GEORGIA DOT RESEARCH PROJECT 17-20

FINAL REPORT

DETECTION TECHNOLOGY TESTBED ON I-475: TECHNOLOGY FEASIBILITY STUDY

OFFICE OF PERFORMANCE-BASED MANAGEMENT AND RESEARCH

**600 WEST PEACHTREE STREET NW
ATLANTA, GA 30308**

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.: FHWA-GA-21-1720	2. Government Accession No.: N/A	3. Recipient's Catalog No.: N/A
4. Title and Subtitle: Detection Technology Testbed on I 475: Technology Feasibility Study	5. Report Date: October 2020	
	6. Performing Organization Code: N/A	
7. Author(s): Angshuman Guin (PI), Ph.D., (https://orcid.org/0000-0001-6949-5126); Michael Hunter (coPI), Ph.D. (https://orcid.org/0000-0002-0307-9127); Han Gyol Kim; and Nishu Choudhary	8. Performing Organization Report No.: 17-20	
9. Performing Organization Name and Address: Georgia Tech Research Corp. 505 Tenth St. Atlanta, GA 30318 Phone: (404) 894-5830 Email: angshuman.guin@ce.gatech.edu	10. Work Unit No.: N/A	
	11. Contract or Grant No.: PI#0015715	
12. Sponsoring Agency Name and Address: Georgia Department of Transportation Office of Performance-based Management and Research 600 West Peachtree St. NW Atlanta, GA 30308	13. Type of Report and Period Covered: Final; August 2017 – October 2020	
	14. Sponsoring Agency Code: N/A	
15. Supplementary Notes: Prepared in cooperation with the U.S. Department of Transportation, Federal Highway Administration.		

16. Abstract:

This project evaluates the feasibility of use and potential benefits of a video-based automatic incident detection (AID) technology relative to existing detection via the Georgia 511 (NaviGAtor) incident reports and transportation management center operators' manual observations. This study proposes a clustering machine learning framework for developing consolidation strategies and filters that will eliminate most noncritical alarms and associate confidence values with the alerts, thereby allowing for a focus on higher confidence alerts during busy periods. The project also investigates the potential of crowdsourced smartphone app-based incident detection and notification, in reducing the time to detection. Finally, the project reviews several of the conventional methods of incident delay estimation, evaluates their accuracy in the presence of noisy data, and develops a new regression-based method to quantify impact of traffic incidents in terms of vehicle delay.

17. Keywords: Automatic Incident Detection, TIRTL, Machine Learning, Incident Delay Estimation	18. Distribution Statement: No Restriction		
19. Security Classification (of this report): Unclassified	20. Security Classification (of this page): Unclassified	21. No. of Pages: 162	22. Price: Free

Form DOT F 1700.7 (8/72)   Reproduction of form and completed page is authorized

GDOT Research Project 17-20

Final Report

DETECTION TECHNOLOGY TESTBED ON I-475:
TECHNOLOGY FEASIBILITY STUDY

By

Angshuman Guin, Ph.D.,
Principal Investigator

Michael Hunter, Ph.D.,
co-Principal Investigator

Han Gyol Kim,
Graduate Research Assistant

Nishu Choudhary,
Graduate Research Assistant

Georgia Tech Research Corporation

Contract with

Georgia Department of Transportation

In cooperation with

U.S. Department of Transportation
Federal Highway Administration

October 2020

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS.

AID	Automatic Incident Detection
ANN	Artificial Neural Networks
ARIMA	Auto-regressive Integrated Moving Average
ATMS	Advanced Traffic Management System
ATR	Automatic Traffic Recorder
CCTV	Closed-Caption Television
CNN	Convolutional Neural Network
CV	Connected Vehicle
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DBTSCAN	Density-Based Temporal and Spatial Clustering Applications with Noise
DR	Detection Rate
DSS	Decision Support System
DTA	Dynamic Traffic Assignment
FHWA	Federal Highway Administration
FSP	Freeway Service Patrol
GDOT	Georgia Department of Transportation
GPRS	General Packet Radio Service
GPS	Global Positioning Systems
GSM	Global System for Mobile communications
GT-MVP	Georgia Institute of Technology's Multi Video Player
HERO	Highway Emergency Response Operator
IoT	Internet of Things

LDA	Latent Dirichlet Allocation
MAD	Mean Absolute Difference
MAPD	Mean Absolute Percentage Difference
MAPE	Mean Absolute Percentage Errors
ML	Machine Learning
MTTD	Mean Time to Detect
NCAR	Noncritical Alarm Rate
RF	Random Forest
SVM	Support Vector Machine
TIRTL	The Infra-Red Traffic Logger
TMC	Transportation Management Center
TTD	Time to Detect
V2X	Vehicle to Everything
VANET	Vehicular Ad hoc Networks
VDS	Video Detection System

EXECUTIVE SUMMARY

Transportation management centers (TMCs) have used automatic incident detection (AID) with varying levels of success in the past. Surveys on the use of incident detection algorithms (Williams and Guin 2007) have indicated a lukewarm response of the industry to AID, primarily because of the false alarms generated by these algorithms. However, AID technology has evolved rapidly in the last several years with significant improvements in video quality and computing resources. Even with the proliferation of mobile phones and the use of smartphone-based applications (apps) to perform crowdsourced incident detection, there is still a relevance for AID under low-volume conditions where there are very few motorists available to make a report, in case the motorists involved in the incident are unable to make a call. Also, AID can significantly cut down on the detection and reporting time, i.e., the time between the actual occurrence of the incident and the time when the TMC is notified about the incident. However, both AID and crowdsourced data have some limitations and challenges related to issues with redundant reports with overwhelming amounts of unusable data.

The key objectives of this project were to:

1. Evaluate the accuracy of the vehicle detection technology deployed in the I-475 testbed.

2. Evaluate the feasibility of using crowdsourced smartphone application–based incident detection for reducing incident detection times.

3. Evaluate the accuracy of the selected AID technology and the feasibility of use of that technology in improving incident management.

1

4. Develop a method to quantify the impact of incidents in terms of vehicle delay.

The project essentially performed four closely related studies. Chapter 2 presents an accuracy evaluation of a vehicle detection technology. Chapter 3 presents an evaluation of the feasibility of using crowdsourced smartphone application–based incident detection for reducing incident detection times. Chapter 4 presents the evaluation of the accuracy of an AID technology and the evaluation of the feasibility of use of that AID technology in improving incident management. Chapter 4 also presents the machine learning–based methodology that was developed for use on top of a base AID algorithm to enable automated identification of potential high-impact incidents. Chapter 5 presents the development of a method to quantify the impact of incidents in terms of vehicle delay to lay the foundations of automated decision support for real-time management of emergency response resources.

Results of the accuracy evaluation of the vehicle detection technology revealed that the count and speed measurements are highly accurate with less than a 2 percent error under normal circumstances. The error in vehicle classification was in the range of 6–7 percent under these conditions, which is typically considered acceptable for most applications. The count errors, however, increase significantly with a downward bias, i.e., the detector fails to detect vehicles under inclement weather conditions, such as heavy rain or snow. The speed measurements had a consistent upward bias when tested at average roadway operating speeds between 40 and 70 mph. However, the errors were typically less than 5 mph (about 10 percent of the average speed).

The evaluation of the feasibility of using crowdsourced smartphone application–based incident detection for reducing incident detection times was performed by comparing detections from the Waze® logs to detections in the Georgia Department of Transportation's NaviGAtor system's incident logs. With the data fusion methodology developed, about 46 percent of the NaviGAtor incidents could be re-identified in the Waze logs in the Atlanta area and about 39 percent in the Macon area. A correlation analysis with the Waze incident attributes confirmed that incidents with lower *report rating* of 0 or 1 in Waze have a slightly lower match rate with NaviGAtor logs. Incidents with a higher *confidence* number in Waze have a higher match rate, and incidents with a *reliability* of 10 in Waze have a higher match rate than the average.

Among the incidents that matched between the two logs, it was observed that in about 57 percent of the cases, the incident appeared in the Waze log before it appeared in the NaviGAtor incident log. In that 57 percent of the cases, the gain in the *time to detection* was largely in the 5–15-minute range. However, in the other 43 percent of the cases, Waze took longer to detect and log the incident than NaviGAtor, with most delays in the range of 0–30 minutes.

The evaluation of the accuracy of an AID technology involved an intensive effort of manual review of videos and images associated with 10,125 incident alarms generated by the AID over a period of 91 days. About 12 percent of the alarms could not be verified to be true because of the lack of evidence based on the videos and images available. About 2.6 percent of the alarms were misplaced in terms of lane assignment. However, there was not enough information available to verify whether the AID missed any incident. A detection rate for the AID technology, therefore, could not be established.

Neither of the unverifiable or misplaced alarm cases, in itself, would likely be a reason for not using AID. The sheer bulk of the "true" alarms generated by the AID, however, consists of very minor incidents that have very little impact on traffic operations or traffic safety. These low-impact alerts, while true alerts, could potentially require significant resources to check and confirm in real time, reducing the efficiency of system usability. A methodology for reducing the high number of noncritical alarms such as shoulder stalls is therefore proposed. The study uses a clustering machine learning framework for developing consolidation strategies along with filters that will eliminate most noncritical alarms and associate confidence values with the alerts, thereby allowing for a focus on higher confidence alerts during busy periods. Clustering evolution patterns of the appearance of multiple alarms, where the basic alarms are generated by the AID system based on traffic anomalies, are used to train the machine learning algorithm to separate potential high-impact incidents from normal congestion or noncritical related stops and slowdowns. The results indicated a significant potential of the framework in consolidating the AID-generated alarms to a small number of high-confidence clusters that can be used in real-time for incident management operations. This methodology might be particularly useful in controlling the number of alarms if AID is deployed over a large coverage area.

In regard to the feasibility of use of AID, it is important to recognize a limitation of the evaluation. The I-475 testbed provided a stretch of freeway with very little recurrent congestion. This helped in the ability to easily confirm the validity of the alerts during the manual review process. However, this also means that the test scenarios did not include recurrent congestion conditions. The performance of the AID technology, when used on freeways with recurrent congestion, has not been evaluated in this study.

In determining AID zone and device placement location, specific attention should be given to merge or diverge points, weaving areas, or other zones with a higher potential for an incident. Should an incident occur outside of the detection zone of the AID device (such as the view of a camera in video-based AID), the AID will not provide feedback until the results of the incident (e.g., spillback) encroach into the detection zones (i.e., come into the view of the cameras). Thus, placement should consider the potential of such lag in receiving information. In addition, for video-based AID, attention should be paid to items such as seasonal growth of vegetation and other potential temporary obstructions in the camera frame, as they may be interpreted as an incident. Finally, for a video-based AID, the same cautions, as required for video-based vehicle detection systems, are recommended. For example, the camera angle should be as steep (overhead) as possible to limit occlusion-related errors (both vertical, i.e., within lane, and horizontal, i.e., across lanes). This can be a particular challenge where a camera angle precludes the AID from being able to distinguish between a vehicle on the shoulder and in the right travel lane, as vehicles on shoulders result in the majority of detections, and filtering out or assigning a lower priority to these alarms is often desirable. During nighttime conditions, flat camera angles can produce views that generate false alarms of wrong-way detection from reflections of headlights on roadside objects such as barrier walls. A balance needs to be achieved between producing a larger area of detection by using flatter angles of the cameras, which will lead to lesser "blind spots," versus a higher quality of detection within a smaller area produced by steeper angles of cameras.

Lastly, the project involved a study to develop a method to quantify the impact of incidents in terms of vehicle delay. Spot speed and vehicle count measurement has been the most

widely accepted performance monitoring method for traffic operations data collection by transportation agencies. Delay estimation methods based on spot speed and cumulative count are typically deployed by practitioners and researchers alike for rapid estimation of delays as a precursor to congestion mitigation. In this report, these commonly used incident-induced delay estimation methodologies, which are based on queuing theory or shockwave analysis models, are reviewed and validated against microscopic simulation of a real-life incident. For the simulation model, NaviGAtor speed–volume data were used. The incident timeline was constructed using NaviGAtor incident logs. The comparison revealed challenges related to noisy data and the failure of spot-speed measurements to adequately capture heterogeneity in congested traffic, which rendered the methodologies impractical for field use. In the absence of any alternative method to accurately quantify delay within the constraints of field observational data, a regression model was developed using data from a non-exhaustive set of incident scenarios simulated using Vissim®, to help obtain rapid estimates of delays for incidents with varying characteristics occurring under varying base conditions. This regression model can aid in resource allocation for efficient incident management and identification of influence factors.

CHAPTER 1.   INTRODUCTION

Transportation management centers (TMCs) have used automatic incident detection (AID) with varying levels of success in the past. The early detection tools used real-time traffic flow data-based algorithms to identify anomalies in traffic. More recent AID tools are based on real-time analysis of video streams. For instance, the Georgia Department of Transportation (GDOT) currently utilizes video analysis technology for the detection of stopped vehicles on shoulders and limited areas of active lanes. However, AID technology has evolved rapidly in the last several years, with significant improvements in video quality and computing resources. In light of the recent evolution in video-based AID technologies, it is necessary to evaluate the feasibility of use of this technology by TMCs.

Surveys on the use of incident detection algorithms (Williams and Guin 2007) have indicated a lukewarm response of the industry to AID, primarily because of the false alarms generated by these algorithms. Similar observations have been made with video-based AID in previous studies (Gillen 2001)regarding the occurrence of false alarms. In addition, with the proliferation of cellular phones, manual detection based on calls from motorists has become the primary method of detection. Crowdsourced methods of detection using smartphone-based applications (apps) are another detection method that has recently made inroads into the detection process. However, there is still a relevance for AID under low-volume conditions where there are very few motorists available to make a report, in case the motorists involved in the incident are unable to make a call. Also, AID can significantly

reduce the detection and reporting time, i.e., the time between the actual occurrence of the incident and the time when the TMC is notified about the incident.

The overarching goal of this project is to study a selected video-based AID technology relative to the existing detection via the Georgia 511 (NaviGAtor) incident reports and the manual observations of TMC operators.

Moreover, crowdsourced methods of incident detection using smartphone-based apps have recently started making inroads into the incident management process. Waze®, owned by Google, has teamed up with state, county, and city departments of transportation in an effort to help with the integration. While the benefits of such a partnership are obvious, crowdsourced data have some limitations and challenges related to issues with redundant reports with overwhelming amounts of unusable data. To investigate the potential of incident detection and notification via crowdsourced smartphone apps, such as Waze, in reducing the time to detection (TTD), a comparative analysis is performed using Waze and the incident logs from NaviGAtor, the GDOT TMC's advanced traffic management system (ATMS).

In addition to the primary goal of evaluating the incident detection technology, this project also evaluates a vehicle detection technology. The project evaluates the detection accuracy and quality of data generated by the vehicle detection technology test-deployed by GDOT on I-475.

The key objectives of the project are to:

1. Evaluate the accuracy of the vehicle detection technology deployed in the I-475 testbed.

2. Evaluate the feasibility of using crowdsourced smartphone application–based incident detection for reducing incident detection times.

3. Evaluate the accuracy of the selected AID technology and the feasibility of use of the AID technology in improving incident management.

4. Develop a method to quantify the impact of incidents in terms of vehicle delay.

Chapter 2 presents the results of the accuracy evaluation of the vehicle detection technology selected by GDOT for testing on the I-475 testbed. The accuracies of vehicle counts and vehicle speed measurements, under different traffic scenarios and different ambient conditions, are evaluated against manual counts and measurements obtained from videos recorded for the test site.

Chapter 3 presents the results of the evaluation of the feasibility of using crowdsourced smartphone application–based incident detection for reducing incident detection times. Data provided by the navigation smartphone application Waze®, through the Connected Citizens Program (now called Waze for Cities), is compared with GDOT's incident management program's incident logs by developing a data fusion approach whereby an incident in one dataset is identified in the other dataset with a high degree of confidence. The data fusion enables the computation of the potential time savings that can be realized by an early detection of an incident by using Waze as compared to the currently employed methods of detection.

Chapter 4 presents the evaluation of the accuracy of the selected AID technology and the use of the AID technology in improving incident management. Evaluations are performed with a manual review of video and image logs of the alarms identified by the AID. The study identified a potential improvement in system efficiency through the separation of incidents with lower impacts on traffic to allow for the prioritization of the high-impact incidents. A methodology based on machine learning (ML) is developed that can be used on top of a base AID algorithm to enable automated identification of potential high-impact incidents.

Chapter 5 develops a method to quantify the impact of incidents in terms of vehicle delay to lay the foundations of automated decision support for real-time management of emergency response resources. This chapter takes a critical look at several of the conventional methods of incident delay estimation and demonstrates their potential failure for accurate estimation of delay in the presence of noisy data and the failure of the homogeneity assumption of traffic under congested conditions. A regression-based model is developed for rapid estimation of incident delay that can be used to produce robust results even in the presence of data noise.

CHAPTER 2.   FLOW AND SPEED DATA VALIDATION

INTRODUCTION

The Georgia Department of Transportation uses a wide array of detection technologies to support NaviGAtor, GDOT's advanced traffic management system. To ensure that the system can take advantage of new and emerging detection technologies, it is essential to run comprehensive field tests on these technologies before deployment. In this chapter, an evaluation of the detection accuracy and quality of data is performed for one of the emerging vehicle detection technologies that is being test-deployed by GDOT on I-475. The vehicle detection technology evaluated is The Infra-Red Traffic Logger, or TIRTL. TIRTL uses non-invasive light-based detection for vehicle count, classification, lane association, and speed measurement (see figure 1) (CEOS 2020). The accuracy evaluation of vehicle counts and speeds consisted of comparisons with data obtained from traditional loop detectors and manual counts under different test scenarios, such as, under light-to-medium traffic flow, during inclement weather conditions, and during ongoing construction that results in heavier traffic on one lane. The Methodology and Results sections below provide details about the method employed to extract the data for analysis and the results of the evaluation, respectively.

