

ABSTRACT

AHMED, ISHTIAK. Analysis of Recurring Freeway Bottlenecks Using Probe Data. (Under the direction of Dr. Nagui Roupail.)

A recurring bottleneck is a location where demand for the usage of a highway section exceeds the section's capacity to handle it on a regular basis. To mitigate recurring bottlenecks, identification, prioritization, and understanding the pattern of activation and impact of the bottlenecks are three essential tasks. This thesis aims to unravel these three issues in the context of freeways using probe vehicle speed data as its primary data source.

To accomplish its goals, the study begins with a critical review of the past studies regarding freeway recurring bottleneck identification and ranking. The review reveals that there are gaps in distinguishing bottlenecks from congested segments, in the proper selection of the thresholds for congestion detection and bottleneck identification, and in filtering the recurring ones from nonrecurring events induced bottlenecks. An already established bottleneck ranking method by RITIS is selected from the past studies for further review and analysis.

Addressing gaps in the literature, the thesis proposes an algorithm that identifies recurring freeway bottlenecks. Although the procedure is built on a prior study, its contribution lies in the justified selection of the thresholds used for congestion detection and filtering the recurring bottlenecks. Moreover, several congestion extent based performance measures are developed in this thesis to rank the bottlenecks (Recurring Bottleneck Impact Factor and Overall Recurring Bottleneck Impact Factor) and to explain their day-to-day variability (Daily Impact). Next, the study proposes a logistic regression model for predicting the probability of activation and a linear regression model for predicting the impact when a

bottleneck activates on a day. Both types of model are based on temporal data fused with the information on non-recurring events.

A case study on freeways in Wake and Durham Counties, NC is conducted by applying the proposed method. The method identified 15 recurring bottleneck locations which are ranked by their overall RBIF. The ranking algorithm by RITIS is also applied to the case study area. The contrasts in the results obtained by these two methods show that the proposed method generates a workable number of recurring bottleneck list rather than mixing them with non-recurring event induced bottlenecks. Detailed investigation of the identified bottlenecks reveals a special case bottleneck, which has its location outside of the study area and what seems to be the bottleneck on the study area is actually a mere spilling back of the queue from another facility.

The development and validation of the logistic and linear regression models encountered several issues regarding sample size. Although several models pass the statistical tests for goodness of fit, validation results for most of the models were poor. Bias in the models due to lack of significant predictors is found responsible for the poor validation results. Nonetheless, the linear regression models for a few bottlenecks show promising results. These models can be used by transportation agencies for applying real-time congestion mitigation strategies as the variation of activation and impact of the bottlenecks on different days of week and times of year is apparent from the models. Moreover, the effect of non-recurring events such as inclement weather and incidents occurring within a bottleneck region is described by the models as well.

In summary, the study demonstrates the development and application of a method for identifying and ranking of recurring freeway bottlenecks. A challenging effort of predicting

the activation and impact of the bottlenecks on a daily basis is illustrated in this study as well. Upon addressing the limitations, the proposed approaches can be used in mitigating freeway congestion.

Analysis of Recurring Freeway Bottlenecks Using Probe Data

by
Ishtiak Ahmed

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Civil Engineering

Raleigh, North Carolina

2017

APPROVED BY:

Dr. Billy Williams

Dr. George List

Dr. Nagui Roupail
Chair of Advisory Committee

DEDICATION

To my father, hopefully residing in heaven, and to the rest of my family, living in three different continents on earth.

BIOGRAPHY

Ishtiak Ahmed was born and raised in Dhaka, Bangladesh and first arrived at North Carolina State University in August 2015. He completed his undergraduate degree in Civil Engineering at Bangladesh University of Engineering and Technology in March 2013. After graduation, Mr. Ahmed got admitted in NCSU in pursuit of his Doctor of Philosophy with an interim Master of Science degree in Civil Engineering.

ACKNOWLEDGEMENTS

Firstly, my sincere gratitude goes to Dr. Nagui Roupail, my academic advisor and source of guidance. I would also like to recognize Dr. Billy Williams as well as Dr. George List; I enjoyed a lot working with them and taking their classes.

I owe big thanks to the North Carolina Department of Transportation and to the many individuals associated with this agency. Their support assisted my graduate study in many ways.

I will be thankful to my mentor Shams Tanvir, my classmate Mohamad Shoaib Samandar, and to all my colleagues in ITRE.

Last but not least, I express my gratitude to the Almighty and to my family. They are the primary source of my mental strength and courage.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER 1: INTRODUCTION.....	1
1.1. Background	1
1.2. Problem Statement, Scope, and Objectives.....	4
1.3. Structure of the Thesis.....	5
CHAPTER 2: LITERATURE REVIEW	7
2.1. Bottleneck Definition and Identification Methods.....	7
2.1.1. Definition of a Traffic Bottleneck.....	7
2.1.2. Bottleneck Identification Methods.....	8
2.1.2.1. Bottleneck Identification Using Loop Detector Data.....	9
2.1.2.2. Bottleneck Identification using Probe Vehicles' Speed Data	13
2.2. Recurring Bottleneck Identification	14
2.3. Performance Measures of Bottlenecks	16
2.3.1. Delay Based Performance Measures	17
2.3.2. Congestion Extent Based Performance Measures	19
2.4. Summary	21
CHAPTER 3: METHODOLOGY	23
3.1. Congestion Detection and Bottleneck Activation Identification.....	23
3.2. Identification of Recurring Bottlenecks	25
3.3. Development of Performance Measures	27
3.3.1. Daily Impact of a Bottleneck	28
3.3.2. Recurring Bottleneck Impact Factor.....	29
3.3.3. Overall RBIF.....	30
3.4. Models for Predicting the Activation and Daily Impact of a Recurring Bottleneck	30
3.4.1. Predicting the Activation of a Bottleneck.....	32
3.4.2. Predicting the DI of a Bottleneck.....	32
3.4.3. Predictor Variables for the Models	33
3.4.4. Development and Validation of the Models	40
3.4.4.1. Time Frame for Model Development and Validation.....	40

3.4.4.2. Assessment of the Models.....	41
3.5. Scope of the Study.....	44
CHAPTER 4: CASE STUDY: METHODOLOGY APPLICATION TO BOTTLENECK IDENTIFICATION AND RANKING	46
4.1. Selection of the Thresholds for Congestion Detection and Bottleneck Identification	46
4.1.1. Threshold for Congestion Index (CI).....	47
4.1.2. Threshold for Average Historic Congestion Index (AHCI)	50
4.2. Identified Bottlenecks and their Performance Measures.....	52
4.3. Visualization of the Bottleneck Locations in the Study Area	56
4.4. Comparison with the Ranking Algorithm Established by VPP	58
4.5. Attributes of the Most Severe Recurring Bottlenecks.....	64
4.5.1. Rank #1: I-540 Eastbound at the On-Ramp from the Falls of Neuse Rd. (MM 14)	64
4.5.1.1. Recurring Bottleneck Region and Daily Impact (DI)	64
4.5.1.2. Attributes of the bottleneck location	67
4.5.2. Rank #2: I-40 EB at the On-Ramp from Harrison Ave. (MM 287)	69
4.5.2.1. Recurring Bottleneck Region and Daily Impact (DI)	69
4.5.2.2. Attributes of the bottleneck location	71
4.6. Special Bottleneck Cases	74
4.6.1. Congestion Created by a Bottleneck on an Arterial Road	74
4.6.2. Bottlenecks Impacted by a Long Term Work Zone.....	76
CHAPTER 5: PREDICTING THE ACTIVATION AND DAILY IMPACT OF BOTTLENECKS 78	
5.1. Data Sources.....	78
5.2. Challenges for Developing and Validating the Models	79
5.2.1. Bottlenecks Near the Fortify Work Zone.....	79
5.2.2. Issue of Sample Size	79
5.3. Logistic Regression Models for Predicting the Activation of a Bottleneck.....	82
5.4. Regression Models for Predicting DI.....	86
5.4.1. Base Models.....	86
5.4.1.1. Results from the Base Model Development.....	86
5.4.1.2. Assessment of the Models.....	89
5.4.2. Full Models	93

5.5. Summary	96
CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS	98
6.1. Recurring Bottleneck Identification and Ranking.....	98
6.2. Predicting the Activation and Daily Impact of bottlenecks	102
6.3. Recommendations for Future Research	105
REFERENCES	107
APPENDIX A: CUMULATIVE RESIDUAL PLOTS FOR BASE LINEAR REGRESSION MODELS	112
APPENDIX B: FULL LINEAR REGRESSION MODELS	116

LIST OF TABLES

Table 4-1: Performance measures and ranking of all the recurring bottlenecks identified in the study area for the weekdays of 2016.....	54
Table 4-2: Comparison of the bottleneck ranking obtained by overall RBIF and VPP algorithm.....	61
Table 4-3: Geometric properties and traffic volumes at top ranked bottleneck	68
Table 4-4: Geometric properties and traffic volumes at the Rank #2 bottleneck location	73
Table 5-1: Number of activations and variance and range of DI for the bottlenecks.....	81
Table 5-2: Results from the logistic regression models.....	83
Table 5-3: Assessment of the logistic models.....	85
Table 5-4: Results from the base models	87
Table 5-5: Assessments of the base models.....	90
Table 5-6: Evaluation of the full models	95

LIST OF FIGURES

Figure 2-1: SCM of average speed from multiple days’ observation. Key on left vertical axis shows speed in miles per hour (Ban et al. (2007)).....	10
Figure 2-2: Binary SCM (Ban et al. (2007)).....	11
Figure 2-3: Projection from volume-occupancy to URS-PUS coordinate system	12
Figure 2-4: Illustration of congested roadways and bottlenecks (FDOT (2011)).....	15
Figure 2-5: Elements and blobs generating from bottleneck activations (RITIS, 2016)	20
Figure 3-1: Congestion Index (CI) contour plot in space-time domain.....	25
Figure 3-2: Recurring bottleneck region using an AHCI=33% criterion	27
Figure 3-3: Daily Impact region of a recurring bottleneck	29
Figure 3-4: Distribution of DI for two different recurring bottlenecks	31
Figure 3-5: Distribution of DI for a typical bottleneck with the corresponding day of week	33
Figure 3-6: Non-recurring events activating at different times within the recurring bottleneck region	38
Figure 3-7: Selected Interstates with the Wake and Durham Counties, NC.....	45
Figure 4-1: RBIF per activation for different CI thresholds on different routes for fixed AHCI = 33%	48
Figure 4-2: RBIF per activation for different threshold of AHCI on different routes	51
Figure 4-3: Location, activation period and travel direction of the recurring bottlenecks in Wake and Durham Counties, NC (2016 weekdays)	57

Figure 4-4: Number of bottlenecks identified by VPP’s algorithm: including and excluding weekends.....59

Figure 4-5: AHCI contour for the Rank #1 bottleneck.....65

Figure 4-6: Cumulative distribution of DI for the Rank #1 bottleneck66

Figure 4-7: Location of the Rank #1 bottleneck: I-540 EB at the Falls of Neuse Road (MM 14)67

Figure 4-8: AHCI contour for the Rank #1 bottleneck.....70

Figure 4-9: Cumulative distribution of DI for the Rank #2 bottleneck71

Figure 4-10: Location of the Rank #2 bottleneck: I-40 EB at the on ramp from N. Harrison Avenue. (MM 287)72

Figure 4-11: Location of the bottleneck on US-70 and the resulting congestion75

Figure 4-12: Location of fortify work zone Phase II. (Courtesy: NCDOT)77

Figure 5-1: (a) Fitted vs. observed plot from trained model (b) Fitted vs. observed plot from validation dataset (c) CURE vs. predicted DI (d) CURE vs. observed DI for validation data92

CHAPTER 1: INTRODUCTION

1.1. Background

Increasing traffic congestion is an inevitable predicament in large and growing cities across the world. Bumper to bumper traffic is a very common scenario during the commuting hours in metropolitan cities. The world is getting faster day by day. To keep pace with that, everyone wants to start their daily jobs on time and wants to go back home as soon as they finish that. Moreover, the way a modern city operates, most of the people must commute almost at the same time of a day. The major problem is, not all the parts of an urban roadway network have the capacity to handle excess demand which can occur during the regular peak hours or due to any irregular happening on the road. Thus, the imbalance between the demand and capacity of a road section results in traffic congestion.

Many techniques have been proposed and applied for mitigating traffic congestion (FHWA, 2017). These techniques can be costly and long lasting like adding lanes (USDOT, 2015) or can be of low-cost such as dynamic traffic management (FHWA, 2009). No matter what the technique is, the first and foremost task for congestion mitigation is finding its origin, also known as the choke point or bottleneck.

A bottleneck is a location where demand for the usage of a highway section exceeds the section's capacity to handle it (Margiotta, Spiller, & Halkias, 2007). This occurrence of demand exceeding the capacity of a section is called the activation of a bottleneck at that section. On a roadway segment, a bottleneck may be activated at approximately the same time almost every day. This type of bottleneck is termed as the recurring bottleneck. If the

activation of a bottleneck is unpredictable and attributed to random events, it is called a non-recurring event induced bottleneck. For the sake of discussion, bottlenecks may also be divided based on the functional class of roadways, such as freeway and arterial bottlenecks.

In the US, about 40% of congestion nationwide has been attributed to recurring bottlenecks (2007). Although these bottlenecks could be mitigated with appropriate operational strategies and roadway improvements, the amount of candidate locations may be so high that some form of prioritization of the bottleneck mitigation schedule is essential. In order to do so, congestion due to recurring bottlenecks need to be distinguished from that due to non-recurring ones and need to be prioritized.

To prioritize the bottlenecks, it is necessary to develop some measures that represent the impact of the bottleneck activation. For example, one can estimate how much delay does the activation of a bottleneck caused every day (Skabardonis, Varaiya, & Petty, 2003). This type of measure represents the severity or intensity of the resulting congestion. A different way could be estimating the resulting queue length or the duration of congestion (Regional Integrated Transportation Information System, 2016). These two are congestion extent based measures. Another way of quantifying a recurring bottleneck could be its frequency of activation. No matter which estimate is used, it must represent the impact of the bottlenecks in a reasonable way so that a prioritizing scheme can be developed.

Another important research field related to traffic bottleneck is understanding the variation in the activation and the magnitude of the impact of a bottleneck. To explain the importance of the field, the current practice of tackling bottlenecks needs to be highlighted.

Nowadays, low-cost countermeasures such as real-time traffic management are gaining popularity over traditional high-cost countermeasures (Latham, F; Trombly, J, 2003). Current technological advancement has allowed rapid dissemination of information which can be used to apply real-time traffic management strategies. Such strategies include ramp metering, opening HOT lanes, dynamic traffic routing and so forth (2017). However, for proper application of these strategies, a clear understanding of the variation of the bottleneck activation and the magnitude of its impact is important. Understanding the effect of different factors related to a bottleneck can help significantly to implement the therapeutic techniques.

Several past studies have identified and rank traffic bottlenecks (CATT Lab, 2016); (Ban, Chu, & Benouar, 2007); (Chen, Skabardonis, & Varaiya, 2004). These studies either used loop detector or probe data. In recent years, probe data is gaining wide popularity because the technology is low-cost, the data collection process is automated, and it does not require any deployment of instruments in the highway right-of-way (Turner, 1998); (Young, 2007).

Of the past studies that identified and ranked bottlenecks, only a few of these focused on filtering recurring bottlenecks only (Skabardonis, Varaiya, & Petty, 2003); (Chen, Skabardonis, & Varaiya, 2004). Even fewer methods were found that advanced toward developing performance measures (Song, 2016); (Regional Integrated Transportation Information System, 2016); (FDOT, 2011). The investigation of the methods developed by these studies revealed that significant research is still needed on establishing a robust

technique for filtering recurring bottlenecks, developing performance measures, and finding the attributes that can explain the activation and variation in the impact of a bottleneck.

1.2. Problem Statement, Scope, and Objectives

Based on the above discussions, it is clear that to analyze recurring traffic bottlenecks, three components must be investigated, including (a) Where and when the recurring bottlenecks activate? (b) How can they be prioritized for scheduling their mitigation? And (c) How the activation and the resulting congestion of the bottlenecks vary day to day? In addition, it is also important to recognize the current practice regarding these three questions. The focus of this study rotates around these three basic questions in the context of freeways.

To answer the first research question, a method is needed that can identify the recurring freeway bottlenecks using probe vehicle speed data. The process must distinguish a bottleneck location from a series of congested road sections and filter the recurring bottlenecks only. The criteria for detecting congestion and filtering the recurring bottlenecks is vital since the subsequent analyses are largely dependent on it. Thus, a robust selection of these criteria is underscored here.

To answer the second question, performance measures that can represent the impact of a bottleneck need to be developed. A single measure or a combination of multiple measures can be used in a reasonable way so that it can properly reflect the impact of a bottleneck.

To answer the third question, each bottleneck needs to be scrutinized so that a pattern of their activation and variation in impact can be found. Factors that can predict such pattern and variation need to be identified with the available data.

In summary, following are the objectives of this study.

- To conduct a critical review of the literature that identifies, ranks, and characterizes recurring freeway bottlenecks.
- To propose a method for identifying and filtering recurring freeway bottlenecks.
- To develop a bottleneck ranking scheme and contrast it with the methods in the literature and in transportation practice.
- To develop a framework for predicting the variation in activation and magnitude of the impact of bottlenecks.

The methods proposed here are applicable to recurring freeway bottlenecks only. For arterial bottleneck analysis, factors such as signal timing, the spacing of intersections, and other information need to be incorporated. The primary data source used in this study is probe vehicle speed data. To illustrate the utility of the approach, a case study was conducted on an urban freeway network, although the applicability of the research is not limited to the case study area.

1.3. Structure of the Thesis

This thesis is divided into 6 chapters. The current chapter presents the background information, motivation, key research questions, and the objectives of the study. Chapter 2 provides a detailed review of the past studies related to recurring freeway bottlenecks. Chapter 3

discusses the mechanism of the proposed methods of recurring bottleneck identification, developing performance measures, and how to predict the variation of bottleneck activation and magnitude of impact. Chapter 4 presents a case study results regarding bottleneck identification, ranking, and comparison of the ranking with an established bottleneck ranking method. Chapter 5 describes the results obtained from the predictive models for bottleneck activation and impact. Chapter 6 summarizes the main conclusions of the analysis as well as recommendations for future study.

CHAPTER 2: LITERATURE REVIEW

This literature review intends to investigate previous studies on recurring freeway bottleneck identification and developing performance measures. The presentation is divided into three parts. The first part discusses the literature on bottleneck definition and freeway bottleneck identification method. The second part focuses on the past research works on recurring bottleneck identification. The third and the final section investigates the literature that quantifies the impact (e.g., delay, duration of congestion, queue length) of recurring and non-recurring bottlenecks.

2.1. Bottleneck Definition and Identification Methods

2.1.1. Definition of a Traffic Bottleneck

According to Margiotta et al. (2007), a bottleneck is a location where demand for usage of a highway section periodically exceeds the section's physical ability to handle it. Cambridge Systematics Inc. (2011) defined it as a roadway feature that causes routine congestion because of the capacity drop, volume surges or both. These definitions of bottlenecks are based on the relationship between the capacity of a roadway section and the demand. However, in reality, it is difficult to measure the physical capacity of a road section. Therefore, there was a need to define bottleneck based on traffic characteristics that are observable and computable. This need created the two classifications of roadway bottlenecks: active and hidden.

Ban et al. (2007) defined an active bottleneck as a location that initiates traffic congestion. Hidden bottlenecks, according to this study, are locations that exhibit traffic

congestion only under a certain demand pattern and usually are located upstream or downstream of an active bottleneck. Bertini and Mayton (2005), and Roupail et al. (2012) defined an active bottleneck as a point upstream of which there is a queue and downstream of which there is freely flowing traffic. Margiotta and Spiller (2012) distinguished between an active and a hidden bottleneck based on the fact that the former can release traffic without any downstream restriction (i.e. queue spill back) where the latter is metered by another upstream bottleneck.

The primary focus of this study is on active freeway bottlenecks. Therefore, from this point onward, the term “Bottleneck” is used to indicate “Active freeway bottleneck” only unless otherwise specified.

2.1.2. Bottleneck Identification Methods

Detecting congested condition is usually the first step of a bottleneck identification process. The congested condition of a road section is the consequence of a bottleneck activation at a downstream section. Cambridge Systematics, Inc. (2005) defined congestion as an excess of vehicles on a portion of a roadway at a time resulting in speeds that are slower than normal or the free flow speed. Previous studies related to freeway bottleneck identification by detecting congestion mainly used either loop detector data (flow, density, and time mean speed) or probe vehicle data (space mean speed). In the next two subsections, these studies are categorized into these two primary data sources.

2.1.2.1. Bottleneck Identification Using Loop Detector Data

This section sheds light on the prior methods that identified freeway bottlenecks using loop detector data. Loop detector data mainly consists of traffic volume, occupancy, and speed at each time interval (Caltrans, 2011). Of the studies that used loop detector data for bottleneck identification, Chen et al. (2004) and Ban et al. (2007) used speed data only. Chen et al. proposed an automated bottleneck identification method by locating the downstream and upstream position of a bottleneck on a roadway segment. The algorithm declares that there is an active bottleneck between any two locations of loop detectors, x_i and x_j , during time interval t if the following four conditions are satisfied:

- Detectors are separated by no more than 2 miles.
- The speed recorded by the upstream detector (x_i) must be at least 20 mph less than the one recorded by the downstream detector (x_j) at the same time t .
- The speed recorded by the upstream detector (x_i) is less than 40 mph, and
- If there were any two detectors at location x_k and x_l between detectors i and j , speed must continuously decrease between these two as well (Hence, $x_i \leq x_k \leq x_l \leq x_j$).

The thresholds (40 mph and 20 mph) were selected to best match the visual evaluations of speed contour plots in space and time. It was found that when a bottleneck is activated, speed drops rapidly from 60 mph (free flow speed at the study area) to below 40 mph. That is why the 20-mph minimum speed differential and the 40-mph congestion speed threshold were selected. The first condition includes locations at which the speed rises from

congested to free flow speed across several detectors, but the speed difference between each adjacent pair of detectors is small. The fourth condition was used to eliminate location where there is more than one active bottleneck or that have more complicated speed profiles. This study is one of the pioneers of automated bottleneck identification method. However, the application of this algorithm is constrained by number and spacing of the loop detectors.

Ban et al. (2007) adopted an approach to identify bottlenecks by plotting a speed contour map (SCM) in a time-speed domain (Figure 2-1). In Figure 2-1, the right vertical axis shows the tag of the detectors and their corresponding post miles (in parentheses). The direction of travel is from Detector 1 to Detector 9. This SCM was plotted using speed data collected from California’s Freeway Performance Measurement System (PeMS). The algorithm detected congested condition based on different percentile speed values from multiple days of observations. A congested condition is flagged if a selected percentile speed drops below 35 mph. Based on this condition, the SCMs were converted to Binary SCMs (0 or 1 as shown in Figure 2-2) which enables to identify the location of the bottleneck as well as the duration and queue length of the resulting congestion.

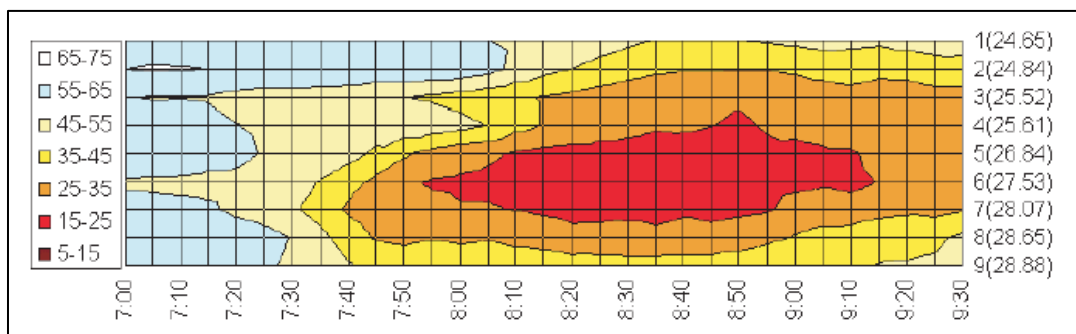


Figure 2-1: SCM of average speed from multiple days’ observation. Key on left vertical axis shows speed in miles per hour (Ban et al., 2007)

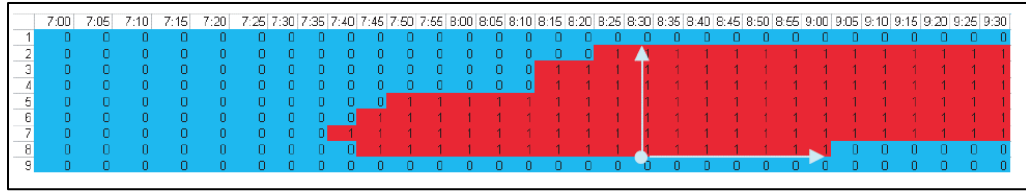


Figure 2-2: Binary SCM (Ban et al., 2007)

Jin et al. (2010) proposed an algorithm for bottleneck identification aiming to better accommodate noisy and inconsistent data at fixed loop detectors. The algorithm starts with a coordinate transformation- all data points in the flow-occupancy diagram were transformed into the URS-PUS (Uncongested Regime Shift - Perpendicular to Uncongested Regime Shift) system as shown in Figure 2-3. The main advantage of the URS-PUS system was that it helped track traffic condition changes more sensibly and descriptively. The PUS values were compared with a site-specific threshold value to identify whether congestion had occurred or not. Both statistical analyses of historical uncongested data and visual inspection of the PUS contour maps were used to determine the congestion threshold. A congestion flag was labeled for each detector location i during each time interval t if the PUS value exceeded the threshold. During the last step, a bottleneck was declared and reported when the frequency of congestion at a location was greater than a pre-defined frequency threshold (e.g., 60%- 80%). The effectiveness and robustness of the proposed algorithm were examined using three loop detector data sets collected in Madison and Milwaukee, Wisconsin. The three datasets represented various levels of data quality.

