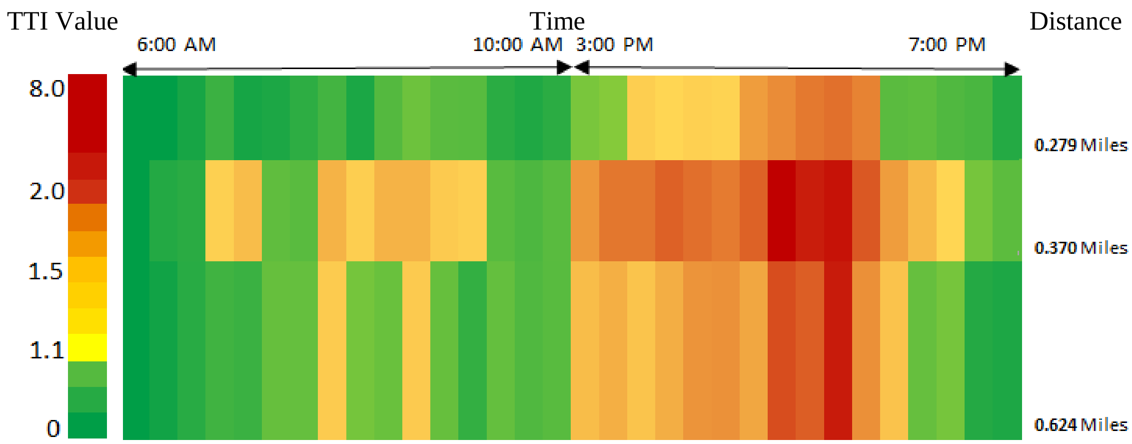

Figure 6. Sample Time-space Map for Segment 01 in July 2015

Figure 7 clearly shows the congested area (over space and time) which encompasses all cells with the TTI value of greater than 1.1.

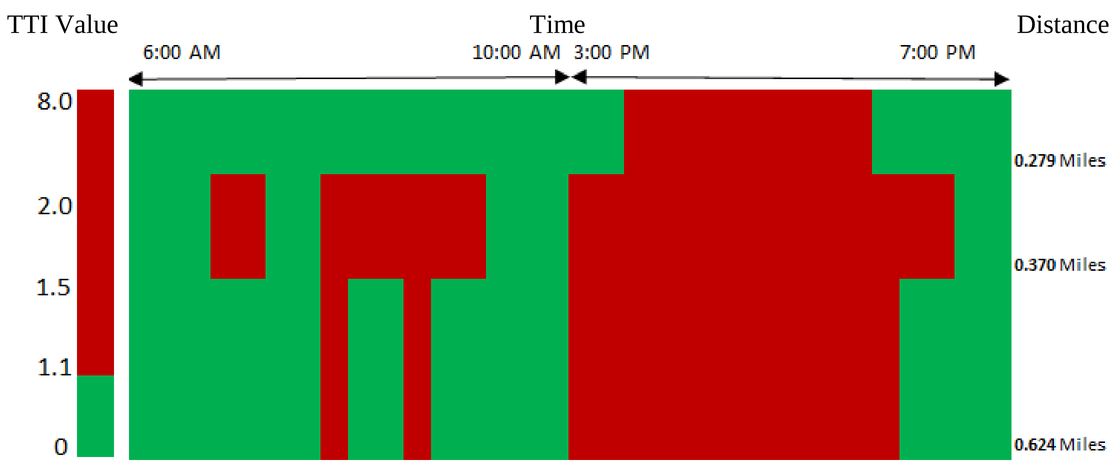

Figure 7. Congested Area for Segment 01 in July 2015

Generating time-space maps in this study provided the information needed to develop Eq. (5) that calculates the daily percentage of Congestion Intensity. Eq. (5) first, multiplies the daily duration of congestion during AM and PM peak to the length of corresponding TMC, then after calculating total congested area divides it by the total area in the space-time map. It is:

$$Congestion\ Intensity_{ij}(\%) = \frac{CongestedArea_{ij}}{TotalArea_i} = \frac{\sum_1^n (DOC_n \times Length_n)_{ij}}{Time \times Length_i} \qquad \textbf{(5)}$$

Where:

i: Segment code, ranging from 1 to 28

j: A work day in 2015 from 1 to 236

n: TMC number along with segment i

DOC: Duration of Congestion in minutes

Time: The Study period (6:00 AM to 10:00 AM and 3:00 PM to 7:00 PM) in minutes

After computing the Congestion Intensity for all work days in 2015, the resulting values range between 0% to 100%. Figure 8 showcases the Congestion Intensity values plotted from the lowest to the highest. The Y axis indicates the percentage of Congestion Intensity and the X axis represents the corresponding work days of year shown as dates (mddyyyy) in Figure 8.

The Congestion Intensity values can be utilized to calculate the 85th Percentile Congestion Intensity that adequately reflects the extent of congestion for the entire year. The 85th Percentile Congestion Intensity is a valuable metric to take into account both the annual variability and reliability. The 85th Percentile of Congestion Intensity simply means that 85 percent of days has a lower value for Congestion Intensity and 25 percent of days has a higher value of Congestion Intensity.

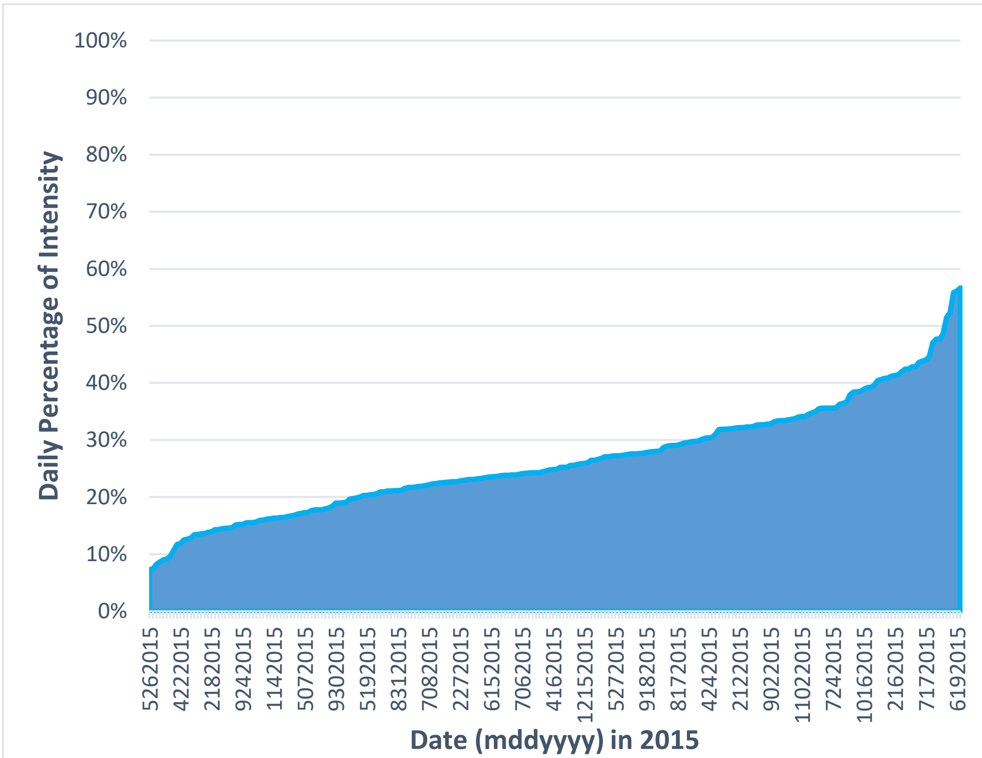

Figure 8. Percentage of Congestion Intensity during all Work Days

3.5.7. Speed-drop

Similar to Congestion Intensity, the Speed-drop is also a two-dimensional measure which accounts for the percentage of deviation from a Cutoff Speed in time-space map. In the case of Speed-drop, each cell in the time-space map represents a reported speed, and when the speed value falls below the Cutoff Speed threshold, the cell is considered as a congested section. Cutoff Speed is the point where the TTI value equals to 1.1 and can be calculated from dividing FFS by 1.1. The daily Speed-drop for each segment can be computed by utilizing Eq. (6). This equation first calculates the percentage of deviation from Cutoff Speed (meaning the difference between the congested speed and the Cutoff speed as a percentage) for each cell. Then, the weighting factor is applied to each cell to get a weighted mean among all congested cells.

$$SpeedDrop_{ij}(\%) = \sum_{1}^{m} \begin{pmatrix} Percentage\ of\ deviation\ from\ Cutoff\ Speed_m \\ \times \\ Weighting\ Factor_m \end{pmatrix}_{ij} \quad \textbf{(6)}$$

$$= \sum_{1}^{m} \left(\frac{(CutoffSP_m - Cng\ SP_m)\times 100}{CutoffSP_m} \times \frac{CellArea_m}{CongestedArea_i} \right)_{ij}$$

Where:

i: Segment code from 1 to 28

j: A work day in 2015 from 1 to 236

m: Number of cells inside the space-time map

Cng SP: Congested Speed

Cutoff SP: Cutoff Speed

CellArea: An area for cell m that is equal to EPOCH * Length of TMC

CongestedArea: Total congested area can be calculated according to the Eq. 5

In the Birmingham case study, calculating the Speed-drop for all work days in 2015 provides a better understanding of the severity of congestion throughout the year. As a sample, Figure 9 showcases the percentage of Speed-drop from the lowest to the highest value for all work days over the course of one year. This is an efficient way of displaying the extent of severity of congestion for the entire year. The resulting values can be utilized to calculate the 85 Percentile of Speed-drop which is another valuable metric related to congestion severity. The 85th Percentile can account for both expected and unexpected circumstances resulting in congestion over the course of the year by focusing on some high values and also avoiding to over-emphasizing the lower values.

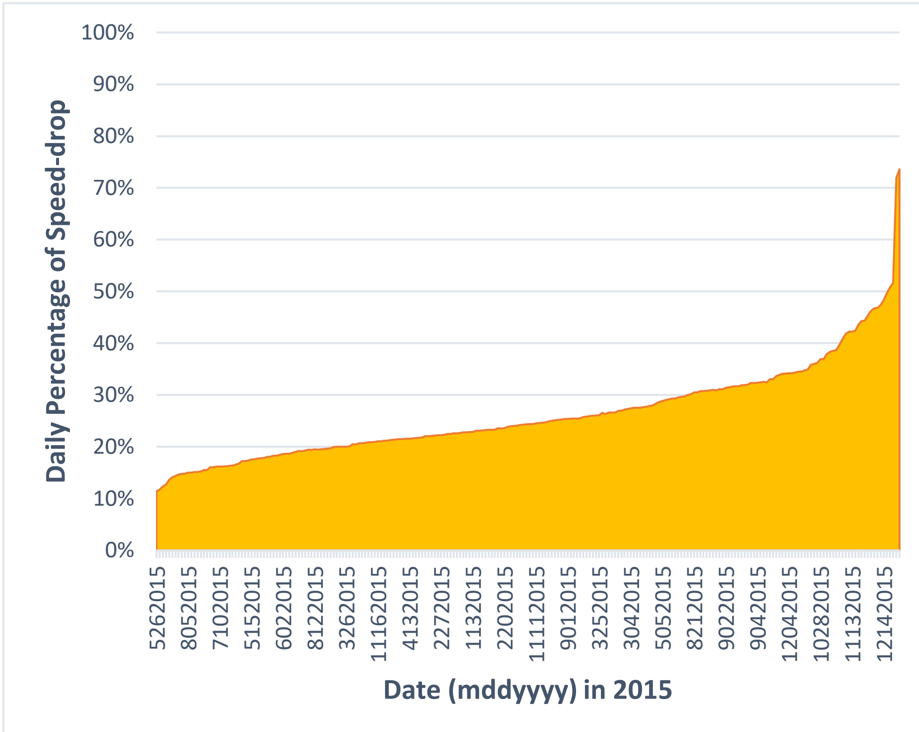

Figure 9. Percentage of Speed-drop during all Work Days

3.5.8. Impact Factor (IF)

Impact factor (IF) is a metric introduced in order to capture the combined effect of both severity and extent of congestion throughout the year. It combines two measures, namely Congestion Intensity and Speed-drop, by multiplying their values for the corresponding day of the year and then computing the 85th percentile for the result values (Eq. 7). Developing Impact Factor is a robust method to identify segments that experience long-lasting and severe congestion throughout the year.

$$IF_i = 85th\ Percentile\ of\ \left\{Congestion\ Intensity_j \times Speed\ drop_j\right\} \qquad \textbf{(7)}$$

Where:

i: Segment code from 1 to 28

j: A work day in 2015 from 1 to 236

CHAPTER 4: ANALYSIS AND RESULTS

4.1. Introduction

Understanding the NPMRDS data characteristics along with their limitations is of critical importance prior to developing performance measures. For this reason, the following subsections first investigate the accuracy, consistency, and competence of NPMRDS dataset. Next, they detail the factors that affect data measurements and discuss ways to address issues related to the data gaps and anomalies. Using the Birmingham testbed, a range of performance measures are computed for identifying the travel time uncertainty on a monthly basis and an efficient method for prioritizing freeway segments is presented.

4.2. Data Gaps

As explained earlier, NPMRDS is a form of vehicle probe-based travel time data set. Observation of at least one passenger car or freight truck is required for data to be reported and any 5-minute period (epoch) with no observations has been removed from the dataset. Therefore, the number of reported epochs can represent the availability of probe points during reporting window, and the missing epochs can show the data gaps in the dataset. By comparing the total number of records for each month in figure 10, it was evident that the total number of probe points increased from January 2015 to December 2015, but still, the availability of data during reporting window can fluctuate noticeably.