METHODOLOGY

For detection accuracy comparison, a data quality check was done for vehicle counts and speed estimates from TIRTL detectors set up along a segment on I-475 (see figure 2, where the TIRTL detector is highlighted by a red rectangle). Subsequent analysis presented in this

section is divided into two parts: first, an evaluation of the vehicle counts generated by TIRTL, and, second, an evaluation of the speed measurements provided by TIRTL.
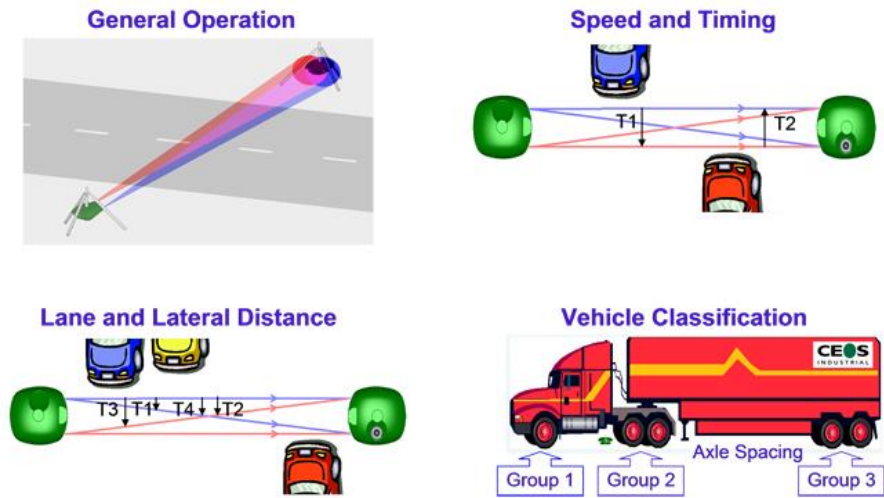
**Figure 1. Illustrations. The working principles of TIRTL.
(Source: TIRTL https://www.ceos.com.au/products/tirtl/)**
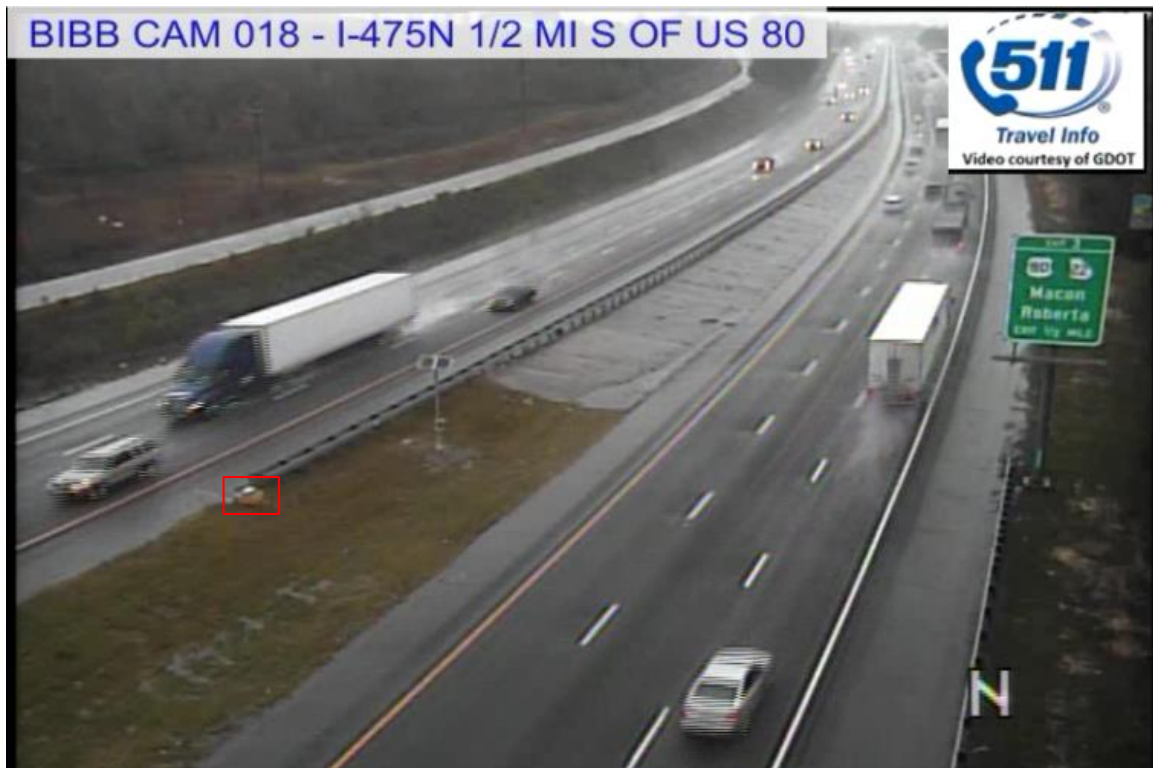
Figure 2. Photo. TIRTL detector setup along I-475 in northbound direction.

Vehicle Count Comparison

To determine the accuracy of the vehicle counts, the data from TIRTL are compared

against the data from an adjacent inductive loop detector, as well as manual counts. The

manual counts for this study were extracted using video recordings of the test segment

obtained from GDOT's traffic monitoring cameras. The videos recorded the traffic from

10/24/2017 to 03/06/2018, with each day's video clip recording from 5:00 AM to 4:00 AM

the next day. Video data were recorded and processed for different traffic conditions: under

inclement weather conditions (i.e., rain and snow), during ongoing construction-related

congestion along the road segment, and under regular light-traffic conditions. The dates and

periods of the video data processed for data extraction are provided in table 1.

Table 1. Schedule of recorded video data for evaluation.

Date of the Recording	Time	Total Duration (minutes)	Driving Conditions
10/30/2017	09:02:03–09:22:03	29	Regular
	17:00:08–17:09:17		
10/31/2017	11:25:06–11:35:09	14	Construction in right-most lane
	12:03:10–12:07:15		
11/09/2017	10:59:00–11:09:00	20	Rainy (inclement weather)
	11:30:04–11:40:49		
12/08/2017	12:57:02–13:17:01	31	Snow (inclement weather)
	13:57:24–14:08:01		

To obtain the vehicle counts, data were manually extracted from the videos using Georgia

Institute of Technology's Multi Video Player (GT-MVP). GT-MVP is a Python®-based

software application developed by the Georgia Tech transportation research group to

provide a user-friendly interface to extract complex traffic data from videos (Saroj et al.

2018). The extracted vehicle count data included the timestamp, lane number, and

classification of each vehicle when it crossed the detector, as observed from the video.

Vehicles were classified according to Federal Highway Administration (FHWA) vehicle

classification guidelines (Office of Research 2014). Table 2 shows a sample of extracted

data for 11/09/2017 with start time as 10:59:00 AM. The two datasets, one given by TIRTL

and the other extracted from the videos, were then combined for verification. Figure 3

shows a screenshot of a typical Excel sheet with the combined datasets, where the left side

of the sheet is manually extracted data and right side is from TIRTL.

The video data was recorded over a 4-month period to ensure that different traffic operation

conditions are captured. However, as described above, the data extraction process was

extremely time consuming, involving frame by frame playback of videos, and a second

pass to review the records in order to ensure high accuracy of the manually extracted data.

Previous studies (Guin et al. 2016; Toth et al. 2013) have shown that attempts at faster data

reduction lead to inaccuracies in the manual data and can lead to errors in the evaluation.

Hence, a small sample size, that allowed for data collection by experienced researchers,

supplemented by review of the datapoints, was chosen to ensure an accurate evaluation.

For data verification, the TIRTL data were compared with the manually extracted data. For

comparisons using small datasets over a limited time period, there is a risk of introduction

of biases because of timestamp mismatches. For example, a network-latency or clock-

offset–related time shift of 15 seconds that results in one vehicle platoon being counted

within one dataset and missed in another will not have a significant effect when the period

of aggregation is 15 minutes, but such a shift will generate large spurious errors when the

aggregation period is short, such as 1 minute. With the resource-intensive nature of the data

collection, the amount of video processed for data extraction was limited to 30 minutes. To

eliminate spurious errors in the comparison, the pattern of arrival of vehicles and the

corresponding vehicle class (in a subset of the data) were used to compute the time offset

adjustment required to ensure that the vehicle arrivals align in both time series. The typical

offset observed was a latency of 8–9 seconds in the TIRTL data.

Table 2. Sample extracted data for 11/09/2017
with start time of 10:59:00 AM.

Time – Video 1	Lane	Vehicle Type
0:00:12	3	11
0:00:15	3	2
0:00:18	3	2
0:00:22	3	2
0:00:25	3	2
0:00:34	3	2
0:00:35	3	7
0:00:40	3	2
0:00:46	3	2
0:00:47	3	2
0:01:06	3	11
0:01:13	3	9
0:01:22	3	2
0:01:28	3	6
0:01:32	3	9
0:01:44	3	10
0:01:51	3	2
0:01:56	3	9
0:01:59	3	2
0:02:02	3	2
0:02:07	3	2
0:02:21	3	2
0:02:23	3	2

(a) Manually extracted data

Cumulative	Lane	Vehicle type	Error Type
10:59:02	1	2	
10:59:02	2	8	Vehicle-class misclassification
10:59:03	1	2	Missed vehicle
10:59:06	1	2	
10:59:07	1	2	
10:59:09	1	2	
10:59:10	2	2	
10:59:12	3	11	Vehicle-class misclassification
10:59:12	2	2	Missed vehicle
10:59:15	3	2	
10:59:16	2	2	
10:59:18	3	2	
10:59:21	2	8	Missed vehicle
10:59:22	3	2	
10:59:25	3	2	
10:59:27	2	2	
10:59:30	1	2	
10:59:30	2	2	
10:59:34	3	2	
10:59:34	2	9	Vehicle-class misclassification
10:59:35	3	7	Missed vehicle
10:59:35	1	2	
10:59:36	2	2	Missed vehicle
10:59:38	1	2	
10:59:40	3	2	
10:59:42	2	8	Vehicle-class misclassification
10:59:45	1	2	
10:59:46	3	2	
10:59:46	2	2	
10:59:47	3	2	
10:59:47	1	3	
10:59:48	2	5	Vehicle-class misclassification
10:59:53	1	2	
10:59:55	1	2	Missed vehicle
10:59:55	2	8	Vehicle-class misclassification
10:59:59	2	2	Vehicle-class misclassification

(b) TIRTL data

time	Latency		lane	Vehicle Ty	Error Type
10:59:34	0:00:32	10:59:02	1	2	
10:59:34	0:00:32	10:59:02	2	5	Vehicle-class misclassification
10:59:38	0:00:32	10:59:06	1	3	
10:59:39	0:00:32	10:59:07	1	3	
10:59:40	0:00:32	10:59:08	1	2	
10:59:42	0:00:32	10:59:10	2	2	
10:59:44	0:00:32	10:59:12	3	3	Vehicle-class misclassification
10:59:47	0:00:32	10:59:15	3	2	
10:59:48	0:00:32	10:59:16	2	2	
10:59:50	0:00:32	10:59:18	3	3	
10:59:54	0:00:32	10:59:22	3	2	
10:59:57	0:00:32	10:59:25	3	2	
10:59:59	0:00:32	10:59:27	2	2	
11:00:02	0:00:32	10:59:30	1	3	
11:00:02	0:00:32	10:59:30	2	2	
11:00:06	0:00:32	10:59:34	2	5	Vehicle-class misclassification
11:00:06	0:00:32	10:59:34	3	2	
11:00:07	0:00:32	10:59:35	1	2	
11:00:09	0:00:32	10:59:37	1	2	
11:00:12	0:00:32	10:59:40	3	3	
11:00:14	0:00:32	10:59:42	2	5	Vehicle-class misclassification
11:00:17	0:00:32	10:59:45	1	3	
11:00:18	0:00:32	10:59:46	2	2	
11:00:18	0:00:32	10:59:46	3	2	
11:00:19	0:00:32	10:59:47	1	3	
11:00:19	0:00:32	10:59:47	3	2	
11:00:19	0:00:32	10:59:47	2	4	Vehicle-class misclassification
11:00:24	0:00:32	10:59:52	1	2	
11:00:27	0:00:32	10:59:55	2	5	Vehicle-class misclassification
11:00:30	0:00:32	10:59:58	2	8	Vehicle-class misclassification
11:00:46	0:00:32	11:00:14	3	3	Vehicle-class misclassification
11:00:54	0:00:32	11:00:22	3	2	
11:01:00	0:00:32	11:00:28	3	5	Vehicle-class misclassification
11:01:02	0:00:32	11:00:30	1	3	
11:01:10	0:00:32	11:00:38	1	2	
11:01:23	0:00:32	11:00:51	3	2	

Figure 3. Datasets. Data for 11/09/2017 with start time 10:59:00 AM.

Speed Estimate Verification

For accuracy evaluation of speed measurements, the TIRTL-generated speeds were compared to speed data extracted manually from videos at the test site. Videos for a particular day (i.e., 10/31/2018) starting from 11:34:00 AM were used for this purpose. During this particular day, construction activity was in progress in the lane farthest from the median, in the northbound direction. To extract ground truth from the selected video, VirtualDub (version 1.10.4), an open-source video editing software, was used (Lee 2000). Figure 4 shows a screenshot of VirtualDub's user interface. The video was played back frame by frame. The speed for any vehicle was estimated by counting the number of frames

it takes a vehicle to cross four skip lines (130 ft), marked by red lines in figure 4. Speed

estimates for 40 vehicles, of different classes, in the two available lanes were obtained

using this approach. The results of this analysis are described in Results below.
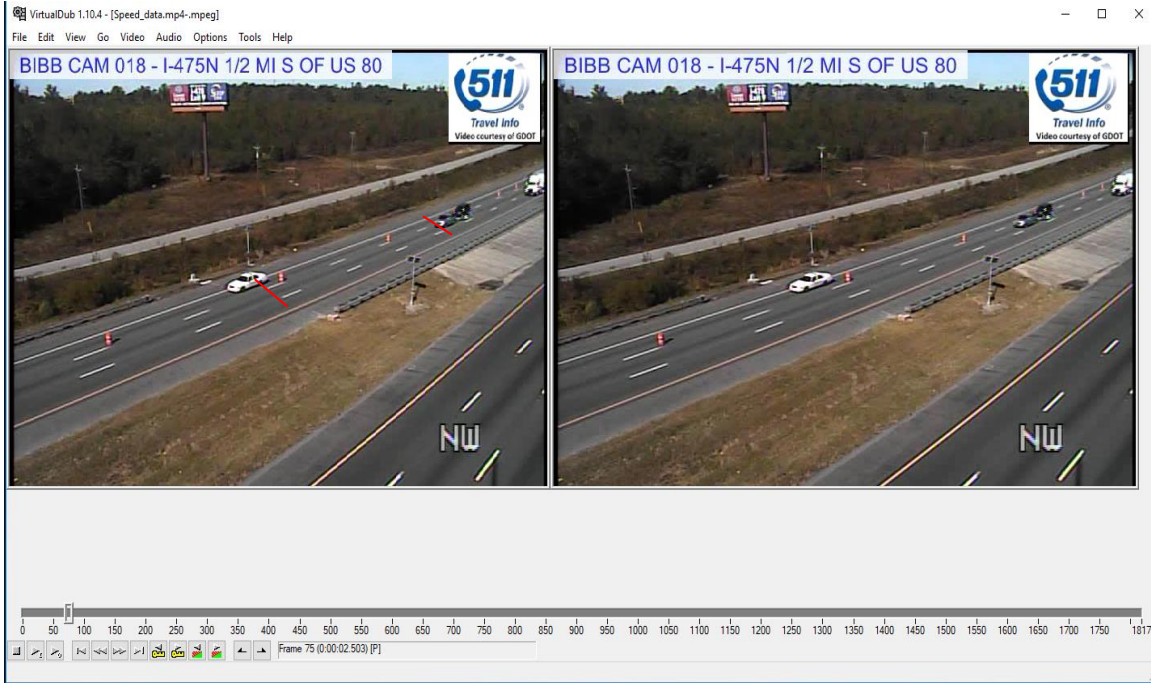
Figure 4. Screenshot. VirtualDub user interface for video-based speed estimates.

RESULTS

There were four types of errors typically observed in the evaluation of the TIRTL data.

These errors types are listed in table 3. Results of the evaluation under different traffic and

ambient conditions are described in the following subsections.

Table 3. Types of errors observed in the TIRTL data.

Type of Error	Description
Vehicle-class misclassification	Detected vehicle with wrong vehicle-class
Missed vehicle	Vehicle not counted by the detector
Lane misclassification	Detected vehicle with incorrect lane number
Lane and vehicle-class misclassification	Both lane and vehicle-class of the detected vehicle are incorrect

Uncongested Flow During Regular Driving Conditions

The base case with normal driving conditions and without any inclement weather conditions was from 10/30/2017. Figure 5 shows a plot of the manual cumulative counts with an overlapping "adjusted cumulative counts" from the TIRTL data, with time on the y-axis and cumulative count on the x-axis. In a cumulative curve plot, a missing vehicle or an extra vehicle would cause a permanent divergence. From such a plot, it is easy to see the overall aggregate effect of the misses/additions; however, it is difficult to separate out the "good" portions of the time series, where there were no errors, from the "bad" portions, where there are errors. Hence the cumulative counts in TIRTL were "adjusted" by the amount of the error at the end of a series of errors to make sure that the divergence was controlled and the lines only showed "gaps" or divergence where there were missing or overcounted data. The two lines overlap to a large extent. The results of the data verification are shown in table 4. The TIRTL vehicle counts match very well with the manual counts. The TIRTL detector missed only one vehicle during the 30-minute analysis period. The error (6.98 percent) was largely in the classification of the vehicles.

Table 5 presents an analysis of the lane-wise variation of accuracy of counts over a total period of 29 minutes (09:02 AM – 09:22 AM and 05:00 PM – 05:09 PM). The error rate of

vehicle misclassification is observed to be higher for the lane farthest from the median. To check if the two datasets are statistically different, a paired t-test was undertaken for the 1-minute aggregated counts. The paired t-test results indicate that there is not enough evidence to reject the null hypothesis that there is no significant difference between the TIRTL counts and manual counts under regular driving conditions, based on data from 10/30/2017.
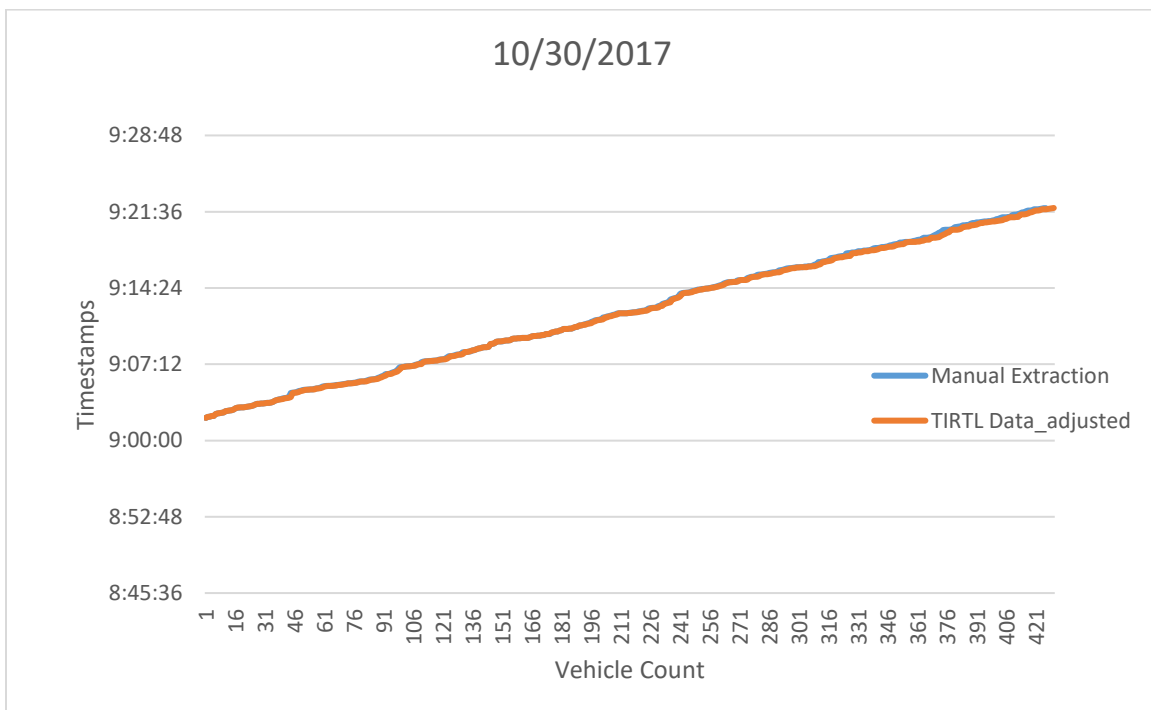
Figure 5. Graph. Data from 09:02:03 AM to 09:22:03 AM, extracted during normal driving conditions.

Table 4. Data verification results for *regular* day – 10/30/2017.

Type of Error	Count	% Error
Vehicle-class misclassification	56/802	6.98
Lane misclassification	11/802	1.37
Lane and vehicle misclassification	1/802	0.12
Missed vehicles	3/802	0.37

Table 5. Data verification results lane-by-lane for *regular* day – 10/30/2017.

Type of Error	Lane 1		Lane 2		Lane 3	
	Count	% Error	Count	% Error	Count	% Error
Vehicle-class misclassification	5/182	0.03%	21/327	6.42%	30/293	10.24%
Lane misclassification	1/182	0.01%	4/327	1.22%	6/293	2.05%
Lane and vehicle misclassification	0/182	0.00%	0/327	0.00%	1/293	0.34%
Missed vehicles	1/182	0.01%	1/327	0.31%	1/293	0.34%
Extra detection by TIRTL[1]	7/182	0.04%	4/327	1.22%	7/293	2.39%

[1]Not observable through video.