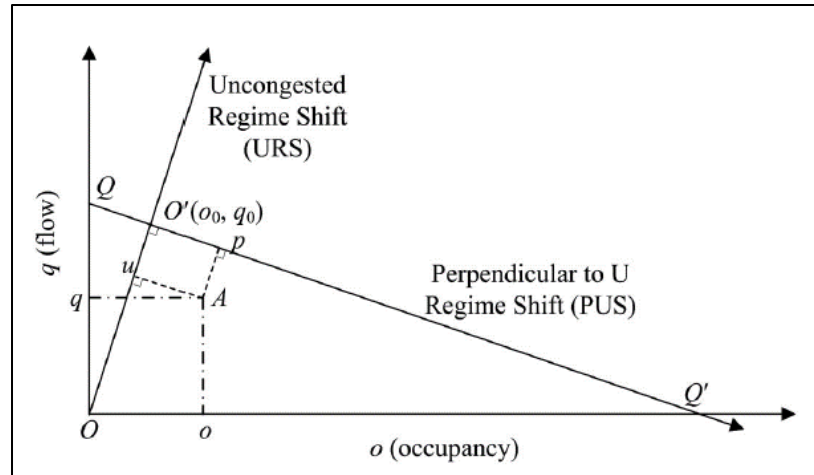


Figure 2-3: Projection from volume-occupancy to URS-PUS coordinate system (Jin et al. 2010; 2012)

Another study by Jin et al. (2012) also used the URS-PUS system. The main difference of this study with the previous one is that the threshold value was selected based on the hit rate (HR) and false alarm rate (FAR) of bottleneck activation compared to ground truth data. The major achievement of this study was its ability to establish an algorithm that is less sensitive to loop detector quality. At the same time, it enabled an evaluation of other methods of bottleneck identification based on loop detector data.

Other studies have identified congested roadway segments but had different objectives than bottleneck identification. Saberi & Bertini (2010) utilized a combination of travel time reliability related measures such as Buffer Time Index (*BTI*), Coefficient of Variation (*CV*), Planning Time Index (*PTI*), Travel Time Index (*TTI*), and Frequency of Congestion (*FOC*). These measures were used to identify and prioritize the unreliable segments along the I-5 freeway in Portland, Oregon. Although the unreliability of a roadway

is related to traffic congestion, this study did not identify the bottleneck locations associated with the unreliability. Skabardonis et al. (2003) used two different thresholds of fixed speed (35mph and 60 mph) to estimate the delays on congested freeways. Its primary objective was to estimate the total, recurrent, and non-recurrent delays on urban freeways. Detail of this study is described in a later section of this chapter.

2.1.2.2. Bottleneck Identification using Probe Vehicles' Speed Data

This section focuses on previous research works on bottleneck identification using probe vehicles' speed data. Most of the past studies that used this data source adopted a common approach. In general, they used a reference speed and defined a cut-off threshold for the ratio of the observed speed to this reference speed to identify congestion. RITIS (2016) used 60% of the free flow speed as the cut-off threshold where the free flow speed is estimated as the 85th percentile observed speed for all time periods. Additionally, RITIS identified the most downstream road segment in a series of congested road segments to locate the physical bottleneck location. A study conducted by the Florida Department of Transportation (FDOT) (2011) used a threshold of 75% of the free flow speed. In this study, the free flow speed is estimated as the 85th percentile speed for all days of a year during the overnight period (10 pm to 5 am). Although this approach was designed to identify bottlenecks, it did not mention whether it identified the actual bottleneck location or it reported the congested road segments only.

Song (2016) used the ratio of speed at capacity to free flow speed to estimate the cut-off threshold for congestion identification. This ratio was calculated using the speed-flow

curves described in the Highway Capacity Manual (HCM) (2010). Assuming a density breakpoint of 45 pc/mi/ln, different cut-off thresholds were selected for different free flow speeds. The free flow speed was selected as 5 mph over the speed limit of a road section.

Cambridge Systematics, Inc. (2011) applied Chen et al.'s (2004) algorithm with probe vehicle speed data and using a fixed speed value of 40 mph to identify traffic bottlenecks. Other studies that used probe vehicle speed data for congestion identification, but they did not focus on bottleneck activation signaling. For example, CDM Smith (2014) assessed the congestion levels along two corridors in Oklahoma (I-40 and US-69). They tested the influence of two congestion threshold values (75% and 85% of the free flow speed) on performance evaluation. The Washington State Department of Transportation (WSDOT) (2014) performed their annual corridor capacity evaluation study using the posted speed limit as their reference speed. Two cut-off thresholds were used (60% and 75% of the reference speed) to estimate the “Percent of state highway system congested” and “Severe Congestion” respectively.

2.2. Recurring Bottleneck Identification

Among the studies discussed above, only a few of them advanced towards the identification of recurring bottlenecks. In this section, the approach adopted by these studies for recurring bottleneck identification are discussed in detail.

FDOT (2011) used the combination of Frequency of Congestion (*FOC*) and Planning Time Index (*PTI*) to identify recurrent congestion. *FOC* was expressed as the percentage of times a roadway segment was congested. *PTI* was calculated as the ratio of the 90th

percentile travel time & free flow travel time. A threshold of 40% for the frequency of congestion and *PTI* values of 3 and 2 for freeways and arterials respectively were used to identify the recurrently congested freeway segments (Figure 2-4).

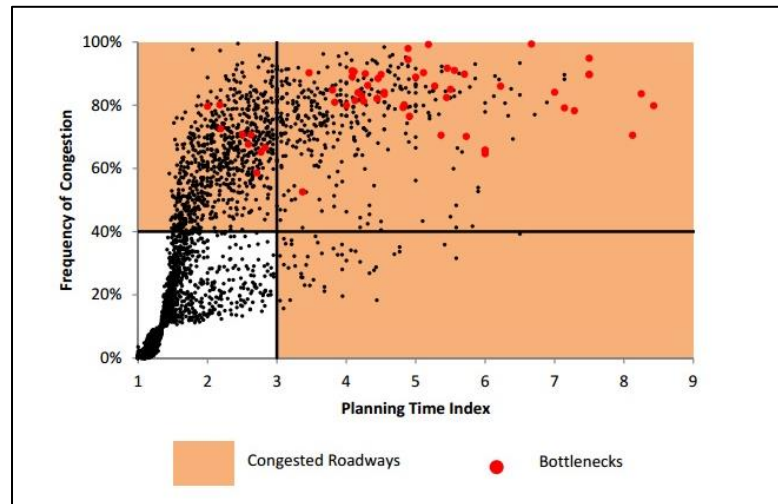


Figure 2-4: Illustration of congested roadways and bottlenecks (FDOT, 2011)

The proposed method provides a logical and straightforward basis to identify congested roadways and bottlenecks. It also incorporates travel time reliability in determining the roadway performance. However, as mentioned earlier, this study did not pinpoint the actual bottleneck location; rather it identified the recurrently congested roadways. Neither did the study indicate the time of occurrence of a recurrent congestion.

Song et al. (2016) introduced a parameter called “Average Historic Congestion Index (*AHCI*)” to differentiate between a recurring and non-recurring bottleneck. This approach determines the probability of bottleneck activation on a road segment at a given time of day. *AHCI* for the road section *i* at time *t* is calculated as the percentage of days in the study

period T where the segment-time pair (i, t) is congested. To identify a recurring bottleneck location, a threshold value for $AHCI(i, t)$ (β) and the ratio of $AHCI$ for two successive (downstream to upstream) congested road segments (γ) were used. The temporal extent of a recurring bottleneck was identified using another threshold (δ). Numerical values for these thresholds (β , γ , and δ) were selected based on several sensitivity analyses. A case study was conducted using $\beta = 50\%$, $\gamma = 2$, and $\delta = 20\%$. A total 95 bottlenecks were identified with the selected thresholds on all Interstates in North Carolina.

Chen et al. (2004) and RITIS (2016) estimated the frequency of bottleneck activation, which could be used to separate recurring bottlenecks from non-recurring ones. However, they used this parameter only as a performance measure and to rank all the identified bottlenecks. Moreover, although the bottleneck activation time was identified in these two studies, they did not use that to separate bottlenecks activating at the same location but at different periods.

2.3. Performance Measures of Bottlenecks

Different performance measures of bottlenecks developed by past studies are discussed in this section. The discussion is divided into two subsections: delay based performance measures and congestion extent based performance measures

In recent years, travel time reliability based performance measures have been widely used for freeway mobility assessment (Saber and Skabardonis (2010), Lint et al. (2008), and Schrank et al. (2015)). However, most of these studies focused on corridor level congestion and unreliability assessment rather than evaluating a single freeway bottleneck.

2.3.1. Delay Based Performance Measures

In the literature, delay based performance measures are found most common for quantifying the impact of traffic bottlenecks (Skabardonis et al. (2003), Hallenbeck et al. (2003), and Chen et al. (2004)). These studies estimated delays as the excess vehicle hours spent by vehicles over the congested section-duration pairs. Hallenbeck et al. (2003) estimated delays due to recurring congestion by contrasting the non-recurring and overall congestion. Speed and volume data from the Puget Sound metropolitan freeway system in Washington State was collected from loop measurements. A non-recurring congestion was declared when the lane occupancy was found to be 5% or higher than the median operating condition. This condition was applied for all days being studied where there was no lane blocking incident during the period of interest. For each day, the spatiotemporal position of non-recurring congestion was recorded. The delays associated with those events were also estimated based on 60 mph reference speed. Using the same reference speed, the delays associated with all days of travel is estimated. Comparing these two values, the delays due to recurring congestion only were estimated for each roadway and time interval.

Skabardonis et al. (2003) described a methodology and its application to estimate the total, recurrent, and non-recurrent delays over multiple days during peak hours on urban freeways. This study measured the random delay $D(s, t)$ over a section-duration pair (s, t) as the excess vehicle-hours spent by vehicles that are traveling at a speed below the threshold speed (V_r). A statistical model was developed to divide the total delay into the recurrent and

non-recurrent delay. The non-recurrent delay was estimated as the excess delay over the recurrent delay. The statistical model showed that:

$$\text{Recurrent delay} = E[D(s, t)|I = 0] \quad \text{Eq. 2-1}$$

$$\text{Non - recurrent delay} = \sum_{I \neq 0} \{E[D(s, t)|I] - E[D(s, t)|I = 0]\}P[I|s, t] \quad \text{Eq. 2-2}$$

Where,

I = Incident type; It can take 0, “acc”, or “non” indicating no incident, accident, and non-accident incident respectively,

$E[D(s, t) | I]$ = Expected delay at section s during time t for incident type i ,

$P[I|s, t]$ = Probability of incident occurrence of type I in section s during time t .

The model was applied to three freeway corridors (I-210, I-880, and I-10). It was assumed that the measurement of delays for the selected sample days are independent and identically distributed. Under this assumption, empirical averages and frequency counts were used to estimate the statistical averages and probability terms in the above equations. Statistical properties of the estimated delays (variance, standard error), the proportion of non-recurrent delay in total delay, and the impact of these delays on travel time distribution were analyzed for each corridor. The possible boundary error due to selecting incidents within a pre-specified congested road section (s) and peak hour (t) was recognized by the authors. The impact of incidents may go beyond the selected road section and peak period. The study suggested using a sufficiently long road section and peak period to minimize the boundary errors.

Chen et al. (2004) used both frequency of congestion and delays to rank the identified bottlenecks. The automated algorithm of bottleneck identification described above was applied to 3 months of data from 270 miles of seven freeways in San Diego, California. Delay was estimated for the identified congested road segments and periods using a 60-mph reference speed.

FDOT (2011) used a combination of two parameters: Frequency of congestion and Planning Time Index to rank the identified bottlenecks. These are the same parameters they used to identify bottlenecks as explained in the previous section. However, it is a critical issue that the report does not indicate how these two measures are combined. Neither does the study provide any indication on how the bottlenecks are ranked.

2.3.2. Congestion Extent Based Performance Measures

Very few studies were found that considered the extent of congestion (duration and queue length) induced by a bottleneck to quantify its impact. RITIS (2016) had developed an algorithm of bottleneck ranking that accounts for the number of bottleneck activation, queue length, and duration of congestion. A performance measure termed “Impact Factor” is estimated as the product of the three aforementioned parameters. Since this parameter was estimated for all the congested TMCs, it was difficult to separate the actual bottleneck TMC from the congested TMCs. Recently, this approach has been modified to highlight the bottleneck TMC only. Upon identifying a congestion occurrence (discussed in Section 2.1.2) during a 5-minute time interval, its impact is estimated as the resulting queue length. When the bottleneck ranking algorithm is applied over a period of time, all occurrences that share

an identical “head” locations (the most downstream location of a series of congested segments) are combined into an “Element”. The total impact of an Element is estimated by summing up the impacts of the associated “Occurrences”. To observe the dynamic nature of the bottlenecks, multiple elements that regularly interact with each other are combined into a “Blob”. A Typical formation of these Elements and Blobs are shown in Figure 2-5. The pink blob consists of two elements, while the purple and the blue blobs consist of a single element each.

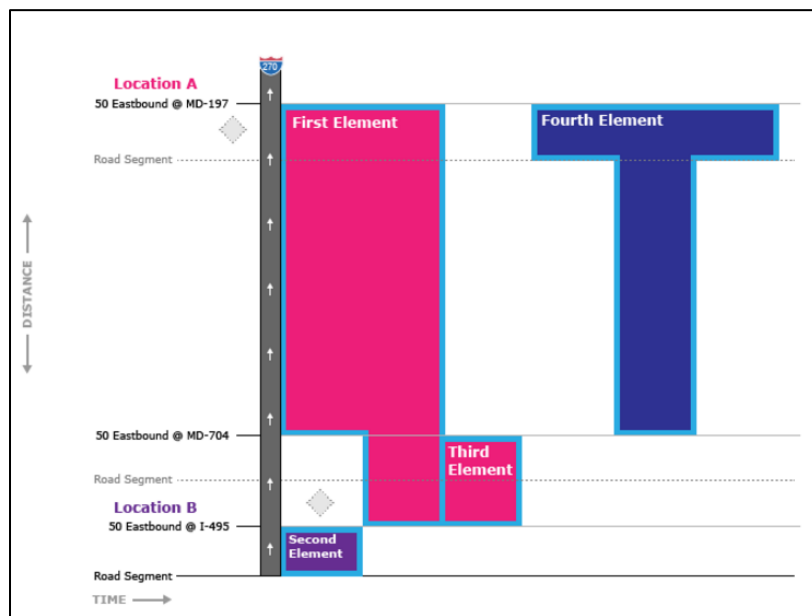


Figure 2-5: Elements and blobs generating from bottleneck activations (RITIS, 2016)

The total impact of an element estimated in this method takes the queue length, duration, and number of activation into consideration. However, this automated method does not allow to rank recurring and non-recurring bottlenecks separately. Moreover, the total

impact of an element does not differentiate the impact of bottlenecks activating at the same location but at different peak periods in a day.

Song (2016) proposed two congestion extent based measures termed as Congestion Spatiotemporal Impact Index (*CSII*) and Recurrent Bottleneck Spatiotemporal Impact Index (*RBSII*). *CSII* represents the impact of a bottleneck activation for a single day, and *RBSII* does the same for multiple days. Both *CSII* and *RBSII* were estimated as the spatiotemporal area of congestion (queue length and duration) induced by a recurring bottleneck. In addition to queue length and duration of congestion, *RBSII* accounted the probability of bottleneck activation. Fusing with crash data, the estimated congestion impact was classified into three categories: i) congestion due to a recurring bottleneck, ii) congestion due to crash, and iii) congestion due to crash within an active bottleneck. The study also discussed the distribution *CSII* and the variation of congestion duration over multiple days to characterize the identified recurring bottlenecks. However, no potential use of these characteristics of a recurring bottleneck was beckoned in this study.

2.4. Summary

The literature review presented above discusses the past research works on freeway bottleneck identification, recurring bottleneck signaling and developing performance measures to rank them. The author pointed out the following noticeable limitations of the reviewed research regarding the literature on freeway bottleneck identification and characterization.

- There are several methods available that can signal traffic congestion. However, only a few of them focused on identifying the bottleneck location and activation time.
- There is a gap in research in justifying the selection of the threshold values and reference speed for congestion detection and recurring bottleneck identification.
- Prior research on developing congestion extent based performance measure is limited in number.
- More research is needed on the day to day variability of congestion impact within a recurring bottleneck region.

CHAPTER 3: METHODOLOGY

This chapter presents a detailed description of the methods developed in this study for analyzing recurring freeway bottlenecks. The first two sections discuss the methods for identification of freeway bottlenecks and filtering of the recurring bottlenecks using probe data. The next section sheds light on developing performance measures for the identified recurring bottlenecks. The last section contains a detailed description of the models developed for predicting the probability of bottleneck activation as well as day to day variation of the impact of the recurring bottlenecks.

3.1. Congestion Detection and Bottleneck Activation Identification

The first step of identifying a bottleneck activation is detecting the resulting congestion. On a weekday m during time interval t , the reported speed at a TMC¹(i) is compared with its free flow speed to detect congestion. This approach is similar to other methods of congestion detection using probe vehicle speed data (Song (2016), RITIS (2016)). Song introduced these two parameters: Congestion Value (CV) and Congestion Index (CI) to detect a congested TMC as shown in Equation (3-1) and (3-2) below.

$$\text{Congestion Value, } CV(i, t, m) = \frac{RS(i, t, m)}{FFS(i)} \quad (3-1)$$

$$\text{Congestion Index, } CI(i, t, m) = \begin{cases} 1 & \text{if } CV(i, t, m) < CI \text{ threshold} \\ 0 & \text{otherwise} \end{cases} \quad (3-2)$$

¹ Traffic Message Channel

Where, $RS(i, t, m)$ = Reported speed (mph) for TMC i during time interval t and weekday m ,

$FFS(i)$ = Free flow speed for TMC i (mph).

The selection of the free flow speed and the CI threshold value is critical to identify congestion. In this study, the reference speed of a TMC adopted in RITIS (2016), which is the 85th percentile observed speed for all time periods is used as the free flow speed. However, the deviation of this reference speed from the posted speed limit weighted over the length of the associated TMC needs to be checked.

In this study, the threshold is selected upon investigating the sensitivity of the proposed performance measure of the bottlenecks to different threshold values. The proposed performance measure of the bottlenecks discussed in a later section (section 3.3.2) of this chapter is estimated for a series of threshold values. The appropriate threshold is the one beyond which a small change in the threshold value results in a drastic change in the performance measure.

Based on the selected CI threshold, a bottleneck activation and the extent of the resulting congestion can be identified. A graphical illustration of a CI contour figure indicating the congested region for a given weekday is shown in Figure 3-1. Here, each row represents a TMC segment (i) and each column is a 15-minute time duration (e.g. $t= 4:00$ PM means 4:00-4:15 PM).

TMC ID	Length (mi)	TMC Tag	4:00 PM	4:15 PM	4:30 PM	4:45 PM	5:00 PM	5:15 PM	5:30 PM	5:45 PM	6:00 PM	6:15 PM
125-17000	3.282	0	0	0	0	0	0	0	0	0	0	0
125N04836	0.751	1	0	1	1	1	1	1	1	1	1	0
125-04836	1.053	2	0	1	1	1	1	1	1	1	0	0
125N04837	0.846	3	0	0	0	0	1	1	1	1	0	0
125-04837	1.313	4	0	0	0	0	0	1	1	1	0	0
125N04838	0.863	5	0	0	0	0	0	0	0	0	0	0

Figure 3-1: Congestion Index (CI) contour plot in space-time domain

In Figure 3-1, congestion starts from TMC 1, and the maximum queue is extended to TMC 4. Since TMC 0 experiences free flow speeds during the congested periods (from 4:15 PM till 6:15 PM), the bottleneck responsible for the congestion is located at this TMC. TMC 1 is called the head of the bottleneck as it is the most downstream of all the congested TMCs.

3.2. Identification of Recurring Bottlenecks

To declare a bottleneck as recurring, it must activate a minimum number of days within a given study period. The parameter “Average Historic Congestion Index (*AHCI*)” was proposed by Song (2016). *AHCI* represents the fraction of days a road segment during a clock time interval is congested relative to the total number of days observed (see Equation (3-3))

$$AHCI(i, t) = \sum_{m=1}^M CI(i, t, m) / M * 100 \quad (3-3)$$

Where M is the number of days in the study period. Everything else is specified as before.

AHCI is the key to identifying recurring bottlenecks’ location and time. For a TMC segment, if the *AHCI* over a time interval is greater than or equal to a threshold value, that segment during that time can be considered to be under recurrent congestion. Like the *CI*

threshold, the threshold for *AHCI* needs to be selected after observing the sensitivity of the proposed performance measure to it. Keeping the *CI* threshold fixed, the proposed performance measure is plotted for different *AHCI* threshold values. The appropriate threshold should be selected in such a way that a small variation of its value does not cause an abrupt change in the performance measure value.

A recurring bottleneck and its spatiotemporal extent can be mapped when the *AHCI* values are plotted in a space-time region (Figure 3-2). In this plot, each cell contains the *AHCI* value of a TMC for each 15-minute duration. Cells meeting the threshold criterion are shaded yellow, and a continuous cluster of these colored cells make up a recurring bottleneck region. Some bottleneck regions may contain multiple bottleneck activations, where the queue from the downstream activation may have merged with an upstream bottleneck activation.

The identification of a recurring bottleneck region using the *AHCI* values proposed in this study is different from what Song (2016) used. Song defined different spatial and temporal patterns of recurring bottleneck regions based on the variation of *AHCI* with TMCs and time periods respectively. Thus, although the terms *CI* and *AHCI* are borrowed from Song's research, the recurring bottleneck identification and the subsequent performance measure development algorithms proposed in this study are unique.

TMC ID	Length (mi)	TMC Tag	3:45 PM	4:00 PM	4:15 PM	4:30 PM	4:45 PM	5:00 PM	5:15 PM	5:30 PM	5:45 PM	6:00 PM	6:15 PM	6:30 PM	6:45 PM
125-17000	3.336	0	5.14	8.70	8.70	13.44	18.18	15.81	14.23	22.13	26.88	27.27	31.62	28.06	18.25
125N04836	0.677	1	24.51	33.99	48.22	60.87	66.80	75.79	83.73	87.75	80.63	73.12	59.29	41.50	25.79
125-04836	1.177	2	24.51	33.99	48.22	60.87	66.80	75.79	83.73	87.75	80.63	73.12	59.29	41.50	25.79
125N04837	0.775	3	22.53	30.43	42.29	58.10	66.01	72.73	81.03	82.61	78.26	65.61	47.83	32.81	14.68
125-04837	1.366	4	22.53	30.43	42.29	58.10	66.01	72.73	81.03	82.61	78.26	65.61	47.83	32.81	14.68
125N04838	0.780	5	20.45	29.55	36.36	52.27	56.82	58.43	63.33	64.44	62.22	55.56	37.78	20.00	7.87
125N04960	0.249	6	20.45	29.55	29.55	29.55	29.55	30.43	30.43	30.43	30.43	29.55	22.53	22.53	22.53

Figure 3-2: Recurring bottleneck region using an AHCI=33% criterion

An important fact about the term “recurring bottleneck” used here needs to be discussed before closing the discussion on the method of identifying recurring bottlenecks. Even after applying the *AHCI* threshold to mark the recurring bottleneck regions, non-recurring events occurring within these regions may induce the activation of the bottleneck on many days. The effect of such events on the total impact of a recurring bottleneck region is discussed in Section 3.4.1. If non-recurring incidents occur very frequently on a freeway, it is expected that the sensitivity analysis for the *AHCI* threshold will change accordingly. If the sensitivity analysis fails to provide a robust answer regarding the selection of the thresholds, it should be left to the judgment of the policy makers.

3.3. Development of Performance Measures

This section illustrates the development of performance measures for the recurring bottlenecks identified using the process described above. Two measures are developed in this study. The Daily Impact (*DI*) accounts for the extent of a bottleneck’s impact on each day. The Recurring Bottleneck Impact Factor (*RBIF*) estimates the expected extent of the impact over multiple days.

3.3.1. Daily Impact of a Bottleneck

A variable termed Daily Impact (DI) is introduced in this study. It represents the impact of a recurring bottleneck activation during a peak period of a day. It is estimated by combining the duration and queue length of the resulting congestion within a recurring bottleneck region. The calculation of this parameter in miles-hours of congestion is expressed in Equation (3-4).

$$DI_{m,B} = \left(\frac{\Delta T}{60}\right) * \sum_{i=1}^I \sum_{t=1}^{N(i)} L_i * CI(i, t) \quad (3-4)$$

Where, $DI_{m,B}$ = Daily Impact of a recurring bottleneck B on day m ,

ΔT = Time aggregate of the reported speed (minutes)

L_i = Length of TMC I (miles)

$i = 1, 2, \dots, I$ is the TMC segment label; I represents the furthest upstream congested TMC within the recurring bottleneck region

$t = 1, 2, 3 \dots N(i)$; 15-minute time label; where $N(i)$ is the latest congested clock time within the recurring bottleneck region on TMC I ,

$m = 1, 2, 3 \dots M$; a weekday within the study period

The black rectangle in Figure 3-3 represents the region of the recurring bottleneck shown in Figure 3-2. Hence, the rectangle extends from 4:00-6:00 PM and from TMC 1 to TMC 5. The shaded cells are the extent of the congestion due to the activation of the bottleneck on weekday m . By applying Equation (3-4) to the data in Figure 3-3, it can be shown that DI for this impact region is 5.18 miles-hours per activation.