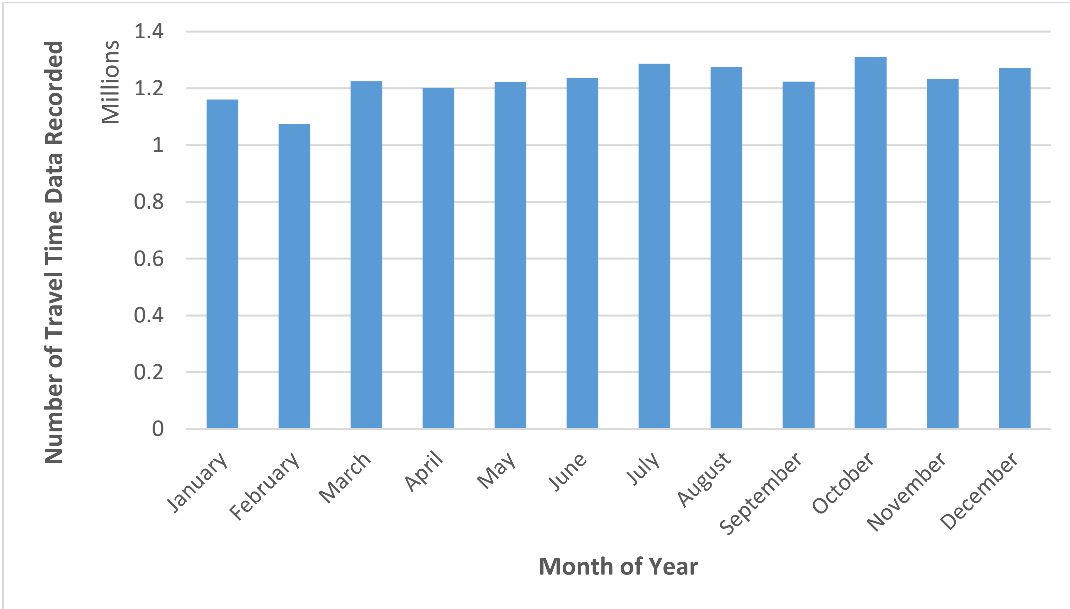

Figure 10. Number of Travel Time Data Recorded per Month in 2015

Time of day, day of week and location of TMC are key factors in the availability of probe points. For instance, the probe points are more expected to be available when the TMC located in urban areas and during peak hours on weekdays. Figure 11 and figure 12 clearly show variance in the average total number of epochs recorded per day during October 2015. The analysis of the Birmingham data revealed that the number of epochs during weekdays were greater than weekends. The mean value of recorded epochs was 232.31 with a standard deviation of 18.85.

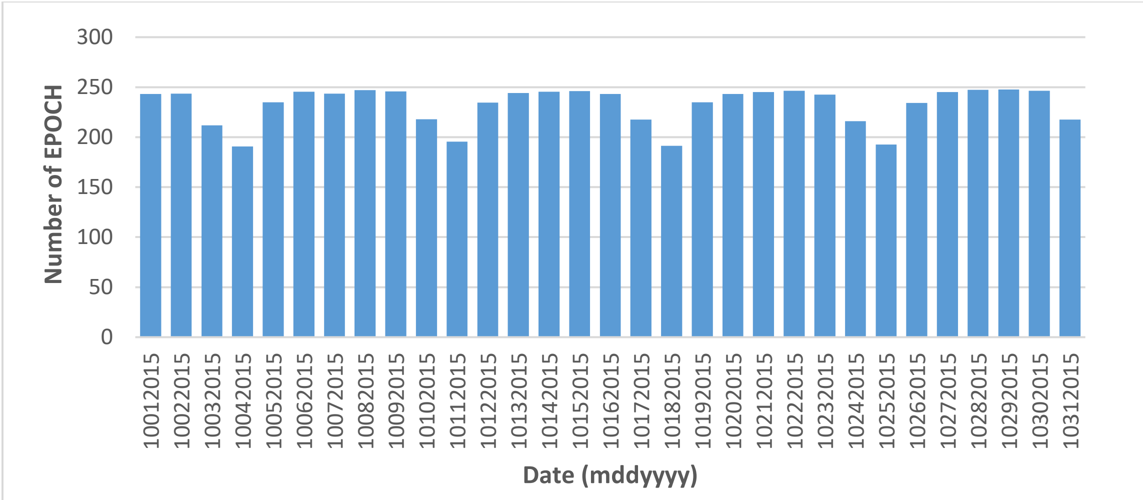

Figure 11. Daily Average Total Number of Epochs Recorded during October 2015

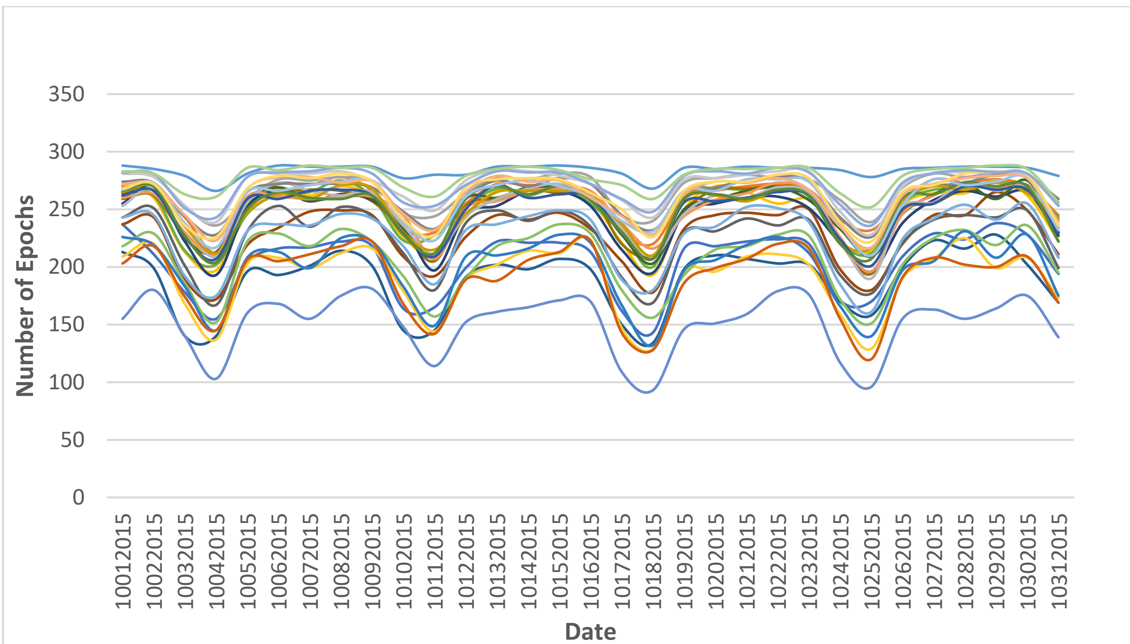

Figure 12. Daily Average Total Number of Epochs per TMC in October 2015

As aforementioned, the availably of probes is also varied depending on the time of day. Figure 13 details the group of TMCs during the peak hours from 6 AM to 7 PM that generated the highest number of epoch over 31 days.

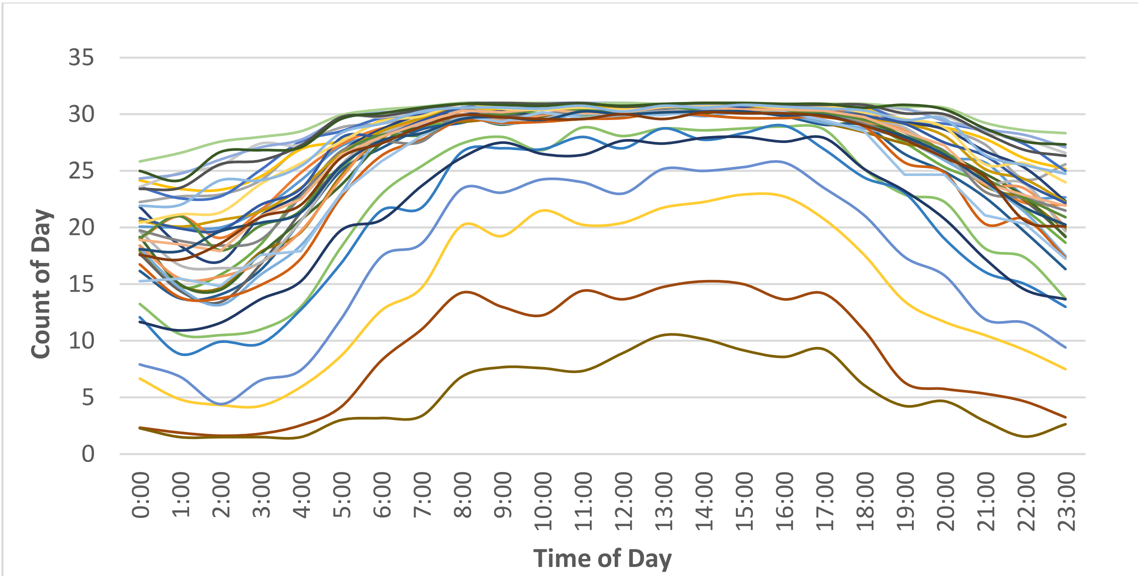

Figure 13. Availability of Epoch per Time of Day per TMC during October 2015

4.3. Data Anomalies

Data anomalies in the database can be defined as illegitimate records resulting from external procedures like inserting, updating and deleting the data. NPMRDS dataset reports the travel time data in seconds as an integer number meaning every record for travel time has been rounded up to the nearest second. This procedure can generate anomalous entries where the length of TMC is short.

Figure 14 illustrates the variability of TMC lengths on interstate roadways in the case study. TMC segments with the length of below 1 mile formed 36 percent of TMCs where 15 percent of them are even shorter than 0.1 miles. This implies that the time granularity of one second might not be precise enough to report the actual speed and it generates anomalous entries where the length of TMC is short.

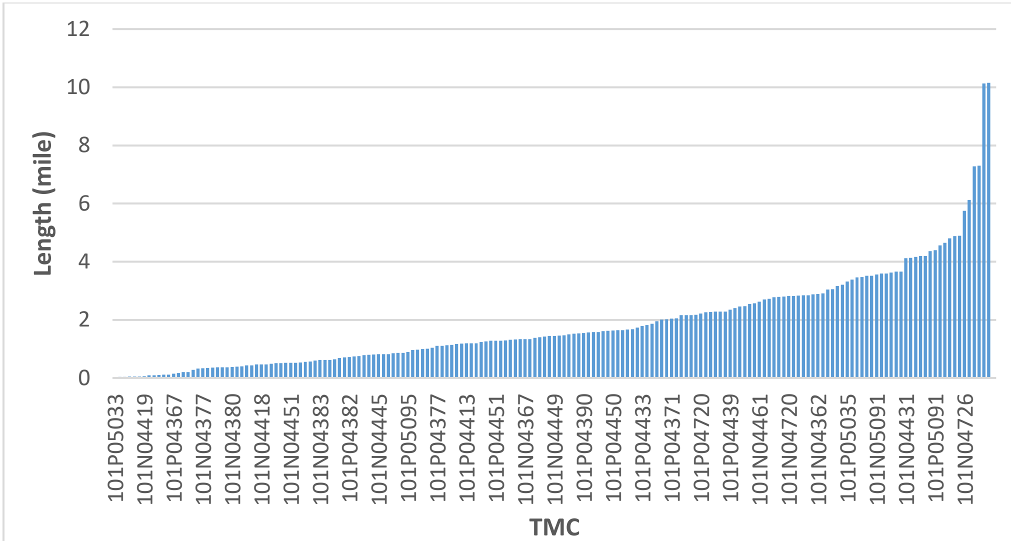

Figure 14. Variability of TMC Lengths

These anomalous entries can be observed by plotting speed calculated from reported travel time in NPMRDS. For example, by plotting TMC "101P05033" speed, it became apparent that only three speeds (105, 52, and 38 miles/hour) were reported during October 2015 as Figure 15 shows. This implies that all the speeds between 105 to 52, and 52 to 38 have been rounded up to the 105 and 52 accordingly as a result of time granularity of one second.

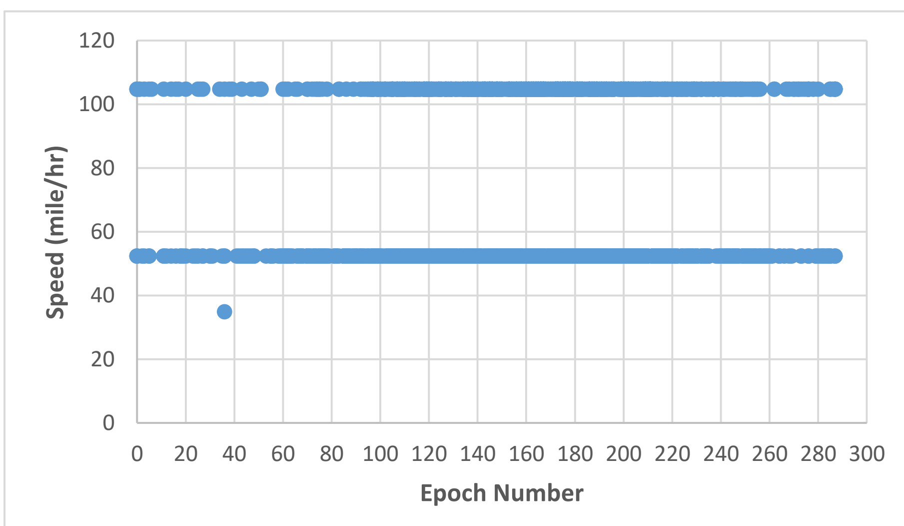

Figure 15. Speed Scattered Plot per Epoch for TMC 101P05033

To investigate the impact of time granularity of one second, the interaction between the actual possible speed and travel time with the granularity of one-second window was drawn. The length of TMC was assumed to be 0.029 miles equal to TMC "101P05033" length. Figure 16 implies that the time granularity of one-second can result in rounding the wide range of actual speed values to the lower number which represents the same travel time in the data set. Particularly for the TMC "101P05033" example, travel time of 2 seconds represents the speeds of 55 miles/hour but, in fact, the actual speed can be a value from 105 miles/hour to 55 miles/hour, as shown in Figure 14.