Congested Flow During Construction

Typically, on this portion of the freeway, there is very little congestion. To test the

performance of TIRTL under low-speed and high-density traffic flow conditions, data from

a day with construction activity were chosen. On this particular day, construction activity

was ongoing in the right-most lane (i.e., the farthest lane from the median in the

northbound direction). The results of this analysis, presented in table 6, show that the

proportion of missed vehicles is marginally higher than under the regular flow conditions,

but overall the error rates are still relatively low. The "adjusted cumulative count" plots

(see figure 6) visually confirm the overlap of the data from TIRTL and the manual

extraction. Table 7 provides an analysis of the lane-wise variation of accuracy. The paired

t-test results conducted on 1-minute aggregates indicate that there is not enough evidence to

reject the null hypothesis that there is no significant difference between the TIRTL counts

and the manual counts under medium congestion.

Table 6. Data verification result for *construction* day – 10/31/2017.

Type of Error	Count	% Error
Vehicle-class misclassification	26/436	5.96
Lane misclassification	0/436	0.00
Lane and vehicle misclassification	0/436	0.00
Missed vehicles	5/436	1.15
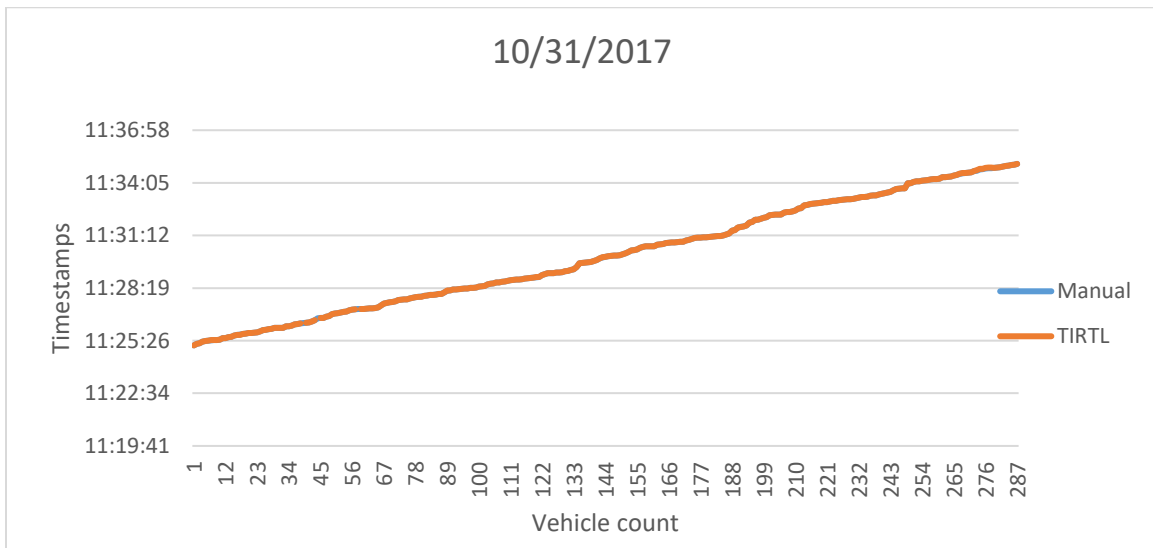
Figure 6. Graph. Data from 11:25:06 AM to 11:35:09 AM, extracted during ongoing construction.

Table 7. Data verification results lane-by-lane for *construction* day – 10/31/2017.

	Lane 1		Lane 2		Lane 3	
	Count	% Error	Count	% Error	Count	% Error
Vehicle-class misclassification	11/258	4.26	15/177	8.47	0/1	0.00
Lane misclassification	0/258	0.00	0/177	0.00	0/1	0.00
Lane and vehicle misclassification	0/258	0.00	0/177	0.00	0/1	0.00
Missed vehicles	3/258	1.16	2/177	1.13	0/1	0.00
Extra detection by TIRTL[1]	0/258	0.00	0/177	0.00	0/1	0.00

[1]Not observable through video but assumed to be present.

Inclement Weather Conditions: Heavy Rainfall

To test the performance of TIRTL under inclement weather conditions, a day with heavy rainfall (i.e., 11/09/2017) was selected. A total of 20 minutes of video data were extracted for the analysis (10:59 AM – 11:09 AM and 11:30 AM – 11:40 AM), beyond which visibility challenges prevented further processing of the video. The results of this analysis are summarized in table 8 and table 9. For this case, the proportion of missed vehicles was significantly higher than for the regular flow and congested flow cases. The results of a paired t-test conducted on 1-minute aggregated data show that there is sufficient evidence to reject the null hypothesis; thus, we conclude that the TIRTL data and the manual data are statistically different. An analysis of the error patterns revealed that when TIRTL missed vehicles, it occurred in batches. This pattern can be seen in figure 7 and figure 8. Note that figure 7 and figure 8 plots the "adjusted cumulative counts" of the TIRTL data over the actual cumulative counts from the manual data, as explained previously in the subsection on Uncongested Flow During Regular Driving Conditions. The majority of missed vehicles during this timeframe were heavy vehicles. It is hypothesized that these misses might be correlated to standing water on the road or water bouncing off the road and water coming off the tires of heavy vehicles. An interesting pattern about the proportion of missed vehicles also emerges from the lane-by-lane analysis of data accuracy given in table 9, where the majority of the missed vehicles were associated with lanes 2 and 3, which are farther from the median.

Table 8. Data verification result for *rainy* day – 11/09/2017.

Type of Error	Count	% Error
Vehicle-class misclassification	79/540	14.63%
Lane misclassification	0/540	0.00%
Lane and vehicle misclassification	0/540	0.00%
Missed vehicles	86/540	15.93%

Table 9. Data verification lane-by-lane on *rainy* day – 11/09/2017.

Type of Error	Lane 1		Lane 2		Lane 3	
	Count	% Error	Count	% Error	Count	% Error
Vehicle-class misclassification	4/152	2.63%	45/223	20.18%	30/165	18.18%
Lane misclassification	0/152	0.00%	0/223	0.00%	0/165	0.00%
Lane and vehicle misclassification	0/152	0.00%	0/223	0.00%	0/165	0.00%
Missed vehicles	16/152	10.53%	35/223	15.70%	35/165	21.21%
Extra detection by TIRTL[1]	1/152	0.66%	2/223	0.90%	1/165	0.61%

[1] Not observable through video but assumed to be present.
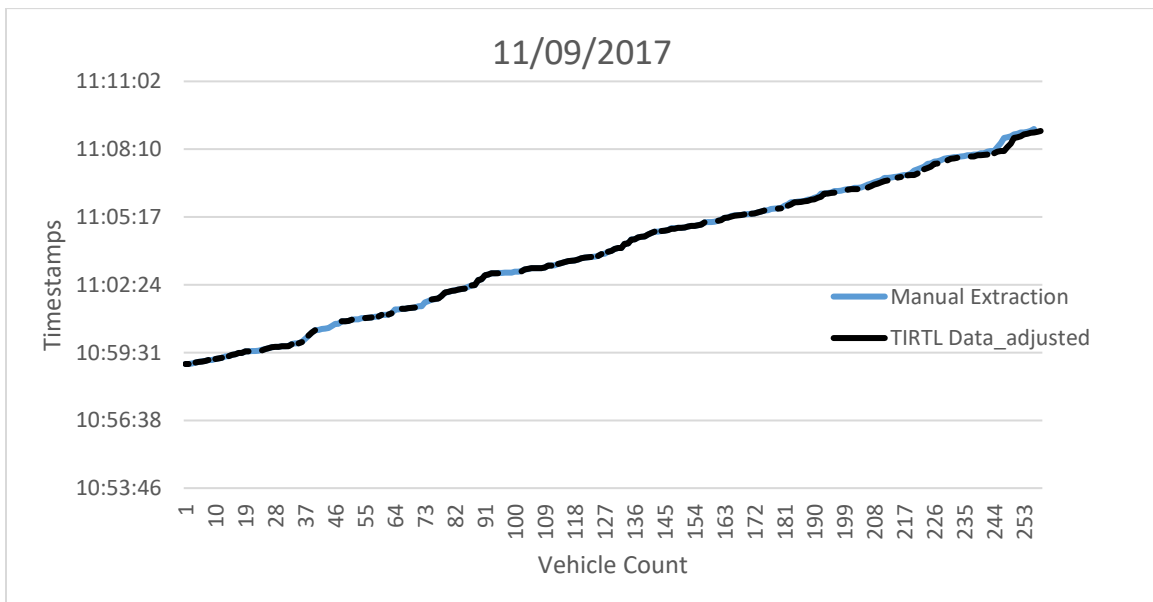
Figure 7. Graph. Data from 10:59:00 AM to 11:09:00 AM, extracted during inclement weather conditions.
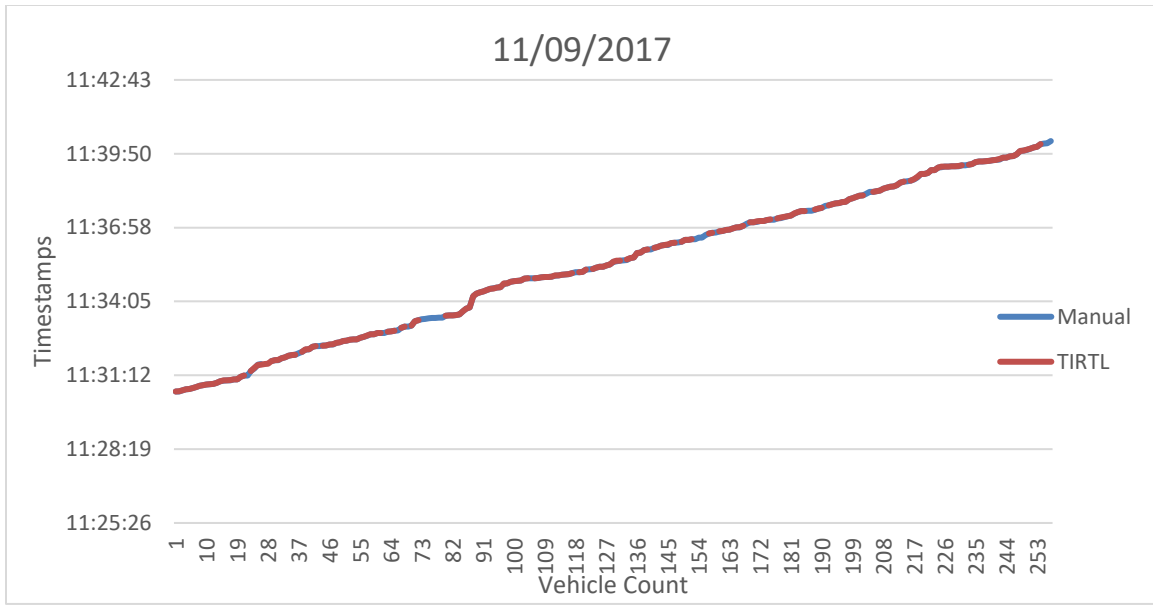
Figure 8. Graph. Data from 11:30:32 AM to 11:40:20 AM, extracted during inclement weather conditions.

Inclement Weather Conditions: Heavy Snowfall

The accuracy of TIRTL data was also evaluated under heavy snowfall conditions. On 12/08/2017 there was heavy snowfall in the region around the test site. Thirty minutes of video data were extracted over two periods to focus on the periods with visible snow accumulation on the roadway, 12:57 PM to 01:17 PM and 01:57 PM to 02:08 PM. The results are presented in table 10 and table 11. The proportion of missed vehicles for this scenario is higher than any other scenario in this study. Again, the majority of these missed vehicles were missed by the detector during a short period of time, as can be seen from figure 9 and figure 10. Results of a paired t-test, conducted on 1-minute aggregated data, show that the two datasets are statistically different. Other errors, such as lane misclassification and vehicle misclassification errors, are at levels similar to the other scenarios.

Table 10. Data verification result for *snowy* day – 12/08/2017.

Type of Error	Count	% Error
Vehicle-class misclassification	71/866	8.20%
Lane misclassification	2/866	0.23%
Lane and vehicle misclassification	0/866	0.00%
Missed vehicles	173/866	19.98%

Table 11. Data verification lane-by-lane on *snowy* day – 12/08/2017.

Type of Error	Lane 1		Lane 2		Lane 3	
	Count	% Error	Count	% Error	Count	% Error
Vehicle-class misclassification	5/235	2.13%	33/376	8.78%	33/255	12.94%
Lane misclassification	0/235	0.00%	1/376	0.27%	1/255	0.39%
Lane and vehicle misclassification	0/235	0.00%	0/376	0.00%	0/255	0.00%
Missed vehicles	35/235	14.89%	81/376	21.54%	57/255	22.35%
Extra detection by TIRTL[1]	5/235	2.13%	9/376	2.39%	9/255	3.53%

[1] Not observable through video but assumed to be present.
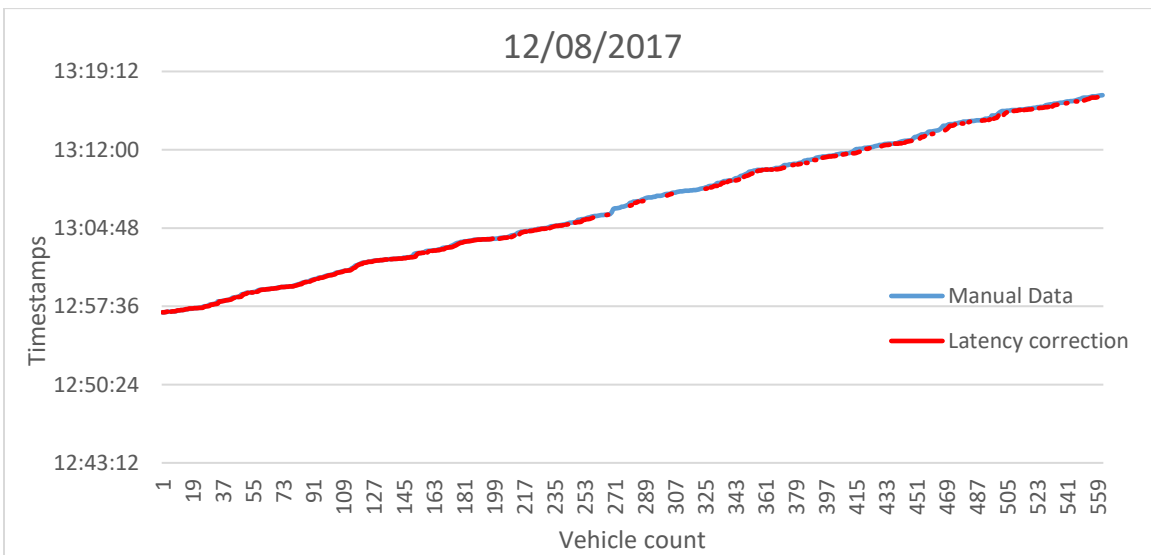
Figure 9. Graph. Data from 12:57 PM to 01:17 PM, extracted during snow conditions.
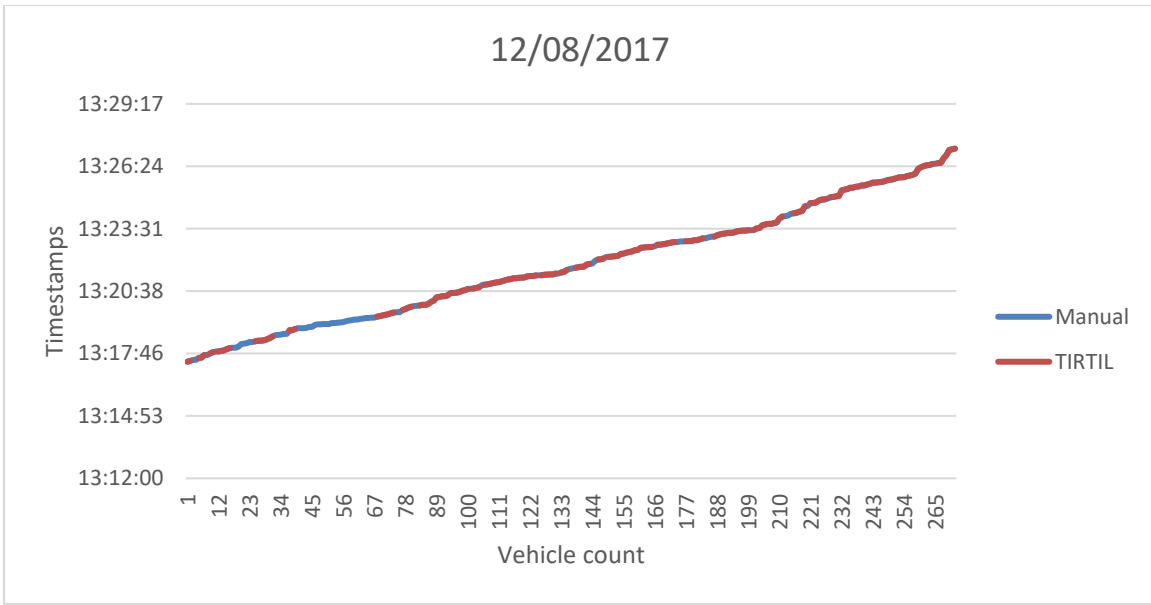
Figure 10. Graph. Data from 01:17 PM to 01:27 PM, extracted during snow conditions.

Results of the analyses for all of the above-described traffic conditions are summarized in table 12. An interesting observation from the table is that during inclement weather conditions, TIRTL's performance deteriorates. This is especially true for the number of vehicles missed and the vehicle-class classification. Other types of errors, such as lane misclassification, remain relatively comparable with and without inclement weather. Even with this high error rate during bad weather, TIRTL can have an advantage over video detection systems (VDSs) when a vehicle is not clearly observable through the camera view, the likelihood of which can increase with an increase in the number of heavy vehicles in the traffic mix because of the additional obscurity from the spray generated by the wheels of the heavy vehicles.

Table 12. Vehicle count data verification results.

Driving Condition	Date and Duration	Vehicle-Class Misclassification[1]	Lane Misclassification	Lane & Vehicle Misclassification	Missed
Regular	10/30/2017 29 min	56/802 (6.98%)	11/802 (1.37%)	1/802 (0.12%)	3/802 (0.37%)
Construction zone (less than 20–30 mph)	10/31/2017 14 min	7/436 (1.60%)	0/436 (0.00%)	0/436 (0.00%)	5/436 (1.10%)
Rain	11/09/2017 20 min	79/540 (14.63%)	0/540 (0.00%)	0/540 (0.00%)	86/540 (15.93%)
Snow	12/08/2017 31 min	81/866 (9.35%)	4/866 (0.46%)	0/866 (0.00%)	145/866 (16.74%)

[1]Vehicle-class misclassification: only the obvious cases, e.g., where 4-axle vehicle was counted as a passenger car, were considered.

Comparison with Loop Detector

The TIRTL data was compared with data from an inductive loop detector station adjacent to the deployment location of the TIRTL detector. While the manual verification of the TIRTL data focused on a small sample, the cross-technology comparison took a higher-level approach. Hourly aggregated data over a 27-day period from 10/23/2017 to 11/18/2017 was used for the comparison. Lane-by-lane data across the 15 FHWA vehicle classes were compared. The results are presented in table x. Since there is not sufficient evidence to consider either of the datasets as ground truth, the difference in magnitude between the vehicle counts in the two datasets is being called "difference" instead of "error". The table presents the Mean Absolute Difference (MAD) and average hourly vehicle count for each lane and each class, as well as the total across all classes. However, several of the 1-hour periods had zero vehicles reported on a lane for one or more of the vehicles classes. This made it impossible to compute a reliable Mean Absolute Percentage Difference (MAPD) because the calculation would create a division by zero error. Hence a defect rate estimation method was used to capture the degree of disagreements between the

27

classification of the vehicles between the two technologies. If for a certain hourly period on a certain day for a certain lane, the TIRTL count is not equal to the inductive loop count, the datapoint is tagged as a point with a disagreement. The ratio of the number of points with disagreements to the total number of points gives the Gross Disagreement Rate. The following observations can be made from the results table:

- Differences in classification of Passenger Cars are uniformly low across all lanes (below 10%)

- There are larger percentage differences in the classification of other categories, even though the absolute values are small.