TMC ID	Length (mi)	TMC Tag	4:00 PM	4:15 PM	4:30 PM	4:45 PM	5:00 PM	5:15 PM	5:30 PM	5:45 PM	6:00 PM	6:15 PM
125-17000	3.282	0	0	0	0	0	0	0	0	0	0	0
125N04836	0.751	1	0	1	1	1	1	1	1	1	1	0
125-04836	1.053	2	0	1	1	1	1	1	1	1	0	0
125N04837	0.846	3	0	0	0	0	1	1	1	1	0	0
125-04837	1.313	4	0	0	0	0	0	1	1	1	0	0
125N04838	0.863	5	0	0	0	0	0	0	0	0	0	0

Figure 3-3: Daily Impact region of a recurring bottleneck

3.3.2. Recurring Bottleneck Impact Factor

To quantify the impact of a recurring bottleneck over multiple days, a parameter termed as Recurring Bottleneck Impact Factor (*RBIF*) is used in this study. This parameter is similar to what Song (2016) called the “Recurrent Bottleneck Spatiotemporal Impact Index”. It can be considered as the average of the Daily Impact of a bottleneck over multiple days. It is estimated as the spatiotemporal impact area of the bottleneck in across weekdays and peaks and can be expressed mathematically according to Equation (3-5).

$$RBIF \text{ per activation} = \left(\frac{\Delta T}{60}\right) * \sum_{i=1}^I \sum_{t=1}^{N(i)} L_i * AHCI(i, t) \quad (3-5)$$

Where *AHCI* (*i*, *t*) is the Average Historic Congestion Index for TMC (*i*) during time interval (*t*). Everything else is specified as before.

Here, the term “activation” requires explanation and a definite guidance for estimation. On a day, a recurring bottleneck is signaled as activated if at least 0.5 miles-hours of congestion is observed within its region (i.e. *DI*=0.5). Therefore, on an average, if a TMC of 1 mile long remains congested for at least for 30 minutes within a recurring bottleneck region, it is declared that the bottleneck was activated on that day. The reason behind

selecting 0.5 miles-hours of congestion is twofold. First, it prevents any discrete congestion (congestion with small duration or queue length which could be the result of erroneous observation) from declaring an activation of a bottleneck. The second reason is to maintain a balance between the number of activation and no activation days within a study period, which is explained in detail in section 3.4 of this chapter.

3.3.3. Overall RBIF

The overall *RBIF* of a recurring bottleneck region is estimated by combining the *RBIF* per activation with the total number of activation within the study period. It serves as a distinct performance measure for the recurring bottlenecks. The Overall *RBIF* (in miles-hours of congestion) for a given number of observation days can be obtained by multiplying the number of activations within those days with the *RBIF* per activation as shown in Equation (3-6).

$$\text{Overall RBIF} = \text{RBIF per activation} * \text{Number of activations} \quad (3-6)$$

3.4. Models for Predicting the Activation and Daily Impact of a Recurring Bottleneck

This section describes the methods for predicting the activation of a recurring bottleneck and the variation of its Daily Impact (*DI*) over multiple days. It is mentioned earlier that the successful implication of low-cost bottleneck mitigation strategies depends significantly on understanding the activation pattern of the bottlenecks and the impact of the bottlenecks when they activate. Thus, the development of the methods for obtaining these two key information is discussed separately.

To further explain these two concepts this thesis is attempting to predict (activation of a bottleneck and the *DI* when it activates) the distribution of the *DI* values for two recurring bottleneck regions over 250 days are shown in Figure 3-4 below.

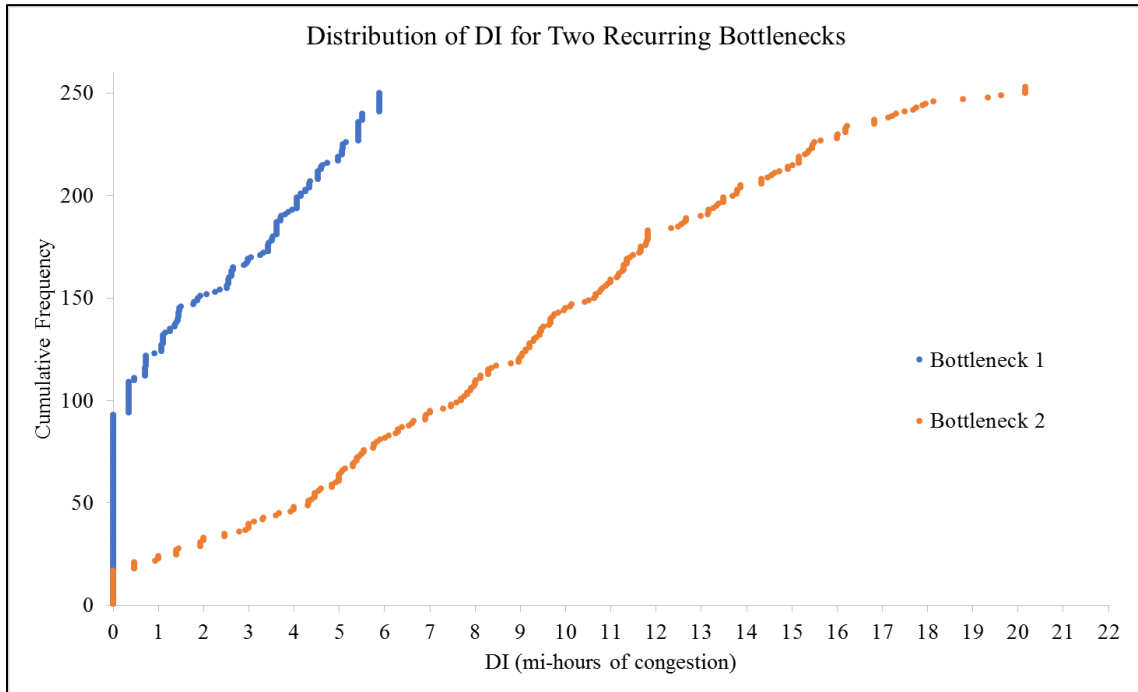


Figure 3-4: Distribution of DI for two different recurring bottlenecks

Figure 3-4 shows that different recurring bottlenecks may have different frequency of activation and different variation of *DI*. In this figure, Bottleneck 1 has a low *DI* range (0 to ≤ 0.6 mi-hours) and small variance of *DI* (4.2 (mi-hours)²) with a significant number of days with *DI* < 0.5. On the other hand, Bottleneck 2 in Figure 3-4 has a wide range (0 to ≤ 21 mi-hours) and high variance (27 (mi-hours)²) of *DI*. The number of days with *DI*<0.5 is also small for this bottleneck. These observations indicate that it is important to

predict both propensity for activation and impact to characterize the recurring bottlenecks since their activation probability and range of *DI* may vary significantly.

3.4.1. Predicting the Activation of a Bottleneck

To predict the activation of a recurring bottleneck on a given day, the day of the week, the time of the year, and occurrence of any non-recurring event are regarded as the critical predictor variables. The justification for selecting these variables are described in a later section. Since most of the predictors are binary, one might think of using descriptive statistics to estimate the probability of activation of a recurring bottleneck. However, it is important to realize that descriptive statistics can only be used to describe the population or data set under study. Hence, the results cannot be applied to any other group of data. (Loether & McTavish, 1980); (Crossman, 2017).

Since the aim is to understand the effect of different predictors and their significance in predicting the activation of a bottleneck, techniques available in inferential statistics are sought in this study. A binary logistic regression model that estimates the probability of a binary response (activation or no activation) is proposed here. The outcome of a logistic regression model is a predicted mean that represents the probability of a bottleneck activation on a day (Hilbe, 2009).

3.4.2. Predicting the *DI* of a Bottleneck

Once the probability of activation of a recurring bottleneck on a day is estimated, the next task is to predict the *DI* of the bottleneck when it activates. A stepwise linear regression technique is proposed in this study. Various binary and numerical predictors are provided

with the model, and the stepwise technique chooses the most significant predictors (level of significance less than 1) based on their t-statistics.

3.4.3. Predictor Variables for the Models

To select the predictor variables of the models, parameters upon which the activation of a bottleneck and variability of the *DI* is dependent need to be identified. Also, the availability of the data is an important issue here. First, the day of the week appears to be a significant variable when the distribution of the *DI* for a typical bottleneck is plotted for 66 weekdays as shown in Figure 3-5.

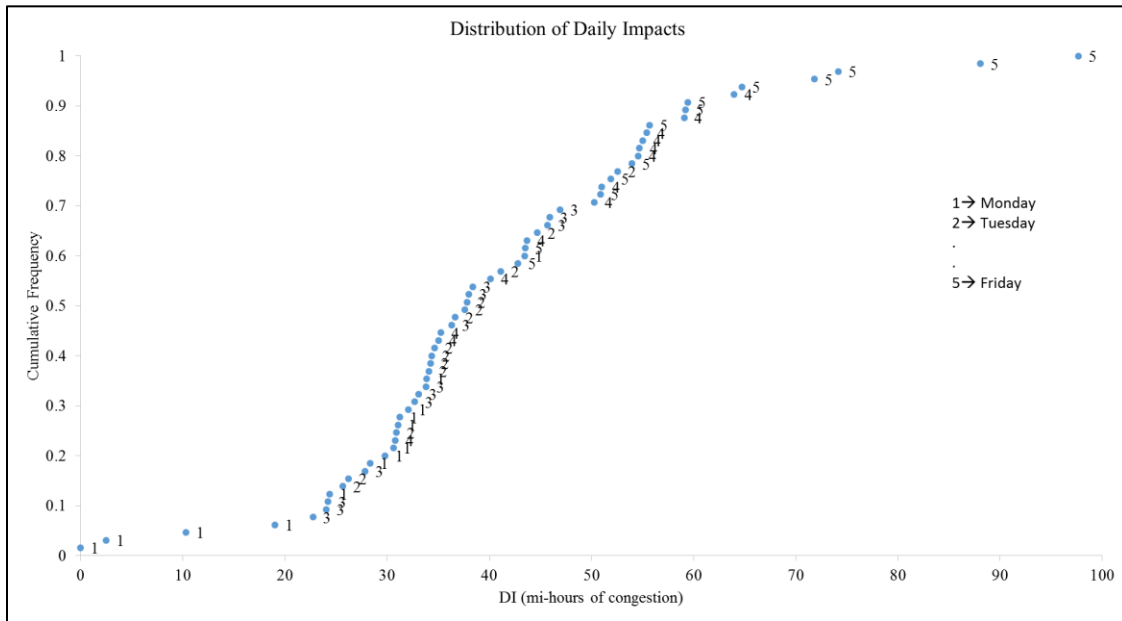


Figure 3-5: Distribution of *DI* for a typical bottleneck with the corresponding day of week

From Figure 3-5, it is apparent that influence the variation of *DI*. During most Fridays, the magnitude of the *DI* was high. In contrast, Mondays yielded a small magnitude

of *DI*. Thus, the days are considered as important predictor variables. Day of week is introduced in the model as a binary variable with Monday as the reference level.

Time of the year also is an important predictor of bottleneck activation and the variability of *DI* on each day. It can be introduced in the model in terms of months, quarters of a year (e.g. Jan-Mar as the first quarter), or seasons. Using months to represent the time of the year includes eleven variables in the model and leads to a significant reduction in the degrees of freedom. It may over-fit the models as well. Between the seasons and quarters of the year, the latter one is selected in this study since quarters of a year are easy to define. There are four quarters in a year where each contains 3 months, and the first quarter is provided as the reference level in the models.

Non-recurring events on the roadways can be a potential predictor that can affect the *DI* of a bottleneck. Its effect on traffic congestion has been recognized in past studies (Dowling, Skabardonis, Carroll, & Wang, 2004). These events can be divided into two major types: weather-related events and vehicular crashes or incidents on the roadways. Both types of non-recurring events are included in the linear regression models.

Weather condition is introduced as a binary variable with two levels: good weather and poor weather. Good weather is selected as the reference level in the model. Considering all these variables, the general forms of the logistic and linear regression model are shown in Equation (3-7) and Equation (3-8) below.

$$\ln\left(\frac{Y_i}{1-Y_i}\right) = \beta_0 + \beta_1 * X_{Tue,i} + \beta_2 * X_{Wed,i} + \beta_3 * X_{Thu,i} + \beta_4 * X_{Fri,i} + \beta_5 * X_{quarter2,i} + \beta_6 * X_{quarter3,i} + \beta_7 * X_{quarter4,i} + \beta_8 * X_{incident\ occurrence,i} + \beta_9 * X_{inclement\ weather,i} \quad (3-7)$$

$X_{inclement\ weather,i}$

$$\begin{aligned} DI_i = & \beta_0 + \beta_1 * X_{Tue,i} + \beta_2 * X_{Wed,i} \\ & + \beta_3 * X_{Thu,i} + \beta_4 * X_{Fri,i} + \beta_5 * X_{quarter2,i} + \beta_6 * X_{quarter3,i} \\ & + \beta_7 * X_{quarter4,i} + \beta_8 * X_{incident\ occurrence,i} + \beta_9 \\ & * X_{inclement\ weather,i} \end{aligned} \quad (3-8)$$

Here,

Y_i is the probability of the dependent variable equaling a "bottleneck activation" against "no activation" of a bottleneck on observation day i ,

β_0 is the intercept term,

β_1, \dots, β_9 are the coefficients for the predictor variables.

$X_{incident\ occurrence,i} = 1$ if there is an incident within the recurring bottleneck region on day i ,

0 otherwise

$X_{inclement\ weather,i} = 1$ if weather is inclement on day i ; 0 otherwise;

$X_{Tue,i} = 1$ if day i is a Tuesday; 0 otherwise;

$X_{Wed,i} = 1$ if day i is a Wednesday; 0 otherwise;

$X_{Thu,i} = 1$ if day i is a Thursday; 0 otherwise;

$X_{Fri,i} = 1$ if day i is a Friday; 0 otherwise;

$X_{quarter2,i} = 1$ if day i is on the 2nd quarter of the year; 0 otherwise;

$X_{quarter3,i} = 1$ if day i is on the 3rd quarter of the year; 0 otherwise;

$X_{quarter4,i} = 1$ if day i is on the 4th quarter of the year; 0 otherwise;

Based on the level of detail of the incidents' information, the linear regression models are further divided into two categories: the base model and the full model. A base model includes a binary variable to indicate whether or not there was an incident within the recurring bottleneck region. Equation 3-8 represents the base form of the linear regression model. A full model replaces the binary variable indicating the occurrence of an incident and includes detailed information on the incident. Four numeric parameters related to an incident are used in the full model. They are: duration of the event, number of lanes closed because of the event, start time of the event, and location of the event. Details about these four parameters are described below.

Duration and the Number of Lanes Closed: These two factors related to an event are assumed to have a significant impact on the DI of a recurring bottleneck on a day. Duration, as a numeric variable is estimated as the difference between the start and end time of an event within the recurring bottleneck region in minutes. Note that only the events that merge with the recurring bottleneck region are selected in the analysis. Although the start (St_e) or end time (End_e) of the selected events can be outside of the activation time of the recurring bottleneck region, St_e and End_e will be estimated only from the beginning and to the ending

of the bottleneck activation respectively. The calculation of St_e , End_e and the duration of the event is expressed in Equation (3-9), (3-10), and (3-11) correspondingly.

$$St_e = \begin{cases} St_b, & \text{if } St_e < St_b \\ St_e, & \text{otherwise} \end{cases} \quad (3-9)$$

$$End_e = \begin{cases} End_b, & \text{if } End_e > End_b \\ End_e, & \text{otherwise} \end{cases} \quad (3-10)$$

$$\Delta t_e = (End_e - St_e) * 60 \quad (3-11)$$

Where, St_e = Start time of an event (in hours)

St_b = Start time of the recurring bottleneck activation (in hours)

End_e = End time of an event (in hours)

End_b = End time of the recurring bottleneck activation (in hours)

Δt_e = Estimated duration of an event (in minutes)

The number of lanes closed is treated as a numeric variable as well. It has a direct relationship to the capacity of a roadway section. Both duration and number of lane closed are zero if there is no event on a day.

Start Time of an Event: When all other variables remain constant, it is hypothesized that the effect of a non-recurring event on the *DI* is largest when it is active at the beginning of the recurring bottleneck activation period. On the other hand, the effect of an event on the *DI* is likely to be reduced if it starts toward the end of the recurring bottleneck activation period.

The statements are illustrated graphically in Figure 3-6 below.

TMC ID	Length (mi)	TMC Tag	3:45 PM	4:00 PM	4:15 PM	4:30 PM	4:45 PM	5:00 PM	5:15 PM	5:30 PM	5:45 PM	6:00 PM	6:15 PM	6:30 PM	6:45 PM
125-17000	3.336	0	5.14	8.70	8.70	13.44	18.18	15.81	14.23	22.13	26.88	27.27	31.62	28.06	18.25
125N04836	0.677	1	24.51	33.99	48.22	60.87	66.80	75.79	83.73	87.75	80.63	73.12	59.29	41.50	25.79
125-04836	1.177	2	24.51	33.99	48.22	60.87	66.80	75.79	83.73	87.75	80.63	73.12	59.29	41.50	25.79
125N04837	0.775	3	22.53	30.43	42.29	58.10	66.01	72.73	81.03	82.61	78.26	65.61	47.83	32.81	14.68
125-04837	1.366	4	22.53	30.43	42.29	58.10	66.01	72.73	81.03	82.61	78.26	65.61	47.83	32.81	14.68
125N04838	0.780	5	20.45	29.55	36.36	52.27	56.82	58.43	63.33	64.44	62.22	55.56	37.78	20.00	7.87
125N04960	0.249	6	20.45	29.55	29.55	29.55	29.55	30.43	30.43	30.43	30.43	29.55	22.53	22.53	22.53

Figure 3-6: Non-recurring events activating at different times within the recurring bottleneck region

The same *AHCI* contour shown in Figure 3-2 is used in Figure 3-6. According to the hypothesis stated above, event on day m_1 is expected to have a more significant impact on the *DI* on that day compared to what the event on day m_2 would have on the *DI* if everything else is held unchanged. This assumption is instinctive since the event on day m_1 has its full duration within the recurring bottleneck region where the event on day m_2 has its duration towards the finishing part of the impact of the bottleneck.

Therefore, the start time of a non-recurring event needs to be expressed in such a way that it has a monotonic relationship with its anticipated effect on the *DI* of the bottleneck. Equation (3-12) is used to express the start time of an event relative to the beginning of the recurring bottleneck activation.

$$St = 1 - \left(\frac{St_e - St_b}{\Delta t_b} * 60 \right) \quad (3-12)$$

Where, St = Start time of an event relative to the beginning of the recurring bottleneck activation (a unit less parameter)

Δt_b = Maximum expected duration of recurrent congestion (in minutes). Everything else is specified as before.

Expressing the start time of the events in this manner makes it monotonic to the hypothetical effect on the *DI* of the recurring bottleneck on a weekday. For example, if the expected activation start and end time of the recurring bottleneck is 4:00 PM and 6:00 PM respectively, and an event is reported to occur at 4:01 PM, St_r will be close to 1. In contrast, if an event occurs just before the end of the bottleneck activation (e.g. 5:59 PM), St_r will be close to 0. When there is no event within the bottleneck region on a day, $St_r = 0$ is reported.

Location of an Event: The location of a non-recurring event occurred within a recurring bottleneck region is also considered to be a critical factor for the *DI* on that day. Intuitively, if an event occurs just at a bottleneck location, it is expected to have a significant increasing effect on the *DI*. On the contrary, if the event occurs towards the end of the queue emanating from the bottleneck, it may have less effect on the *DI* for that day. Equation (3-13) shows how the location of an event is estimated relative to the location of the recurring bottleneck activation so that it has a monotonic effect on the *DI*.

$$Loc = 1 - D_e / Q \quad (3-13)$$

Where, *Loc* = Location of event *k* occurred on day *m* relative to the recurring bottleneck location (a unit less parameter)

D_e = Distance between the event *k* and bottleneck location on day *m* (miles)

Q = Maximum expected length of the queue emanating from the recurring bottleneck (miles)

Thus, the relative location estimated per Equation (3-13) has a monotonic effect on the *DI*. For example, if an event occurs just upstream of the bottleneck, *Loc* will be close to 1. On the other hand, if an event occurs just downstream of the maximum expected queue, *Loc* will be close to 0. When there is no event reported on a day, *Loc* =0 is reported.

3.4.4. Development and Validation of the Models

3.4.4.1. Time Frame for Model Development and Validation

In selecting a sample for developing the models discussed above, it is important to examine the issues regarding the appropriate sample size. First, it is well-known that traffic bottlenecks are very dynamic in nature. Therefore, selecting a very large sample of days may affect the bottleneck identification process if the bottleneck shifts significantly within those days. On the other hand, using a small number of days may affect the degree of freedom of the models. Second, developing a logistic regression model for a bottleneck that does not have a sufficient sample for both cases (activation and no activation) might be troublesome. Bottleneck 2 shown in Figure 3-4 is a good example of such bottleneck where the number of “no activation” cases is very small. The appropriate sample size for a logistic regression model also depends on the number of predictors used. For a logistic regression model with categorical variables, each level of the variables should have both cases (activation or no activation) at least once (University of California, Los Angeles, 2017). Moreover, a rule of thumb about the minimum sample size for a logistic regression model requires sample size for each case to be greater than or equal to about ten times of the number of predictor variables used (Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996). Third, it may be

difficult to develop a linear regression model for the bottlenecks that exhibit a small variation of their *DI* values. Bottleneck 1 shown in Figure 3-4 is a good example of such case.

Considering these issues, the valid weekdays (excluding weekends and holidays) of the calendar year 2016 were selected for developing and validating the models. It is assumed that throughout this period, the bottleneck location and other characteristics remained unchanged. The selected data set includes about 250 weekdays (up to 5 days were found missing for some bottlenecks). To develop the logistic and linear regression models, a “hold-out cross validation technique” (Refaeilzadeh, Tang, & Liu, 2009) is sought in this study. This technique involves partitioning the total data set into two parts: training and validation datasets. In this study, the training dataset is obtained by randomly selecting 75% of the weekdays from the calendar year of 2016, and the models are developed using this dataset. The remaining 25% weekdays are used to validate the models. The selection of the validation dataset (25%) is repeated multiple times to achieve an unbiased estimate of the validation results. The process is further described in the next section

3.4.4.2. Assessment of the Models

Both development and validation of the proposed models need to be evaluated to assess their significance, goodness of fit, and predictability. The logistic regression models can be assessed by common statistical tests. The overall fitness of the model can be evaluated by the McFadden’s R-squared (McFadden & Zarembka, 1974) which is also known as the Pseudo R-squared. It is estimated as shown in Equation (3-14) below. The significance of the parameters can be checked using a statistical Chi-squared test.

$$Pseudo R - squared = 1 - \frac{Log(L_c)}{Log(L_{null})} \quad (3-14)$$

Where L_c denotes the (maximized) likelihood value from the current fitted model, and L_{null} denotes the corresponding value for the null model.

To evaluate the validation of the logistic regression models, it is proposed to use a 2 by 2 contingency table of dependency and conducting a Chi-square test (Everitt, 1992). Such a table displays the frequency distribution of the variables. If the Chi-square test is passed with sufficient significance, it can be said that the predicted and observed activations are dependent, which essentially validates the model.

The linear regression models can be evaluated using their adjusted R-square values. The adjusted R-square is a statistical measure of how close the data are to the fitted regression line. For assessing the validation of the linear regression models, two levels of validations are proposed in this study. A higher level validation involves assessing the overall bias of the DI induced by a linear regression model when it is applied to a test data set. The “K-fold Cross Validation Technique” is applied for this task (Refaeilzadeh, Tang, & Liu, 2009). It involves calculating the mean bias and the root-mean-squared-error (RMSE) for the validation data and repeating the process for k times (Equation (3-15) and (3-16)).

$$Bias_k = mean(DI_{obs,i\epsilon k} - DI_{pred,i\epsilon k}) \quad (3-15)$$

$$RMSE_k = \sqrt{(mean(DI_{obs,i\epsilon k} - DI_{pred,i\epsilon k})^2)} \quad (3-16)$$

Where,

$DI_{obs,i\epsilon k}$, $DI_{Pred,i\epsilon k}$ = Observed and predicted DI respectively on a weekday i that belongs to the test dataset for kth iterations.

To verify that the proposed method is providing an unbiased prediction of the DI , the average bias obtained from all iterations should be close to zero. To inspect whether or not this mean of the bias and its variation cover zero, the following condition needs to be satisfied.

$$Mean\ Bias + 2 * St.\ Error\ of\ Bias > 0 > Mean\ Bias - 2 * St.\ Error\ of\ Bias$$

The average of the RMSE values from all iterations estimated by Equation (3-16) is the best estimate of a model's out-of-sample prediction error. The standard error of these RMSE values also needs to be reported that represents the uncertainty in the estimate of out-of-sample prediction error.

In order to observe the variation of the bias with the predicted DI , a CUMulative RESidual (CURE) plot is suggested in this study (Lin, Wei, & Ying, 2002). This CURE plot provides a more in-depth assessment of the models. CURE plot reveals the bias of the prediction error of a model. In this study, the CURE analysis is conducted to inspect the bias for different values of the fitted and observed DI . To complete the task, the developed linear regression model is applied to a randomly selected 25% validation data, and the residuals are estimated as shown in Equation (3-17).

$$Residual_i = DI_{obs,i} - DI_{Pred,i} \quad (3-17)$$

The residuals are sorted based on the ascending order of *DI* (either fitted or observed *DI*). When the residuals are so cumulated, and these cumulated residuals are plotted against the *DI* (either fitted or observed, the one based on which the sorting was done), plots of this kind are called the CURE plot. Such a plot reveals where the model is overestimating or underestimating the *DI*, where potential outliers exist, and where the model prediction is unbiased.

3.5. Scope of the Study

In this study, all interstate facilities within the Wake and Durham counties of North Carolina are selected as the study site. These interstates are shown in Figure 3-7 below.