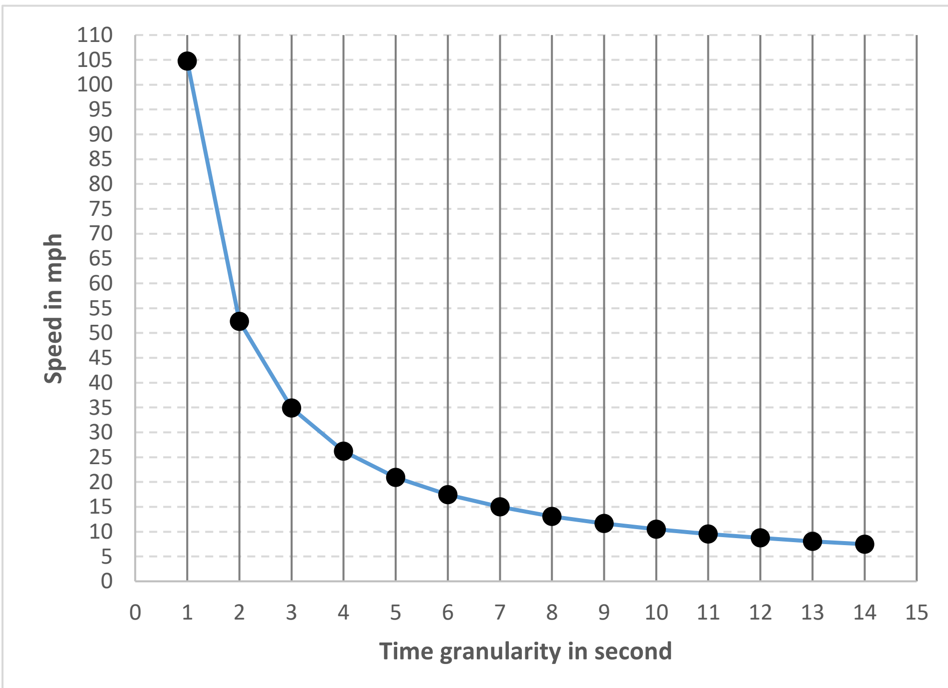

Figure 16. Interaction between Speed and one Second Time Granularity for TMC "101P05033"

Therefore, there is a need for more precise time granularity to be able to report the actual recorded speed and avoid the quantization error while transferring the speed to the travel time in integer seconds.

To find out the percentage of the anomalous entries in the dataset, Eq. (8) is utilized. Eq. (8) calculates the maximum amount of error in the reported speed by NPMRDS, and then the frequency of the anomalies was counted. The desired erroneous speed is set to be equal to 5 miles/hour.

$$Error = \frac{TMC\ Length}{Reported\ Travel\ time - 1} - \frac{TMC\ Length}{Reported\ Travel\ time} \qquad \textbf{(8)}$$

Table 18 shows the percentage of anomalies in the dataset used in the Birmingham case study. The analysis of the Birmingham data for 2015 indicates that a significant amount of anomalous entries had speeds above 60 miles/hour. It was also observed that there were no anomalies in the group of TMCs with the length of greater than 1 mile. At lower speeds (speeds of 50 mph or less), the percentage of anomalies were less than 7 percent which means data records were barely affected. Overall, it was concluded that NPMRDS is adequate enough for the purpose of this study which is congestion detection. It should be mentioned that 0.05 percent EPOCHs with travel time equal to 1 second in the dataset were excluded from the erroneous check. They were all in the group of TMCs with speed above 60 miles/ hour.

Table 18. The Percentage of Anomalies in 2015

	TMC Length (Miles)		
Reported Speed (Miles/hour)	0 to 0.1	0.1 to 1	Above 1
0 to 40	1%	0%	0%
40 to 50	7%	3%	0%
50 to 60	16%	24%	0%
Above 60	32%	45%	0%

4.4. Dataset Validation

Based on documentation provided with NPMRDS data set, data quality analyses have been performed. NPMRDS data were compared to a variety of multiple travel time data sources to ensure the error tolerances is acceptable. Also, Data Quality Validation reports were prepared quarterly. Reference data used to validate the HERE data set includes Weigh-In-Motion (WIM) Data, Microwave and Inductive Loop Data, and Bluetooth Travel Time Data. Average Absolute Speed Error (AASE) and Standard Error of the Mean (SEM) were used to evaluate speed accuracy and error in four-speed ranges (0-30 MPH, 30-45 MPH, 45-60 MPH, and ≥ 60 MPH) in each time period. In the Birmingham case study, statistics of speed bin with data size less than five records in a time period were excluded due to relatively small sample size.

4.5. Travel Time Isochrones

Travel time isochrones were generated in an attempt to visualize the expected average travel time during AM peak periods to downtown along the study corridors. Figure 17 is an example of travel time isochrones generated for the Birmingham region using AM Peak data from April 2015. The red point represents the downtown of Birmingham and the polygons that display the travel times divided the region into 5 different rings. Travelers in the blue area can reach downtown within 10 minutes, in gray within 20 minutes, in yellow within 30 minutes, in orange within 40 minutes, and in red within 50 minutes.

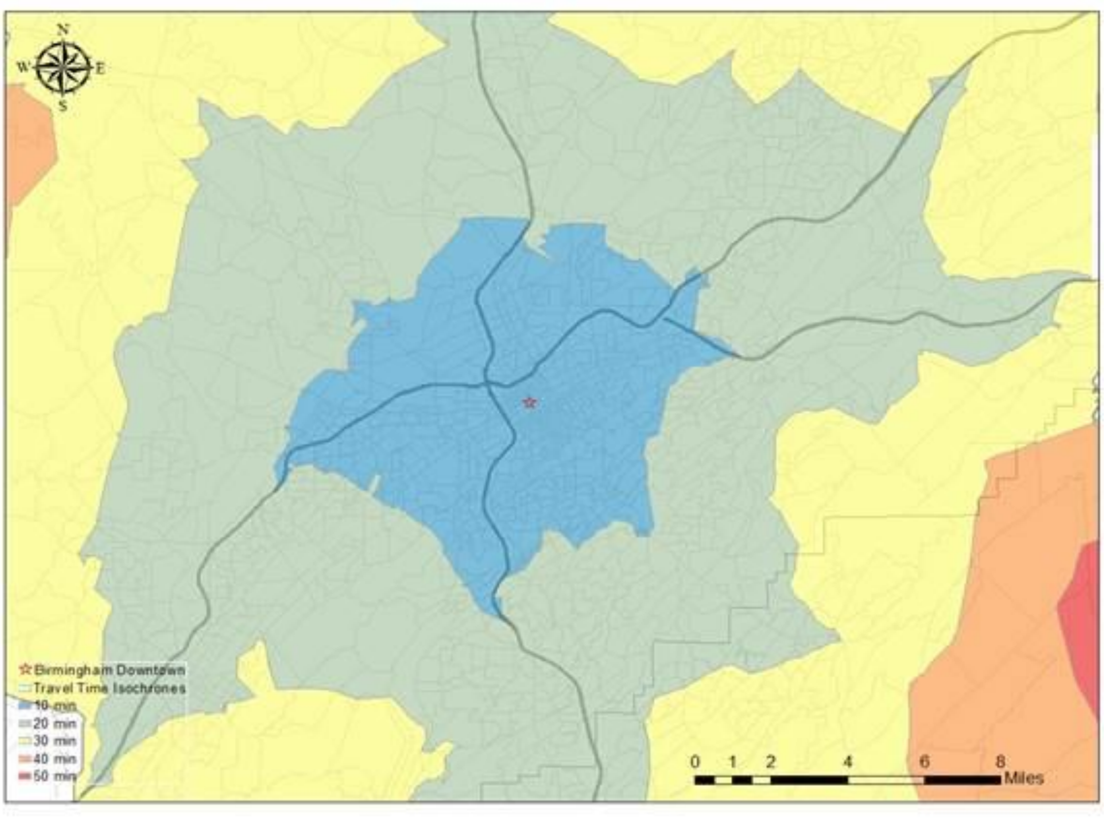

Figure 17. Travel Time Isochrones during AM Peak in April 2015

4.6. Birmingham Case Study Travel Time Reliability Indices

To quantify the reliability of the main corridors on the study area transportation system, Travel Time Reliability measures were generated. As stated earlier, the Travel Time Index (TTI) compares the average travel time to the travel time in free flow condition. On the other hand, Planning Time Index (PTI) is reliability measure that shows the ratio of the 95th percent peak period travel time to the free flow travel time. Buffer Time Index (BTI) calculates the extra time expresses the amount of extra time needed to be on-time 95 percent of the time. In the other word, BTI can be defined as the percentage of the difference between travel time estimated by using TTI and PTI. PTI and BTI both are accounting for the unexpected delay. Therefore, it is expected that such indices are significantly affected by the presence of outliers like crashes, bad weathers, and other unexpected events.

4.6.1. Travel Time Index (TTI)

Based on the analysis of 2015 NPMRDS data for the four Birmingham study sections, TTI values were calculated and summarized in Table 19. Using the TTI threshold values as introduced in Section 3.5.1, Table 19 revealed the variability of travel time experienced by commuters with the range of colors associated with the value of Travel Time Index with green representing best and red representing worst conditions. The lower that the value of TTI is, the closer the travel condition is to free flow travel time.

Table 19. Max TTI Values during AM Peak and PM Peak per Month in 2015

Segment Code	Free Flow Travel Time	Maximum Travel Time Index in AM Peak												Maximum Travel Time Index in PM Peak											
		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	91.63	2.66	1.51	1.43	1.26	1.48	1.30	1.16	1.27	1.24	1.31	1.32	1.19	1.57	1.61	1.54	1.35	1.43	1.81	1.82	1.62	2.08	2.02	2.30	2.59
2	99.86	2.08	1.42	1.57	1.49	1.42	1.83	1.39	1.52	1.75	1.66	1.53	1.45	1.11	2.04	1.75	1.88	1.53	1.42	1.36	1.31	1.37	1.68	2.06	4.91
3	215.78	2.68	1.50	2.17	1.96	2.22	2.19	1.67	2.03	2.20	2.72	1.99	1.99	1.27	1.08	1.53	1.25	1.00	1.15	1.19	1.23	1.30	1.47	2.20	2.49
4	229.40	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.12	1.22	1.26	1.35	1.14	1.15	1.18	1.15	1.27	1.34	1.73	1.48
5	272.86	1.12	1.11	1.10	1.47	1.07	1.08	1.07	1.08	1.07	1.09	1.09	1.08	1.11	1.10	1.08	1.11	1.09	1.07	1.58	1.22	1.07	1.07	1.08	1.10
6	353.41	1.00	1.14	1.00	1.00	1.58	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.40	1.00
7	374.04	1.00	1.09	1.12	1.00	1.03	1.04	1.00	1.03	1.00	1.08	1.33	1.00	1.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.09	1.00
8	307.44	1.12	1.13	1.11	1.11	1.08	1.12	1.06	1.08	1.09	1.09	1.09	1.09	1.10	1.12	1.10	1.08	1.15	1.11	1.06	1.09	1.06	1.08	1.10	1.11
9	330.75	1.10	1.10	3.58	1.10	1.07	1.08	1.06	1.08	1.07	1.46	1.08	1.48	1.13	1.13	1.13	1.18	1.09	1.09	1.08	1.12	1.10	1.10	1.25	1.10
10	404.67	1.96	1.24	1.57	1.48	1.38	1.87	1.21	1.44	1.65	1.51	1.19	1.79	1.33	1.20	1.53	1.08	1.08	1.00	1.18	1.10	1.19	1.08	1.70	1.49
11	362.52	1.06	1.04	1.12	1.03	1.03	1.01	1.01	1.02	1.02	1.03	1.03	1.02	1.42	1.08	1.02	1.32	1.01	1.01	1.10	1.07	1.06	1.05	1.08	1.06
12	396.83	1.85	1.69	1.30	1.28	1.32	1.21	1.03	1.29	1.26	1.74	1.26	1.36	1.06	1.82	1.25	1.10	1.05	1.04	1.08	1.07	1.05	1.03	1.19	1.08
13	438.00	1.00	1.08	1.12	1.01	1.00	1.01	1.06	1.00	1.06	1.00	1.03	1.01	1.27	1.23	1.15	1.12	1.07	1.31	1.08	1.21	1.75	1.93	1.31	1.53
14	422.20	1.54	1.67	1.33	1.74	1.69	1.26	1.55	1.19	1.26	1.13	1.29	1.05	1.64	1.34	2.28	1.48	1.35	1.48	1.31	1.49	1.35	1.67	2.16	2.62
15	429.53	1.34	1.09	1.08	1.07	1.07	1.06	1.07	1.07	1.08	1.09	1.09	1.09	1.08	1.95	1.06	1.05	1.05	1.39	1.04	1.04	1.05	1.06	1.07	1.07
16	457.25	1.31	1.09	1.07	1.07	1.06	1.05	1.04	1.05	1.06	1.06	1.08	1.07	1.09	1.43	1.13	1.07	1.08	1.27	1.16	1.06	1.06	1.06	1.09	1.12
17	578.57	2.14	1.86	1.98	1.88	1.66	1.76	1.63	1.89	2.09	2.22	1.98	1.88	1.28	1.85	1.92	1.77	1.60	1.39	1.50	1.71	1.47	1.40	1.85	2.34
18	513.90	1.86	1.45	1.72	2.00	1.92	1.36	1.24	2.09	2.41	2.18	2.81	2.35	1.09	1.10	1.23	1.37	1.26	1.68	1.49	1.28	1.21	1.34	1.28	1.18
19	516.34	1.08	1.30	1.08	1.06	1.06	1.06	1.10	1.06	1.06	1.07	1.12	1.06	1.12	1.22	1.82	1.41	1.21	1.13	1.86	1.15	1.07	1.15	1.23	1.15
20	537.81	1.11	1.10	1.11	1.35	1.08	1.07	1.08	1.09	1.23	1.26	1.11	1.09	1.39	1.32	1.42	1.33	1.55	1.37	1.32	1.29	1.31	1.43	1.60	1.53
21	648.61	1.12	1.11	1.10	1.06	1.19	1.03	1.10	1.03	1.09	1.11	1.05	1.19	2.23	1.83	2.12	2.23	2.41	1.99	2.06	2.02	2.05	2.14	2.48	2.47
22	560.52	1.77	1.42	1.47	1.74	1.39	1.34	1.17	1.48	1.74	1.75	1.84	1.66	1.12	1.12	1.10	1.10	1.06	1.08	1.61	1.08	1.52	1.81	1.54	1.20
23	595.62	1.07	1.08	1.07	1.06	1.06	1.05	1.04	1.06	1.06	1.06	1.17	1.06	1.71	1.70	1.89	1.62	2.11	1.90	2.05	1.76	1.96	2.11	2.45	1.88
24	636.38	1.62	1.51	1.74	1.66	1.47	1.29	1.17	1.67	1.92	2.07	1.83	2.03	1.14	1.08	1.20	1.39	1.13	1.10	1.29	1.07	1.43	1.18	1.12	1.29
25	622.01	1.40	1.15	1.14	1.20	1.17	1.11	1.08	1.20	1.38	1.28	1.24	1.31	2.36	1.14	1.19	1.10	1.18	1.17	1.11	1.15	1.10	1.10	1.17	1.20
26	655.23	1.12	1.19	1.11	1.12	1.09	1.11	1.09	1.22	1.11	1.20	1.12	1.21	1.27	1.19	1.19	1.35	1.25	1.20	1.12	1.43	1.25	1.15	1.45	1.29
27	790.33	1.30	1.28	1.42	1.42	1.17	1.27	1.28	1.24	1.36	1.42	1.68	1.43	1.57	1.40	1.25	1.15	1.19	1.65	1.27	1.08	1.14	1.10	1.13	1.15
28	836.01	1.35	1.45	1.14	1.33	1.18	1.30	1.15	1.28	1.22	1.21	1.44	1.24	1.19	1.54	1.39	1.28	1.21	1.12	1.18	1.38	1.20	1.21	1.31	1.57