- There are disagreements in all classes across all lanes, except class 7, 14, and 15, are > 10 %

- Lane 1 has the least disagreements

Table 13. Vehicle count and classification comparison between TIRTL and inductive loop detectors.

Vehicle Class	Class Description	Lane 1				Lane 2				Lane 3				All 3 lanes			
		MAD	MAPD	Gross Disagreement Rate	Avg # Vehicles per hour	MAD	MAPD	Gross Disagreement Rate	Avg # Vehicles per hour	MAD	MAPD	Gross Disagreement Rate	Avg # Vehicles per hour	MAD	MAPD	Gross Disagreement Rate	Avg # Vehicles per hour
1	Motorcycles	2	-	63.06%	2	3	84.67%	72.80%	3	6	-	81.30%	6	8	-	89.80%	10
2	Passenger Cars	19	6.74%	96.14%	282	6	1.82%	90.26%	335	10	5.61%	95.83%	189	15	1.83%	95.52%	806
3	Other Two-Axle Four-Tire Single-Unit Vehicles	9	-	90.73%	53	12	11.79%	96.91%	95	9	15.52%	94.74%	63	12	6.43%	95.98%	210
4	Buses	1	-	18.24%	1	1	-	38.95%	5	1	-	43.43%	5	2	-	56.88%	10
5	Two-Axle, Six-Tire, Single-Unit Trucks	7	-	88.87%	2	6	-	93.35%	14	6	-	92.58%	12	17	-	98.92%	26
6	Three-Axle Single-Unit Trucks	1	-	5.41%	1	1	-	51.93%	4	2	-	61.05%	6	2	-	69.55%	9
7	Four or More Axle Single-Unit Trucks	1	-	0.46%	1	1	-	5.26%	1	1	-	6.65%	1	1	-	11.59%	1
8	Four or Fewer Axle Single-Trailer Trucks	4	-	85.63%	1	13	-	99.23%	9	14	-	99.54%	10	30	188.87%	100.00%	19
9	Five-Axle Single-Trailer Trucks	1	-	21.64%	2	5	4.83%	82.69%	90	5	4.35%	83.93%	115	9	3.77%	90.11%	206
10	Six or More Axle Single-Trailer Trucks	1	-	1.85%	1	1	-	44.67%	1	2	-	73.57%	1	3	-	82.69%	2

Vehicle Class	Class Description	Lane 1				Lane 2				Lane 3				All 3 lanes			
		MAD	MAPD	Gross Disagreement Rate	Avg # Vehicles per hour	MAD	MAPD	Gross Disagreement Rate	Avg # Vehicles per hour	MAD	MAPD	Gross Disagreement Rate	Avg # Vehicles per hour	MAD	MAPD	Gross Disagreement Rate	Avg # Vehicles per hour
11	Five or Fewer Axle Multi-Trailer Trucks	1	-	5.26%	1	4	-	87.48%	4	8	-	95.52%	8	11	-	99.07%	11
12	Six-Axle Multi-Trailer Trucks	1	-	1.39%	1	1	-	23.03%	4	1	-	25.04%	5	1	-	32.30%	9
13	Seven or More Axle Multi-Trailer Trucks	1	-	0.31%	0	1	-	9.89%	1	1	-	15.46%	1	1	-	22.87%	1
14	Unused	0	-	0.00%	0	0	-	0.00%	0	0	-	0.00%	0	0	-	0.00%	0
15	Unclassified Vehicle	1	-	0.15%	0	1	-	4.17%	1	1	-	4.48%	1	1	-	8.35%	1
All	All Vehicles	5	1.92%	87.17%	339	8	1.60%	93.20%	560	13	3.65%	97.22%	416	15	1.62%	97.53%	1313

Comparison with Currently Deployed Technologies

For comparison with other currently deployed technologies, the results from a previous

evaluation of accuracy of detectors in the Atlanta Region is provided in table 14 (Guin et

al. 2013). The table contains data for inductive loop detectors used in Automatic Traffic

Recorder (ATR) stations, VDS detectors and RTMS detectors. The lane-by-lane

comparisons showed a wide range of errors varying across deployment locations, with

mean absolute percentage errors (MAPE) in the range of 1-2 percent in ATR, 1-10 percent

in VDS and 32-64 percent in RTMS for the set of detection stations evaluated. It is

important to note that the evaluation presented in table 14 did not consider inclement

weather conditions. In addition, the results in table 14 are presented for 1-hour aggregates.

The average error magnitudes typically tend to be lower at higher levels of aggregation.

The magnitude of errors reported for 1-minute aggregates for TIRTL, therefore, compare

favorably with respect to the 1-hour average errors reported for the other detection

technologies presented in table 14.

Table 14. VDS accuracy evaluation results summary from previous GDOT study (Guin et al. 2013).

No	Type	No of Lanes	Setup Style/Location	All Lanes						Lane By Lane		
				MPE	Lower 95%	Upper 95%	MAPE	Lower 95%	Upper 95%	Lane Number	MPE	MAPE
1	ATR	4	I-285 Northbound near Orchard Road	0.77%	0.38%	1.16%	1.45%	1.18%	1.72%	1	0.44%	1.58%
										2	0.17%	0.94%
										3	1.84%	1.93%
2	ATR	4	I-285 Southbound near Orchard Road	-0.25%	-0.76%	0.26%	1.87%	1.52%	2.22%	1	0.09%	1.85%
										2	-0.84%	1.82%
										3	-1.10%	1.96%
										4	0.71%	1.87%
3	VDS	4	Pole Mounted/ I-285 Northbound near Orchard Road	0.46%	-1.22%	2.45%	5.49%	4.58%	6.41%	1	-4.34%	4.79%
										2	0.44%	5.56%
										3	5.61%	6.17%
4	VDS	4	Pole Mounted/ I-285 Southbound near Orchard Road	0.56%	-0.79%	1.90%	4.36%	3.35%	5.37%	1	-5.86%	6.38%
										2	3.14%	4.43%
										3	4.83%	4.98%
										4	0.61%	1.81%
5	VDS	4	Gantry Mounted Side/ I-285 Northbound near Cumberland Parkway	0.43%	-0.95%	1.80%	4.81%	3.85%	5.77%	1	-3.05%	5.68%
										2	0.05%	5.50%
										3	2.94%	3.60%
										4	1.65%	4.44%
6	VDS	4	Gantry Mounted Median/ I-285 Southbound near Cumberland Parkway	-0.68%	-2.60%	1.23%	6.09%	4.63%	7.54%	1	-4.36%	5.27%
										2	-3.28%	5.00%
										3	3.10%	9.19%
										4	2.26%	5.46%
7	VDS	4	36 feet offset Pole Mounted/ I-285 Northbound near Cascade Road	-1.11%	-3.63%	1.41%	7.24%	5.31%	9.17%	1	-4.98%	8.82%
										2	2.87%	7.69%
										3	-4.19%	6.64%
										4	1.16%	5.53%
8	VDS	4	36 feet offset Pole Mounted/ I-285 Southbound near Cascade Road	-1.34%	-2.93%	0.26%	5.86%	4.92%	6.79%	1	-2.96%	10.18%
										2	-0.59%	4.76%
										3	0.98%	4.67%
										4	-3.58%	4.62%
9	VDS	4	Gantry Mounted Median/ I-285 Northbound near US78	-3.84%	-5.69%	-1.99%	6.97%	5.74%	8.19%	1	-6.84%	7.26%
										2	-4.43%	7.14%
										3	-0.56%	6.52%
10	VDS	4	Gantry Mounted Median/ I-285 Southbound near US78	-4.13%	-5.78%	-2.48%	6.67%	5.43%	7.90%	1	-7.20%	7.59%
										2	-6.67%	7.41%
										3	-2.98%	4.25%
										4	1.08%	7.04%
11	RTMS	3	Pole Mounted/ US-78 Eastbound near Idlewood Road	-2.19%	-3.36%	-1.01%	4.40%	3.25%	5.07%	1	-4.19%	4.19%
										2	0.05%	3.08%
										3	-2.94%	6.57%
12	RTMS	3	Pole Mounted/ US-78 Westbound near Idlewood Road	-3.22%	-4.53%	-1.92%	5.23%	4.30%	6.16%	1	-3.50%	4.09%
										2	-5.92%	7.94%
										3	-0.42%	3.69%
13	VDS	7	Pole Mounted/ I-75/I-85 Near 14th Street	-17.32%	-22.96%	-11.67%	38.92%	35.62%	42.23%	1	13.16%	64.33%
										2	-13.99%	40.63%
										3	-17.92%	34.19%
										4	-23.12%	32.56%
										5	-26.58%	34.26%
										6	-24.95%	32.52%
										7	-24.76%	37.60%

Speed Data Analysis

For evaluation of speed measurement accuracy, data extracted manually from videos for 40 vehicles, as described in the subsection on Speed Estimate Verification, were used for comparison with TIRTL data. Figure 11 shows a plot of the time series of the speeds from TIRTL and the manual extraction. The figure clearly shows that for most datapoints the TIRTL speeds and the manually extracted speeds have a constant difference. A one-way paired t-test was conducted to test if the estimates varied more than 5 mph (about 10 percent of the average speed). The t-test results indicated that there is not enough evidence to reject the null hypothesis and, thus, estimates from the two datasets do not differ by more than 5 mph.
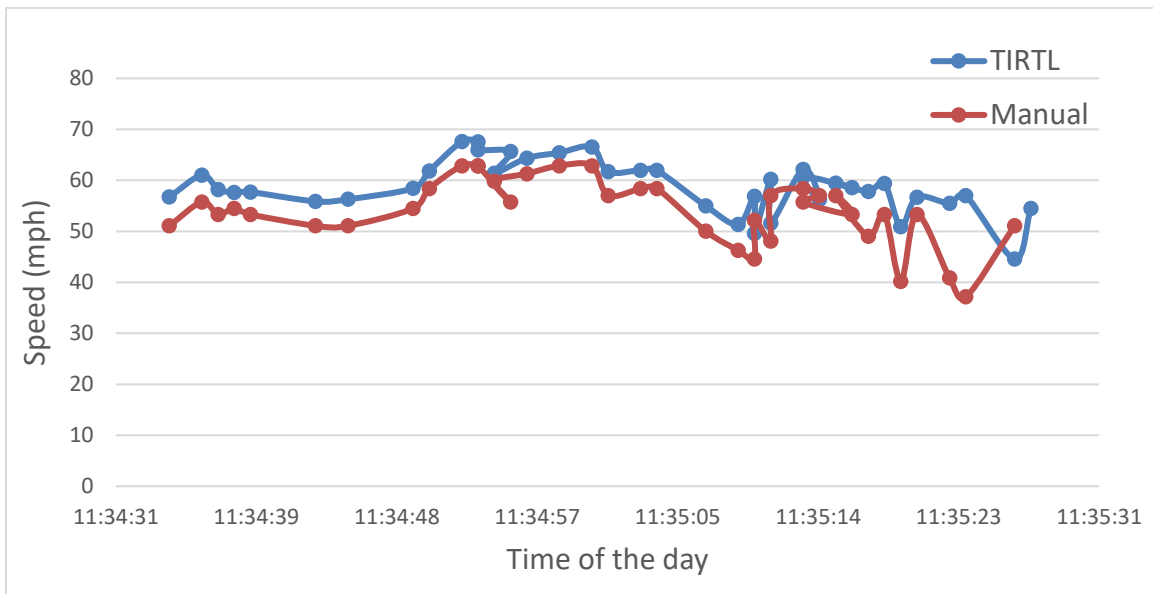
Figure 11. Graph. Comparison of speed estimates (in mph) from TIRTL and manually extracted data.

CHAPTER 3. INCIDENT DATA FUSION FOR WAZE AND NAVIGATOR COMPARISON

BACKGROUND

Crowdsourced methods of incident detection using smartphone-based apps have recently started making inroads into the incident management process. Waze®, owned by Google, has teamed up with state, county, and city departments of transportation in an effort to help with the integration. While the benefits of such a partnership are obvious, crowdsourced data have some limitations and challenges. Amin-Naseri et al. (2018) discussed the problems associated with the lag and spatial inaccuracy associated with trying to report an incident in a mobile application while the vehicle is moving at a high speed, leading to an offset between the location of an incident and the location where it is reported. Vallejos et al. (2020) discussed the issues with redundant reports from Waze that create an overwhelming amount of unusable data. Xavier et al. (2016) pointed out the challenges of merging and consolidating the crowdsourced data with existing incident data sources from DOTs.

To investigate the potential of crowdsourced smartphone app–based incident detection and notification, such as with Waze, in reducing the time to detection, a comparative analysis is performed using Waze and GDOT TMC's NaviGAtor incident logs. An incident data fusion methodology is developed to facilitate the comparison. For this analysis, NaviGAtor event logs and Waze data in the Atlanta metro area were obtained for the period February to April 2018. This analysis focuses on incidents tagged as "accidents," which are likely to have a higher impact on traffic flow.

DATA ACQUISITION

Waze

Through the Connected Citizens Program, Waze makes its data available to traffic management agencies. The incident data from Waze are obtained by polling at a 1-minute frequency. Each record has an associated timestamp, geolocation, and primary and secondary road names, along with other incident description characteristics.

NaviGAtor

The NaviGAtor dataset, generated by the Georgia TMC, consists of information about incidents reported by various sources such as Highway Emergency Response Operator (HERO), police department / 911, mobile operator, motorist call, and Georgia 511 operator. Each incident, identified by a unique incident ID, can have multiple records associated with it, where each record specifies a separate update or response action.

DATA PREPROCESSING

The preprocessing of the Waze data consisted of the following steps:

1. The Waze data were converted from an XML format to CSV format and consolidated into a smaller number of files for ease of use in scripts in the rest of the analysis.

2. The incidents were filtered to exclude everything except the "accident" type.

3. Timestamps were converted from the GMT format to an EDT format for easy comparison with NaviGAtor data.

4. Multiple records for an incident were consolidated into a single record with a start and end time.

The preprocessing of the NaviGAtor data consisted of the following steps:

1. The NaviGAtor data were split into daily files corresponding to the Waze files.

2. The end of the incident was determined using the last updated record for the incident or when the keyword "terminated" was detected.

3. The incident was assigned the highest value of severity observed in any of the records corresponding to the incident in the original data.

DATA FUSION

There were four data dimensions that were leveraged for the data fusion process: timestamp, geolocation (i.e., latitude/longitude), road names (i.e., primary and secondary), and direction of travel. These components are discussed in further detail in the following subsections.

Time Match

The first criterion for matching was a time overlap. The distributions of duration of incidents in the two datasets are shown in figure 12 and figure 13.
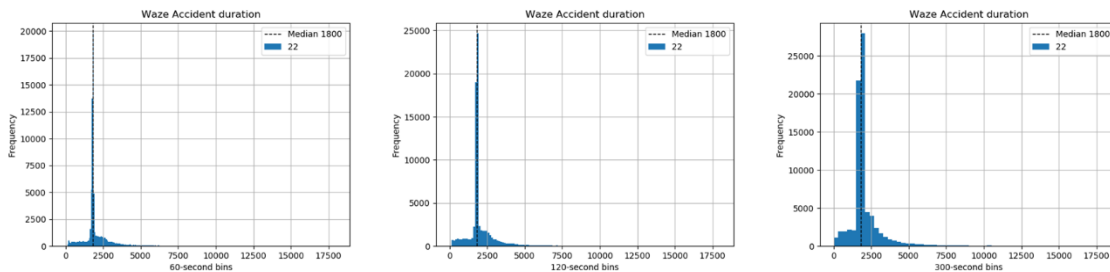
Figure 12. Histograms. Waze accident duration (Feb to Apr 2018).
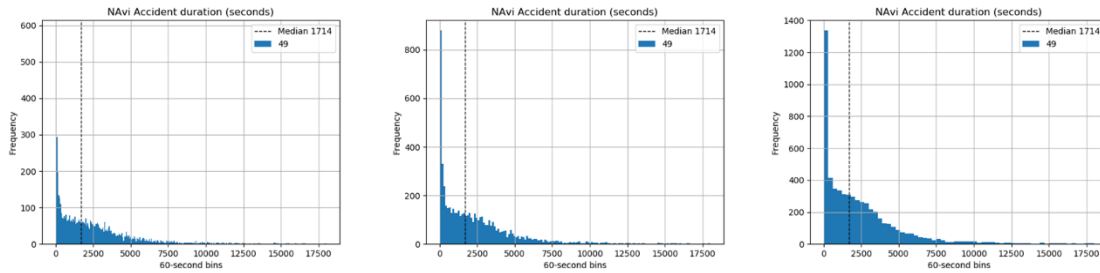
Figure 13. Histograms. NaviGAtor accident duration (Feb to Apr 2018).

The following strategies were explored for identifying candidate incidents in the Waze dataset that could potentially be the same incident as one in the NaviGAtor dataset:

- Matching the start timestamp of a Waze incident to the interval over which a NaviGAtor incident occurs.

- Matching the start timestamp of a Waze incident to the interval over which a NaviGAtor incident occurs plus an additional buffer interval at the beginning of the NaviGAtor incident. Buffer intervals of 30, 60, 90, and 120 minutes were tested. The assumption was that the detection of an incident in NaviGAtor could have some delay due to the limitation of resources.

- Matching the overlaps of the incident duration periods. With the minimum and maximum timestamps in the Waze data, for a given incident, taken as the surrogate for the beginning and ending of a Waze incident, a minimum overlap of time periods was used as the matching criterion. Various overlap percentages of 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50 percent were tested. A 25 percent overlap was selected as the matching criteria. As seen in figure 14, 25 percent provides a reasonable cut-point where the matching rate shows limited decrease in the match rate with a continued increase in the required overlap.

37
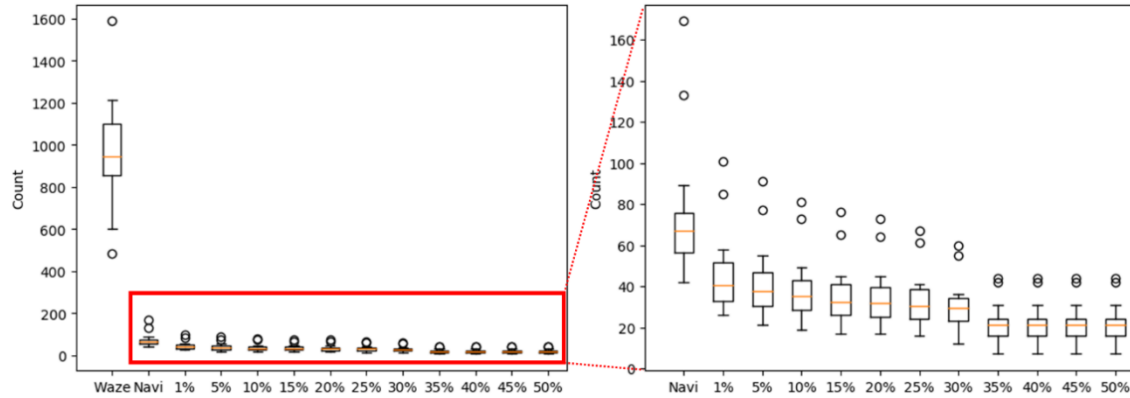
Figure 14. Box plots. Numbers of matched results by overlapping rates for 30 days in April 2018.