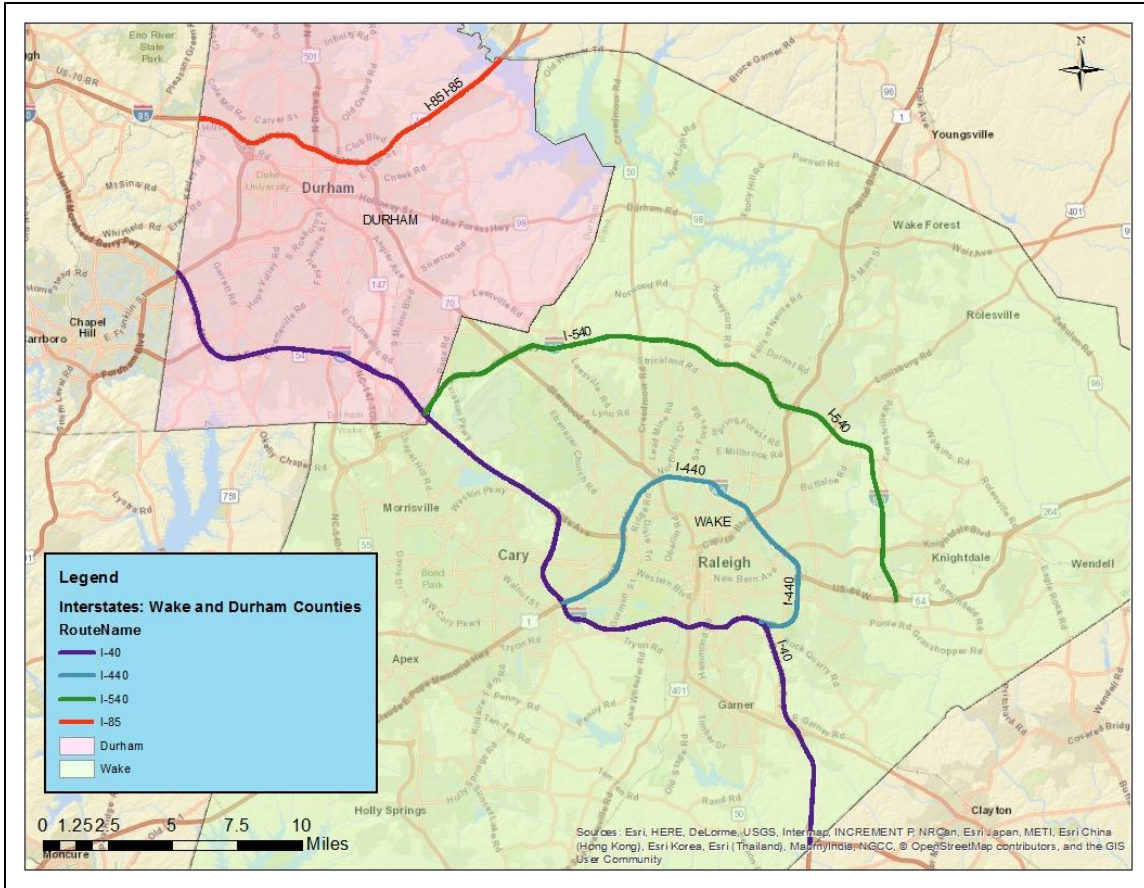


Figure 3-7: Selected Interstates within Wake and Durham Counties, NC

Three out of four interstates within these two counties have east-west alignment. The individual centerline lengths of the interstates are i) I-40: 40.6 mi, ii) I-540: 25.5 mi, iii) I-440: 16.7 mi, and iv) I-85: 13.1 mi, in total 95.9 miles. The next chapter contains results and discussions of the application of the proposed methods in this study area.

CHAPTER 4: CASE STUDY: METHODOLOGY APPLICATION TO BOTTLENECK IDENTIFICATION AND RANKING

This chapter presents an application of the recurring bottleneck identification and ranking method discussed in the preceding chapter. In addition, the contributing circumstances and a comparison of the ranking obtained by different methods for the identified bottlenecks are discussed here. The chapter begins with a sensitivity analysis of the *RBIF* performance measure to the thresholds used in congestion detection and bottleneck identification. The location and ranking of the identified bottlenecks are illustrated in the next section. The next part presents a comparison of the bottleneck rankings obtained by the proposed method and by the method established by the Vehicle Probe Project Suite of the CATT Lab (2016). Following that, a detailed investigation of the attributes and contributing factors for several bottlenecks is presented. Some special bottleneck cases are discussed in the next section.

4.1. Selection of the Thresholds for Congestion Detection and Bottleneck Identification

In the previous chapter, it was proposed that different thresholds used in Equation (3-1) to Equation (3-3) for congestion detection and recurring bottleneck identification need to be selected based on a sensitivity analysis. To investigate the sensitivity of the proposed performance measure of the bottlenecks to the threshold values used, one might consider estimating the number of recurring bottlenecks obtained using different threshold values. Then, simply looking at the thresholds at which the analysis yields an unusual number of

bottlenecks would suffice to understand the sensitivity. However, there are some issues with this approach. Once the congestion detection (*CI* threshold) is raised significantly, the recurring bottleneck regions at different locations of a route start to merge with each other which makes it difficult to ascertain the actual number of bottleneck along that route. Similar phenomena were observed when the threshold for recurring bottleneck identification (*AHCI* threshold) was tested against the number of bottlenecks. To tackle this challenge, rather than considering the number of bottlenecks along a route, the *RBIF* value per activation for the entire length of each roadway is checked against the thresholds for *CI* and *AHCI*. This sensitivity analysis for the two thresholds is presented in the following two subsections.

4.1.1. Threshold for Congestion Index (*CI*)

In Figure 4-1, the sensitivity of *RBIF* (*miles-hours of congestion per activation*) estimated for the four study routes described in the previous chapter to the threshold values of *CI* is plotted. *CI* values are gradually increased from 0.1 to 0.9, and the resulting *RBIF* are estimated using a fixed threshold for *AHCI* (in this case 33%).

In addition, the *RBIF* value corresponding to a fixed speed threshold (45 mph) is shown using the dotted lines. The reason behind testing the effect of a fixed speed threshold of 45 mph is that using a speed threshold of 40 to 45 mph is a common practice for congestion detection on freeways (Day, et al., 2014); (Cambridge Systematics Inc., 2011); (Lynn Peterson, 2014).

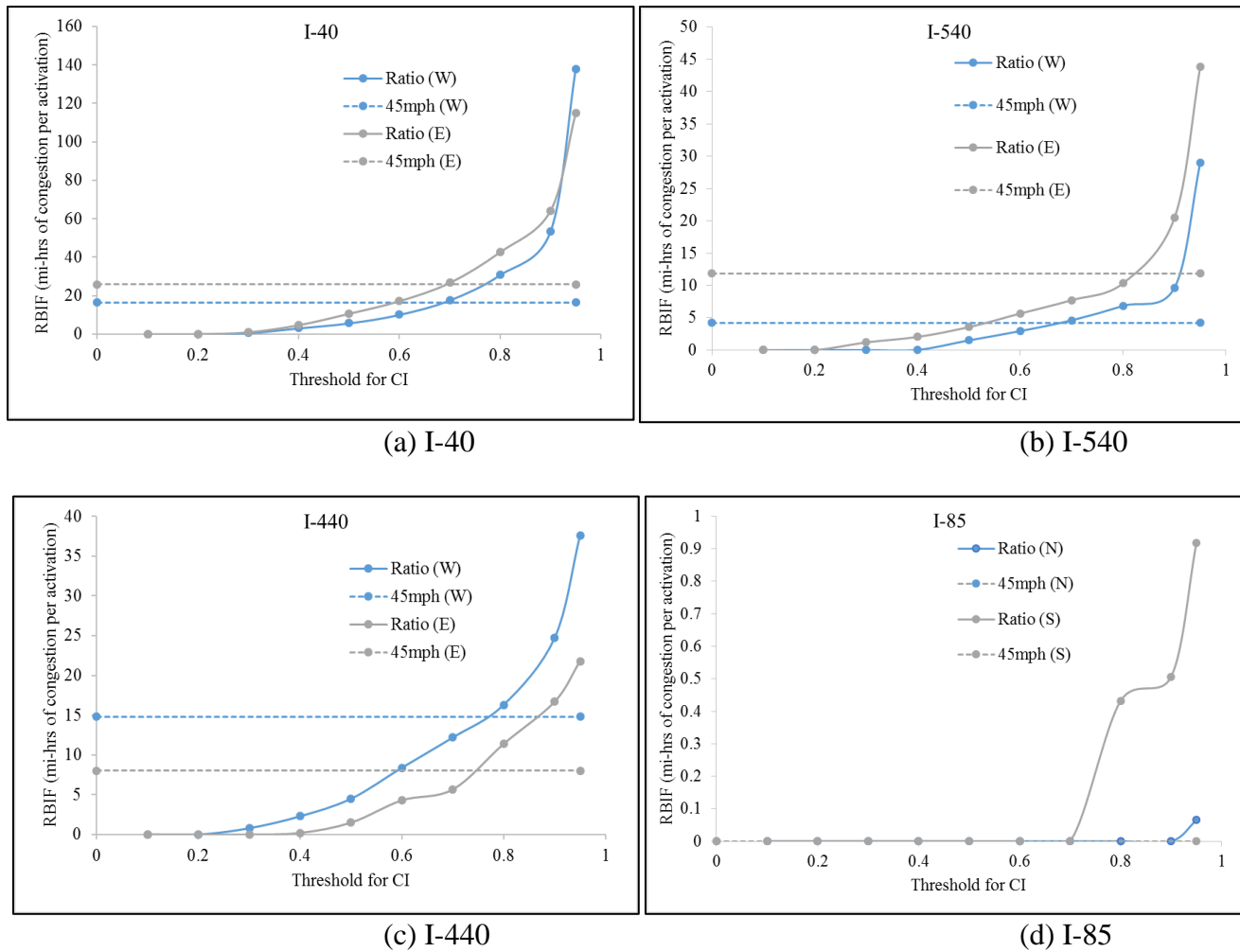


Figure 4-1: RBIF per activation for different CI thresholds on different routes for fixed AHCI = 33%

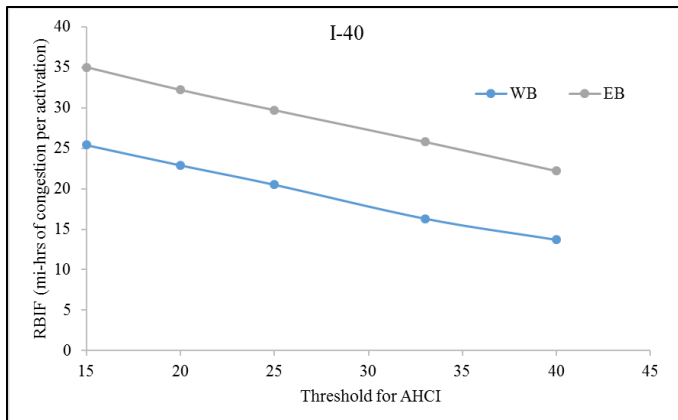
For both directions of I-40 and I-540, the total *RBIF* per activation drastically increases when the threshold for *CI* exceeds 0.8. For I-440 eastbound, such a drastic increase is observed at a threshold greater than 0.7. In case for I-440 westbound, the increase of *RBIF* is almost linear between the threshold values of 0.5 and 0.8. In I-85 southbound, no sign of *RBIF* was found for this fixed *AHCI* threshold before raising the *CI* threshold above 0.7 which is followed by a sharp bend between 0.8 to 0.9. For the northbound route of I-85, very little *RBIF* is generated only when the threshold was as high as 0.95. For these two routes, a lowering the *AHCI* could result in the appearance of *RBIF* value at a much lower threshold of *CI* than 0.7 or 0.9. In other words, if the constraint of being a recurring bottleneck is flexed, more *RBIF* on I-85 would be generated.

For both directions of I-40, use of the fixed speed threshold (45 mph) yields almost the same *RBIF* as what a *CI* threshold value of 0.7 does. For I-540 east and westbound, this fixed speed threshold gives a similar result to what *CI* threshold of 0.82 and 0.65 does respectively. For I-440 east and westbound, 45 mph speed threshold gives a similar result to what *CI* threshold of 0.82 and 0.75 does respectively. No *RBIF* was observed using this fixed speed threshold for both directions of I-85.

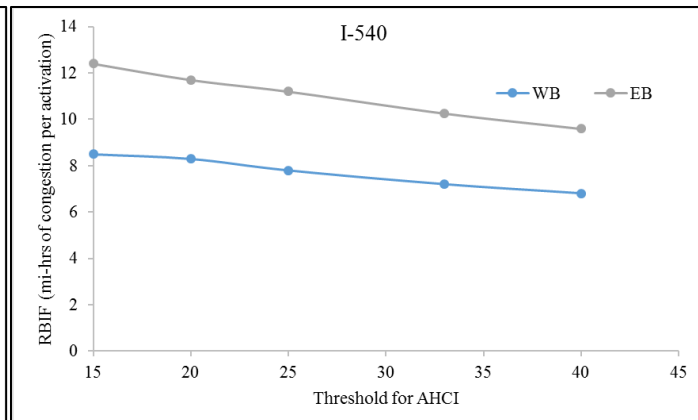
From the above discussion, it is apparent that a *CI* threshold within 0.7 to 0.8 conforms to a 45 mph fixed speed threshold for most of the routes. Moreover, a *CI* threshold within this range does not result in any unusual value of *RBIF* per activation for the routes. Rather, for some interstates it is the limiting threshold range at which the *RBIF* rapidly increases. A *CI* threshold of 0.7 is finally selected for congestion detection in this study.

4.1.2. Threshold for Average Historic Congestion Index (*AHCI*)

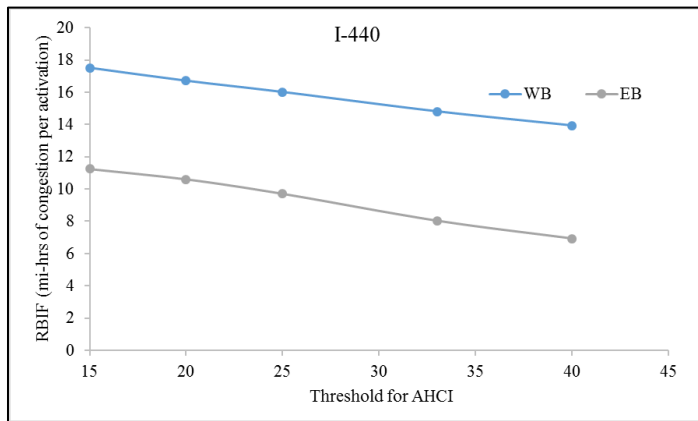
Similar to the prior section, the sensitivity of *RBIF* per activation to the threshold value of *AHCI* is tested in this section. The total *RBIF* per activation for different *AHCI* thresholds (varying from 15% to 40%) are plotted in Figure 4-2. A fixed speed threshold of 45 mph for the *CI* is used for estimating the *RBIF*.



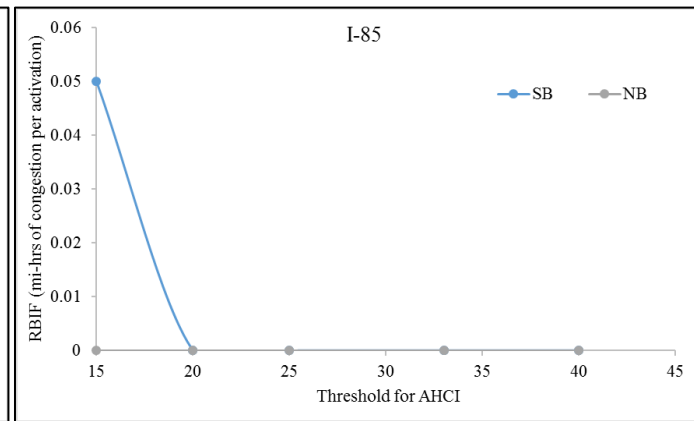
(a) I-40



(b) I-540



(c) I-440



(d) I-85

Figure 4-2: RBIF per activation for different AHCI threshold for fixed CI=0.7

It should be recalled that *AHCI* is the threshold applied to extract the recurring bottlenecks only. Thus, raising the *AHCI* threshold means tightening the constraint resulting in a drop in the magnitude of the *RBIF*. It is evident from Figure 4-3 that unlike the threshold for *CI*, the threshold for *AHCI* within this range does not create any drastic change in the *RBIF* value for all the routes but I-85 southbound. For both directions of I-85, the magnitude of *RBIF* is very small even for such a small value of *AHCI* threshold (0 for the northbound route). The variation of *RBIF* per activation with *AHCI* threshold seems to be almost linear within this range. This makes it difficult to suggest a limiting range for the *AHCI* threshold. Since the sensitivity analysis of *AHCI* threshold for the freeways in question is not showing any critical point, the selection of this threshold becomes a policy question. An *AHCI* threshold of 33% appears a reasonable one for the given study area. In this study, an *AHCI* threshold of 33% is selected for identifying the recurring bottlenecks. The interpretation of this threshold is that on an average, a recurring bottleneck must be activated in at least one out of three weekdays.

4.2. Identified Bottlenecks and their Performance Measures

As the selection of the thresholds for *CI* and *AHCI* is confirmed, the process of recurring bottleneck identification and estimation of the performance measures is applied to the study area and for the weekdays in the calendar year of 2016. The proposed method identified fifteen recurring bottlenecks which are listed in Table 4-1 along with their corresponding peak period, performance measures, and ranking. The ranking is based on the overall *RBIF* value for the entire year.

Table 4-1 shows that out of the fifteen recurring bottlenecks, eight of them are located on I-40. The number of PM peak bottlenecks is significantly higher than the number of the AM peak bottlenecks (9 vs. 6). *RBIF* for these bottlenecks ranges from approximately 0.5 miles-hours of congestion to greater than 12 miles-hours of congestion. Although I-540 has only two bottlenecks with one on each route, they are ranked very high (1st and 4th by the Overall RBIF).

The probability of activation for the bottlenecks varies from 0.52 to 0.95. However, one might ask why the minimum probability of activation being observed from the table is 0.52 where the AHCI threshold was 33%. It should be clarified that AHCI threshold represents the criteria for a TMC and time period to be considered under a recurring bottleneck region. The number of activations is a different parameter than the AHCI. The number of activations of a bottleneck is usually higher than the maximum AHCI within the recurring bottleneck region.

Table 4-1: Performance measures and ranking of recurring bottlenecks in the study area for weekdays in 2016

Rank	Location of Bottleneck	Activation Peak	RBIF (mi-hours of congestion) Per Activation	Probability of Activation	Overall RBIF (mi-hours of congestion)	Contributing Circumstance
1	I-540 EB at Falls of Neuse Rd. MM 14	PM	12.70	0.94	2985	Merging from On Ramp
2	I-40 EB at Harrison Ave. MM 287	PM	11.59	0.95	2758	Merging from On Ramp
3	I-40 WB at I-440 MM 293	AM	10.75	0.94	2516	Merging from On Ramp
4	I-540 WB at Leesville Rd. MM 7	AM	9.10	0.88	1993	Merging from On Ramp
5	I-40 EB at US-70 MM 306	PM	7.43	0.93	1723	Merging from On Ramp
6	I-440 WB at Wade Ave. MM 4	AM	7.20	0.89	1598	Merging from On Ramp
7	I-440 WB at Western Blvd. MM 2	PM	6.17	0.93	1172	Merging from On Ramp
8	I-40 WB at NC-55 MM 278	PM	5.05	0.89	970	Merging from On Ramp
9	I-440 EB at US-70 /Glenwood Ave. MM 7	PM	4.37	0.90	863	Queue Back from Off Ramp of US-70
10	I-40 EB at Rock Quarry MM 300	PM	3.82	0.52	802	Merging from On Ramp
11	I-40 WB at Ramp from I-440 MM 293	PM	1.52	0.78	293	Merging from On Ramp
12	I-440 EB at Melbourne Rd. MM 1D	AM	1.33	0.73	243	Merging from On Ramp
13	I-40 EB at Fayetteville Rd. MM 276	AM	1.29	0.56	179	Merging from On Ramp
14	I-40 WB at N Harrison Ave. MM 287	AM	1	0.66	166	Merging from On Ramp

15	I-440 EB at Wake Forest Rd. MM 10	PM	0.52	0.60	78	Merging from On Ramp
----	-----------------------------------	----	------	------	----	----------------------

Another performance measure, the *DI* of a bottleneck for each observation day is discussed in a later section of the chapter. Recalling the method of estimating the *DI* of a bottleneck on a day, a rectangular box is used that represents the maximum expected duration and the queue length of the recurring bottleneck. However, question may arise that what if the extent of congestion on a single peak extends beyond the maximum queue and duration of the recurring bottleneck region? In response to this question, it is noted here that such phenomena exist within the study period. However, for most bottlenecks, the percentage of days when the extent of congestion goes beyond the recurring bottleneck region varies between 2% to 8%. Although such extension of recurrent congestion may result in the underestimation of the *DI* on these days, since the percentages are small in number for most bottlenecks, it can be safely assumed that the estimation of *DI* for these bottlenecks is almost unbiased. For some bottlenecks that are likely to be impacted by any significant work zones active during 2016, this percentage may be significantly high. These work zone impacted bottlenecks are discussed separately in a later section of this chapter.

4.3. Visualization of the Bottleneck Locations in the Study Area

The locations, activation period, directions, and ranking by the overall *RBIF* of the listed recurring bottlenecks are presented in a map format in Figure 4-3.

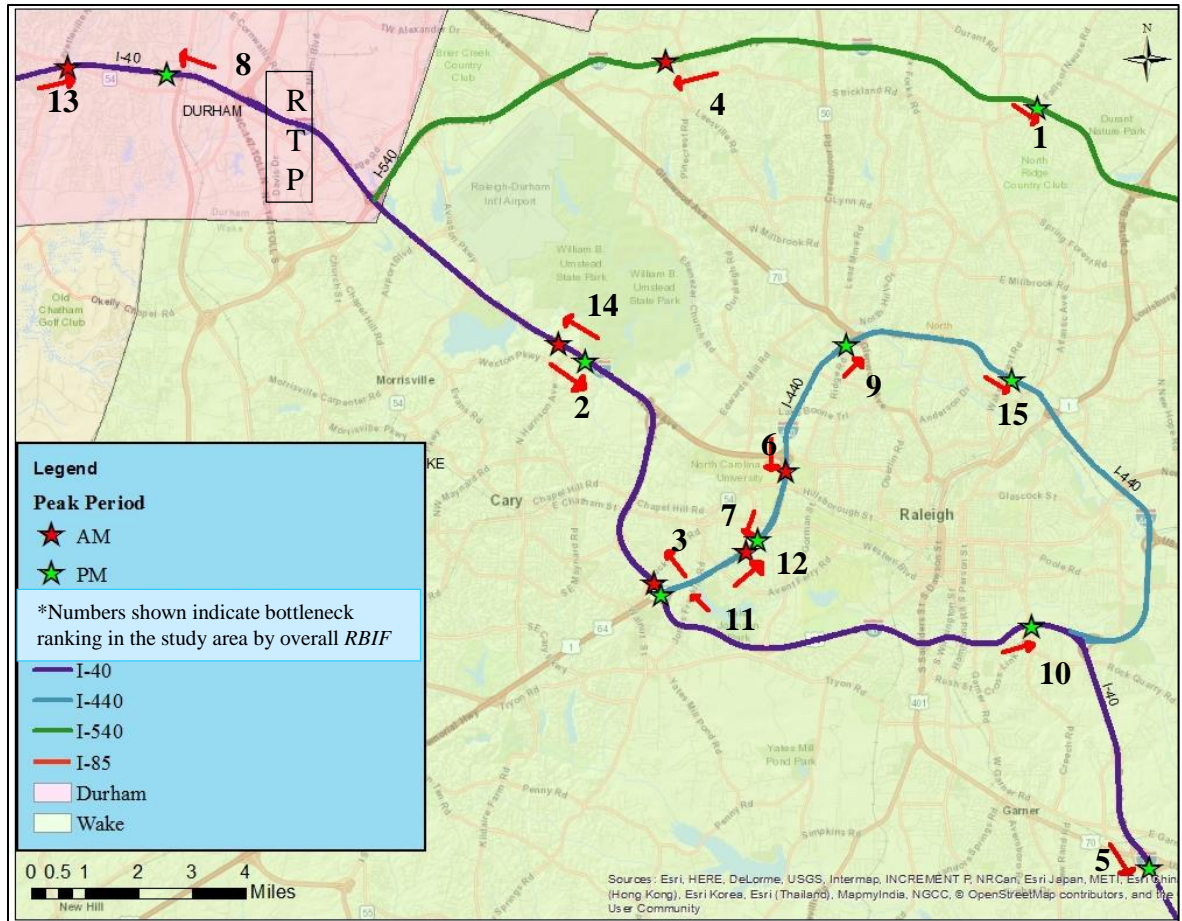


Figure 4-3: Location, activation period and travel direction of the recurring bottlenecks in Wake and Durham Counties, NC (2016 weekdays)

In Figure 4-3, many of the bottlenecks' activation period and the direction of travel conforms to the local traffic patterns of the Wake and Durham Counties. In Durham County, the bottleneck in the eastbound direction (Rank # 13) could be attributed to the heavy traffic during the AM peak generated at the residential areas within the City of Durham towards the Research Triangle Park (RTP) area. The PM peak bottleneck on the opposite direction (Rank #8) is contributed by traffic leaving RTP and returning towards Durham. A similar pattern is

observed at bottlenecks Rank #14 and #2, where the majority of the traffic is going from and coming back to Raleigh during AM and PM peaks respectively. The bottlenecks on the I-440 Beltline in Raleigh are likely to be influenced by the large trip attractors such as NC State University and downtown Raleigh.

4.4. Comparison with the Ranking Algorithm Established by VPP

The Vehicle Probe Project Suite by CATT Lab (2016) published their new algorithm for ranking traffic bottlenecks in 2016 which is available via RITIS (2016). The new algorithm of calculating Impact Factor (in miles-minutes of congestion), a performance measure based on which the bottlenecks are ranked along with their previous method of bottleneck ranking was discussed in CHAPTER 2: LITERATURE REVIEW. This section presents a comparison between the ranking of the bottlenecks obtained by the VPP's new algorithm and by the method proposed in this study. Before applying the VPP's algorithm to the case study area for the calendar year of 2016, one of the issues with the selection of the observation days needs to be discussed. The tool provided in RITIS through which VPP's algorithm is applied to different roads and to different times does not allow the user to select weekdays only from an observation period. However, the bottleneck ranking obtained by the proposed method were obtained by analyzing only the weekdays in 2016. Thus, the options available were either applying the VPP's algorithm for the 52 weeks manually or to ensure that including the weekends is not introducing any significant bias in the ranking and then to apply the method for the whole calendar year of 2016. Since the first option is challenging and time-consuming, the second option was pursued. To make sure that including weekends

is not introducing any bias in the ranking, the algorithm was applied to six different weeks in 2016 with and without including the weekends. The number of bottlenecks identified by the algorithm is shown in Figure 4-4.