The standard deviation also was computed to help understand the variability of values during AM and PM peak periods (Figure 18). By considering the TTI values and associated standard deviation values for each segment, it can be concluded that segment 8 is the most reliable segment in the study area due to having the lowest values for TTI and standard deviation. Also, segments 4 and 7 are the most reliable segments during AM peak and PM peak respectively. On the other hand, the highest values of TTI and standard deviation were obtained for segments 2 and 9 that can be a result of the unexpected delay. The worst TTI value obtained was 4.91 for segment 2 under PM peak conditions in December. A TTI value of 4.91 means that the maximum average travel time during PM in the month of December in this segment is almost five times greater than the free flow travel time or equal to 490.304 seconds corresponding to a speed of just 10.21 miles/hour.

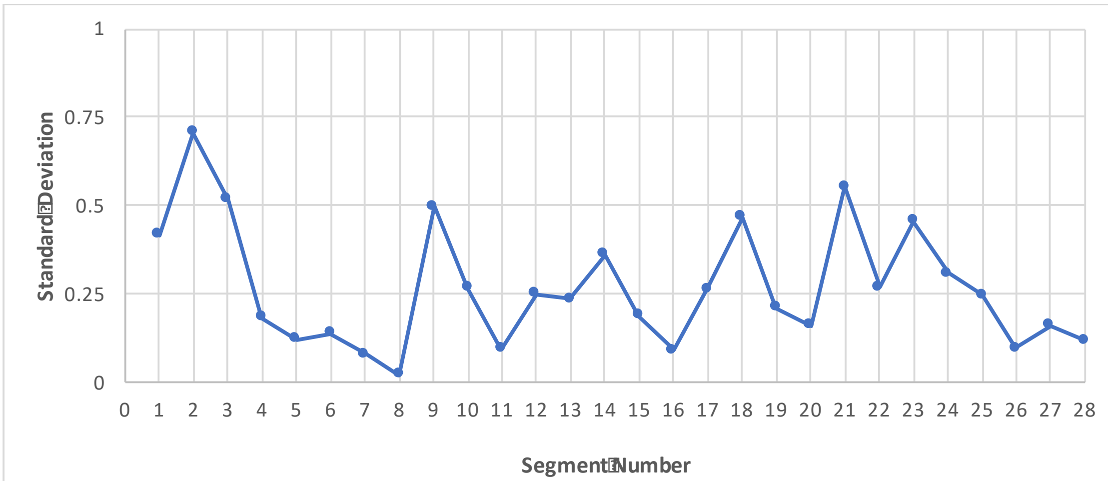

Figure 18. Standard Deviation of TTI per segment in 2015

Furthermore, visual maps were developed by GIS software and used to display the congestion status, areas of congestion, and severity along the study corridors. For demonstration purposes, Figures 19 and 20 show the TTI values for January 2015 during AM peak and PM peak periods. Defined threshold coded the congestion level as Little/None, Moderate, Significant, and Severe.

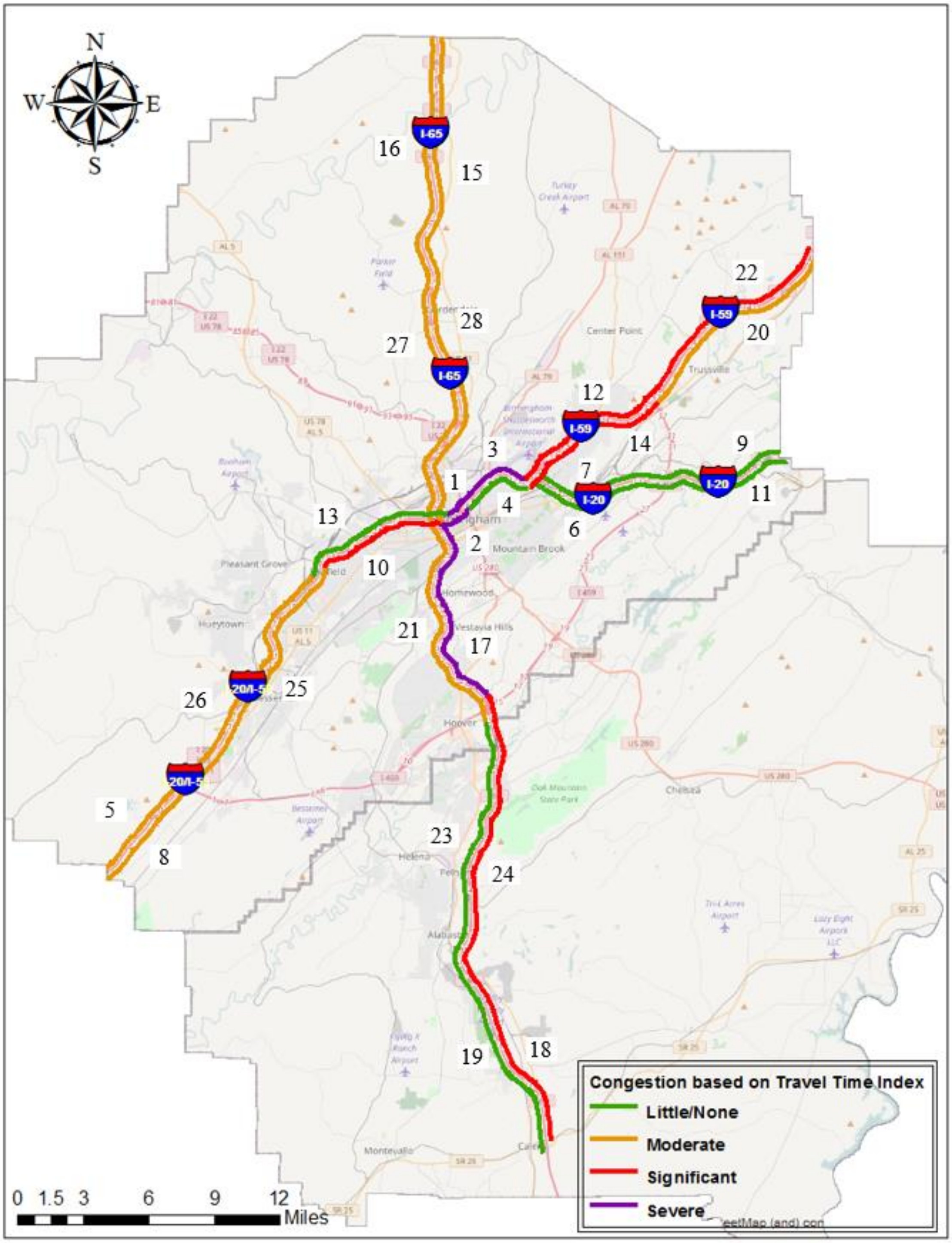

Figure 19. Travel Time Index during AM Peak in January 2015

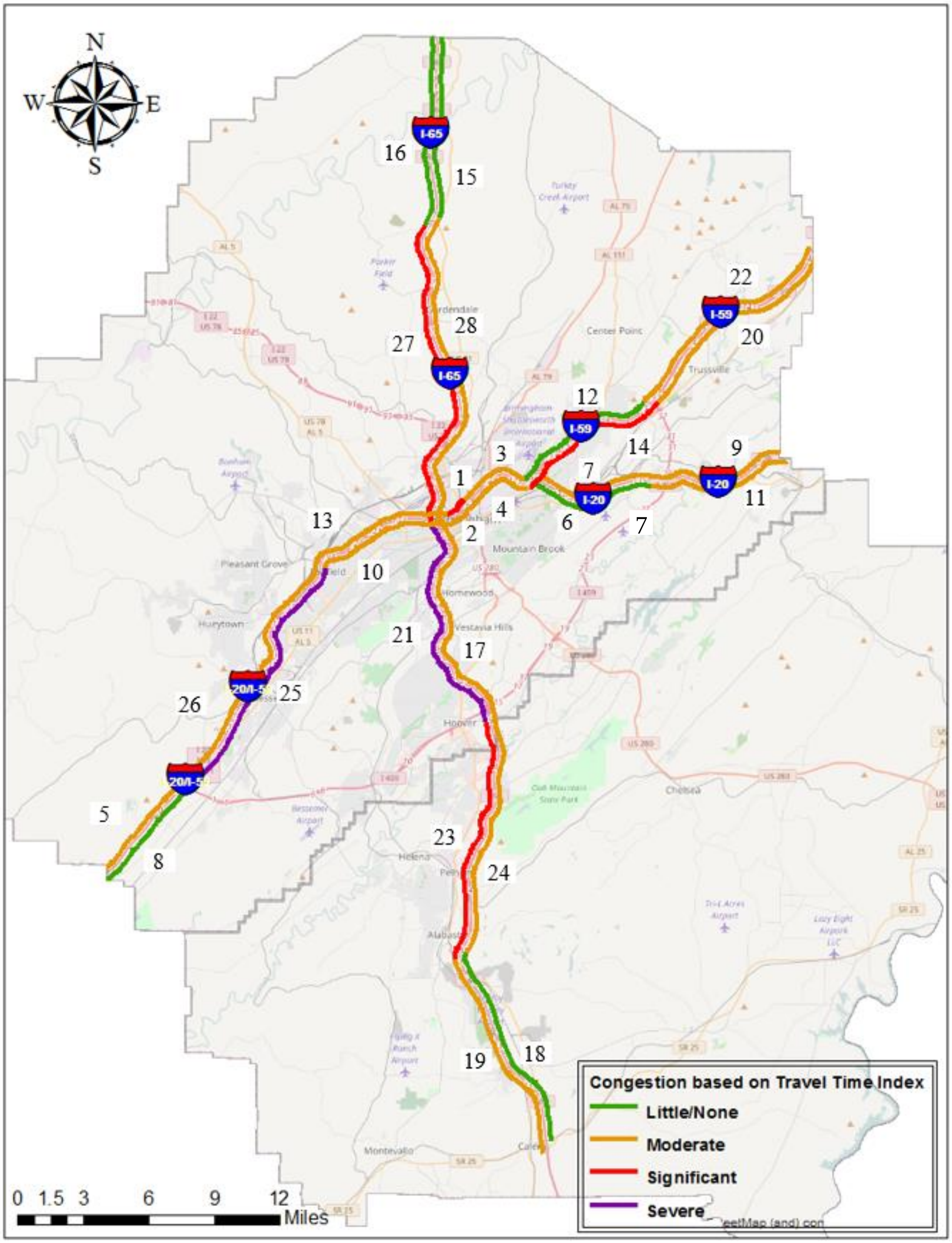

Figure 20. Travel Time Index during PM Peak in January 2015

The results of the analysis show that the level of congestion on the study segments that provide primary access to the Birmingham downtown area highly depends on the time of day. For instance, segments 17 and 21 are part of I65 from I459 to I20/59. Segment 17 representing the northbound and section 21 the southbound direction. During the AM peak, severe congestion occurred in segment 17 that has the direction of travel toward the downtown whereas segment 21 showed the moderate congestion. During the PM peak, the most significant congestion occurred in the opposite direction along segment 21 that travels from the downtown and outwards.

4.6.2. Duration of Congestion (DOC)

Travel Time Index values were used to measure the duration of congestion and findings for all study segments by month were summarized in Table 20. Also, Figures 21 and 22 were developed to help visualize the duration of congestion on the study corridors during the AM and PM peak periods in January 2015.

It can be seen that congestion is persistent, continuing for more than 1 hour during the peak periods. Congestion was also found to be persistent along I- 20/59 in downtown Birmingham. It should be noted that a high value for TTI does not necessary accompany a high value for congestion duration, since the congestion durations represent the persistence of congestion during peak hours and TTI shows the worst congested 15 minutes during peak hours. For instance, as shown in Table 20, segment 20 with TTI around 1.2 is moderately congested during AM peak in September and October 2015 but the duration of congestion is 240 minutes. This implies that commuters using segment 20 any time from 6:00 AM to 10:00 AM should adjust their travel plans as travel is expected to take almost twice of the amount of travel time in ideal condition. On the other hand, segment 2 with TTI equal to 2.08 experienced severe congestion in January 2015 but with duration of 90 minutes (Table 20).