Location Match

For the location match, the Euclidean distance between the latitude/longitude coordinates of the Waze incident and the NaviGAtor incident was computed, and thresholds of 0.5, 1, 2, and 3 miles were tested. The gains in matches were not significant when the strictness of the threshold was reduced; therefore, the 0.5-mile threshold was used.

Road Name Match

Since the majority of the incident responses by the GDOT TMC are on freeways, there is usually a numerical portion of the road name (e.g., GA 400, I-75, etc.). The first pass of the filter matches the road name attributes in Waze and NaviGAtor using the numerical portion of the road name. If the numerical portion of the road name does not result in a match, then the text portion is used to detect a match. If a match is not detected on the primary roadway, the search is performed on the cross-street text. Where the text match was the cross street, the location match criterion was reduced to 0.5 mile to increase confidence in the match.

Direction of Travel

The direction of travel, if specified within the road name attributes of both datasets, is used to further refine the matches. Where the direction of travel is not specified in Waze, or differs from that in NaviGAtor, the location match is further constrained to 0.5 mile to increase confidence in the match.

Final Match

The algorithm for the data fusion process is shown in figure 15. As seen, a match requires satisfying the time, roadway name or number, and distance criteria.
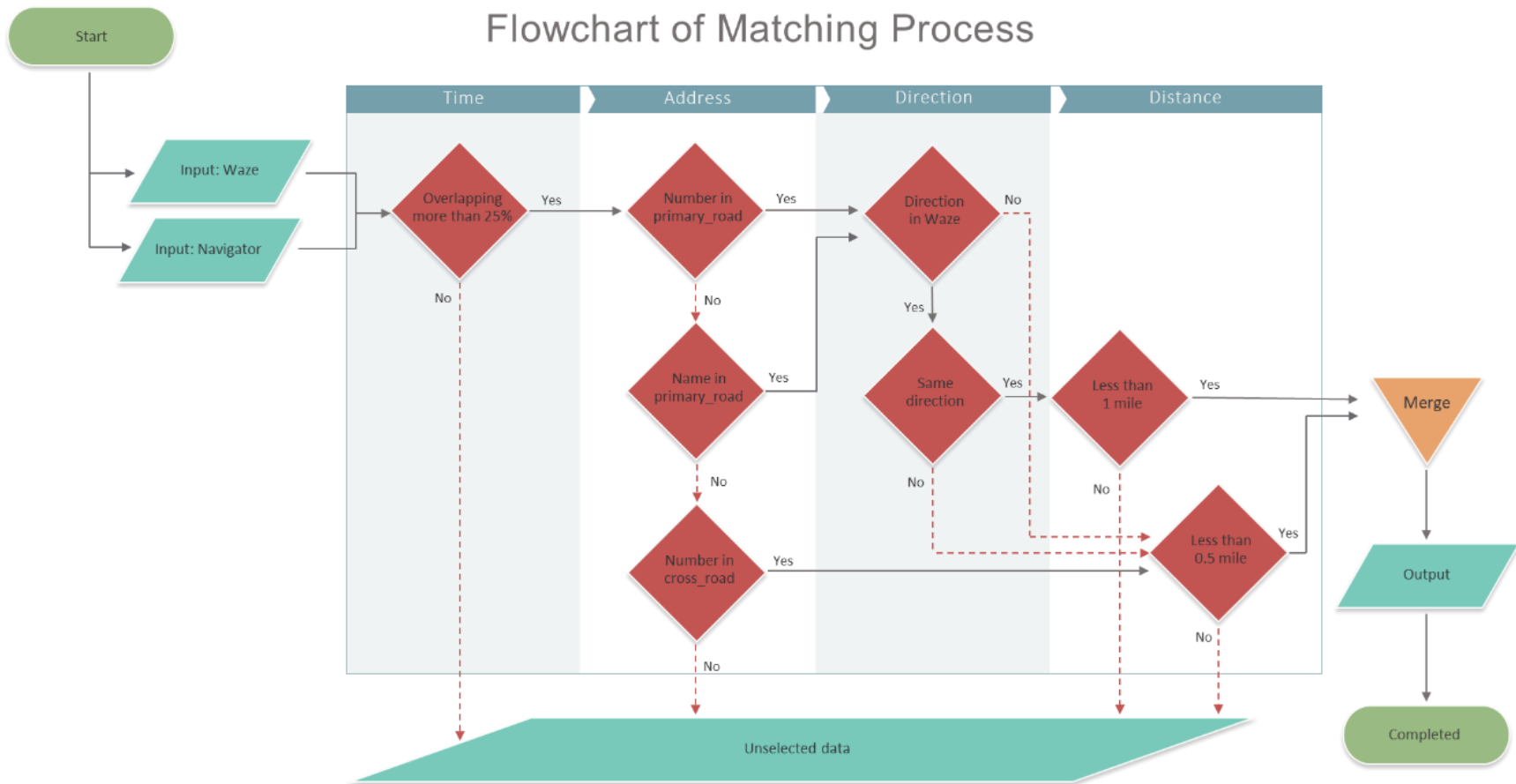
Figure 15. Flowchart. Data matching process.

RESULTS AND DISCUSSION

Detection Rate and Time-to-Detection Comparison

This analysis is based on incidents that are tagged as "accidents." This restriction reduces the probability of spurious matches that can result from the sheer volume of the large number of low-impact incidents, such as shoulder stalls, in the datasets., The results of the comparison between the Waze data and the NaviGAtor data are shown in table 15.

Table 15. Detection rate by the matching methodology.

	NaviGAtor	Waze
Total number of records (original)	5,923	78,735
Number of matched records	2,475 (41.79%)	2,475 (3.14%)
Number of records with Waze detection preceding NaviGAtor detection	—	1,429 (57.74%)
Number of records with NaviGAtor detection preceding Waze detection	1,046 (42.26%)	—

— = not applicable

During February through April 2018, 2,475 accidents were matched between the two datasets. This represents 41.79 percent of the total accidents in the NaviGAtor dataset, and 3.14 percent of the total accidents in the Waze dataset. Among the 2,475 accidents, in 1,429 cases, the Waze detection preceded NaviGAtor detection. In 1,046 cases, the NaviGAtor detection preceded Waze detection. Figure 16 shows the histograms by month from February through April as well as all 3 months together, for the detection time savings provided by Waze as compared to NaviGAtor. A negative number of seconds indicates that Waze detected the incident before NaviGAtor, while a positive number indicates that NaviGAtor detected it earlier. The patterns are fairly consistent over these months.

41
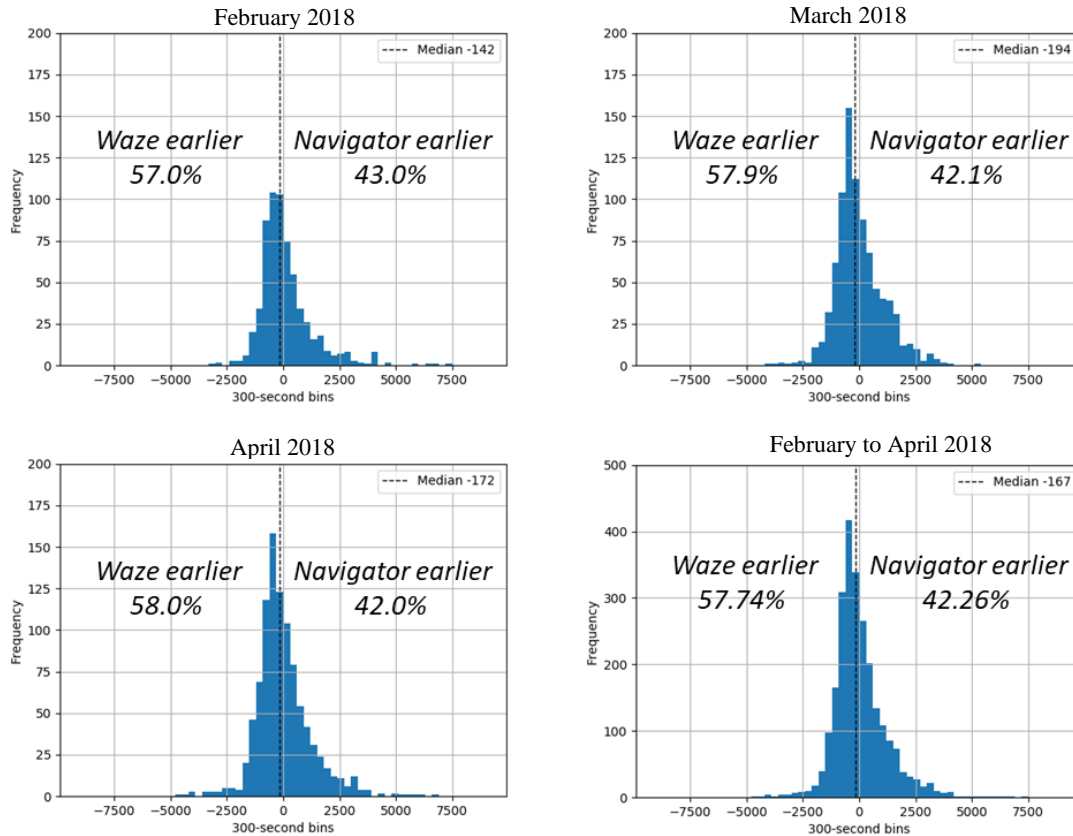
Figure 16. Histograms. Detection time savings for matched incidents (February through April 2018).

To investigate the presence of any relationship of incident characteristics with the likelihood of earlier detection of an incident by Waze, the comparison data were plotted along multiple incident characteristic dimensions. Figure 17 shows the histograms of the time savings in detection split by the occurrence time of incidents, weekday vs. weekend. Figure 18 shows the histograms of the time savings in detection separated by incident severity as indicated in Waze. Figure 19 shows the split by incident severity as indicated in NaviGAtor. Figure 20 shows the split by the time of occurrence of incidents within the day, AM peak, PM peak, and off-peak. Figure 21 shows the split by peak vs. off-peak. Figure 22 and figure 23 show the split by incident durations as measured in Waze and

NaviGAtor, respectively. Figure 24 and figure 25 use an alternative categorization based on

NaviGAtor incident duration. However, across all these plots, no conclusive evidence of

meaningful correlation between the earlier detection by either method and the incident

characteristics was observed.
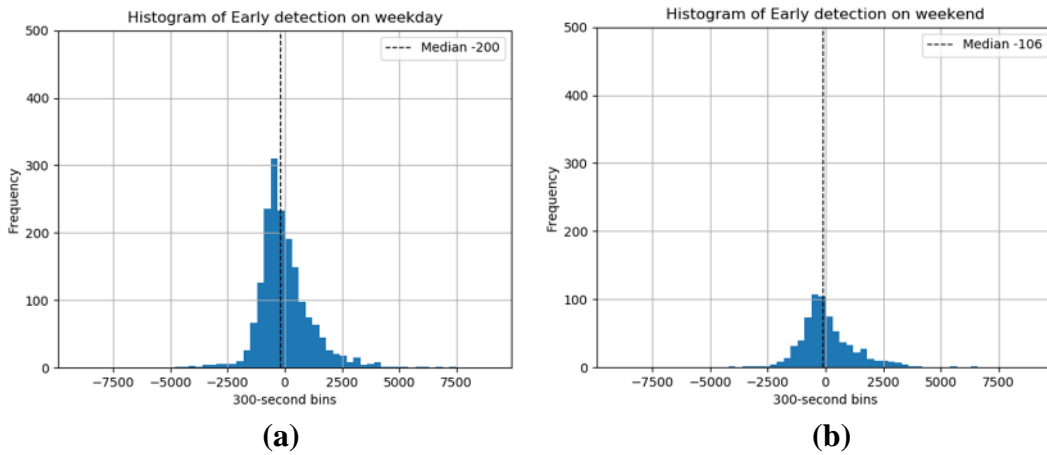
Figure 17. Histograms. Detection time savings of matched accidents separated by: (a) weekdays, (b) weekends.
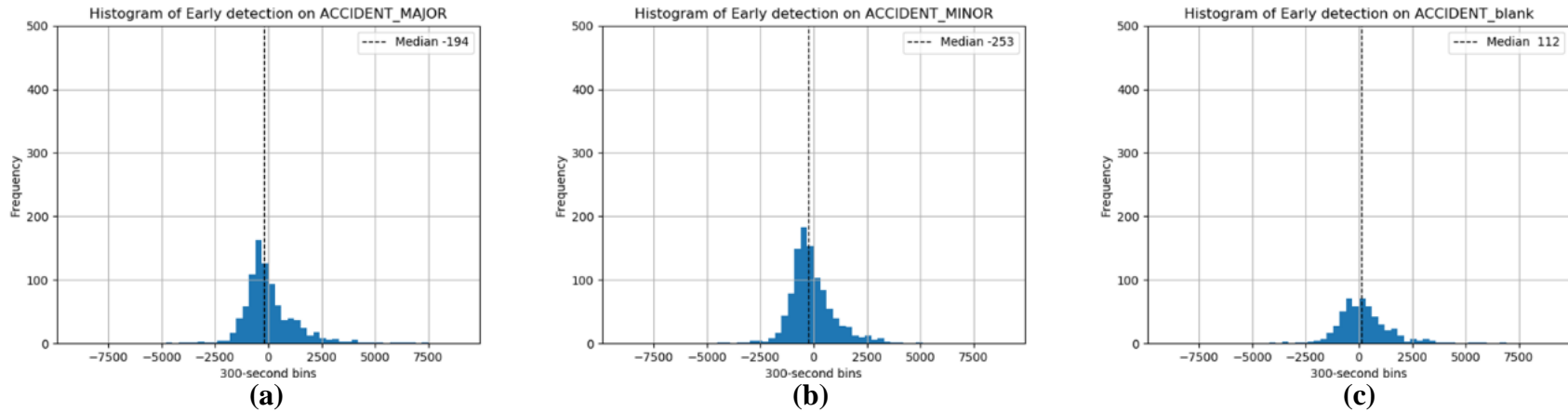
Figure 18. Histograms. Detection time savings of matched accidents separated by Waze accident severities: (a) major, (b) minor, (c) unspecified.
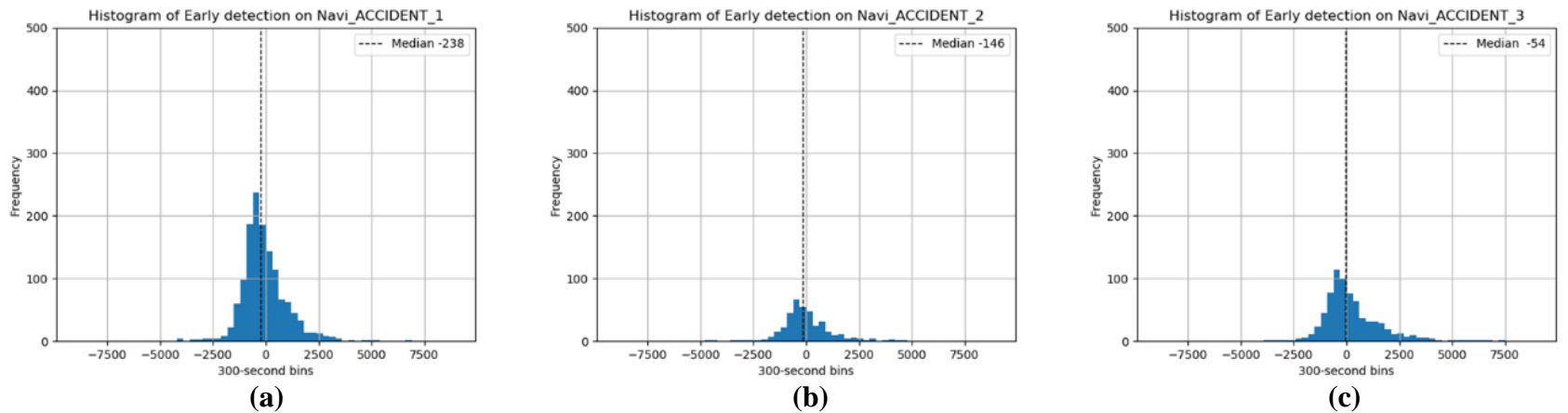
Figure 19. Histograms. Detection time savings of matched accidents separated by NaviGAtor accident severities: (a) severity 1, (b) severity 2, (c) severity 3.