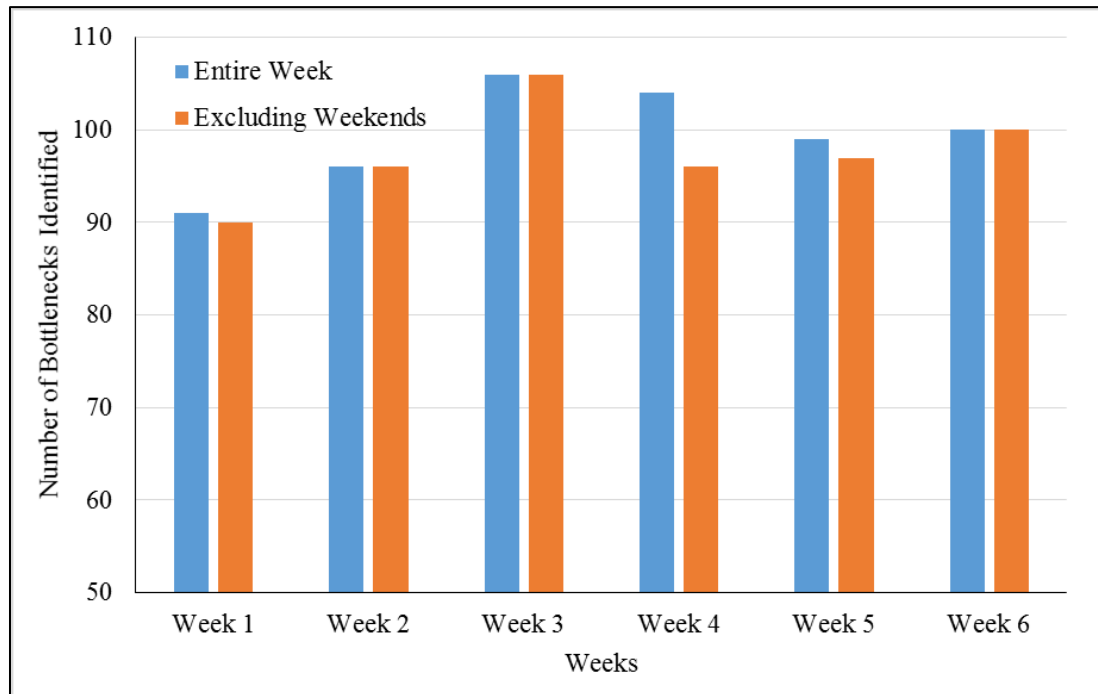


Figure 4-4: Number of bottlenecks identified by VPP’s algorithm: including and excluding weekends

Figure 4-4 indicates that in week 4, excluding the weekends caused a reduction of the bottleneck numbers by 8 (about 7%). However, for the remaining 5 weeks, the decrease in the number of bottlenecks is insignificant (0~2%). Moreover, one important finding not shown in Figure 4-12 is that the total number of different bottlenecks identified from all these 6 weeks is 108 and 105 when weekends are included and excluded respectively (a difference

of less than 3%). Thus, it can be safely assumed that analyzing all the days in 2016 is not generating any significant bias in identifying the total number of bottlenecks.

The VPP algorithm was applied to the study area and period. It identified 130 recurring bottleneck locations. The primary reason behind such a huge difference in the number of bottlenecks identified by the two methods lies in the filtering applied by the *AHCI* threshold to extract recurring bottlenecks only. According to the VPP's algorithm, even a single activation of a bottleneck which could be triggered by a non-recurring event will find its place in the final ranking list.

Another reason behind the high number of bottlenecks identified by VPP's algorithm is their aggregation of time. VPP's algorithm uses a 1-minute time aggregate. Thus, if a bottleneck creates a congestion that lasts for 1 minute, it is captured by the VPP's algorithm. Since the proposed algorithm uses an aggregate of 15-minutes, those short duration bottleneck activation would not be captured by it.

Table 4-2 lists the bottlenecks identified in this study along with their overall *RBIF*, ranking based on the overall *RBIF*, the Impact Factor calculated by VPP's new algorithm, and their ranking. Among the 130 bottlenecks identified by VPP, rankings of the recurring bottlenecks identified by the method proposed in this study varied from 1 to 68. From the list of 130 bottlenecks, the ranking of the fifteen identified bottlenecks within themselves (relative ranking from 1 to 15) based on their corresponding Impact Factor is listed in the last column of Table 4-2.

Table 4-2: Comparison of the bottleneck ranking obtained by overall RBIF and VPP algorithm

Rank by Overall RBIF	Location	Overall RBIF (mi-hours of congestion)	VPP Impact Factor (mi-hours of congestion)	VPP Rank	Relative Ranking by VPP
1	I-540 EB at Falls of Neuse Rd. MM 14	2985	343	19	8
2	I-40 EB at Harrison Ave. MM 287	2758	508	11	4
3	I-40 WB at I-440 MM 293	2516	465	14	6
4	I-540 WB at Leesville Rd. MM 7	1993	649	5	2
5	I-40 EB at US-70 MM 306	1723	1181	1	1
6	I-440 WB at Wade Ave. MM 4	1598	555	9	3
7	I-440 WB at Western Blvd. MM 2	1172	498	12	5
8	I-40 WB at NC-55 MM 278	970	293	23	9
9	I-440 EB at US-70/NC-50/Glenwood Ave. MM 7	863	174	37	12
10	I-40 EB at Rock Quarry MM 300	802	172	38	13
11	I-40 WB at Ramp from I-440 MM 293	293	465	14	6
12	I-440 EB at Melbourne Rd. MM 1D	243	78	68	15
13	I-40 EB at Fayetteville Rd. MM 276	179	235	30	10
14	I-40 WB at N Harrison Ave. MM 287	166	203	34	11
15	I-440 EB at Wake Forest Rd. MM 10	78	155	42	14

From Table 4-2, the following salient observations can be noted regarding the comparison of the ranking of the bottlenecks by these two methods.

- The first observation from the above table is that the VPP ranking does not match the rank obtained by the proposed method. However, it can be seen that the bottlenecks that are ranked from 1st to 6th according to the relative ranking of VPP have their *RBIF* ranking within 2 to 6.
- Unlike *RBIF* analysis, VPP's analysis is not confined within a single peak. Rather, Impact Factor is analyzed by considering 24-hours of data. This could be the primary reason behind the large number of bottleneck generation by VPP algorithm
- Another possibility stated earlier is that *RBIF* is estimated only for the recurring bottleneck region. By contrast, VPP does not apply any filter to extract the recurring bottlenecks only.
- The inherent difference in the process of congestion detection by the two methods may also cause this difference in ranking of the bottlenecks (see the congestion detection by the two methods in Chapter 2 and Chapter 3)
- The relative ranking of VPP for these fifteen bottlenecks appears to have a fewer difference with the Overall *RBIF* ranking. Except for the three shaded rows, the relative ranking by VPP and Overall *RBIF* ranking for the other 12 bottlenecks differ by less than or equal to 3 units.

- The Rank #1 bottleneck according to the Overall *RBIF* is 8th in the relative ranking by VPP. It has the highest difference in ranking between the Overall *RBIF* and relative ranking by VPP.

The results and the discussions above highlight the difference in the two methods although the estimation of their performance measures may appear similar. However, one may ask that what are the other top ranked bottlenecks in VPP's 130 bottleneck list and where are they located? To answer this question, those top-ranked bottlenecks were searched and found that they are regarded as recurrently "congested" road segments by the proposed method. In other words, the other top-ranked bottlenecks in VPP's list are located upstream of these fifteen recurring bottlenecks. This could be one of the reasons that VPP's Impact Factor value is much lower compared to the corresponding overall *RBIF* value although the basis of estimation of these two parameters is synonymous (see Chapter 2 and Chapter 3). The overall *RBIF* of a single bottleneck identified by the proposed method is divided into multiple bottlenecks with smaller Impact Factor values.

From the above discussion, it is apparent that the proposed method is more focused and marking the recurring bottlenecks only, where the VPP algorithm generated a very high number of bottlenecks that may include non-recurring events induced bottleneck. From the point of view of applying countermeasures or treatments for mitigating the bottlenecks, it is necessary to have a short list of the truly recurring bottlenecks only.

4.5. Attributes of the Most Severe Recurring Bottlenecks

This section contains a detailed investigation of the top two bottlenecks listed in Table 4-1. The respective recurring bottleneck region and the distribution of the *DI* are discussed for the bottlenecks to provide a sense of their expected extent and variability. Following that, the characteristics of their locations, traffic volumes, and other properties are summarized in an attempt to point out the possible contributing factors of the bottlenecks.

4.5.1. Rank #1: I-540 Eastbound at the On-Ramp from the Falls of Neuse Rd. (MM 14)

4.5.1.1. Recurring Bottleneck Region and Daily Impact (DI)

This bottleneck is ranked #1 in the list of identified recurring bottlenecks in terms of both the Overall *RBIF* and *RBIF* per activation. It activates during the PM peak hour in 235 out of 250 weekdays (94%) in the 2016 calendar year. Figure 4-5 shows the *AHCI* contour plot of this recurring bottleneck.

Time	16:15	16:30	16:45	17:00	17:15	17:30	17:45	18:00	18:15	18:30	18:45	TMC ID	TMC Tag	Cumulative Length (mi)
	0.79	0.40	1.19	0.79	0.00	0.40	0.00	0.79	1.19	1.19	1.19	1.19	125+05083	0.0
4.74	9.49	19.76	28.85	34.78	39.92	41.90	27.67	23.72	12.65	5.95	12.65	125P05082	1.0	2.49
4.74	9.49	19.76	28.85	34.78	39.92	41.90	27.67	23.72	12.65	5.95	12.65	125+05082	2.0	4.50
4.74	13.83	24.90	39.92	58.89	66.01	66.40	53.36	35.97	18.58	4.35	18.58	125P04081	3.0	4.96
4.74	13.83	24.90	39.92	58.89	66.01	66.40	53.36	35.97	18.58	4.35	18.58	125+04081	4.0	6.19
2.77	16.60	35.97	47.83	68.38	76.68	73.52	62.85	51.38	31.62	4.74	31.62	125P04080	5.0	6.75
2.77	16.60	35.97	47.83	68.38	76.68	73.52	62.85	51.38	31.62	4.74	31.62	125+04080	6.0	8.85
2.77	81.03	87.75	90.12	89.72	90.12	86.96	83.00	64.82	43.87	4.74	43.87	125P04079	7.0	9.42
2.77	81.03	87.75	90.12	89.72	90.12	86.96	83.00	64.82	43.87	4.74	43.87	125+04079	8.0	11.06
3.95	80.63	88.93	90.12	90.12	90.51	86.56	77.87	55.73	39.92	3.57	39.92	125P04899	9.0	11.84
3.95	80.63	88.93	90.12	90.12	90.51	86.56	77.87	55.73	39.92	3.57	39.92	125+04899	10.0	12.51
3.95	26.09	59.29	69.57	77.47	75.49	66.40	44.66	21.34	10.28	3.57	10.28	125P04898	11.0	12.87
3.95	26.09	59.29	69.57	77.47	75.49	66.40	44.66	21.34	10.28	3.57	10.28	125+04898	12.0	13.59
1.98	3.95	13.44	23.32	35.18	41.90	31.62	17.79	7.51	4.35	1.98	4.35	125P04897	13.0	14.94
1.98	3.95	13.44	23.32	35.18	41.90	31.62	17.79	7.51	4.35	1.98	4.35	125+04897	14.0	15.30
0.40	0.40	0.40	1.19	1.58	1.19	0.40	0.40	0.40	0.00	0.00	0.00	125+04896	15.0	15.90

Figure 4-5: AHCI contour for Rank #1 bottleneck

From Figure 4-5, it is seen that the maximum expected queue length and maximum duration of the resulting congestion for this bottleneck are 13.2 miles and 2 hours 15 minutes respectively. Such a large extent regarding both queue length and duration, along with a high frequency of activation makes this bottleneck as the top one in the list.

Another important observation from the *AHCI* contour for this bottleneck is that some spatiotemporal cells outside the recurring bottleneck region were congested for several days. For example on TMC tag 2.0, the time duration 18:45-19:00 was congested for 5.95% days, which approximates to 15 days out of 250 weekdays. Since the *DI* of a bottleneck is estimated only within the rectangle representing the maximum queue and duration of a recurring bottleneck (Figure 3-3), this extension of congestion outside the region may introduce bias in the estimation of the *DI*. However, it is seen that if any long-term work zone is not present within the recurring bottleneck region, the percentage of days where the

congestion extends beyond the recurring bottleneck region is very small (e.g. 5.95% for this recurring bottleneck).

The distribution of the *DI* for this bottleneck is shown in Figure 4-6. It highlights the day to day variability of the impact of the bottleneck in terms of its extent.

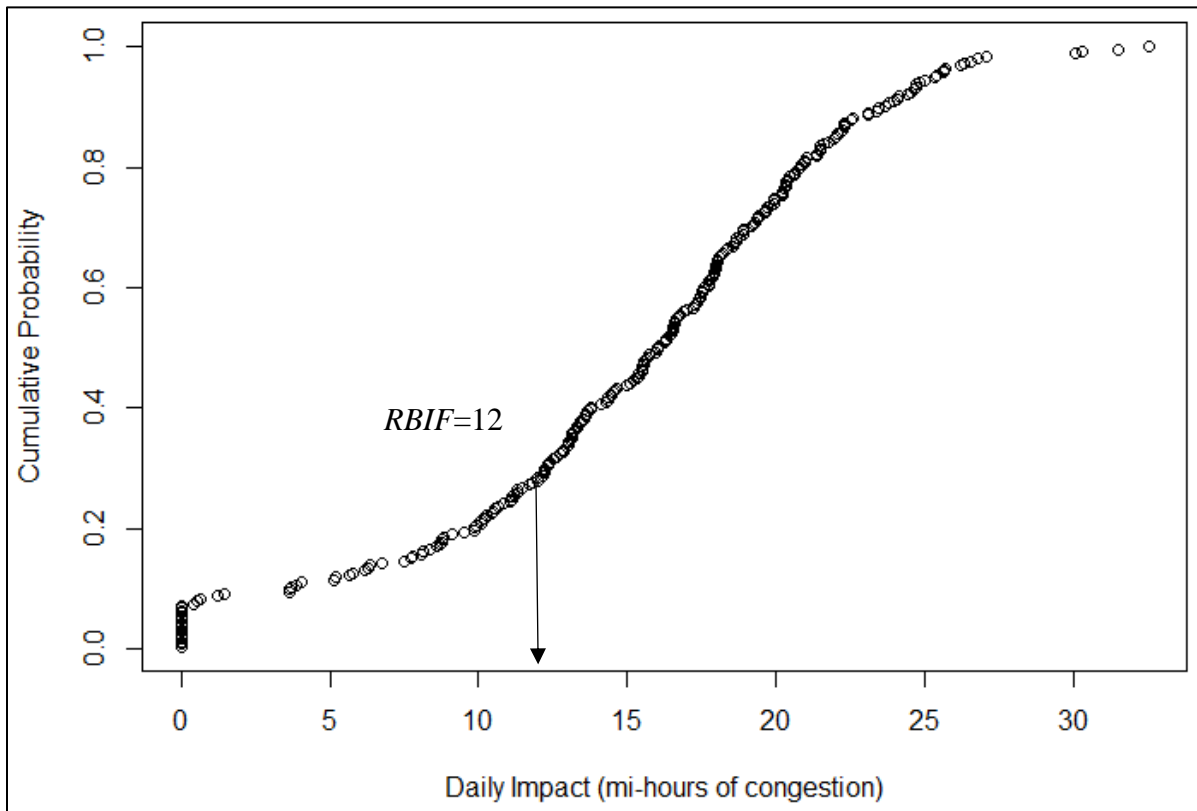


Figure 4-6: Cumulative distribution of DI for Rank #1 bottleneck

It is evident from Figure 4-6 that the *DI* of this bottleneck expands over a wide range (0 to about 33 units). The variance of *DI* for this bottleneck is also very high (51.4). The four observation days with a very high value of *DI* on the upper right corner (>30 unit of *DI*)

could have experienced non-recurring events (inclement weather or some type of incidents). An analysis of the variation in *DI* is addressed in the next chapter.

4.5.1.2. Attributes of the bottleneck location

A close look at the bottleneck location and other factors is needed to point out the possible contributing circumstances that are activating the bottleneck. A detailed geographic location of the bottleneck is shown in Figure 4-7.

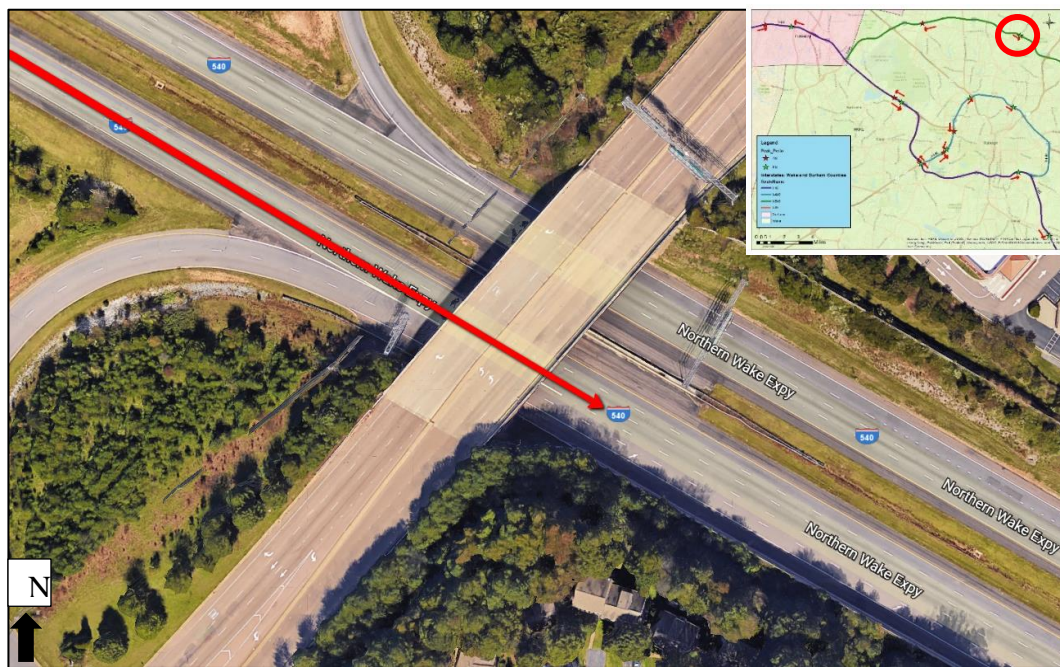


Figure 4-7: Location of the Rank #1 bottleneck: I-540 EB at the Falls of Neuse Road (MM

14)

The red arrow shows the travel direction and the location of the bottleneck. The on-ramp just upstream of the bottleneck is the one coming from the Falls of Neuse Road.

Different geometric properties and the AADT values at this roadway are provided in Table 4-3.

Table 4-3: Geometric properties and traffic volumes at top-ranked bottleneck

Number of Mainline Lanes	3
Number of Ramp Lanes	2
Segment Type D/S of the Bottleneck	Merging
Lane Drop	No
AADT on I-540 x1000	101
Possible Impact of Work Zone	No
Mainline Speed Limit (mph)	70
Ramp Speed Limit (mph)	Unknown
Falls of Neuse AADT (x1000)	38

The following observations can be pointed from the above table.

- The AADT on the mainline road is high given that there were only 3 lanes. As an arterial, the AADT on the Falls of Neuse Road is not insignificant as well. This high volume may cause the demand during peak hour to exceed the capacity of the freeway segment. Assuming a directional distribution (DD) of 0.5, hourly factor (k) of 10% (typical for PM peak in an urban area), and capacity (c) of each lane as 2000 pc/mi/ln, the demand to capacity ratio is roughly calculated as shown in Equation (4-1).

$$\frac{Demand}{Capacity} = \frac{AADT * DD * k}{3 * c} = \frac{101,000 * 0.5 * 0.1}{3 * 2000} = 0.85 \quad \text{Eq. (4-1)}$$

It appears that the demand on the main line of I-540 eastbound is close to its capacity. When the traffic from the on-ramp starts merging with the mainline lanes, the friction can reduce the capacity significantly. At the same time, this additional demand from the ramp can cause the overall demand to exceed capacity. Thus, the bottleneck at this merging segment may activate.

- The speed limit on the mainline road is high. Although the speed limit on the on-ramp is unknown, it can be assumed less than 35 mph considering the small radius of the curve
- Vehicles from the two lanes of the ramp with such a slow speed are merging with a high-speed freeway with heavy traffic volume during the PM peak. This may cause a lot of side friction to the flow and subsequent reduction in the capacity because of the merging maneuver by the vehicles.

4.5.2. Rank #2: I-40 EB at the On-Ramp from Harrison Ave. (MM 287)

4.5.2.1. Recurring Bottleneck Region and Daily Impact (DI)

Figure 4-8 shows the *AHCI* contour plot of this recurring bottleneck.

Time →	16:00	16:15	16:30	16:45	17:00	17:15	17:30	17:45	18:00	18:15	18:30	TMC	TMC Tag	Cumulative Length (mi)
	5.93	7.91	12.25	19.76	28.85	46.64	63.24	65.22	54.55	47.43	29.64	125N04860	1	0.59
	5.93	7.91	12.25	19.76	28.85	46.64	63.24	65.22	54.55	47.43	29.64	125-04860	2	2.26
	12.65	25.30	49.80	74.70	85.38	87.75	89.72	87.75	81.03	71.15	56.52	125N04861	3	2.88
	12.65	25.30	49.80	74.70	85.38	87.75	89.72	87.75	81.03	71.15	56.52	125-04861	4	3.52
	7.11	13.44	34.39	62.06	79.84	86.96	85.38	83.79	76.68	63.24	45.06	125N04862	5	4.06
	7.11	13.44	34.39	62.06	79.84	86.96	85.38	83.79	76.68	63.24	45.06	125-04862	6	4.76
	3.56	5.53	11.86	33.60	56.52	73.52	77.08	68.77	57.71	37.55	25.69	125N04863	7	5.45
	3.56	5.53	11.86	33.60	56.52	73.52	77.08	68.77	57.71	37.55	25.69	125-04863	8	5.78
	30.83	80.63	86.56	86.96	86.17	86.96	81.82	68.38	51.78	35.57	24.90	125N04864	9	6.08
	30.83	80.63	86.56	86.96	86.17	86.96	81.82	68.38	51.78	35.57	24.90	125-04864	10	6.27
	37.15	88.14	88.93	89.33	86.96	84.98	79.05	65.22	46.25	30.43	21.34	125N04865	11	6.52
	37.15	88.14	88.93	89.33	86.96	84.98	79.05	65.22	46.25	30.43	21.34	125-04865	12	6.95
	29.64	81.03	87.75	88.54	81.03	76.28	69.57	53.75	35.18	22.92	14.23	125N04866	13	7.27
	29.64	81.03	87.75	88.54	81.03	76.28	69.57	53.75	35.18	22.92	14.23	125-04866	14	7.62
	17.00	41.11	62.06	61.26	52.17	46.25	45.06	32.02	18.97	13.83	11.07	125N04867	15	8.48
	17.00	41.11	62.06	61.26	52.17	46.25	45.06	32.02	18.97	13.83	11.07	125-04867	16	9.40
	12.25	21.34	39.13	41.50	35.97	30.04	26.09	21.34	13.83	8.30	7.11	125N04868	17	9.68
	12.25	21.34	39.13	41.50	35.97	30.04	26.09	21.34	13.83	8.30	7.11	125-04868	18	11.50

Figure 4-8: AHCI contour for the Rank #1 bottleneck

The region of this Rank #2 bottleneck is even more extensive than that for the Rank #1 bottleneck. However, the queue length for this bottleneck is significantly shorter compared to that in the Rank #1 bottleneck region. Nonetheless, the expected queue length emanating from this bottleneck can be up to 11.5 miles. Another noteworthy observation is the variability of the duration of congestion for this bottleneck. The congestion start time varies between 16:00 to 17:15 hours and the end time ranges from 17:00 to 18:30 hours respectively depending on the location of the TMCs. The maximum AHCI value within the bottleneck region is 89.72%. Therefore, the probability of activation for this bottleneck must be this percentage or more.

The distribution of the *DI* for this bottleneck is shown in Figure 4-9.