Segment Code	Congestion Duration in AM Peak (minute)												Congestion Duration in PM Peak (minute)											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	**165**	**75**	**135**	**45**	**45**	**45**	**30**	**75**	**60**	**90**	**90**	**30**	**105**	**165**	**180**	**165**	**150**	**210**	**180**	**150**	**195**	**210**	**195**	**240**
2	**90**	**90**	**135**	**120**	**105**	**105**	**60**	**105**	**90**	**135**	**105**	**105**	**0**	**195**	**105**	**150**	**60**	**60**	**120**	**135**	**135**	**195**	**195**	**180**
3	**150**	**75**	**150**	**90**	**105**	**75**	**60**	**90**	**120**	**135**	**165**	**105**	**75**	**0**	**90**	**30**	**0**	**15**	**30**	**45**	**45**	**75**	**240**	**195**
4	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**15**	**45**	**60**	**60**	**15**	**15**	**30**	**15**	**15**	**90**	**135**	**90**
5	**15**	**30**	**15**	**105**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**30**	**15**	**0**	**15**	**0**	**0**	**45**	**15**	**0**	**0**	**0**	**0**
6	**0**	**0**	**0**	**0**	**15**	**15**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**15**	**0**
7	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**60**	**0**	**30**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
8	**90**	**45**	**15**	**15**	**0**	**15**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**60**	**0**	**0**	**60**	**15**	**0**	**0**	**0**	**0**	**30**	**15**
9	**75**	**15**	**195**	**30**	**0**	**0**	**0**	**0**	**0**	**105**	**0**	**120**	**135**	**135**	**30**	**120**	**15**	**0**	**0**	**15**	**15**	**45**	**90**	**60**
10	**75**	**60**	**120**	**105**	**105**	**105**	**45**	**75**	**90**	**105**	**90**	**180**	**135**	**60**	**75**	**15**	**0**	**0**	**45**	**15**	**135**	**0**	**165**	**165**
11	**0**	**0**	**15**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**45**	**0**	**0**	**120**	**0**	**0**	**15**	**0**	**0**	**0**	**0**	**0**
12	**105**	**60**	**45**	**60**	**60**	**45**	**0**	**75**	**90**	**105**	**45**	**45**	**0**	**15**	**30**	**15**	**0**	**0**	**0**	**15**	**15**	**0**	**60**	**0**
13	**0**	**0**	**45**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**60**	**30**	**120**	**15**	**0**	**30**	**0**	**60**	**150**	**105**	**135**	**225**
14	**30**	**45**	**60**	**45**	**45**	**15**	**15**	**15**	**90**	**165**	**45**	**0**	**120**	**180**	**210**	**135**	**135**	**105**	**135**	**105**	**135**	**180**	**150**	**135**
15	**45**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**120**	**0**	**0**	**0**	**165**	**0**	**0**	**0**	**0**	**0**	**0**
16	**60**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**135**	**0**	**0**	**0**	**15**	**15**	**0**	**0**	**0**	**0**	**15**
17	**195**	**150**	**210**	**165**	**105**	**120**	**120**	**165**	**210**	**165**	**165**	**135**	**135**	**210**	**225**	**210**	**180**	**210**	**210**	**135**	**105**	**180**	**225**	**210**
18	**165**	**135**	**135**	**150**	**120**	**135**	**135**	**240**	**165**	**150**	**180**	**180**	**0**	**30**	**135**	**210**	**120**	**195**	**195**	**90**	**90**	**210**	**180**	**60**
19	**0**	**90**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**30**	**0**	**60**	**90**	**210**	**120**	**90**	**15**	**135**	**120**	**0**	**90**	**150**	**90**
20	**45**	**60**	**30**	**30**	**0**	**0**	**0**	**0**	**240**	**240**	**15**	**0**	**180**	**225**	**165**	**135**	**135**	**120**	**135**	**135**	**195**	**135**	**150**	**135**
21	**0**	**0**	**0**	**0**	**30**	**0**	**0**	**0**	**0**	**0**	**0**	**15**	**165**	**195**	**210**	**195**	**195**	**150**	**195**	**165**	**180**	**180**	**180**	**195**
22	**150**	**105**	**90**	**105**	**105**	**135**	**30**	**90**	**150**	**120**	**120**	**120**	**90**	**90**	**30**	**15**	**0**	**0**	**240**	**0**	**240**	**240**	**75**	**15**
23	**0**	**15**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**30**	**0**	**195**	**240**	**240**	**210**	**240**	**225**	**240**	**195**	**240**	**240**	**240**	**225**
24	**120**	**120**	**150**	**165**	**135**	**90**	**60**	**135**	**135**	**135**	**165**	**180**	**15**	**0**	**75**	**135**	**45**	**0**	**180**	**0**	**105**	**60**	**15**	**195**
25	**150**	**150**	**120**	**105**	**105**	**0**	**0**	**45**	**120**	**135**	**90**	**120**	**180**	**150**	**135**	**15**	**75**	**90**	**30**	**60**	**15**	**30**	**165**	**135**
26	**195**	**120**	**135**	**135**	**30**	**15**	**0**	**60**	**45**	**120**	**150**	**150**	**240**	**240**	**210**	**240**	**225**	**150**	**75**	**165**	**165**	**135**	**225**	**240**
27	**165**	**150**	**165**	**75**	**45**	**135**	**60**	**75**	**105**	**120**	**180**	**120**	**75**	**195**	**30**	**30**	**60**	**60**	**60**	**0**	**45**	**15**	**75**	**120**
28	**135**	**90**	**135**	**135**	**60**	**45**	**45**	**90**	**75**	**165**	**210**	**165**	**135**	**240**	**105**	**180**	**135**	**0**	**90**	**120**	**30**	**180**	**150**	**180**

Table 20. Congestion Duration during Peak Periods in 2015

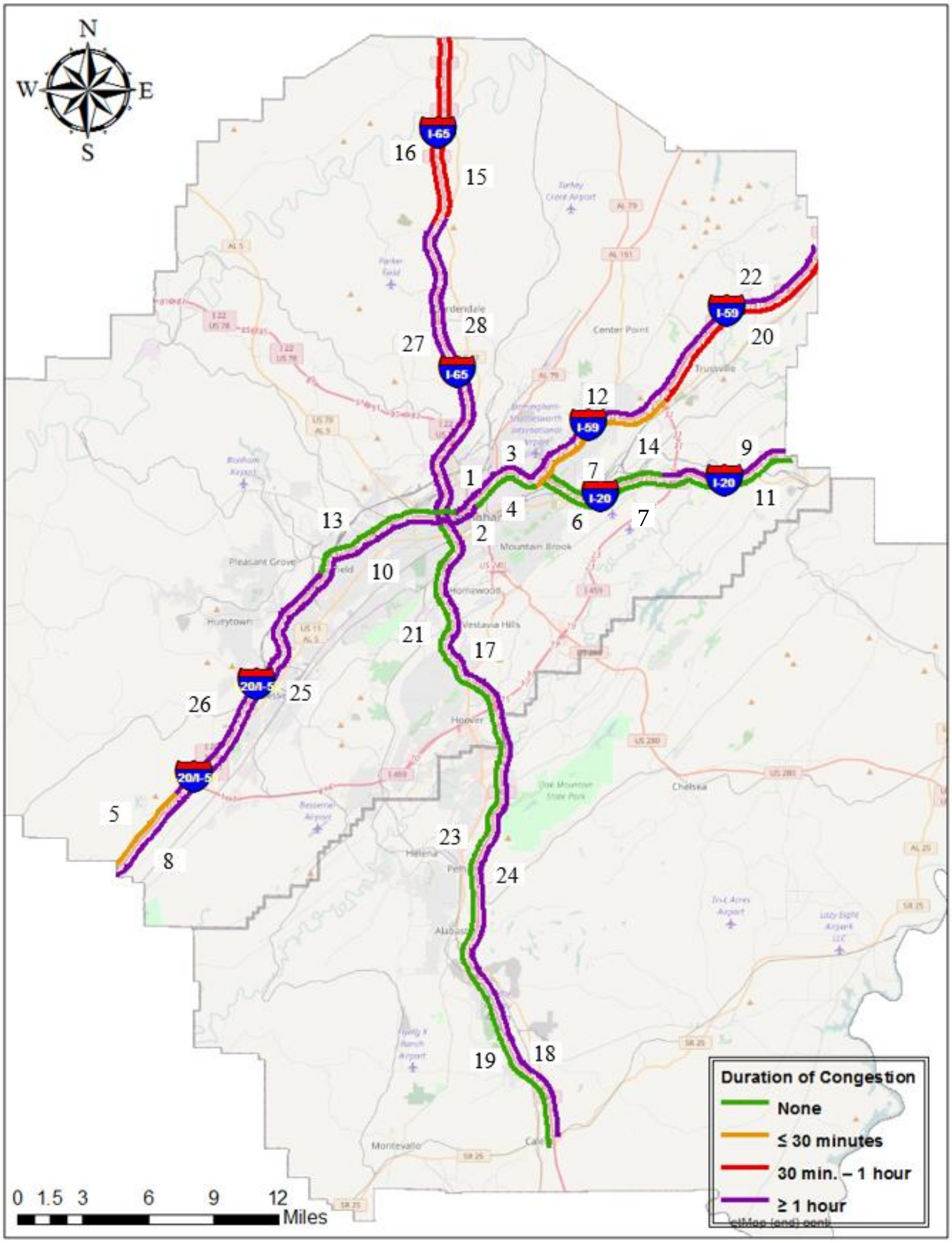

Figure 21. Duration of Congestion during AM Peak in January 2015

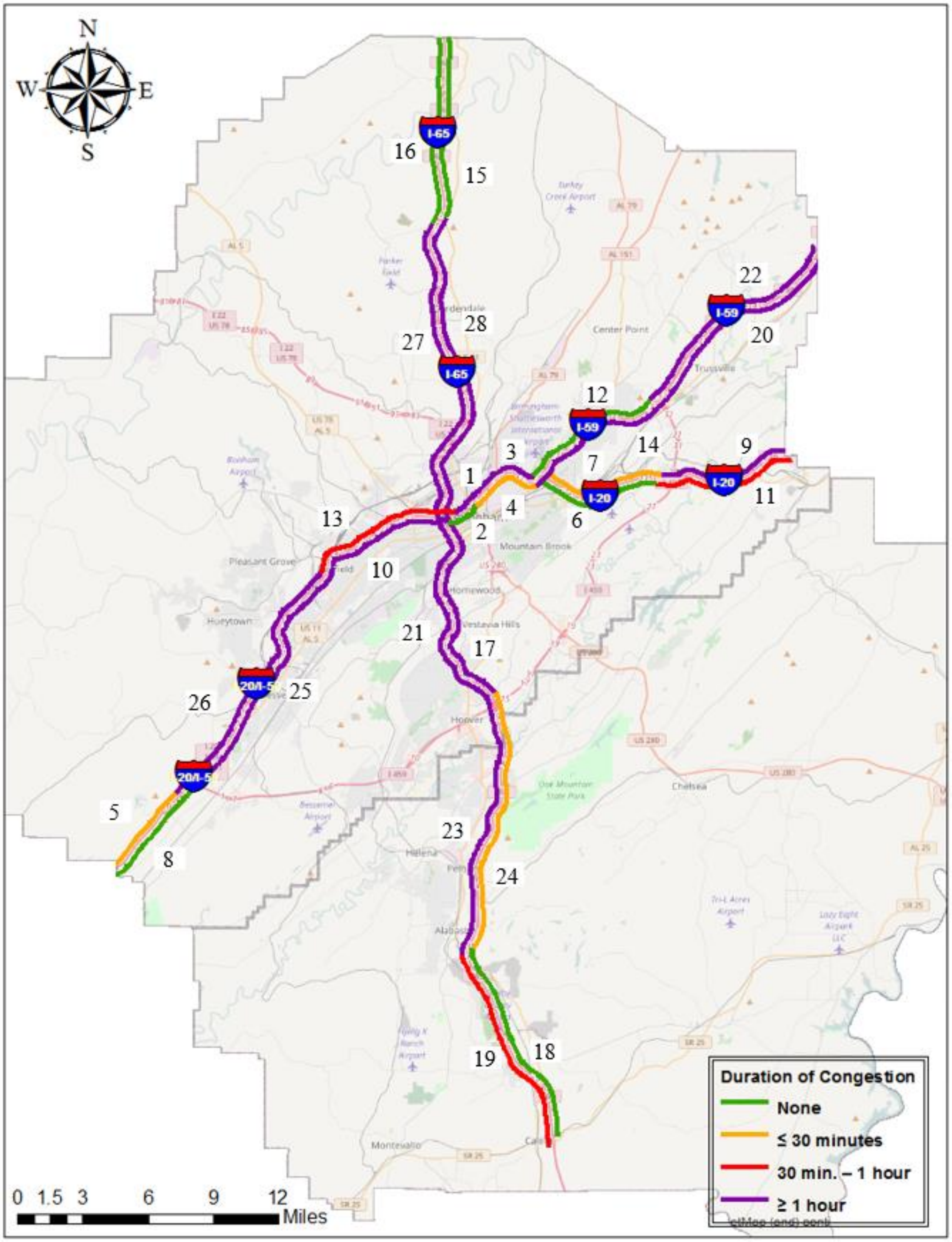

Figure 22. Duration of Congestion during PM Peak in January 2015

4.6.3. Planning Time Index (PTI)

The Planning Time Index values for the Birmingham case study were summarized in Table 21. As it can be seen, that the range of PTI values in the findings varies widely from 1.05 to 23.03. The lower values of PTI represent traffic conditions close to the ideal condition meaning travelers can travel along these segments with free flow speed in that particular time period. The higher PTI values show the presence of expected and unexpected delays which means that commuters in those segments usually experience an average time of travel that is greater than ideal condition, and must also be prepared for unexpected congestion to occur during this particular time period. For instance, segment 2 shows a PTI value equal to 23.03, which shows that despite a 99.86 second free flow travel time the planned travel time should be 2,299.74 seconds, which is equal to the travel speed of 2.175 miles/hour.