44
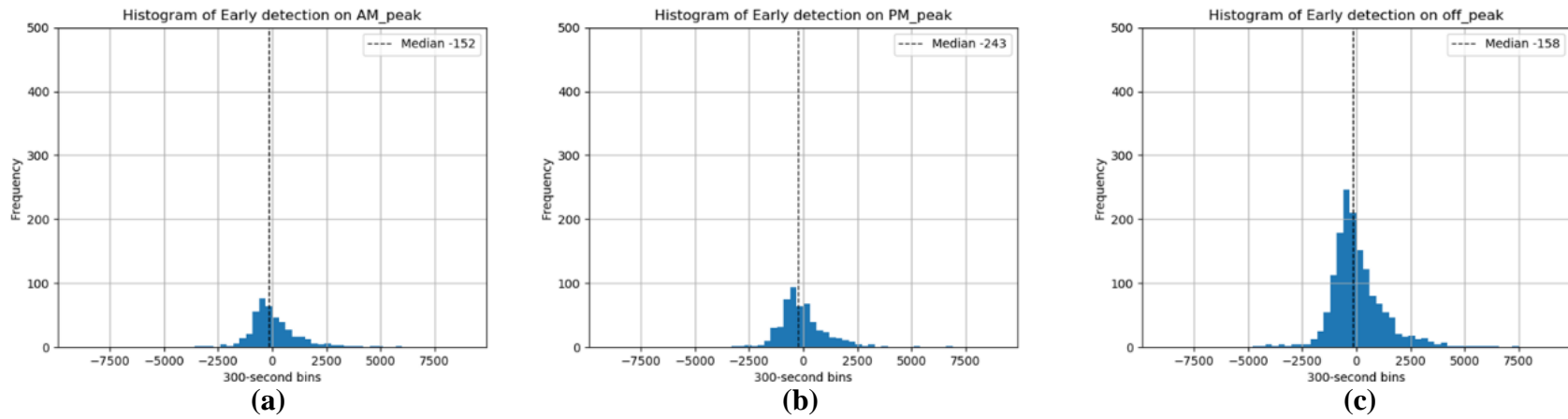
Figure 20. Histograms. Detection time savings of matched accidents separated by time of day: (a) AM peak, (b) PM peak, (c) off-peak.
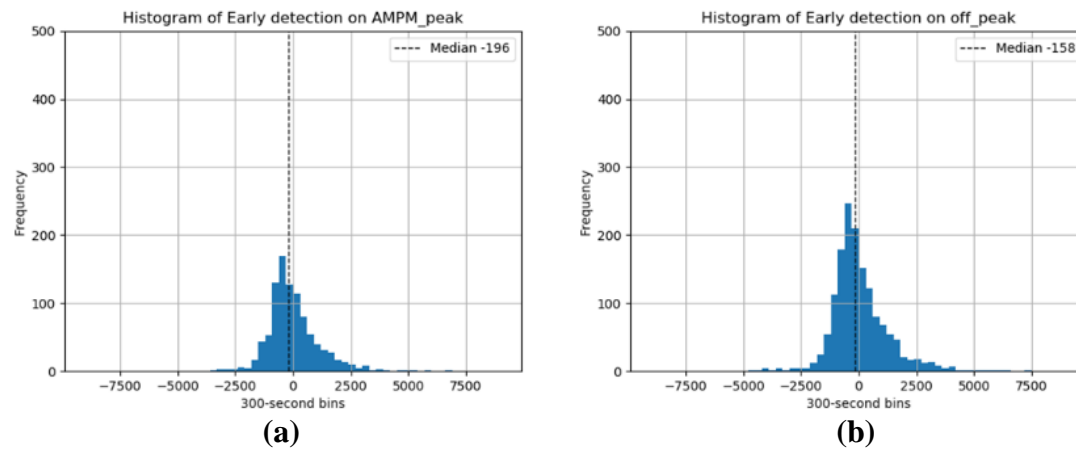
Figure 21. Histograms. Detection time savings of matched accidents separated by traffic flow conditions: (a) peak, (b) off-peak.
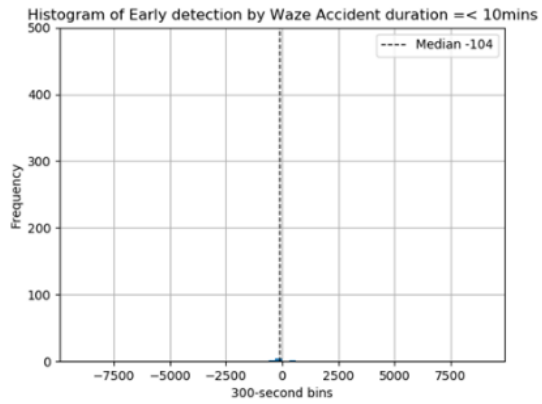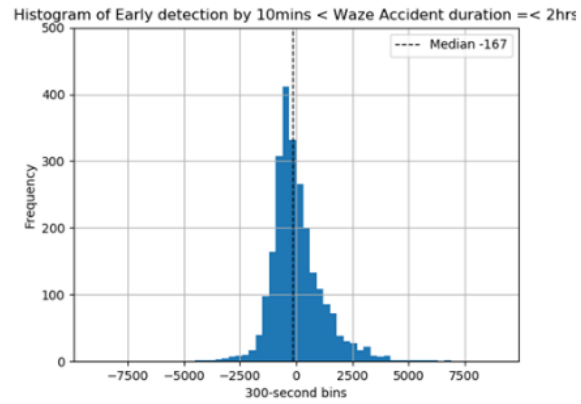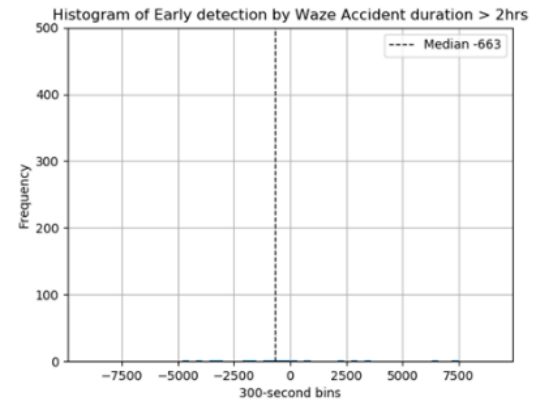
Figure 22. Histograms. Detection time savings of matched accidents separated by Waze accident durations: (a) less than 10 minutes, (b) more than 10 minutes but less than 2 hours, (c) more than 2 hours.
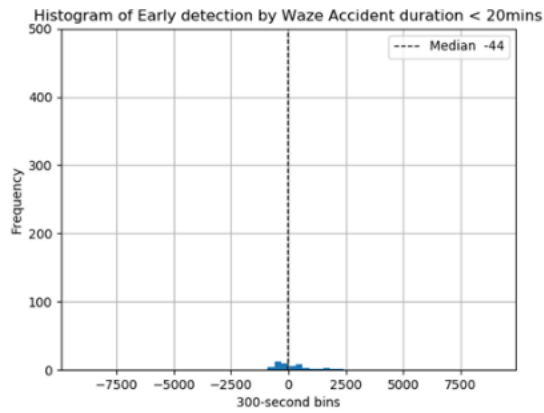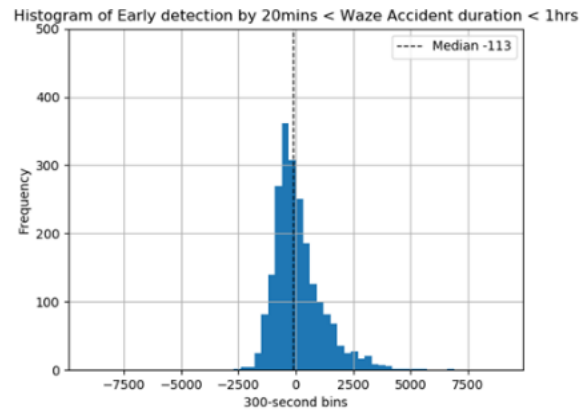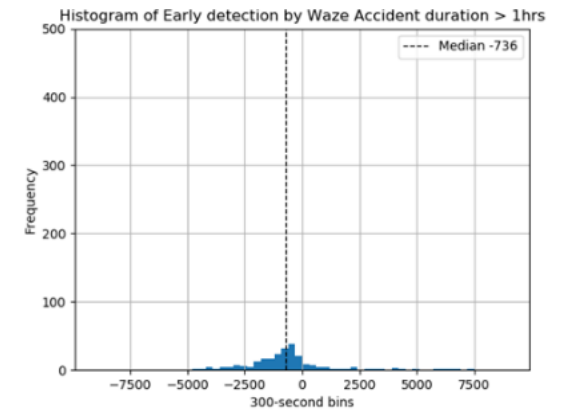
Figure 23. Histograms. Matched accidents separated by NaviGAtor accident durations: (a) less than 20 minutes, (b) more than 20 minutes but less than 1 hour, (c) more than 1 hour.
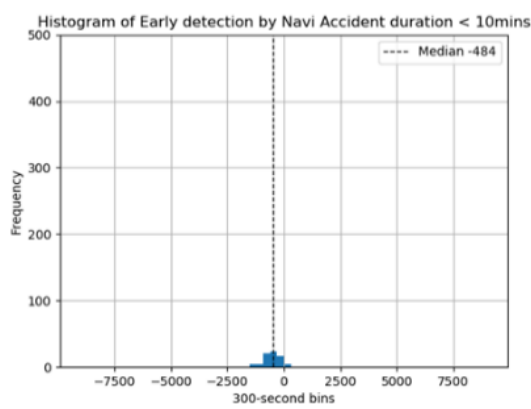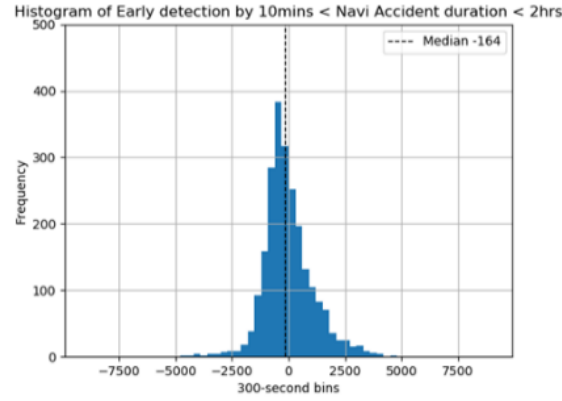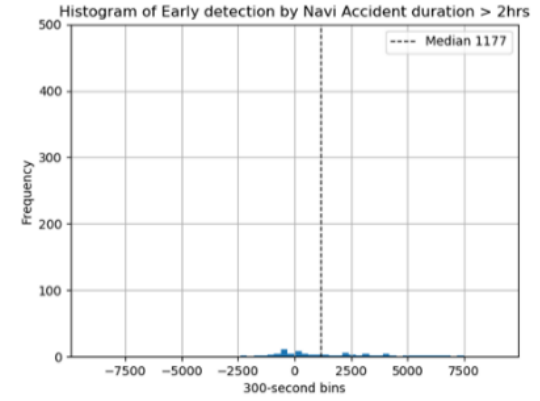
Figure 24. Histograms. Matched accidents separated by NaviGAtor accident durations: (a) less than 10 minutes, (b) more than 10 minutes but less than 2 hours, (c) more than 2 hours.
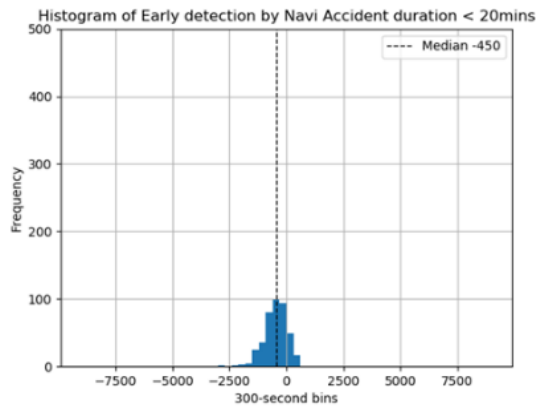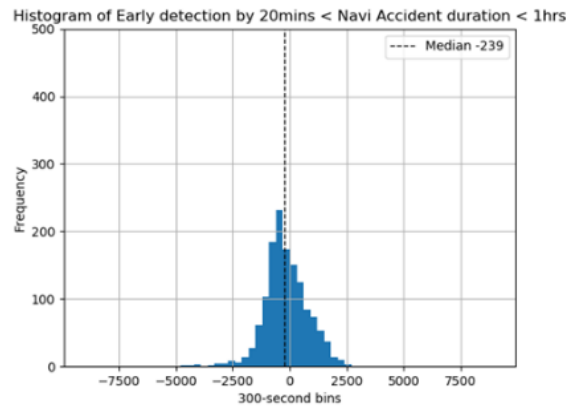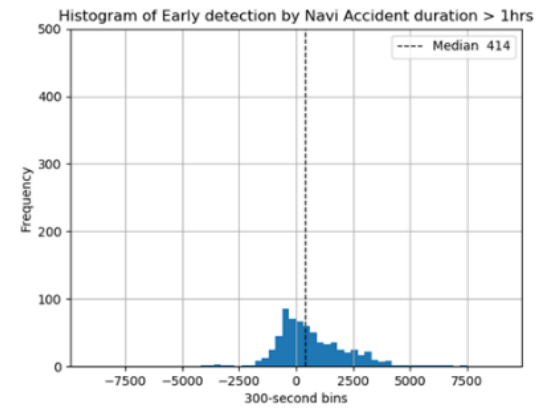
Figure 25. Histograms. Matched accidents separated by NaviGAtor accident durations: (a) less than 20 minutes, (b) more than 20 minutes but less than 1 hour, (c) more than 1 hour.

WAZE AND NAVIGATOR DATA ANALYSIS IN METRO ATLANTA AND MACON

To investigate the presence of any location-specific differences in the matching rates of

Waze and NaviGAtor incidents, a comparative analysis was performed between the Atlanta

Metro area and Macon, Georgia. The incidents were geo-filtered using a bounding box

defined as [(34.55, −85.4), (33.15, −83.5)] for Metro Atlanta and [(32.97, −83.83),

(32.67, −83.33)] for Macon. There were 54,237 recorded Waze accidents and 5,005

recorded NaviGAtor accidents in Metro Atlanta and 921 recorded Waze accidents and 127

recorded NaviGAtor accidents in Macon during the period February through April 2018,

within the defined geo-fences. Figure 26 shows the original points plotted on maps in the

top row and the matched points plotted on the maps in the bottom row. The results of the

matching process are presented in table 16. The number of matched incidents separated by

incident duration category are presented in table 17.
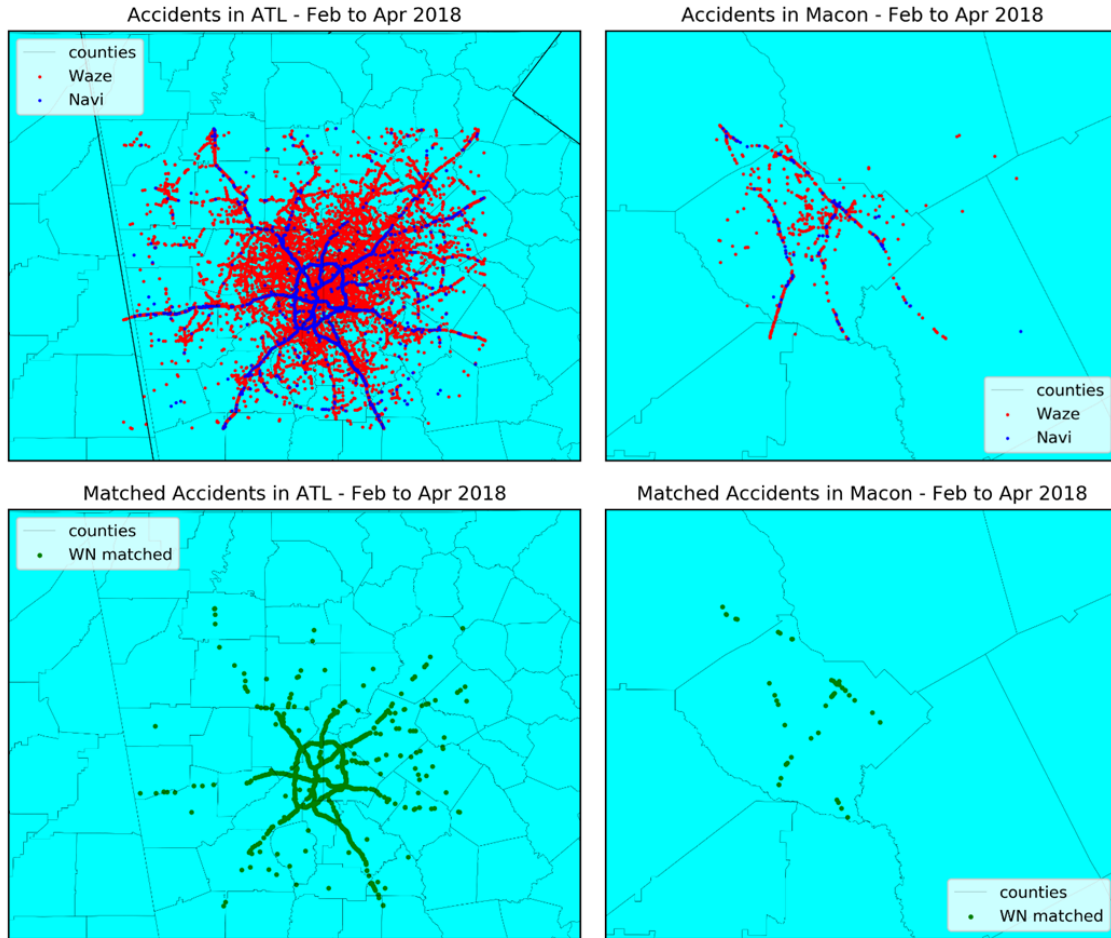
Figure 26. Plots. Waze and NaviGAtor accident plots and matched accident plots in Metro Atlanta and Macon areas.

Table 16. Matched results by Metro Atlanta and Macon.

	NaviGAtor	**Waze**
Total number of records (original)	5,923	78,735
Number of matched records	2,475 (41.79%)	2,475 (3.14%)
Total number of records (Metro Atlanta)	5,005	54,237
Number of matched records (Metro Atlanta)	2,286 (45.67%)	2,286 (4.21%)
Total number of records (Macon)	127	921

Table 17. Matched numbers of accidents by NaviGAtor accident duration (by second).

	$t \le 300$	$300 < t \le 900$	$900 < t \le 1800$	$t > 1800$	Total
Total	1,220	667	858	2,260	5,005
Matched	3	222	523	1,517	2,286
Matching rate (%)	0.25	33.28	60.96	67.12	45.67

Dataset Reduction to HERO Coverage Area

Since the detection of incidents in NaviGAtor is restricted largely to the freeways, further filtering was employed to screen out the nonfreeway detections in Waze, so that the match percentages for Waze incidents are easier to interpret. The results of the filtering are shown in figure 27. To further refine this dataset, the incident datasets are then filtered with respect to the bounds of the HERO route coverage areas. The resulting dataset is plotted in figure 28.
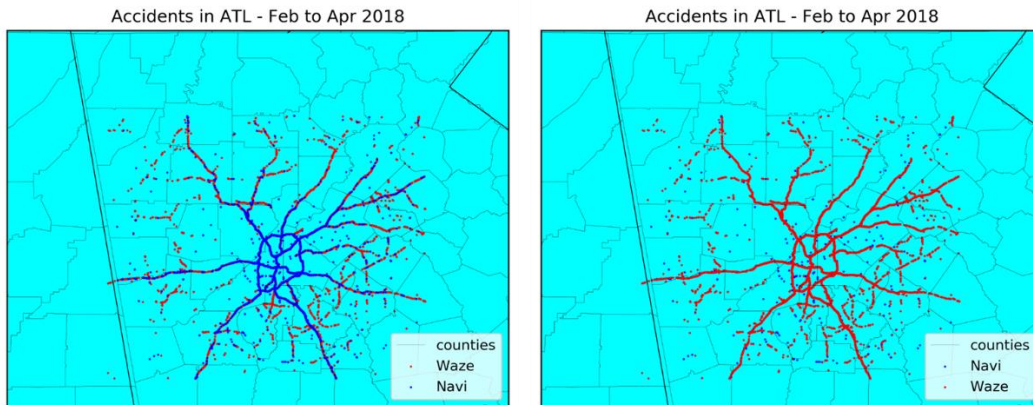
Figure 27. Plots. Waze and NaviGAtor accidents on Interstates and State Routes only.
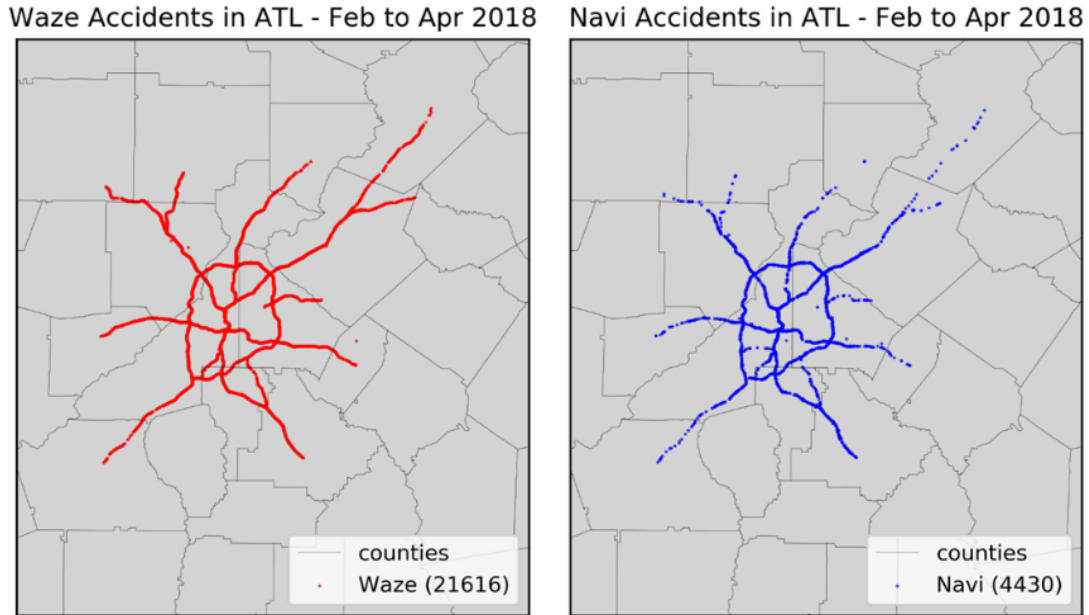
Figure 28. Plots. Waze and NaviGAtor accidents on selected Interstates only.

By applying the HERO route extent filter, the number of Waze accidents is reduced to 21,616 from 54,237 and the number of NaviGAtor accidents is reduced to 4,430 from 5,923.