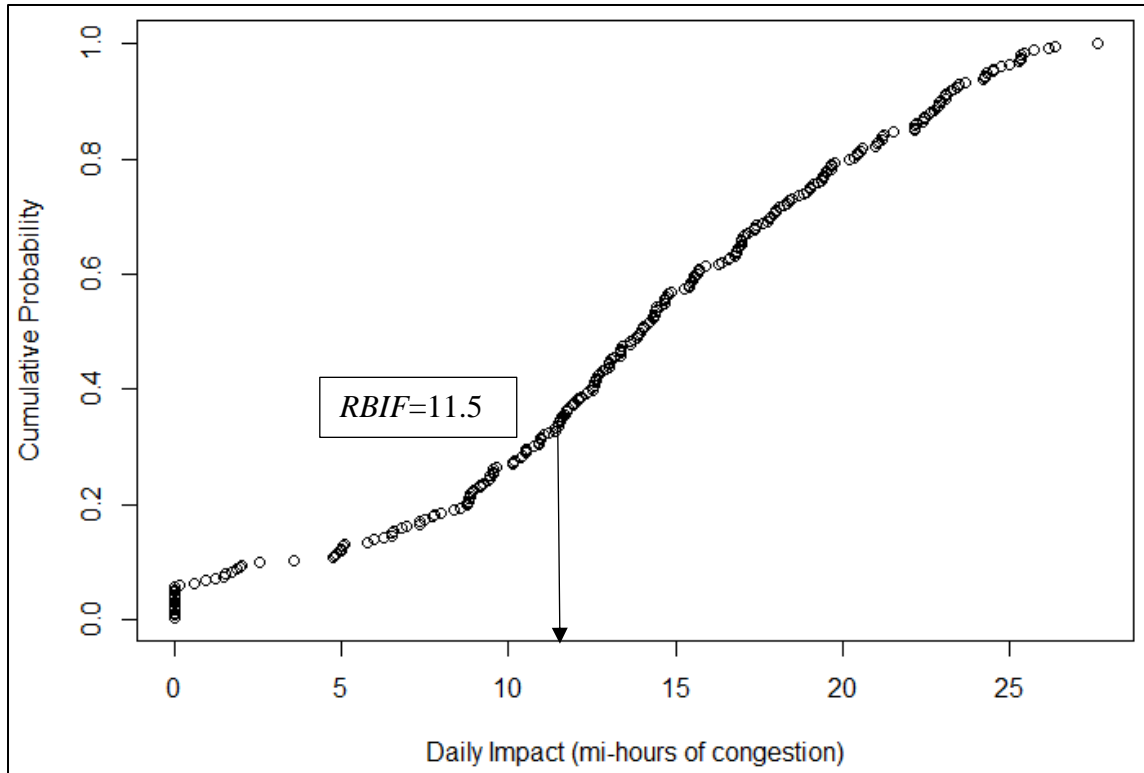


Figure 4-9: Cumulative distribution of DI for the Rank #2 bottleneck

The *DI* for this bottleneck over the 250 observation days spreads over a wide range (0 to >27 units). The variance of the *DI* is estimated as 47.2. The bottleneck activated ($DI \geq 0.5$ unit) on 238 out of 250 weekdays (95%) in the 2016 calendar year, which is even higher than that for the Rank #1 bottleneck.

4.5.2.2. Attributes of the bottleneck location

A detailed geographic location of the bottleneck is shown in Figure 4-10

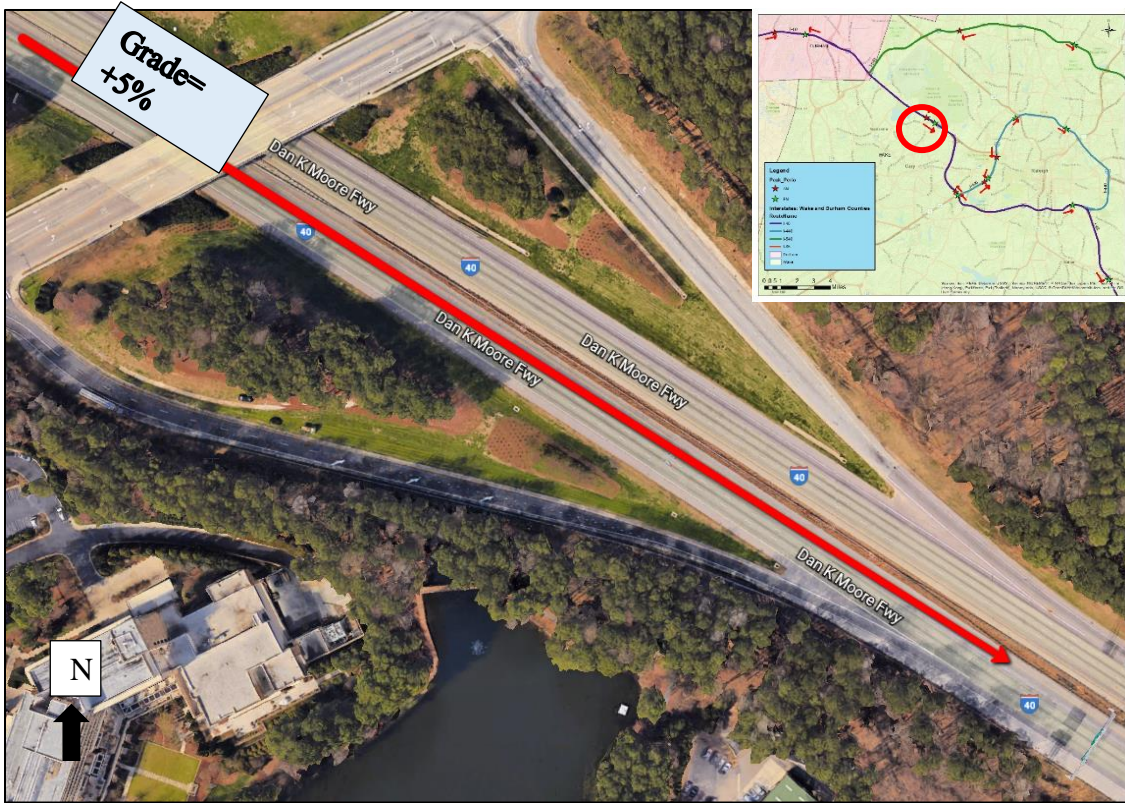


Figure 4-10: Location of the Rank #2 bottleneck: I-40 EB at the on ramp from N. Harrison Avenue. (MM 287)

Different geometric attributes and the AADT values at this roadway is provided in Table 4-4.

Table 4-4: Geometric properties and traffic volumes at the Rank #2 bottleneck location

Number of Mainline Lanes	4
Number of Ramp Lanes	1
Segment Type D/S of the Bottleneck	Weaving
Lane Drop	No
AADT on I-40 (x1000 vpd)	174
Possible Impact of Work Zone	No
Mainline Speed Limit (mph)	65
Ramp Speed Limit (mph)	Unknown
Harrison Ave. AADT (x1000 vpd)	45

From the geographic location and the tabulated information, the following points need be mentioned.

- This segment of the roadway has an extremely high AADT. The traffic from the connected arterial, which also has a high value of AADT is entering into the weaving segment that may cause the activation of the bottleneck. The weaving maneuver is associated with the traffic entering in I-40 and exiting the Wade Avenue which is a major arterial located further downstream of this bottleneck.
- As mentioned earlier, the returning traffic from the RTP to the residential areas located within the beltline during the PM peak hour seems to be the most significant contributing traffic for this bottleneck.

- To estimate the demand to capacity ratio (d/c) at this bottleneck location, a directional distribution value $k = 0.5$ is selected. Assuming an hourly factor of 10%, the ratio is calculated as shown in Equation (4-2)

$$\frac{\text{Demand}}{\text{Capacity}} = \frac{AADT * DD * k}{4 * c} = \frac{174,000 * 0.5 * 0.1}{4 * 2000} = 1.09 \quad \text{Eq. (4-2)}$$

Thus, it is evident that the demand alone on the mainline segment is exceeding the capacity of the roadway segment. Additionally, the entering demand from the ramp and weaving maneuver can exacerbate the condition.

- The upgrade (5%) of this segment of I-40 eastbound upstream of the bottleneck as shown in Figure 4-9. It may reduce the capacity of the segment. Moreover, it may make the merging maneuver for the entering traffic more difficult.

4.6. Special Bottleneck Cases

A number of the bottlenecks listed in Table 4-1 are found to have some unusual characteristics. These characteristics are mainly attributed to local phenomena (e.g. presence of a long-term work zone project) or influence of major arterials connected to the freeway segment. These two special cases are discussed in the following subsections.

4.6.1. Congestion Created by a Bottleneck on an Arterial Road

In Table 4-1, Rank #10 bottleneck (I-440 EB at the off-ramp to US-70 during PM peak) is the only bottleneck that is associated with an off-ramp, where traffic demand is being transferred to another road. Due to this factor, attention was drawn toward this bottleneck.

In order to find why this bottleneck is located just upstream of an off-ramp, the arterial road US-70 connected to the other end of the ramp is analyzed using the data for 2016 calendar year to identify any recurring bottleneck activation. The analysis results showed that there is a recurring bottleneck just at the location where the off-ramp connects to US-70. This recurring bottleneck also activates during the PM peak period. Figure 4-11 shows the resulting congestion due to this bottleneck (see red arrow), downstream of which free flow conditions prevail. The congested I-440 and the ramp are shown by the yellow line.

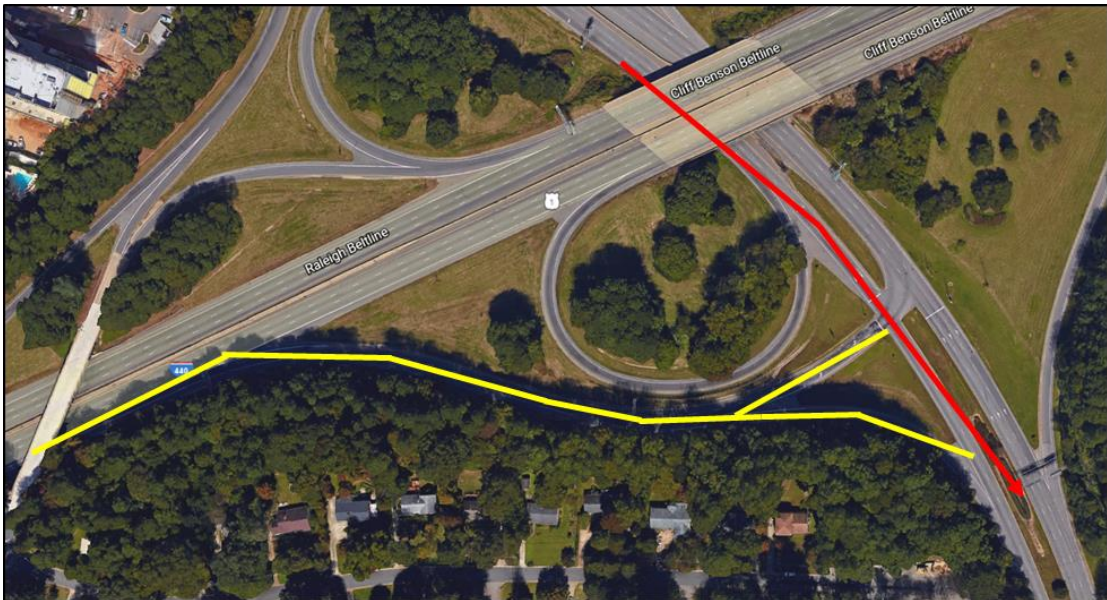


Figure 4-11: Location of the bottleneck on US-70 and the resulting congestion

The presence of the recurring bottleneck on US-70 as shown in Figure 4-11 is evidence that the congestion on I-440 EB upstream of the off-ramp is likely to be the result of the queue spilling back from that bottleneck. US-70 is a major arterial with an AADT of 77,000 vpd. This bottleneck on US-70 activated for 229 days in 2016 and out of those days

the queue spilled back to I-440 in 226 days. Since the target of the study is to identify the actual location of the bottleneck, the congestion on the I-440 should not be termed as “bottleneck”. To characterize this bottleneck, the activations of this bottleneck on US-70 need to be focused. In estimating the performance measures, the queue both on US-70 and I-440 needs to be considered. It should be noted that for other bottlenecks identified in this study, the queue may go back to the connected arterials from the freeways. Like this special case bottleneck, if the queue on the connected road is significantly long, the proposed algorithm needs to be modified to consider that additional queue length.

4.6.2. Bottlenecks Impacted by a Long Term Work Zone

One of the issues of selecting the calendar year 2016 for analyzing the study roads for bottleneck identification is that a major work zone project supervised by the North Carolina Department of Transportation was active during this time (2017). This project, also known as the Fortify I-40/440 Rebuild Project Phase II was active on the portion of road segment as shown in Figure 4-12.

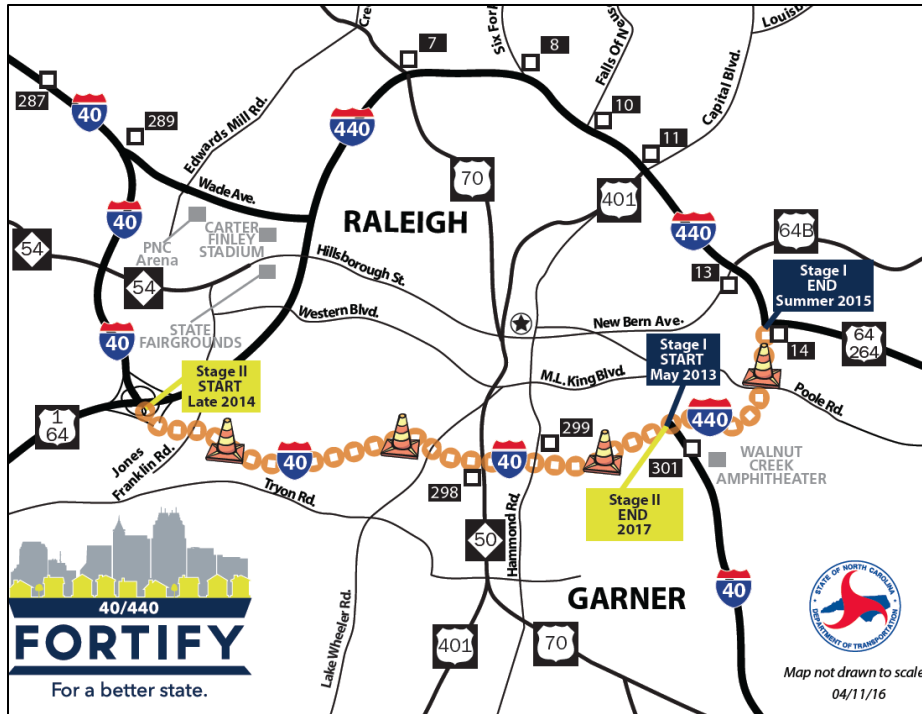


Figure 4-12: Location of fortify work zone Phase II. (Courtesy: NCDOT)

Considering the location of this work zone, three bottlenecks listed in Table 4-1 are likely to be impacted by this work zone: Rank # 3, Rank #10, and Rank #11.

The presence of this work zone needs to be underscored for these three bottlenecks and the rank of all the bottlenecks. Without the presence of this work zone, the rank of the bottlenecks based on Overall RBIF could be changed significantly. It is hard to tell to what extent the work zone is impacting these three bottlenecks. Since detailed information on the schedule of the work zone (time and location on each day) is not available, its impact could not be excluded from the analysis.

CHAPTER 5: PREDICTING THE ACTIVATION AND DAILY IMPACT OF BOTTLENECKS

This chapter presents the development and validation of models for predicting the activation and Daily Impact (*DI*) of the recurring bottlenecks identified in the previous chapter. Starting with an introduction to the data sources, it discusses the different challenges for developing such models. Then, the development and assessment of the logistic models for predicting the activation of the bottlenecks are described. The development and validation of the daily impact models are presented in the last section.

5.1. Data Sources

As mentioned and illustrated in the previous chapter, the estimation of the performance measures of a bottleneck (*DI* and *RBIF*) was carried out using the probe vehicle speed data obtained from RITIS (2016). To develop the models discussed in this chapter, non-recurring events occurring within a recurring bottleneck region need to be identified. The non-recurring events are classified into two types: inclement weather conditions and roadway incidents. The reasons behind dividing the events into these two categories are that they came from two different data sources and the structure of their data sets are entirely different.

For the study area, weather data is obtained from an online data source (Weather Underground, 2017) which stores the raw data collected at the nearest airport (Raleigh-Durham Intl. Airport). This airport is located close to the middle of the study area.

Incident data is obtained from NCDOT's Travel Information website (TIMS, 2017). The original data are obtained from observations of the traffic management unit of the agency.

5.2. Challenges for Developing and Validating the Models

5.2.1. Bottlenecks Near the Fortify Work Zone

Of the recurring bottlenecks obtained from the case study, ones that are likely to be impacted by NCDOT's fortify work zone (North Carolina Department of Transportation, 2017) were already discussed in the previous chapter. Developing a predictive model for these bottlenecks is a challenging task because of the complexity associated with the work zone location. The lack of data regarding the schedule of the work zone makes the task even more challenging. Moreover, it was already discussed in the previous chapter that the estimation of *DI* for these bottlenecks might contain some bias since the congestion went beyond the region of the corresponding recurring bottleneck for a significant number of days. Considering these issues, the three bottlenecks pointed in the previous chapter that are likely to be impacted by the fortify work zone are excluded from this analysis. It should be noted that for the remaining twelve bottlenecks, minor work zone activity could be present during the selected study period. However, the effects of such short-term work zones are ignored.

5.2.2. Issue of Sample Size

It has already been discussed in Chapter 3 that the development of the logistic model requires sufficient sample size (minimum $10 * \textit{number of predictors}$) for both activation and no activation cases (Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996). On the

other hand, a linear regression model does not give reasonable fit if the variation of the dependent variable (in this case, DI) is small.

Table 5-1 shows the number of activations and the variance and the range of DI for the twelve recurring bottlenecks.

Table 5-1: Number of activations and variance and range of DI for the bottlenecks

Rank by Overall <i>RBIF</i>	Bottleneck Location	Number of Activations in 250 Days	Variance of <i>DI</i>	Range of <i>DI</i> (mi-hours of congestion)
1	I-540 EB On Ramp from Falls of Neuse Rd. MM 14	235	51.4	32.5
2	I-40 EB On Ramp from Harrison Ave. MM 287	238	47.2	27.6
4	I-540 WB On Ramp from Leesville Rd. MM 7	219	44.1	27.6
5	I-40 EB On Ramp from US-70 MM 306	232	27.2	20.2
6	I-440 WB On Ramp from Wade Ave. MM 4	222	33.9	23.4
7	I-440 WB On Ramp from Western Blvd. MM 2	232	14.2	18.8
8	I-40 WB On Ramp from NC-55 MM 278	222	11.3	11.7
9	I-440 EB Off Ramp to US-70/NC-50/Glenwood Ave. MM 7	226	11.3	13.2
12	I-440 EB On Ramp from Melbourne Rd. MM 1D	183	1	4
13	I-40 EB On Ramp from Fayetteville Rd. MM 276	139	4.2	5.8
14	I-40 WB On Ramp from N Harrison Ave. MM 287	166	9.8	14.2
15	I-440 EB On Ramp from Wake Forest Rd. MM 10	151	1.5	6.3

It is evident from Table 5-1 that the top eight bottlenecks in this list did not have sufficient numbers of no activation days (shaded blue). Due to this issue of sample size, developing a logistic regression model for predicting the activation of these bottlenecks is challenging. However, since these eight bottlenecks have a considerable variation and wide range of *DI*, the proposed linear regression model can be developed for them.

Bottlenecks ranked #12, #13, and #15 (shaded yellow) have a significant number of both activations and non-activation cases. However, their variation of *DI* is minimal. Thus, the proposed linear regression models are not developed for these bottlenecks. Only the logistic regression models are developed for these three bottlenecks.

The bottleneck ranked #14 has a sufficient number of both activation and no activation cases as well as a wide range and variation of *DI*. Therefore, both logistic and linear regression models are developed for this bottleneck.

5.3. Logistic Regression Models for Predicting the Activation of a Bottleneck

Table 5-1 shows that 5 bottlenecks have a sample size within the study period that allows developing a logistic regression model for each of them to predict their activation. The general form of the logistic regression model is shown in Equation (5-1)

$$\ln\left(\frac{Y_i}{1-Y_i}\right) = \beta_0 + \beta_1 * X_{Tue,i} + \beta_2 * X_{Wed,i} + \beta_3 * X_{Thu,i} + \beta_4 * X_{Fri,i} + \beta_5 * X_{quarter2,i} + \beta_6 * X_{quarter3,i} + \beta_7 * X_{quarter4,i} + \beta_8 * X_{inclement\ weather,i} \quad (5-1)$$

Here, interpretation of all the terms are explained in Chapter 3 (Eq. (3-7)). All the predictors listed above (days of the week, quarters of the year, and inclement weather) are

provided into the model as binary variables. However, as discussed in Chapter 3, a limitation of using binary variables in logistic models is that each level of a binary variable must have both cases (activation or no activation) at least once. For all the logistic regression models developed here, the “occurrence of incident” variable does not satisfy this constraint, because of which this variable is not included in the model.

Table 5-2 shows the results of the logistic regression analyses. It contains the estimates of the parameters and their significances.

Table 5-2: Results from the logistic regression models

Rank by O. RBIF	Estimate/ Significance	Intercept	Day of a Week				Quarter of a Year			Severe Weather
			Tue	Wed	Thu	Fri	2 nd	3 rd	4 th	
Rank 12	Est.	1.13	2.64	1.51	1.13	-0.41	-0.22	-0.17	-0.09	-1.28
	Sig.	0.05	0.05	0.05	0.1	x	x	x	x	x
Rank 13	Est.	2.92	0.91	0.38	0.48	-1.17	-1.19	-1.37	-1.62	-1.32
	Sig.	0.001	x	x	x	x	x	x	x	x
Rank 14	Est.	1.60	1.64	0.77	0.91	-0.79	-0.65	-0.37	-0.07	-0.88
	Sig.	0.01	0.1	x	x	x	x	x	x	x
Rank 15	Est.	1.87	1.33	0.92	0.50	-0.56	-0.38	-0.48	-0.86	-1.39
	Sig.	0.01	x	x	x	x	x	x	x	0.1

*x indicates significance level >0.1%

The salient points that can be made from the results listed above are discussed below:

- The numbers in Table 5-3 can be used to estimate the log-odds of the activation of each bottleneck on a day by inserting them into Eq. (5-1). For example, the intercept term for the bottleneck Rank #12 shows that on a Monday during the 1st quarter of the

year and good weather condition, the log-odds of the activation of the bottleneck is 1.13. The log-odds increases by 2.64 on Tuesday and decreases by 0.41 on Friday relative to Monday given that everything else remains same. This information can be useful to transportation agencies when they need to predict the activation of a bottleneck for applying real-time congestion mitigation strategies.

- Agencies could also predict the impact of non-recurring events such as inclement weather condition from this model. For all the four bottlenecks in question, the log-odds of activation decreases during inclement weather condition compared to good weather condition. However, the effect of inclement weather is insignificant only for all bottlenecks. Thus, it is difficult to make a solid statement regarding the effect of weather condition for the other bottlenecks using these results. A higher significance of the inclement weather variable is needed to comment on its effect.
- Overall, all these models suffer from a lack of significant predictor. Models for Rank #13, #14, and #15 bottlenecks show that the intercept term is the only parameter that is significant at the level of 0.01. However, there are other variables significant for the models of these bottlenecks but only at a level of 0.1.

To assess the fit of the trained models discussed above, the p-value and the McFadden R-squares (McFadden & Zarembka, 1974) are provided in Table 5-3 below. The p-value is obtained from a Chi-square test that tests the predictive capability of the model compared to a null model (a model with no predictor).

To validate the trained models, they were applied to their corresponding test data set. A 2 by 2 contingency table was developed using the observed and predicted test data. A Chi-square test is performed to verify the dependency between them as shown in Table 5-3. The Chi-square value is assessed by the corresponding p-value for a *degree of freedom* = 1 (for a 2 by 2 table, degree of freedom is 1).

Table 5-3: Assessment of the logistic models

Rank by Overall RBIF	Trained Model		Validation using Test Dataset	
	p-value	McFadden R-square	Chi-square statistics for dependency test	p-value for dependency test
Rank 12	0.002	0.14	2.39	0.12
Rank 13	0.012	0.13	0.54	0.46
Rank 14	0.013	0.12	0.69	0.4
Rank 15	0.016	0.11	6.2	0.012

Table 5-3 shows that the McFadden R-square values for all the models are almost similar and varies within 0.11-0.14. Also, the p-values indicate that the overall effect of the parameters used in the models is significant compared to a null model. However, these two values are not large enough to conclude that the models are robust. When the models were applied to the corresponding test data set, the Chi-square values for the first three models in the list were found to be very small. It results in a high p-value which results in failing to reject the null hypothesis that the predicted probabilities of activation and observed activations are independent. Validation result for the bottleneck Rank #15 shows that the null hypothesis can be rejected at a level of 0.05. Thus, only this model passes the validation test.

Lack of more significant parameters in the trained models may cause the failure of the validation of the test models.

5.4. Regression Models for Predicting *DI*

The regression models are intended to predict the *DI* of a bottleneck for each day as a function of the day of a week, a quarter of a year, weather condition, and parameter(s) related to roadway incidents occurred within the bottleneck region. Table 5-1 shows that there are 9 bottlenecks which exhibit sufficient variation of *DI* that allows developing a linear regression model to predict their *DI*. A stepwise linear regression technique was used to identify the significant predictors for each bottleneck. Results from the two types of the linear regression models: the base model and full model are described in the next subsections.

5.4.1. Base Models

The form of the base linear regression model and the interpretation of the terms are already discussed in Chapter 3, Equation 3-8.

5.4.1.1. Results from the Base Model Development

The results obtained from the base linear regression models are provided in Table 5-4 below.