Segment Code	Free Flow Travel Time	Maximum Planning Time Index in AM Peak												Maximum Planning Time Index in PM Peak											
		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	91.63	12.89	4.88	4.12	2.55	4.57	2.52	2.22	2.12	2.33	3.35	3.01	1.96	2.91	2.82	4.39	2.22	2.52	5.31	3.64	3.11	6.79	5.67	6.25	8.31
2	99.86	6.87	4.80	2.73	2.64	1.77	3.48	2.07	2.10	2.70	2.98	2.39	2.15	1.59	8.71	7.03	6.04	5.06	4.85	2.81	3.33	2.67	6.59	5.95	23.03
3	215.78	4.85	2.46	5.57	3.22	4.73	3.95	2.78	3.05	3.96	5.08	6.16	2.88	3.33	1.89	5.65	3.78	1.30	2.60	3.10	2.94	3.49	3.92	9.58	7.08
4	229.40	1.03	1.07	1.13	1.12	1.00	1.03	1.12	1.03	1.22	1.03	1.01	1.00	1.39	3.43	2.10	2.99	1.72	1.83	1.94	1.49	2.15	2.18	4.93	2.99
5	272.86	1.26	1.24	1.18	4.63	1.14	1.18	1.14	1.15	1.14	1.16	1.18	1.14	1.21	1.19	1.18	1.46	1.26	1.18	5.75	2.62	1.15	1.15	1.15	1.34
6	353.41	1.28	2.36	1.36	1.05	6.55	1.79	1.13	1.26	1.03	1.13	1.03	1.02	1.46	1.19	1.17	1.55	1.00	1.25	1.08	1.00	1.36	1.09	4.65	1.22
7	374.04	1.22	2.09	2.20	1.08	1.97	2.07	1.00	1.94	1.17	2.53	4.21	1.07	3.17	1.22	1.14	1.12	1.13	1.13	1.05	1.17	1.21	1.03	2.25	1.15
8	307.44	1.20	1.33	1.30	1.30	1.17	1.52	1.14	1.15	1.38	1.17	1.20	1.20	1.18	1.24	1.20	1.20	1.71	1.53	1.23	1.36	1.14	1.24	1.16	1.26
9	330.75	1.19	1.18	22.81	1.43	1.15	1.21	1.13	1.19	1.18	4.52	1.16	4.00	1.22	1.23	1.57	1.97	1.25	1.26	1.19	1.38	1.21	1.30	2.53	1.19
10	404.67	4.17	2.17	3.28	2.51	2.50	4.38	1.90	2.75	3.20	2.74	2.13	4.93	3.44	2.66	5.78	1.78	2.05	1.31	2.75	2.48	2.86	2.12	6.57	3.61
11	362.52	1.13	1.13	1.86	1.12	1.10	1.10	1.07	1.08	1.08	1.08	1.09	1.09	3.78	1.39	1.16	3.47	1.09	1.19	1.98	1.62	1.47	1.38	1.42	1.20
12	396.83	5.37	6.18	2.43	1.92	2.98	2.04	1.31	1.99	2.71	7.05	1.90	2.22	1.28	7.52	3.12	1.78	1.52	1.36	2.03	1.69	1.58	1.22	2.23	1.53
13	438.00	1.19	1.86	2.17	1.25	1.06	1.50	2.11	1.28	1.72	1.20	1.29	1.19	2.27	1.80	2.15	1.65	1.58	3.71	2.11	3.28	7.46	8.90	2.59	4.19
14	422.20	4.22	5.81	3.54	6.00	6.25	3.20	5.61	2.57	3.14	1.93	2.92	1.17	2.76	2.76	8.25	3.09	2.24	4.03	2.75	2.83	2.30	4.67	5.58	9.11
15	429.53	2.89	1.23	1.16	1.14	1.16	1.16	1.29	1.21	1.15	1.16	1.20	1.27	1.16	8.43	1.15	1.13	1.26	4.20	1.18	1.16	1.15	1.13	1.13	1.19
16	457.25	2.87	1.31	1.16	1.17	1.14	1.13	1.14	1.14	1.14	1.13	1.22	1.15	1.19	4.07	1.74	1.17	1.29	3.04	2.19	1.19	1.16	1.17	1.17	1.55
17	578.57	4.49	3.43	3.14	2.79	2.53	2.95	2.53	3.81	3.52	3.74	3.19	3.15	2.48	5.15	6.00	6.29	5.37	3.77	3.21	4.70	4.45	2.44	3.16	5.53
18	513.90	4.67	3.23	2.64	6.00	3.22	1.95	2.97	5.77	5.22	3.08	6.54	4.44	1.22	1.23	2.50	3.09	2.51	5.52	4.79	2.72	2.48	2.96	2.64	1.91
19	516.34	1.18	3.21	1.16	1.17	1.13	1.16	1.80	1.20	1.13	1.13	1.52	1.17	1.25	2.01	6.42	4.13	2.13	1.61	7.93	1.83	1.24	1.89	1.82	1.49
20	537.81	1.21	1.27	1.25	3.53	1.19	1.22	1.18	1.36	1.82	1.85	1.37	1.22	1.88	2.07	2.07	1.74	5.21	2.86	1.65	2.20	1.98	2.05	2.67	2.05
21	648.61	1.75	2.19	1.44	1.56	2.56	1.26	2.24	1.50	1.68	1.71	1.33	2.08	4.21	4.68	3.99	4.34	7.70	3.83	3.84	4.14	4.14	4.49	4.34	6.05
22	560.52	4.40	3.76	2.21	3.44	2.45	2.68	1.58	2.58	3.84	2.67	3.43	2.71	1.31	1.43	1.22	1.23	1.21	1.16	5.95	1.19	3.02	4.68	4.54	2.09
23	595.62	1.18	1.36	1.21	1.20	1.32	1.16	1.16	1.40	1.20	1.33	1.77	1.18	2.69	2.95	3.78	2.54	4.60	2.89	5.42	2.97	3.10	3.61	5.90	2.78
24	636.38	2.48	3.23	3.35	2.97	2.45	1.85	1.51	2.74	3.14	4.34	3.78	7.41	1.70	1.25	1.86	4.32	1.76	1.52	2.63	1.35	4.28	1.64	1.38	2.99
25	622.01	2.66	1.63	1.31	1.51	1.88	1.44	1.21	1.95	3.23	1.92	2.36	2.62	10.18	1.29	1.92	1.32	1.88	2.08	1.42	1.78	1.20	1.21	1.78	1.94
26	655.23	1.31	1.83	1.20	1.22	1.16	1.42	1.28	2.32	1.29	2.14	1.23	1.98	1.91	1.58	1.52	2.90	2.29	2.07	1.30	3.76	2.47	1.27	3.68	2.54
27	790.33	2.18	2.84	3.41	2.68	1.58	2.82	2.96	2.08	3.31	3.76	2.79	2.31	3.88	3.93	2.50	1.92	2.15	6.15	2.78	1.50	1.83	1.49	1.37	1.75
28	836.01	2.46	3.55	1.87	3.17	1.95	3.21	1.63	3.12	2.47	1.98	3.63	2.07	1.92	5.09	4.07	2.53	1.96	1.74	1.81	3.32	2.01	1.75	2.14	4.21

Table 21. Max Planning Time Index during Peak Periods in 2015

4.6.4. Buffer Time Index (BTI)

The Buffer Time Index (BTI) is the percentage of average travel time that travelers should add to their average time of travel to ensure on-time arrival for the 95 percent of the time. The resulting values from BTI analysis can be found in Table 22. Some segments show high BTI values and are marked with red. For instance, segment 9 with a high BTI value of 518 during AM peak in March 2015 is shown in Table 22 in red.

Table 22. Max Buffer Index during Peak Periods in 2015

Segment Code	Maximum Buffer Index in AM Peak												Maximum Buffer in Index PM Peak											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	160	165	153	96	134	95	92	85	89	116	120	61	91	110	150	87	80	192	121	113	250	166	147	221
2	187	156	124	78	41	79	44	46	60	76	60	49	49	273	212	223	158	170	101	150	98	252	189	307
3	100	62	191	83	143	102	146	86	113	137	235	60	134	76	263	179	32	120	160	135	165	163	341	195
4	13	15	22	17	11	15	23	15	33	13	11	9	33	180	77	131	53	65	68	33	70	91	205	96
5	16	13	9	207	8	13	9	9	9	8	11	8	11	9	10	31	17	11	262	77	8	8	7	21
6	24	44	30	13	198	70	21	27	10	19	11	9	35	26	22	47	10	28	17	10	35	17	105	21
7	24	77	90	17	64	93	12	75	25	110	197	16	119	27	22	19	21	18	13	28	28	12	86	21
8	8	18	17	19	9	36	8	9	26	8	11	11	9	11	11	11	49	37	16	25	9	16	8	14
9	10	10	518	29	8	14	9	13	11	166	8	134	10	13	13	63	14	15	11	24	10	19	92	10
10	129	74	105	82	98	139	49	93	88	95	71	155	122	105	251	54	70	33	119	110	108	86	212	128
11	9	11	59	9	7	9	8	8	7	8	8	7	133	28	13	144	9	17	78	53	35	32	34	13
12	175	190	80	60	94	60	26	65	105	191	47	55	22	192	118	58	41	30	72	50	44	21	68	39
13	17	57	79	22	11	39	72	30	54	22	19	15	77	44	86	43	45	134	67	139	147	227	90	147
14	51	68	99	108	75	68	122	60	61	61	58	15	63	113	204	98	67	128	97	99	69	133	128	157
15	46	13	9	7	10	9	19	14	9	10	10	19	9	253	10	9	20	150	14	13	12	8	8	13
16	90	20	9	9	10	10	11	10	10	8	13	9	10	156	48	13	12	124	71	13	10	11	9	31
17	85	78	102	76	67	108	59	111	150	86	78	93	78	142	201	226	205	177	118	161	147	97	77	148
18	140	124	63	152	60	54	120	220	153	59	127	116	13	14	101	121	90	187	202	99	98	120	93	56
19	12	99	11	11	9	11	53	14	9	7	34	10	11	53	205	190	66	40	220	55	15	62	45	28
20	11	15	9	111	11	13	10	25	40	40	23	11	29	67	50	24	155	80	30	72	59	36	70	25
21	32	53	31	45	94	22	78	36	34	50	27	73	81	159	97	94	187	86	111	105	102	115	85	112
22	127	148	43	85	82	112	32	61	145	50	79	58	16	27	11	15	13	8	190	11	85	122	123	60
23	11	25	14	15	23	11	13	30	16	25	48	13	66	112	117	60	136	77	160	79	102	74	157	50
24	52	107	96	129	64	43	28	93	63	112	122	138	40	17	47	148	50	37	97	26	165	38	23	102
25	33	39	14	33	51	27	13	56	123	49	69	86	167	14	51	19	46	70	26	44	9	10	42	52
26	16	25	9	11	7	25	17	77	16	60	12	52	47	31	28	89	65	65	15	140	83	12	98	81
27	58	84	119	80	31	104	74	68	114	79	52	63	55	92	55	46	48	109	89	32	50	28	20	41
28	51	63	45	65	44	75	29	67	56	38	68	40	36	145	79	86	44	51	48	92	52	35	60	83

In Table 22, the red cells with high values represent a poor reliability rating for corresponding segments during that specific time period. Segment 8 with all cells in green represents a good reliability rating during AM and PM peak in 2015.

To investigate the inter-relationships among the Travel Time Reliability indices, the values for TTI, BTI, and PTI were used to generate scatter plots as shown in Figures 23 through 25. According to Figures 23 and 24, it can be clearly seen that there is a strong correlation between the PTI and TTI as well as PTI and BTI values with the R-squares equal to 0.73 and 0.84 respectively.