ANALYSIS USING WAZE REPORTING RELIABILITY ATTRIBUTES

To investigate if there is evidence of impact of incident attributes recorded in Waze incidents, such as *road type, report rating, confidence*, and *reliability* on the matching process, bar charts of incident counts for each of these attributes were plotted and compared for "all" Waze incidents (in the HERO route–filtered dataset) versus the "matched" Waze incidents. From figure 29, figure 30, figure 31, and figure 32, the following observations can be made:

- Incidents with lower *report rating* of 0 or 1 have a slightly lower match rate.
- Incidents with a higher *confidence* number have a higher match rate.

51

- Incidents with a *reliability* of 10 have a higher match rate than the average.

- Relationship of the *road type* attribute with match rate was nonconclusive.
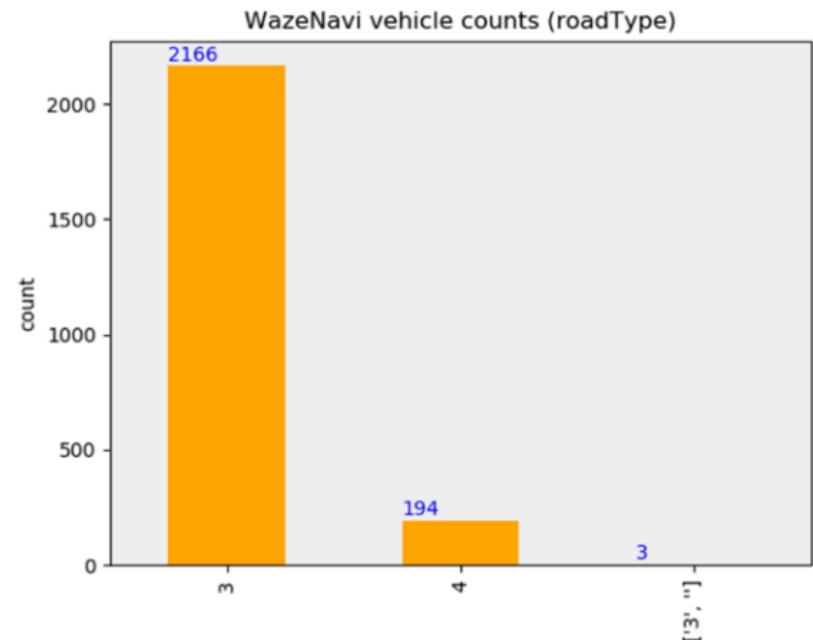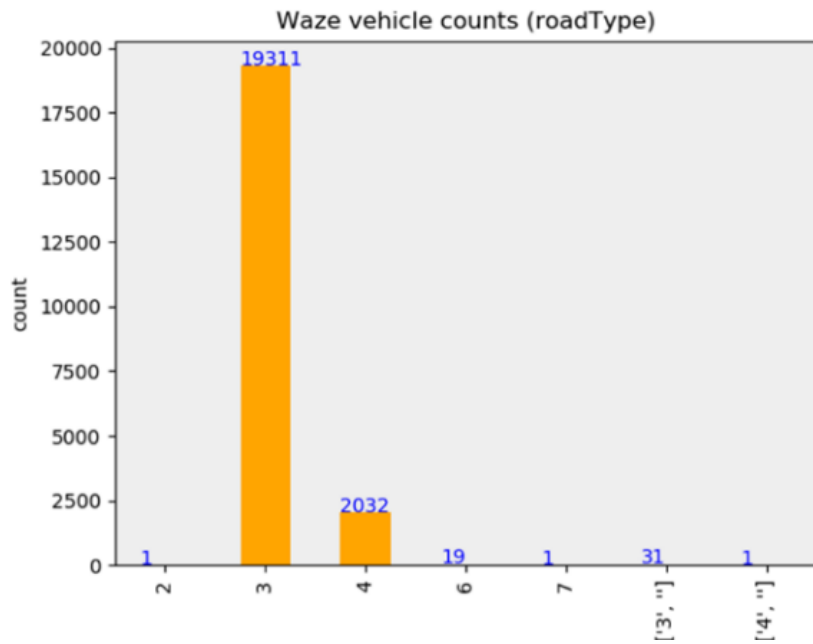
Figure 29. Graphs. Accident counts by *road type* (Waze versus NaviGAtor Matched Waze).
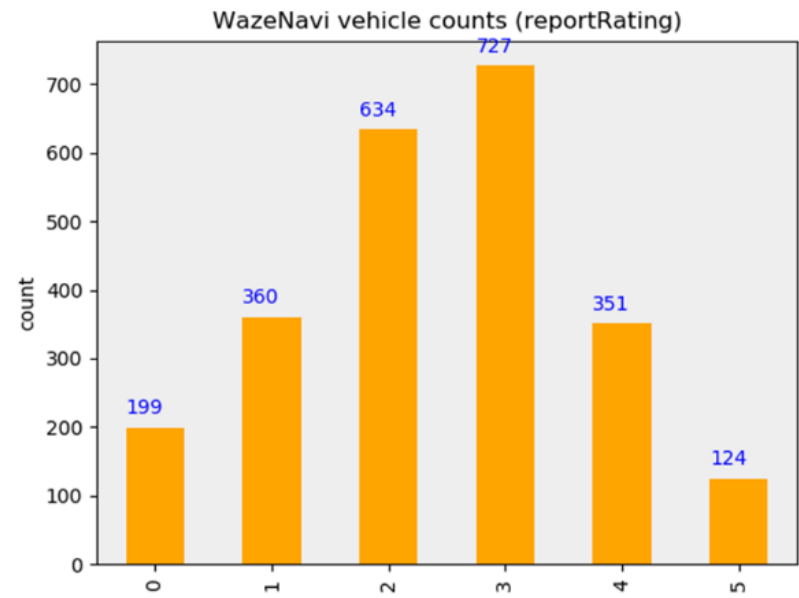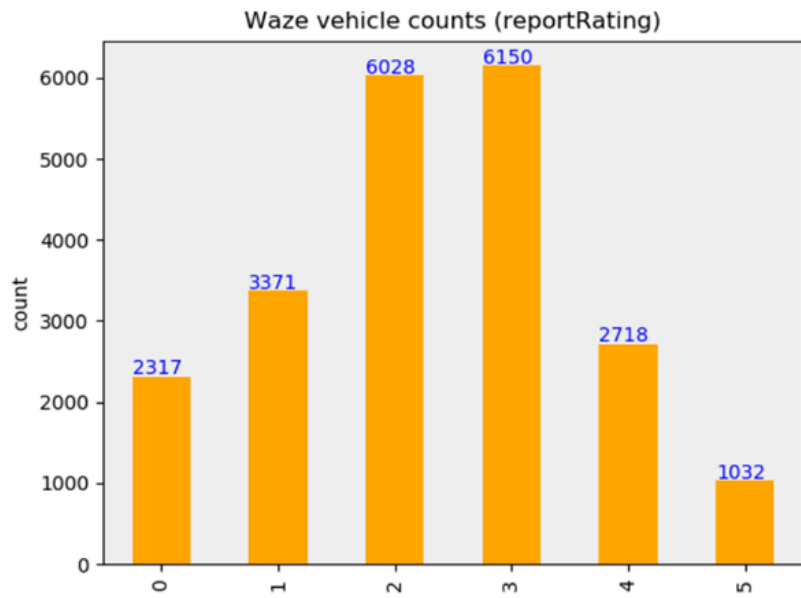
Figure 30. Graphs. Accident counts by *report rating* (Waze versus NaviGAtor Matched Waze).
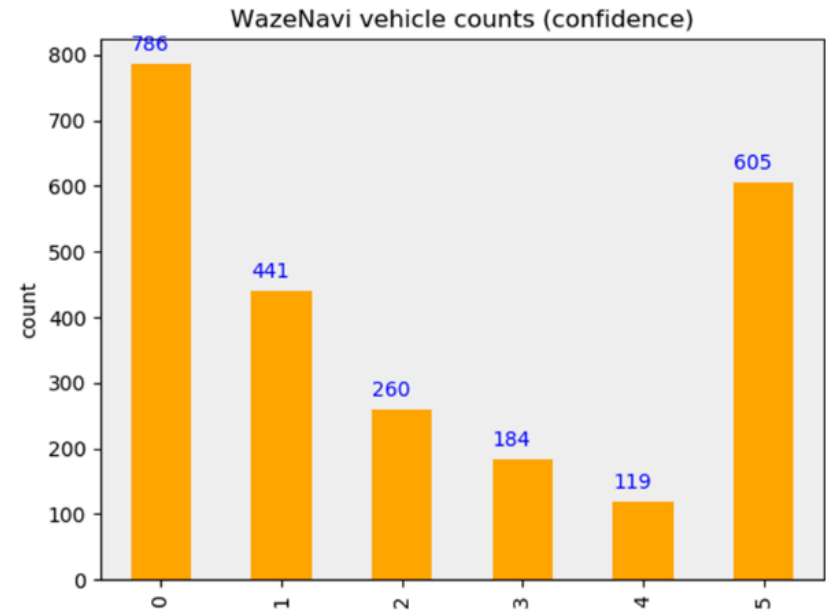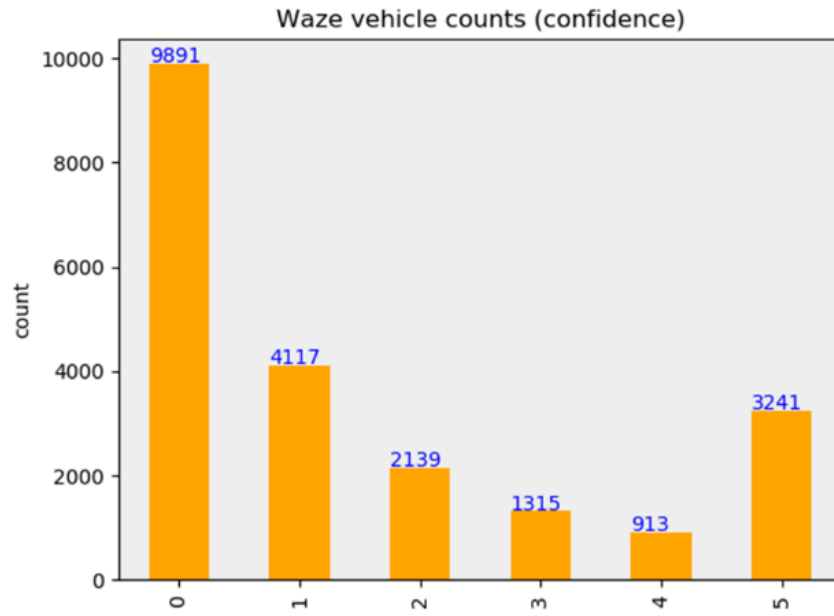
Figure 31. Graphs. Accident counts by *confidence* (Waze versus NaviGAtor Matched Waze).
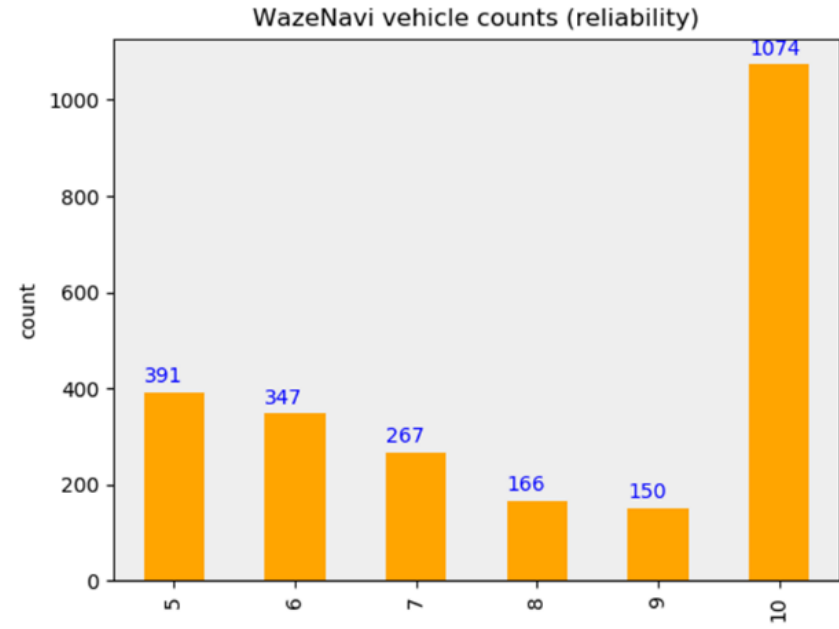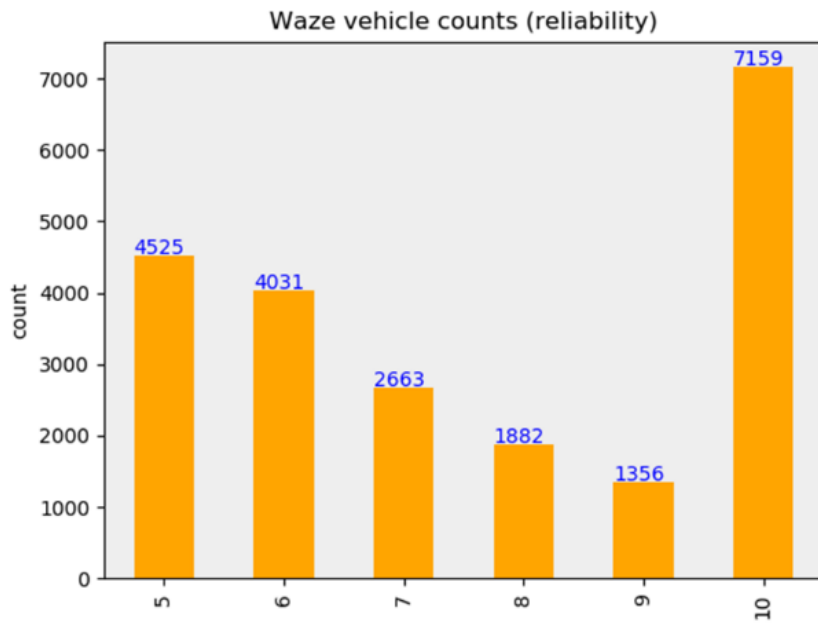
Figure 32. Graphs. Accident counts by *reliability* (Waze versus NaviGAtor Matched Waze).

CHAPTER 4.   TRAFFIC VISION ON I-475

This section presents the results of the investigation of the feasibility of use and the potential benefits of a video-based AID technology relative to the existing detection technologies, such as incident reports via Georgia 511, closed-caption television (CCTV) monitoring by TMC operators, and incident reports by mobile HEROs.

In order to evaluate the AID technology, detection rate (DR), noncritical alarm rate (NCAR, a combination of false, unverifiable, and noncritical alarms), and time to detection are used as performance measures. This chapter presents the development of the AID data analysis methodology to compare the performance of the video detection technology with the existing detection technologies by utilizing the Georgia NaviGAtor incident logs on I-475. The results of the investigation of how effective and efficient the AID technology is in detecting incidents is also presented in this chapter. The last section of the chapter is dedicated to the development of a methodology for filtering out incident alarms with low-traffic impacts to aid in focusing on the high-impact incidents, thereby improving the efficiency of AID technologies in incident management.

INTRODUCTION

AID technologies seek to automate the detection of incidents and unusual conditions using real-time traffic data, such as those from vehicle detection equipment, infrared detectors, and inductive loop detectors. TMCs have employed AID technologies with varying degrees of success. With a desire to detect incidents as quickly as possible, such systems may be

perceived as a necessity where the number of detectors exceeds the reasonable ability to manually monitor devices.

However, surveys on the use of incident detection technologies have indicated a lukewarm response by the industry to AID, primarily due to high false alarm rates (Williams and Guin 2007) and the inability of AID systems to distinguish between noncritical incidents, such as stopped or slowed traffic, and highly critical incidents, such as crashes. This survey is from 2007 and there has been significant technological advances in recent years. Also, the advent of mobile phones and crowdsourced data collection has significantly reduced TTD from means other than transportation agencies or public safety assets. However, the need for AID remains. For instance, under low volume conditions, few motorists may be available to make a report. Also, 911 is often not directly linked to traffic management, with incident management largely dependent on Georgia 511 calls, which may not be as frequent. Even in higher volume scenarios, it may take many valuable minutes for a call to be received and processed.

Video-based AID has evolved rapidly over the last several years, with significant improvements in video quality and computing resources, resulting in an improved potential for automating the detection process. Even with ongoing technology improvements, however, it is recognized that AID technologies continue to struggle to separate those vehicles stopped on the road due to recurrent congestion from vehicles stopped due to a crash. The effort reported in this chapter seeks to help address the noncritical alarm challenge with a focus on identifying those incidents with a high likelihood of being crashes or high-impact, enabling more robust AID.

BACKGROUND

Many AID technology development studies have been conducted over the last few decades, seeking to improve detection rates while decreasing false alarms. Improvements have incorporated algorithmic advances, improved sensor technology, and application of advanced analysis, such as machine learning, deep learning, data fusion, etc. Each of these analyses is discussed in the following subsections.

Detection Algorithms

Incident detection may be categorized into four basic detection algorithm types, based on: comparative, statistical, time series, and traffic theory. Comparative algorithms, also known as pattern-based algorithms, compare current traffic parameters with historical or known conditions to identify potential abnormal or incident conditions. Example comparative algorithms include the decision tree algorithm (Payne and Tignor 1978) and the pattern recognition algorithm (Collins et al. 1979). Statistical algorithms utilize statistical methods to determine whether the observed data sufficiently deviates from that expected. Such unexpected changes in traffic characteristics are identified as an incident. Example statistical algorithms include the standard normal deviate algorithm (Dudek et al. 1974) and the Bayesian algorithm (Levin and Krause 1978). Time-series algorithms assume non-incident traffic follows a predictable pattern over time, using sufficient deviations from this time-series pattern to indicate likely incidents. A common example of this approach is the auto-regressive integrated moving average (ARIMA) model (Ahmed and Cook 1979). Traffic theory–based algorithms utilize traffic flow theory to detect traffic behavior under incident conditions. Incident detection is triggered based on a comparison of observed traffic measurements and traffic model–based estimated traffic measurements.

Representative traffic theory–based algorithms include the dynamic model (Willsky et al. 1980) and the catastrophe theory model (Gall and Hall 1989).

Detection Sensor Technologies

Historically, roadway detection has been passive, with in-roadway devices, such as inductive loops, and nonintrusive devices, such as video, microwave radar, ultrasonic, and passive infrared sensors. However, of these, video detection remains the dominant sensor technology utilized in incident detection. Various studies have been conducted to improve the overall performance of video-based AID. For instance, to overcome difficulties related to varying lighting conditions, reflections of the sun, fog, or snow, systems have been developed to combine video and thermal imaging cameras. Vermeulen (2014) demonstrated that thermal cameras experience minimal impact due to sun glare, headlights, shadows, underpasses, wet streets, snow, and many other conditions.

Beyond these traditional sensors are emerging alternatives that take advantage of communication advances and connected vehicles. Smartphone technology, cellular networks, V2X (vehicle-to-everything), internet of things (IoT), global positioning systems (GPS), general packet radio service (GPRS), global system for mobile communications (GSM) modem, crowdsourced social media monitoring, etc. are all contributing to a vast expansion of the available detection technologies and approaches.

Of these, V2X and crowdsourcing are receiving some of the most significant attention. For example, to address data collection delays and inaccuracy, Iqbal and Khan (2018) suggested using multiple traffic flow parameters via V2X communications. In their model, typical traffic characteristics, such as speed and lane changes, are collected, as well as

60

nontraditional parameters, such as acceleration, orientation, and deviation factors. The parameters are utilized for incident detection and determining the confidence in the detection. Other examples of utilizing V2X include Chen, Xu et al. (2016), who utilized vehicular ad hoc networks (VANETs) to collect traffic data and develop an approach for AID based on support vector machine (SVM), and Dogru and Subasi (2018), who simulated data collected from VANETs and utilized machine learning algorithms to detect incidents.