Table 5-4: Results from the base models

Rank by Overall RBIF	Est./ Sig.	Intercept	Inclement Weather	Quarter of Year			Incident Occurrence	Day of Week			
				2 nd	3 rd	4 th		Tuesday	Wednesday	Thursday	Friday
Rank 1	Est.	15.21	3.59					3.04	1.84	3.62	-3.84
	Sig.	0.001	0.01					0.01	0.1	0.001	0.001
Rank 2	Est.	8.13	3.28	2.64	1.77	3.06	2.35	3.06	3.26	7.08	5.08
	Sig.	0.001	0.01	0.05	0.1	0.01	0.01	0.01	0.1	0.001	0.001
Rank 4	Est.	12.05		0.34	1.34	2.03		1.28	0.93	-0.12	-8.4
	Sig.	0.001		x	x	0.05		x	x	x	0.001
Rank 5	Est.	6.26	2.2	-0.4	0.56	2.66	1.14	0.44	2.18	3.02	6.36
	Sig.	0.001	0.05	x	x	0.001	x	x	0.01	0.001	0.001
Rank 6	Est.	6.74	2.83	1.06	3.01	3.15		2.97	2.74	3.05	-3.83
	Sig.	0.001	0.01	x	x	0.001		0.01	0.01	0.001	0.001
Rank 7	Est.	5.03	2.37	-0.17	0.35	3.59	2.14	1.2	1	1.94	-0.22
	Sig.	0.001	0.01	x	x	0.001	0.001	0.05	0.1	0.01	x
Rank 8	Est.	3.61	3.2	-0.43	-0.39	1.08	3.52	1	1.47	3.17	4.82
	Sig.	0.001	0.001	x	x	x	0.001	x	0.05	0.001	0.001
Rank 9	Est.	3.44	2.82	0.6	-0.14	2.02		1.85	1.58	2.62	0.78
	Sig.	0.001	0.001	x	x	0.001		0.05	0.01	0.001	x
Rank 14	Est.	3.55	2.24				2.8				
	Sig.	0.001	x				0.001				

*x indicates a significance level >0.1

Key observations from Table 5-4 are listed below

- The intercept for each model shows the predicted *DI* of the corresponding bottleneck on a day with the reference conditions. For example, the intercept value of 15.21 for bottleneck Rank #1 indicates that on a Monday during the 1st quarter of the year, and with no occurrence of nonrecurring events, if the bottleneck activates it is predicted to create a *DI* of 15.21 miles-hours of congestion.
- The coefficients along with the intercept of the model can be inserted into Equation (3-8) to predict the change in *DI* relative to the reference condition. For example, the coefficient of inclement weather for Rank #1 bottleneck indicates that the *DI* on a day with inclement weather is predicted to be increased by 3.59 miles-hours of congestion compared to a good weather day given that everything else remains same.
- The variation of effects of non-recurring events across different bottlenecks is another important observation from Table 5-5. For example, unlike all other 8 bottlenecks, inclement weather does not have any effect on the *DI* for bottleneck Rank #4. Incident occurrence affects the *DI* of Rank #2, #5, #7, #8, and #14 bottlenecks. These information can be used by transportation agencies in dynamic traffic management in case of occurrence of such non-recurring events within a recurring bottleneck region. Based on these results, they can decide on which bottleneck region a real-time congestion mitigation strategy should be applied.
- The days of a week are selected as a predictor by the stepwise linear regression for 8 out of the 9 bottlenecks. Except for Friday, the remaining three weekdays (Tuesday,

Wednesday, and Thursday) have an increasing effect in *DI* compared to the reference day (Monday). *DI* for bottleneck Rank #1, #4, #6, and #7 are predicted to be less on a Friday compared to a Monday.

- The significance of the coefficients listed in Table 5-5 is also important for the users of these models. For example, regardless of the value and sign of the coefficients, the last quarter of the year was found most significant compared to the reference (1st) quarter. Some other coefficients that are found significant for most bottlenecks are inclement weather condition and Friday as the day of the week.

5.4.1.2. Assessment of the Models

In Table 5-5, the goodness of fit of the nine models described above is discussed based on their adjusted R-square values. In addition, results from the application of the models to their corresponding validation datasets are discussed.

Table 5-5: Assessments of the base models

Rank by Overall RBIF	Adjusted R-square of Trained Model	Validation Results			
		Mean Bias (mi-hours of congestion)	Mean Bias \pm 2*St. Error (mi-hours of congestion)	Mean RMSE (mi-hours of congestion)	St. Error of RMSE (mi-hours of congestion)
Rank 1	0.2	-0.090	(-0.33,0.14)	5.255	0.074
Rank 2	0.26	-0.069	(-0.3,0.16)	5.308	0.075
Rank 4	0.34	-0.111	(-0.31,0.09)	4.738	0.052
Rank 5	0.32	-0.054	(-0.26,0.16)	4.028	0.060
Rank 6	0.31	0.063	(-0.13,0.25)	4.428	0.053
Rank 7	0.3	-0.070	(-0.2,0.06)	2.898	0.037
Rank 8	0.38	-0.077	(-0.21, 0.06)	3.169	0.041
Rank 9	0.21	-0.043	(-0.16,0.08)	2.652	0.028
Rank 14	0.1	0.059	(-0.08,0.2)	2.918	0.061

It is evident from Table 5-5 that the adjusted R-square of the models vary over a wide range, where the model for Rank #14 bottleneck has the lowest value (0.1) and Rank #8 bottleneck has the highest value (0.38). From Table 5-5 it was shown that the linear regression model for the Rank #14 bottleneck only has the nonrecurring events (inclement weather and incident occurrence) as its predictor variables. The presence of only two predictors in the model may cause the poor fit.

Several models listed in Table 5-5 have reasonable R-square value (R-square>0.3). Although R-square value shows the overall goodness of fit of the models, the prediction capability of the models is revealed when they are applied to the validation data. Once a model is applied to its validation dataset, the resulting mean bias should equal to zero

signaling that the model predictions are unbiased. In the “mean bias” column of Table 5-5, although all the models exhibit a non-zero value of mean bias, some of them are close to zero (Rank #9, Rank #14, Rank #5). Moreover, the ranges of the mean bias \pm two times of the standard error cover zero for all the models. Thus, it can be said that some of the model validation results are promising.

The means of the RMSE values along with their standard errors are reported in Table 5-5 as the best estimate of the “out of sample” prediction error of the model. Considering the range of *DI* the bottlenecks exhibit (0-35 miles-hours of congestion), the mean RMSE values listed in Table 5-6 (approximately 2.9 to 5.3) are high.

Results from Table 5-5 indicate that the validation results for most of the bottlenecks are poor. Regardless, it is important to investigate the variation of bias created by the models. The proposed CURE plot (CUMulative REsiduals vs. predicted *DI* described in Chapter 3) was drawn for each bottleneck to complete this task. Figure 5-1 (a) shows the fitted vs. observed *DI* plot for the trained model of Rank #2 bottleneck. Figure 5-1 (b), (c), and (d) show the fitted vs. observed *DI*, CURE vs. predicted *DI*, and CURE vs. observed *DI* respectively when the trained model is applied to a validation data set. As described in Chapter 3, these CURE plots show the variation of cumulative residuals of *DI* against the *DI*. The dataset is selected randomly from one of the fifty iterations of validation sets. Prior to cumulating the residuals, the data set is sorted by the increasing order of predicted and observed *DI* for Figure 5-1 (c) and (d) respectively. It appears from Figure 5-1 (c) that some data points have multiple cumulative residual values for a single predicted *DI* (e.g., there are

four data points for predicted $DI=8$). Average of the cumulative residuals are used in such cases to plot the line.

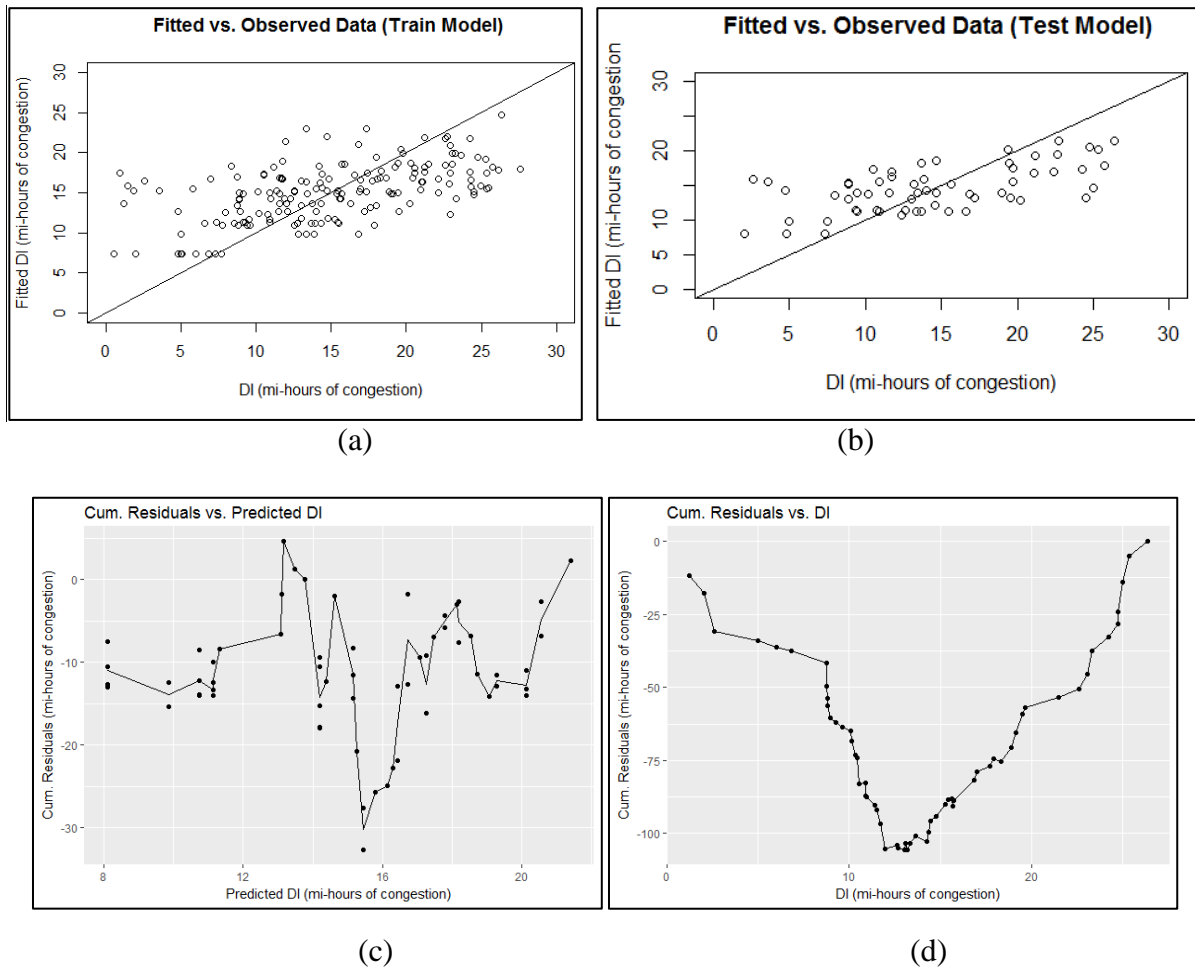


Figure 5-1: (a) Fitted vs. observed plot from trained model (b) Fitted vs. observed plot from validation data (c) CURE vs. predicted DI (d) CURE vs. observed DI for validation data

In Figure 5-1 (a) and (b), the solid line (a reference line) pass through the origin with a slope of 1. These two plots evidence that the model is overestimating the DI when the observed DI is of small magnitude (data points lie above the reference line) and underestimating when the observed DI is high (data points lie below the line). The CURE

plot in Figure 5-1 (c) shows the variation of this bias more clearly. It should be noted that an increase in the cumulative residual value indicates an underestimation of the *DI* by the model. Since the CURE plot has upward and downward trends for different values of the fitted *DI*, it can be said that the model is biased at almost all magnitudes of predicted *DI*. Although between 10 to 12 units of predicted *DI* the trend of the CURE plot is somewhat horizontal, it is followed by a precipitous drop when *DI* exceeds 15 units and a sudden increase when it exceeds 16 units. These large jumps indicate the presence of potential outliers. The trend is again random between 17 to 20 units of predicted *DI*, but the upward trend from 20 to the maximum predicted *DI* value tells that the model is underestimating the *DI* in this region. Figure 5-1 (d) illustrates the similar scenario to what Figure 5-1 (c) depicts. The model is overestimating the *DI* when the observed *DI* is of small magnitude and underestimating when it is of high magnitude. This figure serves as another way of illustrating the trend of the residuals of a model.

The estimates of the bias values shown in Table 5-6 and the CURE plot for the Rank #2 bottleneck prove that the models need significant improvement to exhibit a better validation result. CURE plots for other models show that none of the models are unbiased and these plots are provided in Appendix A.

5.4.2. Full Models

A full model includes detailed information (duration, the number of lanes closed, start time, and location) related to the incidents occurring within a recurring bottleneck region. Numerical variables related to these information replace the binary variable used in the based

model that simply indicates the presence of any incident. The “duration” variable represents the duration of an incident in minutes (Equation (3-11)). The “start time” predictor represents the reported start time of incident relative to the start time of the bottleneck activation (Equation (3-12)). The “location” variable is the distance of incident occurrence from the bottleneck location (Equation (3-13)). The way these variables are provided with the model, it is expected that a higher value of these variables will result in a higher magnitude of DI on a day.

The parameters regarding the details of an incident are provided in the full models with an aim to improve the fit and validation of the models. As an example, the full model for the Rank #2 bottleneck is described in Equation (5-1). Next, the summary of the validation of the full models for all bottlenecks is described. The description of the full models for the other eight bottlenecks is provided in Appendix B.

$$\begin{aligned}
 DI_i = & 8.05 + 3.42 * X_{inclement\ weather,i} + 3.1 * X_{Tue,i} + 3.33 * X_{Wed,i} + 7.25 \quad (5-1) \\
 & * X_{Thu,i} + 4.87 * X_{Fri,i} + 0.03 * X_{Duration1,i} + 3.06 * X_{quarter2,i} \\
 & + 2.25 ** X_{quarter3,i} + 3.49 * X_{quarter4,i}
 \end{aligned}$$

Here, $X_{Duration1,i}$ = Duration (minutes) of the 1st incident occurring within the recurring bottleneck region in question on day i .

It is apparent from Equation (5-1) that the binary variable indicating the occurrence of any incident in the base model is replaced by the duration of the incident. The coefficient of the duration term implies that with each minute increase of the duration of an incident

occurring within the recurring bottleneck region, the *DI* is expected to be increased by 0.03 mi-hours of congestion. Table 5-6 summarizes the assessment of the nine full models.

Table 5-6: Evaluation of the full models

Rank by Overall RBIF	Adjusted R-square of Trained Model	Validation Results			
		Mean Bias	Mean Bias \pm 2*St. Error	Mean RMSE	St. Error of RMSE
Rank 1	0.2	-0.090	(-0.33,0.14)	5.255	0.074
Rank 2	0.26	-0.040	(-0.29,0.21)	5.300	0.055
Rank 4	0.41	0.124	(-0.07,0.32)	4.470	0.060
Rank 5	0.43	-0.317	(-0.75,0.11)	5.230	0.850
Rank 6	0.31	-0.110	(-0.32,0.09)	4.400	0.055
Rank 7	0.41	0.040	(-0.08,0.16)	2.700	0.034
Rank 8	0.42	0.110	(-0.009,0.23)	3.170	0.036
Rank 9	0.21	-0.018	(-0.15,0.12)	2.720	0.030
Rank 14	0.18	0.011	(-0.14,0.16)	2.910	0.050

Table 5-6 shows that for the Rank #2 bottleneck, although the inclusion of the duration of incident does not increase the R-square, the validation result is improved as the mean bias gets closer to zero (from -0.069 in the base model to -0.04 in the full model). The mean and standard error of the RMSE value remain almost same. The full model for the Rank #1 bottleneck does not show any change from its base form. However, for other models, a significant increase in the R-square value is apparent from this table. Nonetheless, the validation results for most of the models degrade compared to their corresponding base form. This phenomenon may be contributed to the overfitting of the model by the additional

parameters related to incidents. Lack of incident occurrence in the validation data set may also be responsible for the poor validation results.

5.5. Summary

An overview of the key findings from the model development and validation efforts is provided below.

- The logistic regression models that predict the activation of a bottleneck yielded reasonable McFadden R-square and Chi-square values. However, the models suffer from lack of significant predictors. It is assumed that there are other omitted variables playing a key role in the activation of the bottlenecks. This could also be the reason that only one (Rank #15) logistic model passed the validation test.
- The base linear regression models predicting the *DI* of the recurring bottlenecks exhibit R-square value ranging from 0.1 to 0.38. Although none of the validation results show a mean bias of zero, the results are promising as the mean bias is close to zero for several models. These models provide useful information about the factors affecting the *DI* and their significance. The variation of the effect of various parameters across different bottleneck region is another important information that is delivered by the models. Nonetheless, the models contain significant bias and potential outliers as diagnosed by the CURE plot. These need to be eradicated before applying the models in real-life decision-making instances.

- Detailed information about the incidents occurring within the recurring bottleneck regions is provided in the full linear regression models with an expectation of improving the model fit and validation. However, although the R-square value of several full models increased, the validation result degraded for the models as the inclusion of additional parameters is probably over-fitting the model.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

Based on the analysis of the results obtained and discussed in Chapter 5 and Chapter 6, the key findings obtained from this study are summarized here. The first section of this chapter discusses the findings obtained from the recurring bottleneck identification and rankings in the case study area. Next, key results from the predictive models are discussed. The limitations of the study and recommendations for future research are presented in the last section

6.1. Recurring Bottleneck Identification and Ranking

This study presents a method for identifying and ranking recurring freeway bottlenecks using probe vehicle speed data. While selecting the thresholds for congestion detection and recurring bottleneck identification, a sensitivity analysis for the proposed performance measure of bottlenecks to the thresholds is offered. Three performance measures for the identified bottlenecks are developed based on the extent of the resulting congestion. These three measures focus on different characteristics of a bottleneck. The Daily Impact (*DI*) of a bottleneck estimates the extent of congestion in miles-hours as a result of its activation on a single day. Recurring Bottleneck Impact Factor (*RBIF*) represents the average impact of a bottleneck over multiple days in miles hours of congestion per activation. The overall *RBIF* is similar to *RBIF* per activation, except for the fact that it takes the total number of activation of a bottleneck into account over a user pre-specified time period.

The proposed method of bottleneck identification and ranking was applied to all interstate freeways in Wake and Durham Counties, NC for all weekdays in calendar year 2016. The key results obtained from this case study are.

- The selection of the threshold for congestion detection and recurring bottleneck identification is based on a sensitivity analysis. The sensitivity for the *RBIF* per activation for the study roads to the Congestion Index (*CI*) threshold showed a drastic increase in *RBIF* at a threshold between 0.7 and 0.8. Moreover, using a fixed speed threshold of 45 mph for congestion detection results similar magnitude of *RBIF* to what a *CI* threshold value of 0.65~0.82 does. In this analysis, the threshold of *AHCI* was kept constant (33%). These results justify the final selection of *CI* threshold of 0.7.
- The sensitivity of *RBIF* per activation to the *AHCI* threshold (for filtering recurring bottlenecks only) did not reveal any critical point. Rather, the variation of *RBIF* with *AHCI* threshold was found to be almost linear within the threshold range of 15% to 40%. A fixed speed threshold of 45 mph was used in this analysis. Finally, an *AHCI* threshold of 33% was selected. Its interpretation is that on an average, a recurring bottleneck must be activated in at least one out of three weekdays. However, the method can be applied using any user-desired threshold.
- The application of the proposed method to the case study area (using a *CI* threshold of 0.7 and *AHCI* threshold of 33%) identified 15 recurring bottleneck locations along

- with their activation clock times. Two-thirds of the bottlenecks activated during the PM peak period. Eight out of the 15 bottlenecks are located on the I-40 facility.
- Many of the bottlenecks' activation periods and the direction of travel conform to the local traffic pattern of the Wake-Durham Counties. For example, two AM peak bottlenecks on I-40 were located in the direction toward the Research Triangle Park (RTP) area which is a major trip attractor. On the other hand, two PM peak bottlenecks on I-40 is located in the direction away from the RTP area. The bottlenecks on the I-440 Beltline in Raleigh are likely to be influenced by the large trip attractors such as NC State University and downtown Raleigh.
 - The probability of activation for the fifteen bottlenecks varied within a broad range of about 50% to 95% (139 to 238 out of 250 weekdays). The *RBIF* per activation ranges from about 0.5 to 12.7 miles-hours of congestion. The range of these numbers illustrates the diverse characteristic of recurring bottlenecks.
 - The top two bottlenecks according to the overall *RBIF* are scrutinized in this study, and their contributing circumstances are identified. Rank #1 bottleneck is located on I-540 eastbound at the on-ramp from Falls of Neuse Rd. (MM 14). A detailed investigation of this bottleneck location revealed that the merging vehicles approaching from the curved two-lane on-ramp with low-speed limit might cause much friction to the mainline traffic on I-540. Moreover, a rough estimate of the demand to capacity ratio during the bottleneck activation on I-540 in the peak hour

appears to be high (0.85). These factors can trigger the recurrent activations of the bottleneck.

- A similar estimate of the demand to capacity ratio for the road segment for Rank #2 bottleneck (I-40 EB at the on-ramp from N. Harrison Avenue. MM 287) reveals that during that peak hour, the ratio may exceed 1. Some potential factors behind the activation of this bottleneck are high demand on I-40 eastbound, weaving traffic between I-40 and the downstream off-ramp to Wade Avenue, and the upward grade of the road upstream of the bottleneck.
- The distribution of *DI* for these two bottlenecks ranges from 0 to about 30 miles-hours of congestion. This wide range of *DI* indicates the dynamic nature of the top ranked bottlenecks regarding the queue length and duration of congestion.
- Some special case bottlenecks are observed from the case study. One of them is an actual bottleneck located outside the study freeway. Detailed inspection of the Rank #10 bottleneck (I-440 eastbound at the off-ramp to US-70) revealed a bottleneck activation on I-440 is actually a queue spill back from a bottleneck located on US-70. Moreover, some bottlenecks are located on the section of road where NCDOT's I-40/440 fortify work zone was active in 2016. These bottlenecks are assumed to be impacted by this long-term work zone.

The ranking obtained by the proposed method was compared to an already established bottleneck ranking tool, known as the VPP bottleneck ranking tool (CATT Lab, 2016). The main findings from the comparison are listed below

- The application of the VPP bottleneck ranking method to the case study resulted in 130 bottlenecks. Although the relative ranking of several bottlenecks by the VPP algorithm were found to be similar to the ranking by overall *RBIF*, in general, the VPP algorithm results do not conform to the proposed ranking.
- The most significant difference between the two methods is that the VPP algorithm does not identify the recurring bottlenecks only. A non-recurring event induced bottleneck activating only once but creating a heavy congestion can be ranked highly according to VPP algorithm. This is also the reason VPP generates such a large number of bottlenecks. However, from the perspective of applying countermeasures for mitigating regular congestion, it is necessary to have a workable list of the truly recurring bottlenecks only.
- There are other key differences between the two methods. The VPP algorithm analyzes speed data for 24 hours period rather than a single peak. Also, the speed threshold used for congestion detection is different in the two methods.

6.2. Predicting the Activation and Daily Impact of bottlenecks

This study presents an attempt to predict the activation of and resulting *DI* for recurring bottlenecks. Two models are proposed: a logistic regression model to predict the activation and a linear regression model to predict the *DI* of a bottleneck. Due to the issue of sample size within the case study period, logistic regression models are developed only for four bottlenecks. Moreover, since many of the bottlenecks' *DI* did not exhibit significant day to day variation in the Daily Impact (*DI*), linear regression models are developed for nine

bottlenecks. Chi-square test of dependency and K-fold cross validation technique are used to validate the logistic and linear regression models. Key findings from the model development and validation are listed below.

- Weekday type, weather condition, quarter of year are provided as binary predictors in the logistic regression models. In three out of the four logistic models, none of these predictors showed much significance. Only for the bottleneck Rank #12, three days of the week (Tuesday to Thursday) were found to be significant. Based on the sign of the coefficients, it was found that the log odds of the activation of this bottleneck were higher on these three weekdays than the reference weekday (Monday).
- The assessment of the all the models shows a reasonable McFadden R-square (0.11~0.14) and p-value (<0.016). However, only one model (for Rank #15 bottleneck) passed the Chi-square test of dependency conducted for validation purpose.
- In general, it can be said that all the logistic regression models suffered from a lack of significant predictors. However, it is difficult to comment on the validation with the presence of more significant variables, because including more variables may results in overfitting the model and in degrading the validation.
- In addition to the binary predictors used in the logistic regression model, incident occurrence was included in the linear regression model. Linear regression models were developed for a total of nine bottlenecks. The significance of variables

varied for different models. However, some similarities are observed. For example, inclement weather was found to be significant, and the expected *DI* increased during an inclement weather day compared to a good weather condition for most bottlenecks. The 4th quarter of the year has a significant increasing effect on the expected *DI* compared to the reference quarter (1st quarter). It is assumed that the holiday traffic and excessive travel demand on the road before the weekend of holidays during this time of the year are attributing to this finding.

- The presence of incident indicator used in the base model has a significant increasing impact on the expected *DI* for several models (bottleneck Rank #2, #5, #7, #8, and #14). The full model for Rank #2 was scrutinized. The duration of the 1st incident was found to be significant for this model. This variable had an increasing effect on *DI*, which implies that a higher duration of incident results in a higher estimated *DI* for the bottleneck.
- Assessment of the several models show reasonable values of R-square (0.3~0.4) except for the Rank #1, #9, and #14 bottleneck (R-square = 0.1~0.2). However, the mean bias obtained from the validation results and the CURE plots show that there is a significant bias in the models. The CURE plot for Rank #2 bottleneck revealed that the model is overestimating *DI* when the observed value is small and underestimating the *DI* when it is of high magnitude. Including detailed information regarding incidents in the full model slightly improves the validation for Rank #2 bottleneck but degrades the validation result for several models.

6.3. Recommendations for Future Research

The study presents a method and its application for identifying and ranking recurring freeway bottlenecks. In addition, it demonstrates an attempt to predict the activation and the day-to-day performance measure of the bottlenecks. Although the study was conducted with careful consideration of all the factors related to the goals and tasks, there are several critical issues that deserve further attention. Following are the limitations of this study that are recommended to be addressed in any future study related to this topic.

- In selecting the threshold for *AHCI*, the sensitivity analysis does not show any critical point based on which the threshold should be selected. For such cases, it is recommended to conduct the sensitivity analysis against a different parameter. Testing the sensitivity of the performance measure against the number of bottlenecks could be a possibility, but it requires an algorithm to separate the bottleneck locations when their impact areas merge with each other.
- The algorithm used in this study for bottleneck identification cannot automatically distinguish between a bottleneck activation and a mere queue spilling phenomena from a connected road. It is necessary to analyze the major arterials connected to the study freeway to ensure that the observed congestion is not the result of a bottleneck activation on the connected roads.
- The developed logistic regression models to predict bottleneck activation suffer from a lack of significant parameters. Moreover, most of the linear regression models, despite having a reasonable fit have had exhibited poor validation results.

Several factors can play a significant role for these predicaments. First, the predictors already used in the models may not reflect the pattern of bottleneck activation or the *DI* on a day. For example, bottleneck activation may not be sensitive to the quarter of the year. In that case, a robust way of combining the months needs to be used so that each group of months becomes a significant predictor. Second, there could be other variables that are being omitted from the models. For instance, roadway incidents occurring upstream or downstream of a bottleneck region may affect the activation or impact of the bottleneck. It is highly recommended to collect information on these parameters and to inspect whether they have any significance in the model or not.