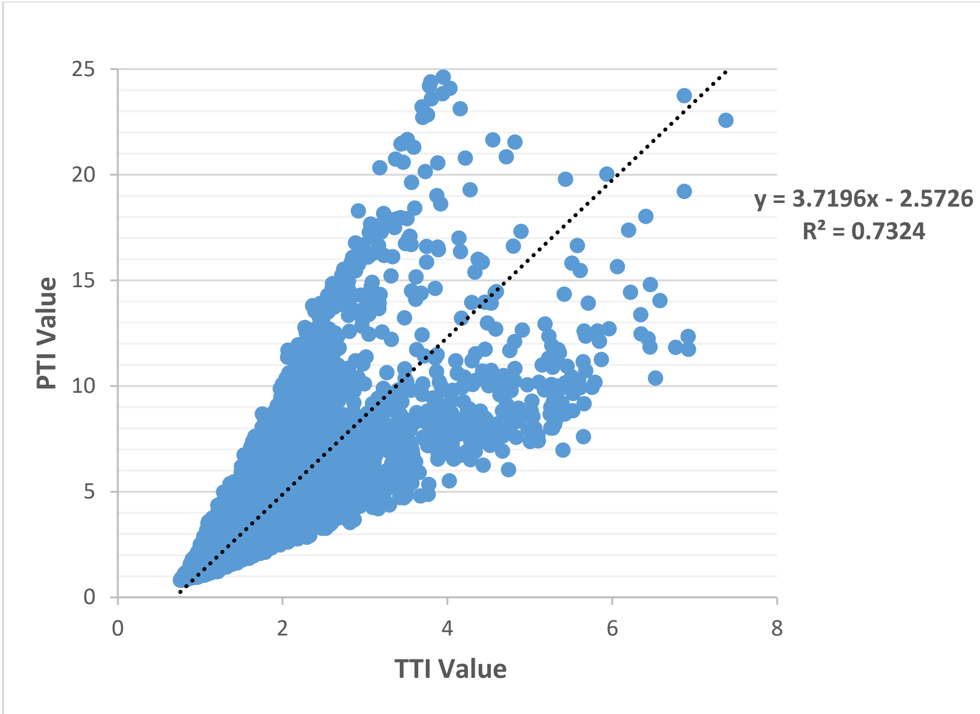

Figure 23. Inter-Relationship between TTI and PTI

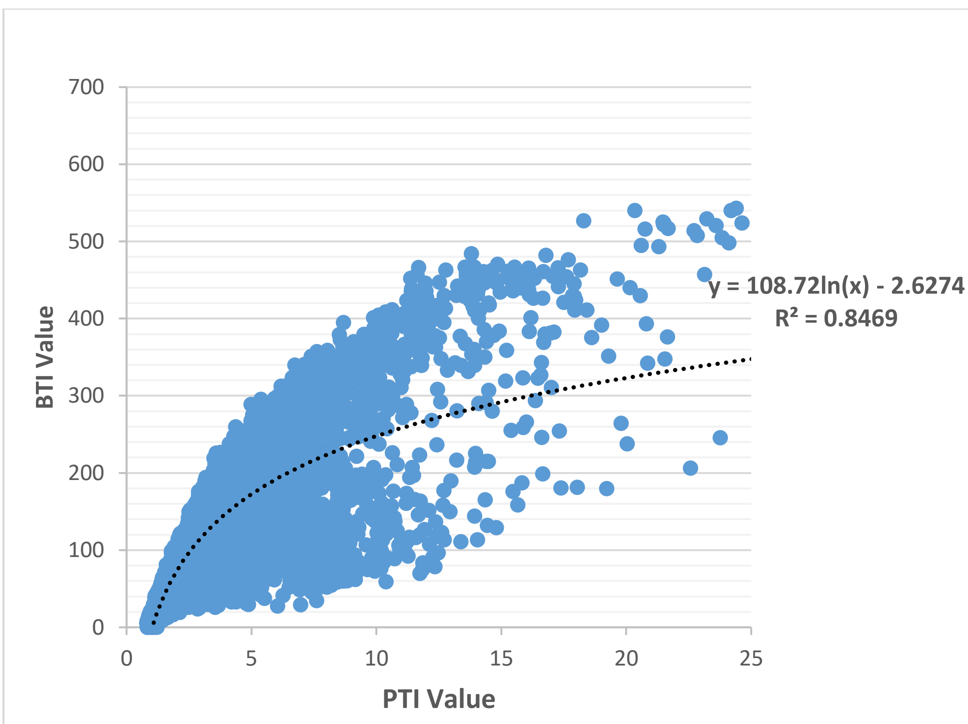

Figure 24. Inter-Relationship between BTI and PTI

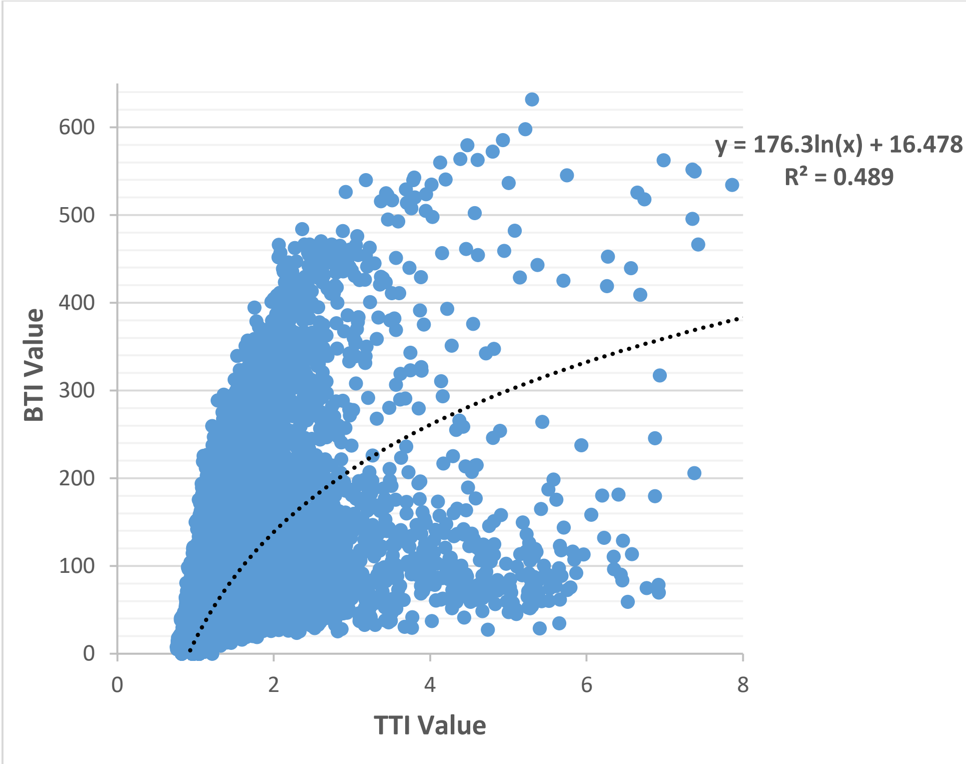

Figure 25. Inter-Relationship between TTI and BTI

4.7. 85th Percentile of Congestion Intensity and Speed-drop

Congestion Intensity and Speed-drop represent annual measures that reflect the extent and severity of all circumstances occurring over the year respectively. It is believed that the 85th Percentile Congestion Intensity and Speed-drop to be an effective way to represent both expected and unexpected circumstances over the year.

Table 23 shows the comparison among the 28 study segments based on their 85th Percentile Congestion Intensity and Speed-drop values, and Figures 26 and 27 display their location in the study area. Table 23 implies that segment 26 has the highest value in the 85th Percentile Congestion Intensity meaning during the AM and PM peak hours more than 50 percent of this segment is congested. However, the value of 6.31 percent for 85th

Speed-drop reveals that most of the area along this segment should be moderately congested, which can also be proved by looking at the range of associated TTI values throughout the year. On the other hand, segment 3 shows the highest value for 85^{th} Percentile Speed-drop which implies that a high level of delay occurring on this segment over the year. It can clearly be seen that a high value for 85^{th} Percentile Congestion Intensity is not necessarily accompanied by a high value for 85^{th} Percentile Speed-drop. Therefore, both measures alone cannot be representative of a metric which enables ranking and prioritizing segments.

Table 23. Segments 85^{th} Percentile Congestion Intensity and Speed-drop

Segment Code	85 Percentile of Congestion Intensity	85 Percentile of Speed-drop
26	52.67%	6.31%
25	50.44%	6.62%
23	48.40%	30.70%
17	45.15%	31.99%
20	42.92%	13.86%
8	41.30%	3.66%
22	41.15%	17.96%
9	40.88%	5.04%
18	39.43%	26.10%
24	38.74%	22.28%
27	37.84%	15.02%
1	36.12%	34.17%
28	36.06%	13.73%
14	35.87%	17.67%
2	34.43%	34.51%
21	34.36%	34.52%
5	34.26%	3.74%
19	31.33%	8.73%
16	30.19%	4.74%
3	27.88%	40.67%
15	24.71%	4.00%
12	23.84%	16.42%
11	21.12%	4.99%
10	16.68%	37.35%
4	16.56%	21.65%
6	13.97%	25.17%
13	12.04%	21.31%
7	7.00%	23.63%

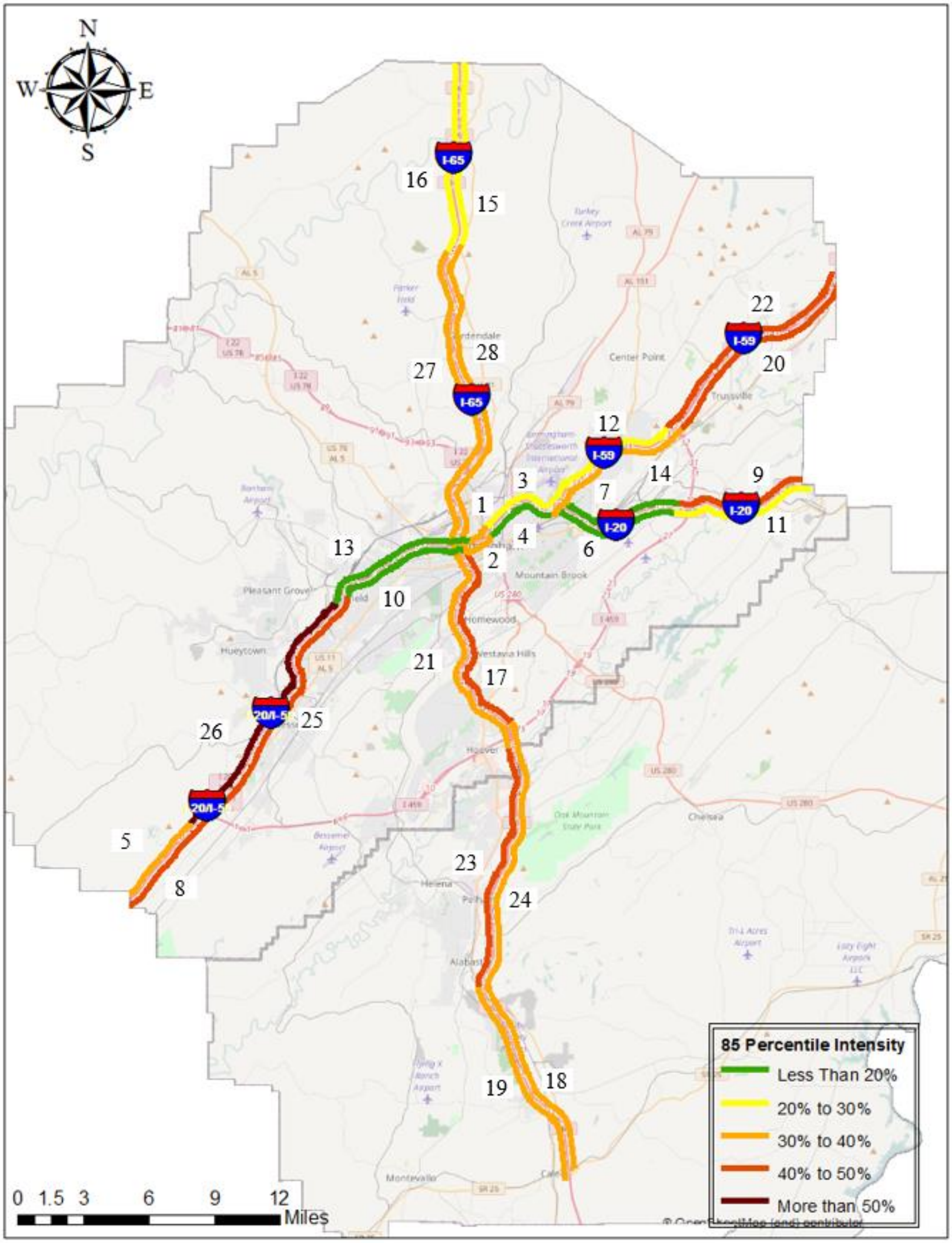

Figure 26. 85th Percentile Congestion Intensity in 2015

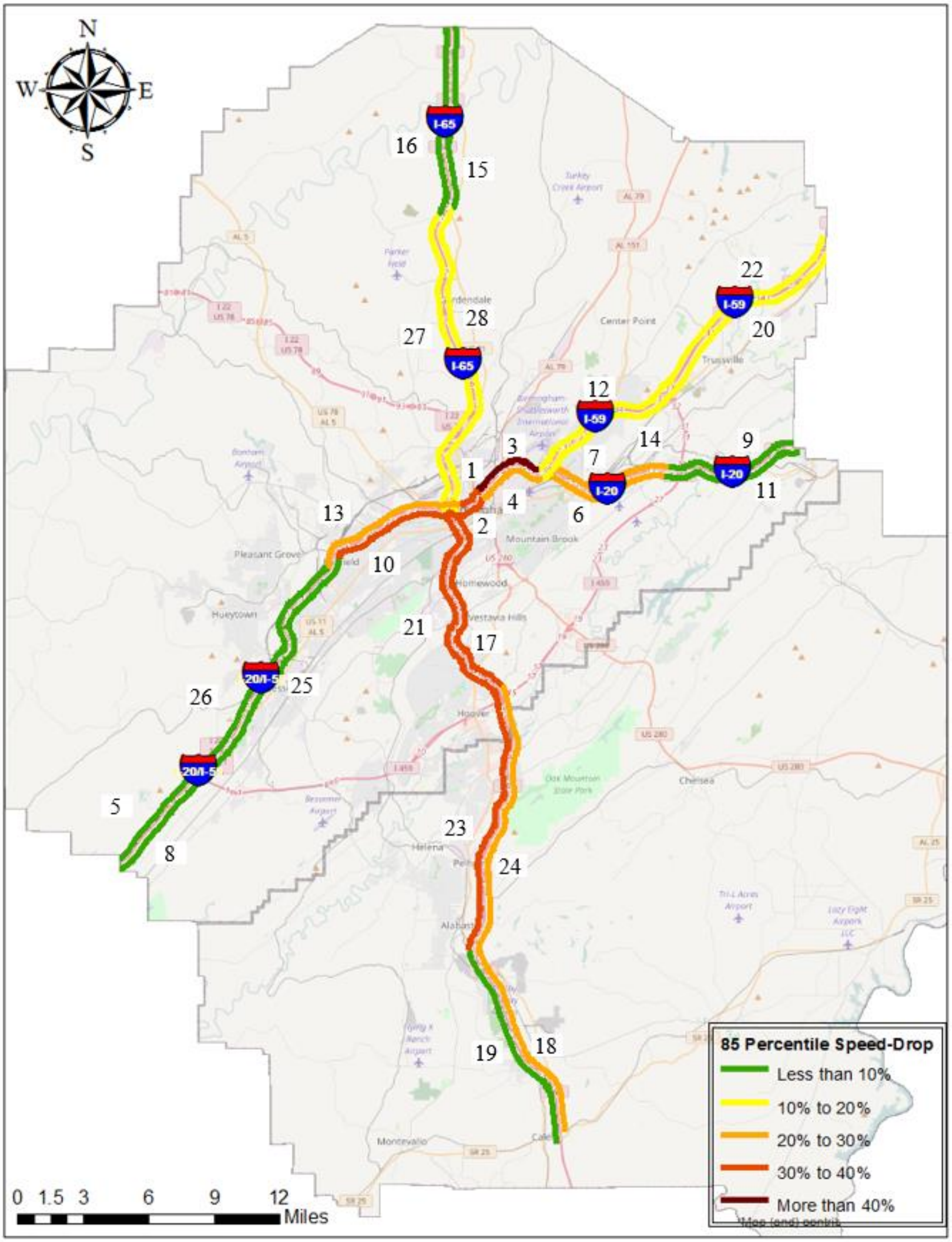

Figure 27. 85th Percentile Speed-drop in 2015

4.8. Impact Factor (IF)

To compare and rank segments in a comprehensive way that captures the effect of both Congestion Intensity and Speed-drop, the Impact Factor is determined. The Impact Factor metric is accounting for both reliability and variability of congestion throughout the year. Table 24 shows the Impact Factor for all study segments from the highest to the lowest value.

Table 24. Segments Impact Factors in 2015

Segment Code	Impact Factor
17	14.35%
23	14.05%
1	11.64%
21	11.40%
2	11.07%
3	10.63%
18	8.49%
24	7.85%
22	6.30%
14	6.26%
10	5.43%
20	5.18%
27	4.96%
28	4.17%
12	3.38%
4	3.21%
26	3.18%
25	2.88%
6	2.82%
19	2.68%
13	2.25%
9	1.90%
8	1.55%
7	1.33%
16	1.26%
5	1.15%
11	1.01%
15	0.90%

According to the Impact Factor values in Table 24, segments 17 and 23 with the highest values are the least reliable sections and segments 15 with the lowest value is the most reliable section throughout the year 2015.

Figure 28 displays the Impact Factor values from highest to the lowest accompanied by the 85th Percentile of Congestion Intensity and Speed-drop for corresponding segments. It reveals that segments with the relatively high value for both 85th Percentile Congestion Intensity and Speed-drop result in a high Impact Factor as well.

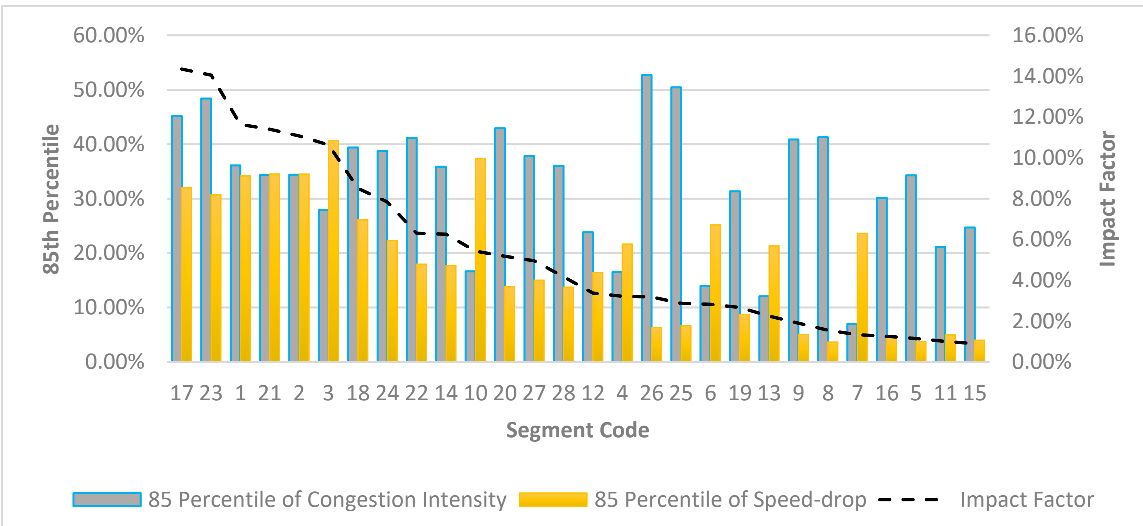

Figure 28. Impact Factor along with 85th Percentile of Congestion Intensity and Speed-drop for all segments in 2015

Figure 29 that displays the location of segments and their Impact Factor value indicates part of I-65 located in the Southside of Birmingham should be considered as of highest priority for receiving investment toward the operational improvements.

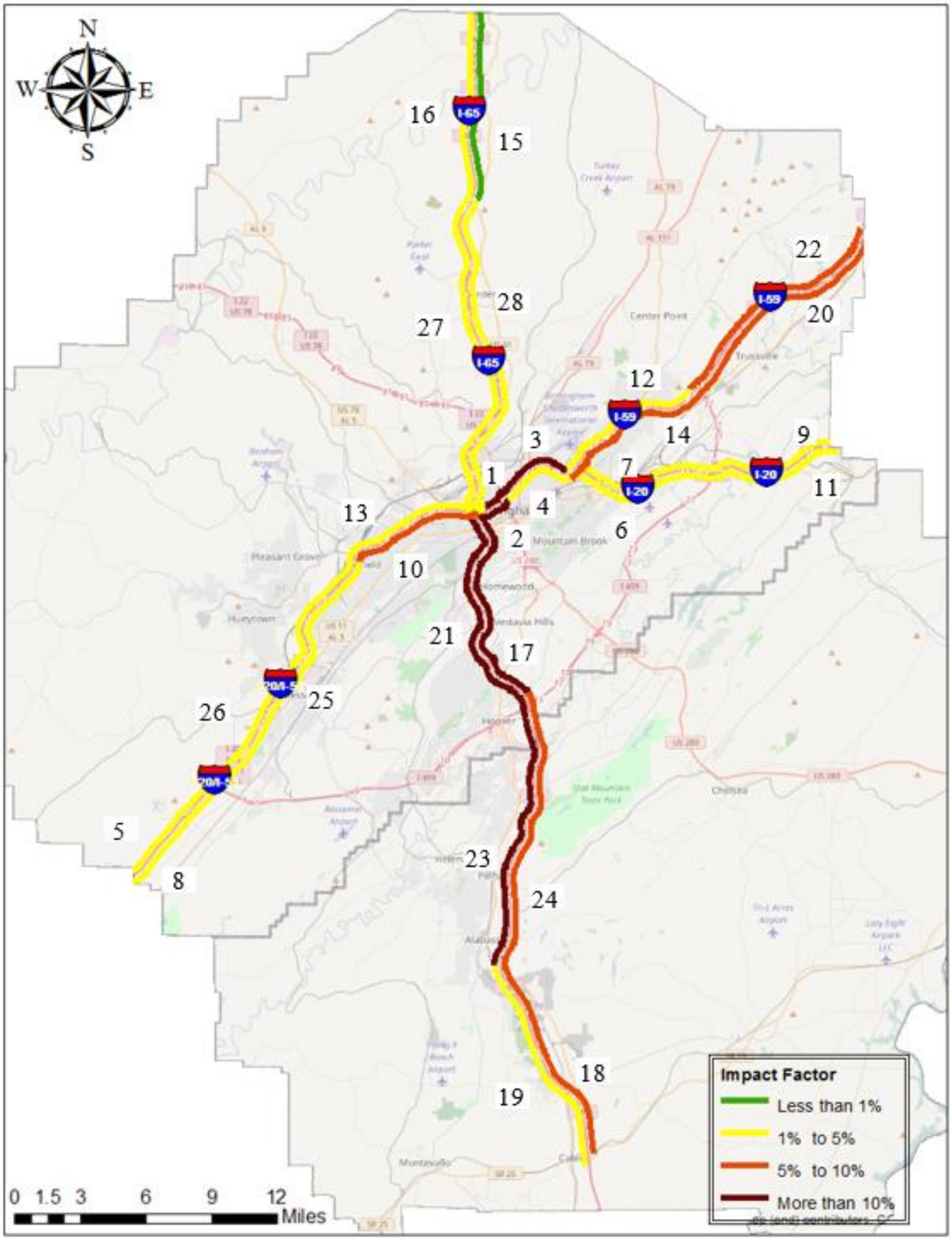

Figure 29. Impact Factor in 2015

CHAPTER 5: CONCLUSION AND RECOMMENDATION

This study was undertaken to (a) showcase the development of an automated process to facilitate management, storage and processing of big transportation datasets such as NPMRDS for congestion monitoring, (b) use traffic data analytics and statistical analysis to extract travel time reliability performance measures in a Birmingham case study, and (c) use reliability performance measures to determine the congestion extend and severity and guide optimization of traffic operations in the Birmingham region.

The case study utilized the NPMRDS dataset in order to quantify congestion in the Birmingham region over an one-year period (2015). Four major freeways were considered for data analysis namely I-65, I-20, I-59, and I-20/I-59 extending over two counties from Jefferson/Blount County line on the North to the Shelby/Chilton County line on the South and from the Tuscaloosa/Jefferson County line on the East to the Jefferson/St. Clair County line on the West.

In order to understand the issues associated with the management of the NPMRDS dataset and determine ways to overcome the challenge of handling “big data”, this thesis conducted first a detailed literature review and built on recommendations from existing researches on methods that show promise, followed by an assessment of the dataset to determine its potential and limitations. Then, the Relational Database Management System (RDBMS) was employed as an efficient and economical tool for data management and Structured Query Language (SQL) were used to extract data and perform the analysis. Next, a range of performance measures was calculated for quantifying the congestion location, level, and extent and prioritizing freeway segments with respect to congestion. A comprehensive study of dataset characteristics, including influencing variables that affect data measurements was presented.

It was found that the NPMRDS dataset offers detailed information for quantifying congestion, however, the availability of samples varies spatially and temporarily. A process for identifying anomalies was developed as part of this thesis. It was found that the variability of segment length can cause unreliable data points while the time granularity is one second. Furthermore, recommendations for improving data accuracy and easing data anomalies were reported.

The thesis also performed a detailed analysis of the variation of reliability indices such as Travel Time Index (TTI), Planning Time Index (PTI), Buffer Time Index (BTI), and Duration of Congestion (DOC). It became evident that different mobility performance measures represent different reliability aspects. The 85^{th} Percentile Congestion Intensity measure along with 85^{th} Speed-drop were proposed as valuable tools reflecting the extent and severity of all circumstances occurring over the year. In addition, the Impact Factor was developed for ranking the congested segments. Such rankings can be used as a systematic and data-driven method for prioritizing resource allocations for operational improvements. The analysis revealed that the segments 17 and 23 with relatively high values for 85^{th} Percentile Intensity and Speed-drop are the most unreliable segments in the study area due to the level congestion severity that extent throughout the year, and thus need attention.

Overall, the study findings can be valuable in guiding transportation professionals and agencies on how to make best use of big transportation databases such as NPMRDS to quantify the level and extent of congestion, generate performance-based reports and produce performance measures. Such performance measures can, in turn, be used as an initial screening process for congestion management purposes as well as in setting agency priorities for implementing congestion mitigation initiatives with the best potential return for the investment.

Future work can consist of validating the proposed approach using a larger sample size. It is recommended that future analysis can take place to understand the issues and evaluate the improvements. In addition, it is recommended that further studies be conducted that investigate in greater depth the effect of outliers on Travel Time Reliability measures. It is also desirable to extend the work to include consideration of additional data sources such as volume data, incident data, weather events and work zone presence information in order to improve the understanding of the causes of uncertainty in travel time and more accurately quantify recurrent and non-recurrent congestion in the future.

REFERENCES

[1] D. Schrank, B. Eisele, and T. Lomax, "Urban Mobility Report," Texas A&M Transportation Institute, 2015.

[2] "Traffic Congestion and Reliability: Linking Solutions to Problems," Federal Highway Administration, Washington, DC, 2004.

[3] D. Nam and P. Dongjoo, "Estimation of Value of Travel Time Reliability," *Journal of Advanced Transportation*, 2004.

[4] E. Lindquist, "Assessing Effectiveness Measures in the Istea Management Systems " Southwest Region University Transportation Center, Texas Transportation Institute, College Station, 1999.

[5] M.G.H. Bell and Y. Iida: "The Network Reliability of Transport," *Proceedings of the 1st International Symposium on Transportation Network Reliability (INSTR)*, 2003.

[6] T. Lomax, D. Schrank, S. Turner, and R. Margiotta, "Selecting Travel Reliability Measures," 2003.

[7] D. M. R. Margiotta, and D. McLeod, "Travel Time Reliability as a Service Measure for Urban Freeways in Florida," presented at the Transportation Research Board 94th Annual Meeting, 2013.

[8] "Highway Capacity Manual," Transportation Research Board, Washington DC, 2010.

[9] A. S. C. Chen, and P. Varaiya, "Travel Time Reliability as a Measure of Service," *82nd Annual Meeting Transportation Research Board*, 2002.

[10] W. Kittelson and M. Vandehey, "Incorporation of Travel Time Reliability into the HCM," Cambridge Systematics, ITRE, Texas A&M Research Foundation, 2013.

[11] "Introduction to the National Performance Management Research Data Set (NPMRDS)," HERE and the Volpe Center, 2013.

[12] P. Rafferty and C. Hankley, "Applying National Probe Data for Performance Measures: Crafting and Visualizing Mobility from the NPMRDS," Wisconsin Traffic Operations and Safety Laboratory, 2014.

[13] C.-F. Liao, "Using Truck GPS Data for Freight Performance Analysis in the Twin Citi es Metro Area," Minnesota Department of Transportation Research Services & Library, 2014.

[14] "Cost of Congestion to the Trucking Industry," ATRI, 2014.

[15] "I-95 Corridor Coalition Probe Data Comparison," UMD CATT Lab, Maryland, 2014.

[16] *Travel Time Reliability Reference Manual*, Upper Midwest Reliability Resource Center, 2014.

[17] "The iPeMS MAP-21 Module: Producing the Information you Need from the National Performance Management Research Data Set (NPMRDS)," Iteris, 2015.

[18] FHWA, "2013 Urban Congestion Trends," Federal Highway Administration, Washington DC, 2015.

[19] FHWA, "2015 Urban Congestion Trends: Communicating Improved Operations with Big Data," Federal Highway Administration, Washington DC, 2016.

[20] E. F. Codd, *Relational Database: a Practical Foundation for Productivity*, Communications of the ACM, 1982.

[21] H. Tu, J. W. v. Lint, and H. J. v. Zuylen, "Travel Time Reliability Model on Freeways," *87th Annual Meeting of Transportation Research Board*, 2008.