- Another important factor that is directly related to the activation and impact of the bottlenecks is the traffic demand variations at the bottleneck segments. Since obtaining information on demand is a difficult task, including a surrogate measure for demand, could improve the model significantly. Such a potential measure can be the flow data on the ramps within the bottleneck regions. However, since this data is not available for this study, it is recommended for future study to use a surrogate measure of demand in the models.
- To address the issue regarding the development of the logit models, a larger sample would be a possible solution which may have a sufficient number of observations for both bottleneck activation and non-activation cases.

REFERENCES

- Ban, X. J., Chu, L., & Benouar, H. (2007). Bottleneck Identification and Calibration for Corridor Management Planning. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1999, *Transportation Research Board of the National Academies*, Washington, D.C., DOI: 10.3141/1999-05, 40-53.
- Bertini, R. &. (2005). Use of performance measurement system data to diagnose freeway bottleneck locations empirically in Orange County, California. *Transportation Research Record: Journal of the Transportation Research Board*, 48-57.
- Caltrans. (2011, April). *Caltrans Inside7*. Retrieved April 4, 2017, from <http://www.dot.ca.gov/dist07/Publications/Inside7/story.php?id=629>
- Cambridge Systematics Inc. (2005). *Traffic congestion and reliability: Trends and advanced strategies for congestion mitigation*. Texas Transportation Institute. Retrieved 2 16, 2017, from http://ops.fhwa.dot.gov/congestion_report_04/index.htm.
- Cambridge Systematics Inc. (2011). *Bottleneck Performance in the I-95 Corridor: Baseline Analysis Using Inrix Vehicle Probe Data*. I-95 Corridor Coalition.
- CATT Lab. (2016, April 20). *Vehicle Probe Project Suite*. Retrieved April 21, 2017, from CATT Lab: <http://www.cattlab.umd.edu/?portfolio=vehicle-probe-project-suite>
- CDM Smith. (2014). *Travel Time Based Oklahoma Congestion Analysis: Pilot Study*. Oklahoma Department of Transportation.
- Chen, C., Skabardonis, A., & Varaiya, P. (2004). Systematic identification of freeway bottlenecks. *Transportation Research Record: Journal of the Transportation Research Board*, 4, 45-62.
- Chen, C., Skabardonis, A., & Varaiya, P. (2004). Systematic Identification of Freeway Bottlenecks. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1867, *TRB, National Research Council*, Washington, D.C., 46-52.
- Crossman, A. (2017, March 02). *Understanding Descriptive vs. Inferential Statistics*. Retrieved May 10, 2017, from ThoughtCo.: <https://www.thoughtco.com/understanding-descriptive-vs-inferential-statistics-3026698>
- Day, C., Remias, S., Li, H., Mekker, M., McNamara, M., Cox, E., . . . Bullock, D. (2014). *2013-2014 Indiana Mobility Report: Summary Version (Joint Transportation Research Program Indiana Mobility Reports)*. West Lafayette, IN: Purdue University.

- Dowling, R., Skabardonis, A., Carroll, M., & Wang, Z. (2004). Methodology for Measuring Recurrent and Nonrecurrent Traffic Congestion. *transportation Research Record: Journal of the Transportation Research Board*, Vol. 1867, No. 1, 60-68.
- Everitt, B. S. (1992). *The analysis of contingency tables*. CRC Press.
- FDOT. (2011). *Bottlenecks on Florida SIS*. Florida Department of Transportation.
- FHWA. (2009). *Combating Bottlenecks: Georgia*. Retrieved April 30, 2017, from https://ops.fhwa.dot.gov/bn/successful_efforts/example/ga.htm
- FHWA. (2017, April 30). *Traffic Bottlenecks*. Retrieved from 21st Century Operation Using 21st Century Technologies: https://ops.fhwa.dot.gov/bn/successful_efforts/
- Florida Department of Transportation. (2011). *Bottlenecks on Florida SIS*. FDOT.
- Freedman, D. (2009, April 22). *Statistical Models: Theory and Practice*. Cambridge University Press. Retrieved from <https://onlinecourses.science.psu.edu/stat504/node/150>
- Hallenbeck, M. E., Ishimaru, J. M., & Nee, J. (2003). *Measurement of Recurring Versus Non-recurring Congestion, Report No. WA-RD 568.2*. Seattle: Washington State Transportation Center (TRAC).
- Hilbe, J. M. (2009). *Logistic regression models*. CRC press.
- Hillis, D. M., & Bull, J. J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic biology*, 42(2), 182-192.
- Jin, J., Yu, W., Fang, J., & Ran, B. (2010). A Robust Bottleneck Identification Method using Noisy and Inconsistent Fixed-Point Detector Data. *89th Transportation Research Board Meeting*.
- Jin, P. J., Parker, S., Fang, J., Ran, B., & Walton, C. M. (2012). Freeway Recurrent Bottleneck Identification Algorithms Considering Detector Data Quality Issues. *Journal of Transportation Engineering*, No. 10, 138.
- Latham, F; Trombly, J. (2003). *Low Cost Traffic Engineering Improvements: A Primer , No: FHWA-OP-03-078*. Office of Transportation Management, FHWA.
- Lin, D., Wei, J., & Ying, Z. (2002). Model-checking techniques based on cumulative residuals. *Biometrics*, 58(1), 1-12.
- Loether, H. J., & McTavish, D. G. (1980). *Descriptive and inferential statistics: An introduction*. Allyn & Bacon.

- Lynn Peterson. (2014). *WSDOT's Handbook for Corridor Capacity Evaluation*. Washington State Department of Transportation.
- Margiotta, R. A., & Spiller, N. C. (2012). *Recurring traffic bottlenecks: A primer: Focus on low-cost operational improvements (No. FHWA-HOP-12-012)*. FHWA.
- Margiotta, R., Spiller, N., & Halkias, J. (2007). *Traffic Bottlenecks: A Primer Focus on Low-Cost Operational Improvements, Report No. FHWA-HOP-07-130*. Washington, D.C.: U.S. Department of Transportation, Federal Highway Administration/ Office of Transportation Management.
- McFadden, D., & Zarembka, P. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in econometrics*, 105-142.
- North Carolina Department of Transportation. (2017, April 25). *NCDOT I-40/440 Fortify project*. Retrieved 11 19, 2015, from <http://www.ncdot.gov/fortifync/maps/>
- North Carolina Department of Transportation. (2017, April 25). *Fortify I-40/440 Rebuild Project*. Retrieved from NCDOT: <https://www.ncdot.gov/fortifync/>
- Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996, April 10). A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology*, 49(12), 1373-1379. Retrieved from Statistical Horizon: <http://statisticalhorizons.com/logistic-regression-for-rare-events>
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. In *Encyclopedia of database systems* (pp. pp. 532-538). Springer US.
- Regional Integrated Transportation Information System*. (2016). Retrieved 3 20, 2017, from Regional Integrated Transportation Information System: <https://www.ritis.org/traffic/>
- Rouphail, N., Petty, K., Eads, B., & McDermott, J. (2012). *Low-Cost Improvements for Recurring Freeway Bottlenecks: NCHRP 3-83*. Washington D.C.: Transportation Research Board of the National Academies.
- Saberi, M., & Bertini, R. L. (2010). Beyond Corridor Reliability Measures: Analysis of Freeway Travel Time Reliability at the Segment Level for Hot Spot Identification. *Transportation Research Board 89th Annual Meeting*, No. 10-2341).
- Schrank, D., Eisele, B., Lomax, T., & Bak, J. (2015). *2015 URBAN MOBILITY SCORECARD, Appendix A*. Texas A&M Transportation Institute, The Texas A&M University System.
- Skabardonis, A., Varaiya, P., & Petty, K. F. (2003). Measuring Recurrent and Nonrecurrent Traffic Congestion. *Transportation Research Record 1856, Paper No. 03-4261*.

- Song, T. J. (2016). Recurrent Freeway Bottlenecks Identification and Applications.
- TIMS. (2017, April 26). *Traveler Information Management System*. (Traveler Information Management System) Retrieved 11 19, 2015, from <http://tims.ncdot.gov/tims/IncidentDetail.aspx?id=460239>
- Transportation Research Board. (2010). *Highway Capacity Manual*. Washington, DC.: Transportation Research Board.
- University of California, Los Angeles. (2017, April 4). *Logit Regression / R Data Analysis Examples*. Retrieved from <http://stats.idre.ucla.edu/r/dae/logit-regression/>
- USDOT. (2015). *I-77 HOT Lanes, Charlotte, NC*. Retrieved April 30, 2017, from I-77 HOT Lanes
- Van Lint, J. W., Van Zuylen, H. J., & Tu, H. (2008). Travel time unreliability on freeways: Why measures based on variance tell only half the story. *Transportation Research Part A: Policy and Practice*, 42(1), 258-277.
- Weather Underground. (2017, April 28). *Weather Underground*. Retrieved from <https://www.wunderground.com/>

APPENDICES

**APPENDIX A: CUMULATIVE RESIDUAL PLOTS FOR BASE LINEAR
REGRESSION MODELS**

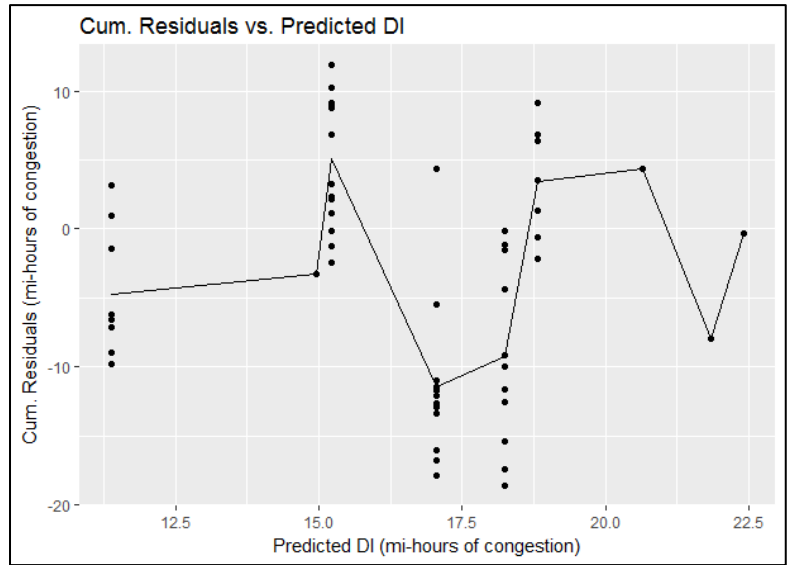


Figure A-1: CURE plot for the Rank #1 bottleneck

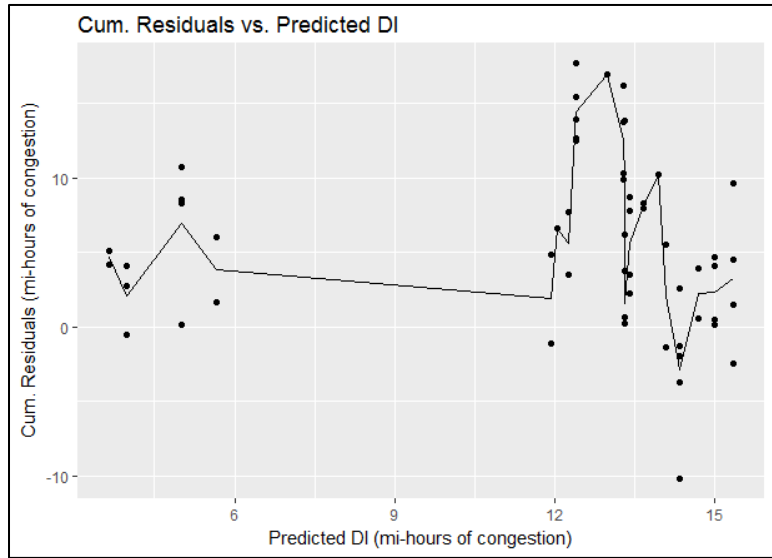


Figure A-2: CURE plot for the Rank #4 bottleneck

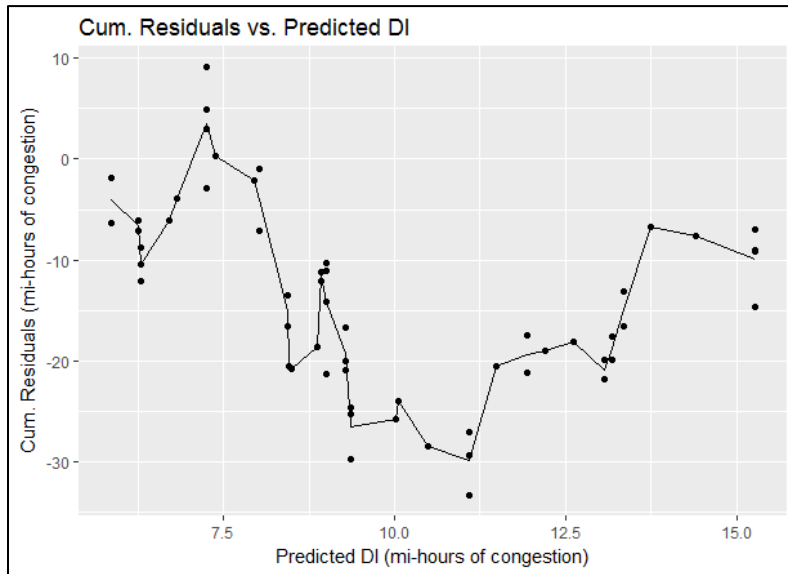


Figure A-3: CURE plot for the Rank #5 bottleneck

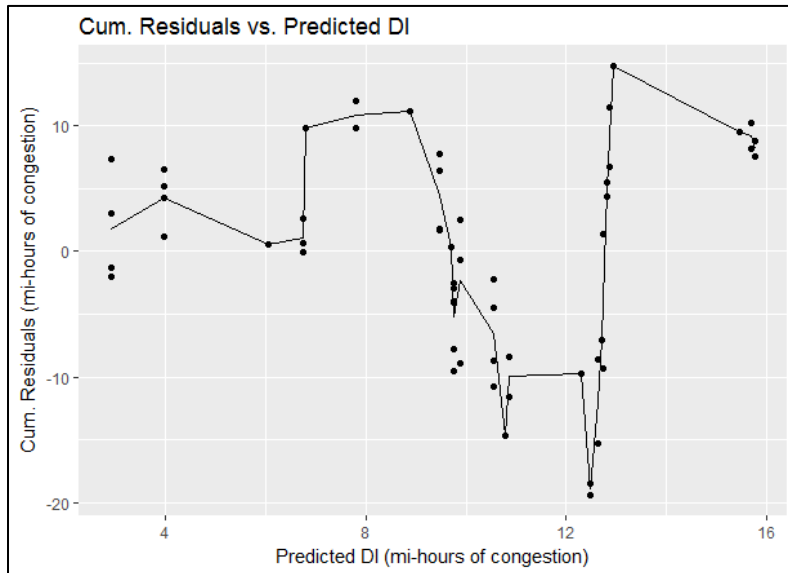


Figure A-4: CURE plot for the Rank #6 bottleneck

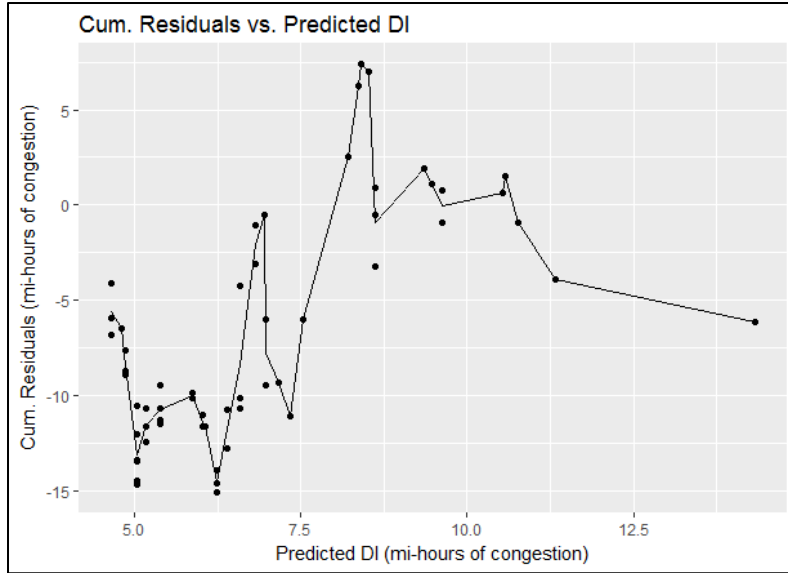


Figure A-5: CURE plot for the Rank #7 bottleneck

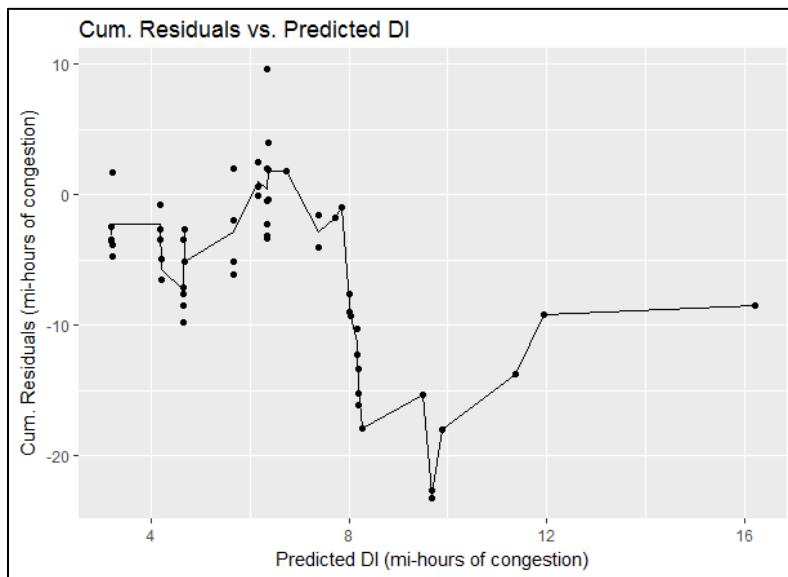


Figure A-6: CURE plot for the Rank #8 bottleneck

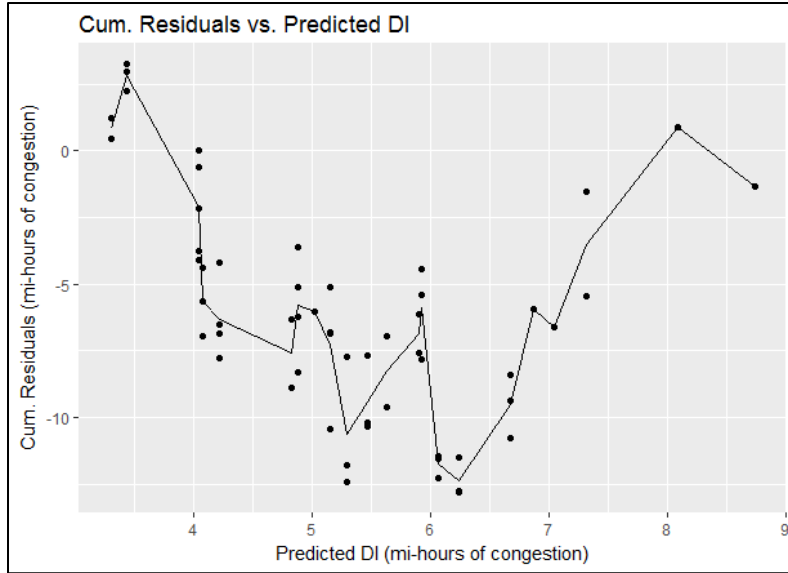


Figure A-7: CURE plot for the Rank #9 bottleneck

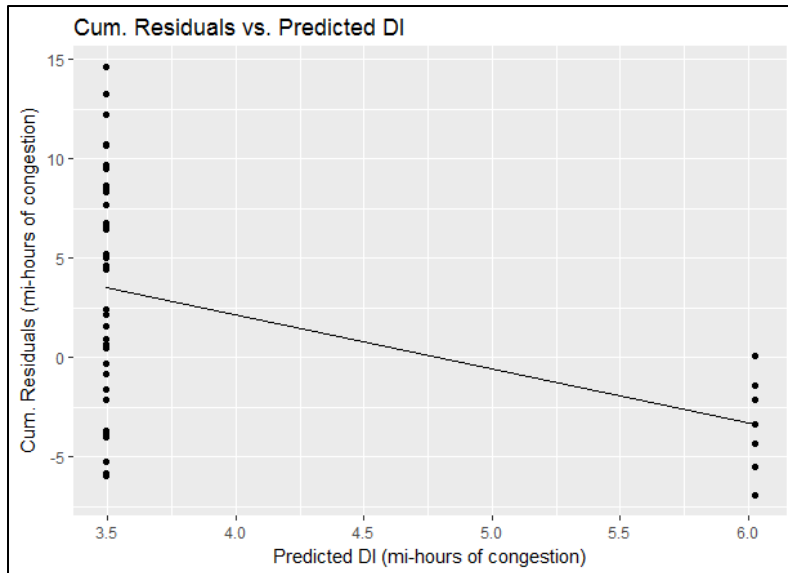


Figure A-8: CURE plot for the Rank #14 bottleneck

APPENDIX B: FULL LINEAR REGRESSION MODELS

Table B-1: Results from full linear regression model for the Rank #1 bottleneck

Parameters	Coefficients	Significance
Intercept	15.21	0.001
Incident weather	3.59	0.01
Tuesday	3.04	x
Wednesday	1.84	x
Thursday	3.62	x
Friday	-3.84	0.001
Residual St. Error	5.15 on 225 Degrees of freedom	
Adjusted R-square	0.2	

*x indicates a significance level >0.05

Table B-2: Results from full linear regression model for the Rank #4 bottleneck

Parameters	Coefficients	Significance
Intercept	12.05	0.001
2nd Quarter of Year	0.34	x
3rd Quarter of Year	1.34	0.05
4th Quarter of Year	2.03	0.01
Tuesday	1.28	x
Wednesday	0.93	x
Thursday	-0.12	x
Friday	-8.4	0.001
Residual St. Error	4.508 on 202 Degrees of freedom	
Adjusted R-square	0.41	

*x indicates a significance level >0.05

Table B-3: Results from full linear regression model for the Rank #5 bottleneck

Parameters	Coefficients	Significance
Intercept	6.71	0.001
Inclement weather	4.02	0.01
2nd Quarter of Year	-0.68	x
3rd Quarter of Year	-0.17	x
4th Quarter of Year	2.65	0.01
Tuesday	-0.42	x
Wednesday	1.23	x
Thursday	2.66	0.01
Friday	5.6	0.001
Start time of 2nd event	13.75	0.01
Location of 2nd event	9.36	x
Duration of 2nd event	-0.37	0.05
Location of 1st event ²	3.52	x
Duration of 2nd event ²	0.002	x
Residual St. Error	3.37 on 141 Degrees of freedom	
Adjusted R-square	0.43	

Table B-4: Results from full linear regression model for the Rank #6 bottleneck

Parameters	Coefficients	Significance
Intercept	6.52	0.001
Inclement weather	2.68	0.01
2nd Quarter of Year	1.21	x
3rd Quarter of Year	3.11	0.001
4th Quarter of Year	3.43	0.001
Tuesday	2.88	0.01
Wednesday	2.64	0.01
Thursday	3.05	0.001
Friday	-3.86	0.001
Location of 1st event ²	4.02	x
Residual St. Error	4.25 on 202 Degrees of freedom	
Adjusted R-square	0.31	

Table B-5: Results from full linear regression model for the Rank #7 bottleneck

Parameters	Coefficients	Significance
Intercept	4.88	0.001
Inclement weather	2.31	0.01
2nd Quarter of Year	-0.31	x
3rd Quarter of Year	0.19	x
4th Quarter of Year	4.13	0.001
Tuesday	1.36	0.05
Wednesday	1.46	0.05
Thursday	2.26	0.001
Friday	0.18	x
Start time of 1st event ²	4.89	0.001
Residual St. Error	2.58 on 203 degrees of freedom	
Adjusted R-square	0.41	

Table B-6: Results from full linear regression model for the Rank #8 bottleneck

Parameters	Coefficients	Significance
Intercept	3.33	0.001
2nd Quarter of Year	0.4	x
3rd Quarter of Year	-0.04	x
4th Quarter of Year	1.72	0.01
Tuesday	1.03	x
Wednesday	1.57	0.05
Thursday	3.21	0.001
Friday	4.79	0.001
Start time of 1st event	-12.8	0.05
Duration of 1st event	1.54	0.001
Start time of 1st event ²	12.8	0.05
Duration of 1st event ²	-0.0008	0.01
Location of 2nd event ²	1.48	0.001
Lane closed during 2nd event	7.04	0.001
Residual St. Error	3.03 on 195 Degrees of freedom	
Adjusted R-square	0.42	

Table B-7: Results from full linear regression model for the Rank #9 bottleneck

Parameters	Coefficients	Significance
Intercept	3.17	0.001
Inclement weather	2.95	0.001
2nd Quarter of Year	0.38	x
3rd Quarter of Year	-0.18	x
4th Quarter of Year	1.78	0.001
Tuesday	1.95	0.001
Wednesday	1.9	0.001
Thursday	2.64	0.001
Friday	1.1	x
Location of 1st event ²	3	0.01
Lane closed during 1st event ²	-0.47	x
Residual St. Error	2.61 on 215 Degrees of freedom	
Adjusted R-square	0.21	

Table B-8: Results from full linear regression model for the Rank #14 bottleneck

Parameters	Coefficients	Significance
Intercept	3.37	0.001
Inclement weather	1.38	x
Duration of 1st event	0.01	x
Lane closed during 1st event	1.96	0.05
Duration of 2nd event	0.05	0.01
Start time of 1st event ²	2.11	x
Residual St. Error	2.85 on 164 Degrees of freedom	
Adjusted R-square	0.16	