Crowdsourcing information with mobile technologies offers great potential for better engaging the general public in transportation management. For example, Villela et al. (2018) researched a smart, interoperable decision support system (DSS) for emergency and crisis management based on mobile crowdsourcing information. Zuo et al. (2018) used a latent Dirichlet allocation (LDA) model to automatically classify incident-related tweets and incident types using Twitter data, including messages and geolocation information.

Machine Learning Algorithms for Incident Detection

Most incident detection algorithms focus on identifying traffic incident patterns derived from various sensor streams. Machine learning techniques allow the algorithms to "learn" (i.e., improve AID performance) as additional data are collected and analyzed, and tend to be highly applicable for the real-time needs of AID.

For example, Dogru and Subasi (2018) exploited supervised ML algorithms, such as artificial neural networks (ANN), Support Vector Machine (SVM), and random forests (RFs) to develop models to distinguish incident from non-incident traffic data. Chen and Wang (2009) applied the decision tree technique of supervised ML algorithms in a

simulated environment based on traffic characteristics, including volume, speed, time

headway, and occupancy. Liu et al. (2014) presented a random forest algorithm to reduce

the noise of false incident alarms and to overcome training dataset overfitting problems.

Wang et al. (2013) presented a hybrid approach to AID in transportation systems by

combining time series analysis with ML techniques. Zhu et al. (2018) proposed the

convolutional neural network (CNN) model for automatic detection of traffic incidents

using traffic flow data in Central London, UK.

However, as previously mentioned, even with these efforts and technological advances,

deployed AID systems still tend to suffer from high alarm rates, where it is difficult to

distinguish the small subset of critical alarms from noncritical alarms, creating, creating

significant challenges in their field deployment and use. Thus, in this effort, the research

team seeks to layer a cluster-based ML framework on top of an existing AID system with

the intent of reducing low-impact noncritical incident alarm rates.

AID SYSTEM (CASE STUDY)

This study utilizes video-based AID system data for a 15.83-mile section of I-475

(figure 33) during a 3-month period from 6/27/2018 through 9/25/2018. A total of

186 cameras are located along this stretch of roadway with two to three cameras typically

mounted on a single traffic pole, with each camera surveilling either the northbound or

southbound traffic. When the AID system identifies an incident, it generates video clips

and images, as well as records the following information: date; time; incident type as

identified by the AID technology (stopped, congestion, slow, wrong-way, or pedestrian; see

figure 34); camera name (according to GDOT TMC naming convention); roadway

direction (northbound or southbound); and location (shoulder, ramp, or lanes).

<div align="center">(a)   (b)</div>
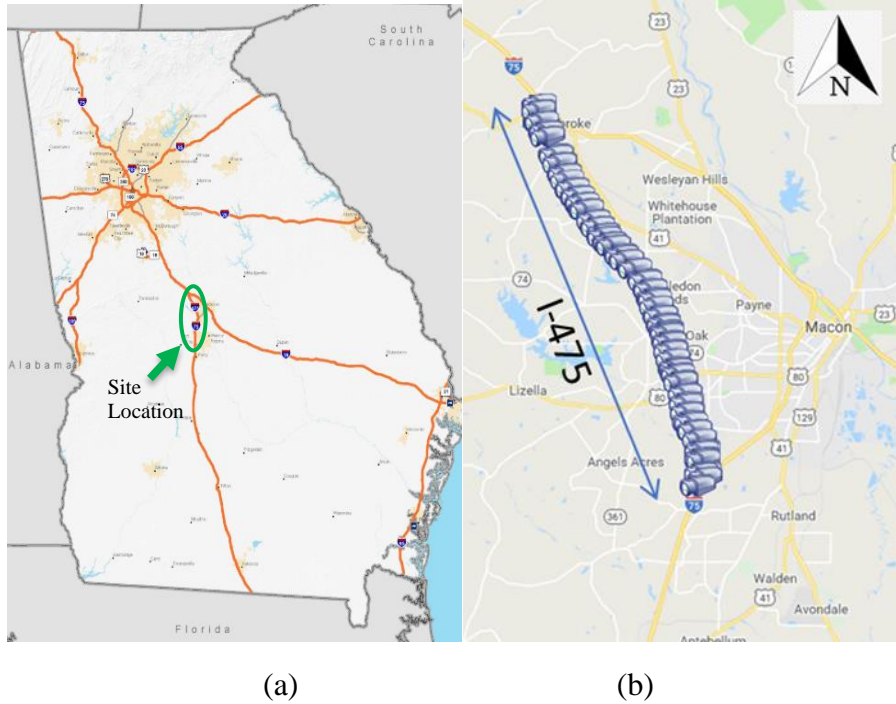
Map credits:
  (a) https://ops.fhwa.dot.gov/freight/infrastructure/ismt/state_maps/states/images/nhfn_map/ga_georgia.jpg,
  (b) TrafficVision software, trafficvision.com, 2019 GDOT application.

Figure 33. Maps. (a) I-475 location, (b) camera locations with AID system.

Figure 34. Photos. Examples of incident types (stopped, congestion, slow, wrong-way, and pedestrian).

The five incident alarm types and thresholds are defined within the AID system settings as follows:

- **Stopped incident:** indicates stopped vehicle/debris in the roadway if a vehicle or object is stationary for [30][1] seconds.

- **Congestion incident:** indicates an incident if the congestion index score is above [30] or if the congestion index score changes from below [20] to above [80] in [30] seconds.

- **Slow incident:** indicates an incident if:

  - Yellow alert: speeds stay below [25] mph for [30] seconds.

  - Red alert: speeds stay below [10] mph for [30] seconds.

---

[1] Threshold values shown in "[ ]" are the default values and may be changed by a system operator. Unless otherwise stated, the analysis presented in this report utilizes the default values.

- **Wrong-way incident:** indicates wrong-way vehicle movement where vehicle movement is detected in a direction opposite that expected in the given location.

- **Pedestrian incident:** indicates pedestrian movement detected in the roadway.

Data Overview

Figure 35 provides the daily incident alarm counts for the study area over the 3-month study period. There are 91 days of the dataset for the study period with a total of 10,125 incident alarms. The average number of daily incidents is 111.26, and the standard deviation is 27.25. The maximum daily incident count is 208 on 8/6/2018, and the minimum daily incident count is 40 on 9/25/2018.
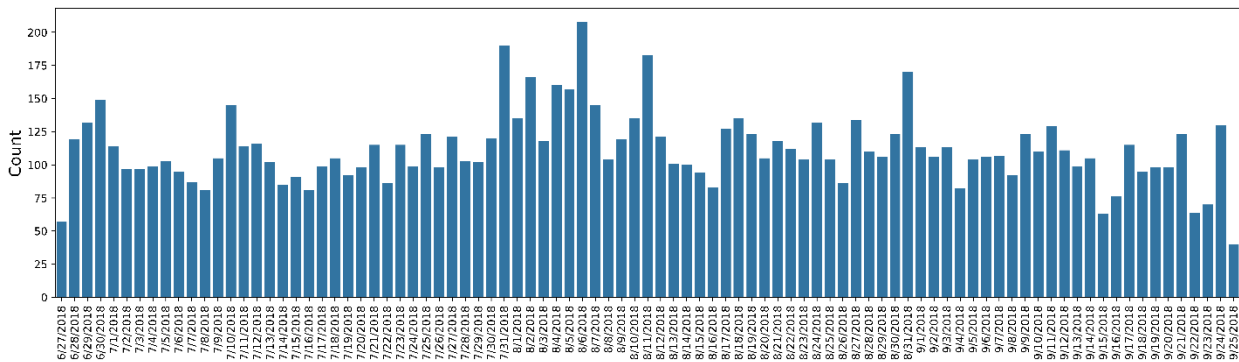
Figure 35. Graph. Daily incident alarm count for studied I-475 section, from 6/27/2018 through 9/25/2018.

Figure 36 provides the overall hourly distribution of incidents over 24 hours, as well as the distribution by direction, i.e., southbound in Figure 36 (b) and northbound in Figure 36 (c).

65
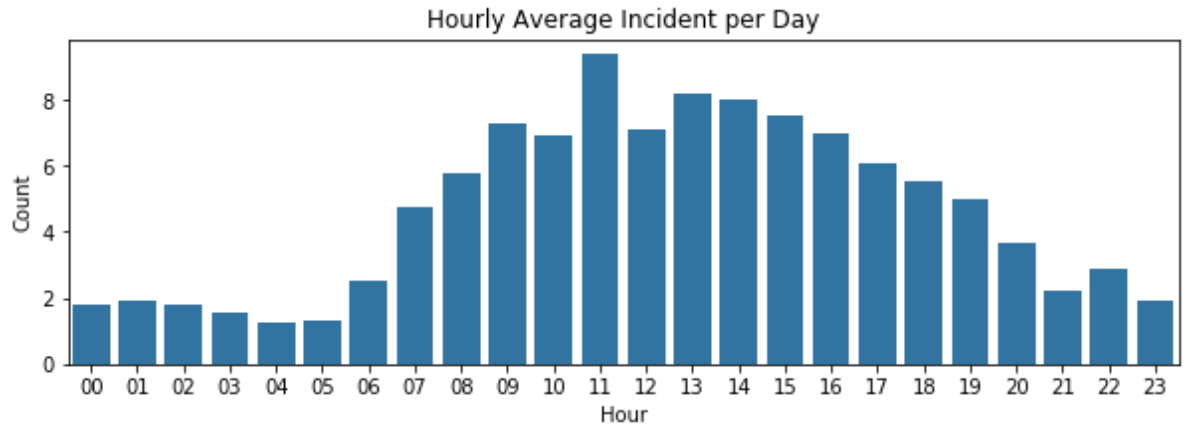
(a)
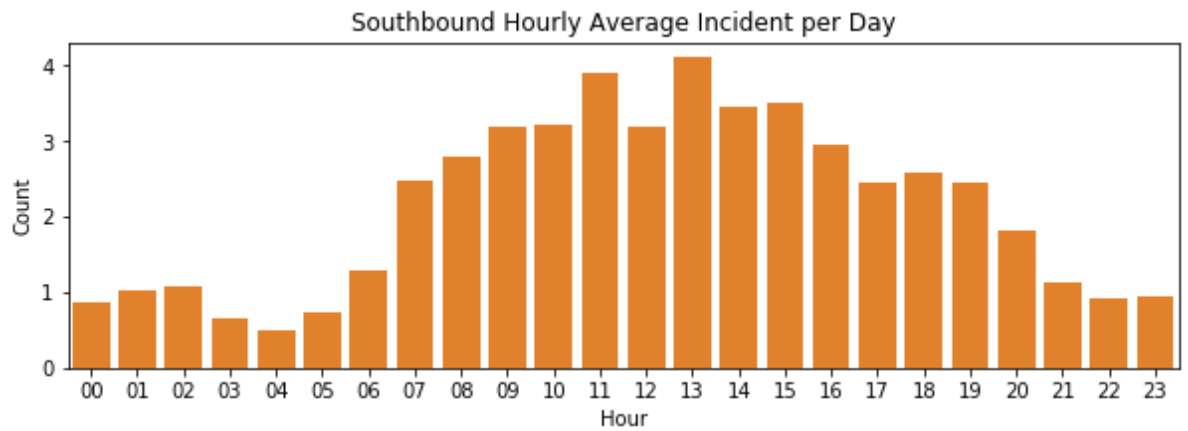
(b)
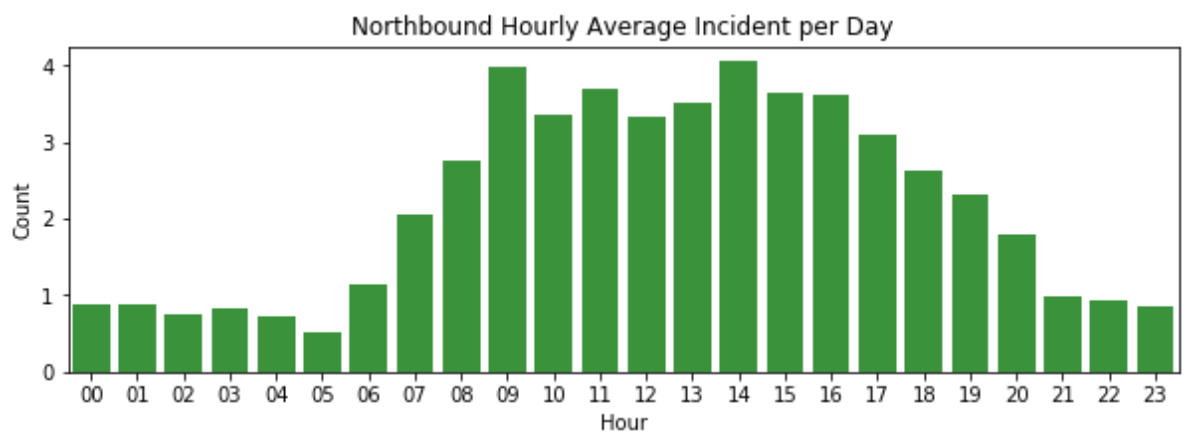
(c)

Figure 36. Graphs. Hourly average incident alarm count for studied I-475 section, from 6/27/2018 through 9/25/2018: (a) total, (b) southbound, (c) northbound.

The average hourly incident per day on I-475 is 4.31, and the standard deviation is 2.36. Incident alarms are more frequent from 7:00 AM to 7:00 PM than the average hourly incident rate, with a fairly consistent peak rate between 9:00 AM and 4:00 PM. Again, it is seen that a system covering a larger portion of the Interstate system is not easily scalable given the number of alarms.

The vast majority of reported AID alarms are stopped incidents (92.7 percent), followed by congestion (4.6 percent), slow (2.4 percent), and wrong-way (0.3 percent) incidents. For each reported alarm type, figure 37 provides the location categorization: shoulder, ramp, or active lanes.
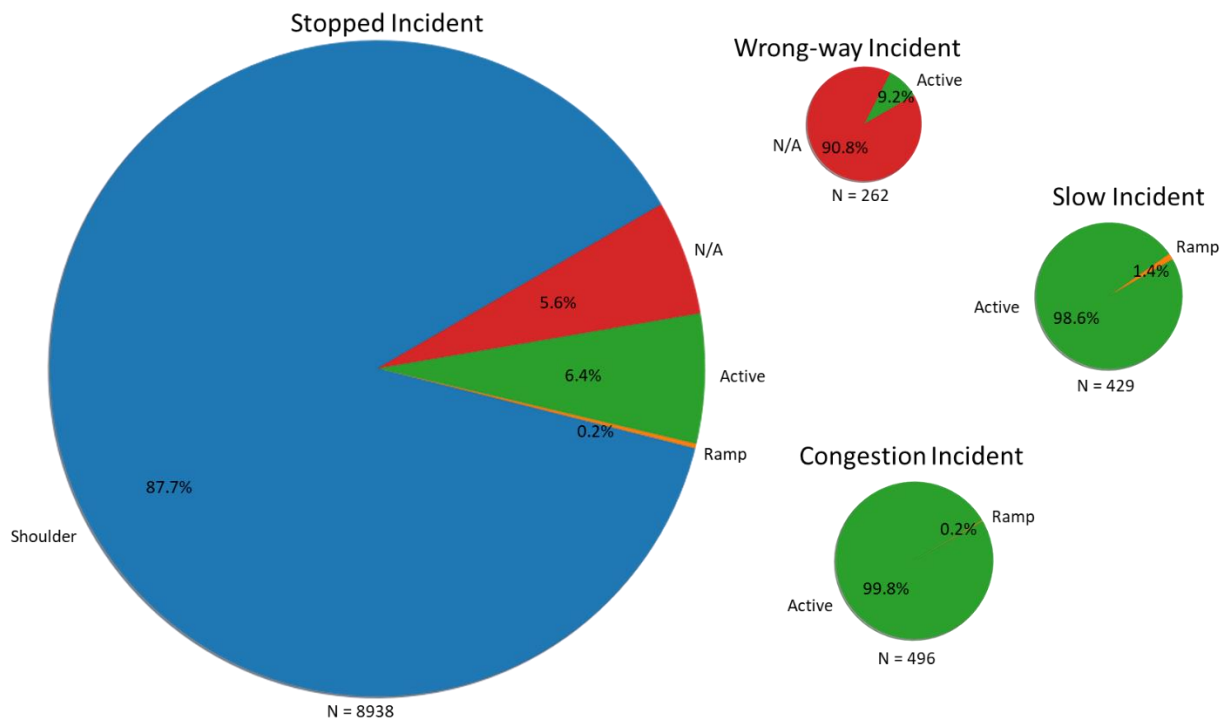
Figure 37. Chart. Location categorization per AID incident alarm type for studied I-475 section, from 6/27/2018 through 9/25/2018.

In figure 37 it is readily seen that the most prevalent alarm type is stopped incidents located on shoulders, comprising 77.4 percent of all alarms. The remaining 22.6 percent of alarms are distributed over the remaining categories.

AID False Alarm Evaluation

A key component of the undertaken study is an evaluation of the AID system's false alarm rate. This evaluation focuses on false alarms, as an independent data source is not available to identify missed incidents, i.e., false negatives.

For the AID false alarm rate evaluation, two analyses were conducted. First, alarms were manually reviewed, using the images and video clips provided by the video-based AID software. Second, these alarms were compared with incident logs from the existing incident management operations, based on Georgia 511 calls, public safety reports, and TMC operators identifying an incident.

Manual Review of Alarm Video and Image Data

For AID evaluation analysis using the images and video clips provided by the video-based AID software, two aspects of false alarms were considered: incident presence or type error, and incident area error.

The incident type analysis included manually viewing the video and images to determine if an incident was identified where none occurred or if an incident was misidentified, e.g., a slow incident was reported where traffic was stopped. Identified errors consisted of predominantly two types: (1) other objects (e.g., lights, leaves, reflection on the camera lens) within the camera view triggered the AID false alarm; and (2) camera settings,

particularly camera angle, resulted in misidentification (e.g., traffic appeared slow-moving but was not). As reviewer judgement was required in identifying an alarm as false, the error had to be obvious to be considered. For instance, from visual inspection of the photo and video it is often difficult to differentiate between "congestion" and "stopped"; thus, such an error was rarely identified. Figure 38 shows some examples of scenarios that result in false alarms.

(a) Incident alarms for wrong-way driving are likely to be generated when there is a shift in camera view in a Pan-Tilt-Zoom camera.

(b) Alarm was generated by movement of foliage.

(c) Alarm for stopped incident on an active lane generated when a heavy-duty vehicle is stopped on the shoulder.

Figure 38. Photos. Examples of false alarms due to: (a) a shift in camera view in a Pan-Tilt-Zoom camera, (b) movement of foliage, (c) a heavy-duty vehicle stopped on the shoulder.

Incident area analysis confirmed the incident location (i.e., ramp, active lane, or shoulder), as well as confirmed if the incident roadway direction was correctly identified (i.e., northbound or southbound). Incident roadway location is critical in the evaluation of AID performance, as it is a key component in determining the incident severity and response. For example, a vehicle stopped in the active lanes may require a rapid emergency vehicle response, whereas a vehicle stopped in the shoulder may be sufficiently served by roadside assistance vehicles. Figure 39 provides the findings from the manual review of incident alarms.

Due to the limitations of the data available for the manual verification process, a clear distinction must be made between false and nonverified alarms. The evaluation was dependent on the images and short video clips archived by the AID as part of the detection process logs. The view was limited to a small section of the roadway as seen by the AID camera. It is not possible to conclude with certainty that an alarm is false, if there is no evidence of a stopped vehicle, or debris on the roadway within this limited viewport. There is always a possibility that the disabled vehicle moved outside the view following the occurrence of the incident. While a few cases could be identified with a high degree of certainty that they are false, such as those illustrated in figure 38**Error! Reference source not found.**, the percentage of such cases is small. The cases where no evidence of an incident could be found are therefore referred to as *unverifiable* rather than *false* for the rest of this analysis.
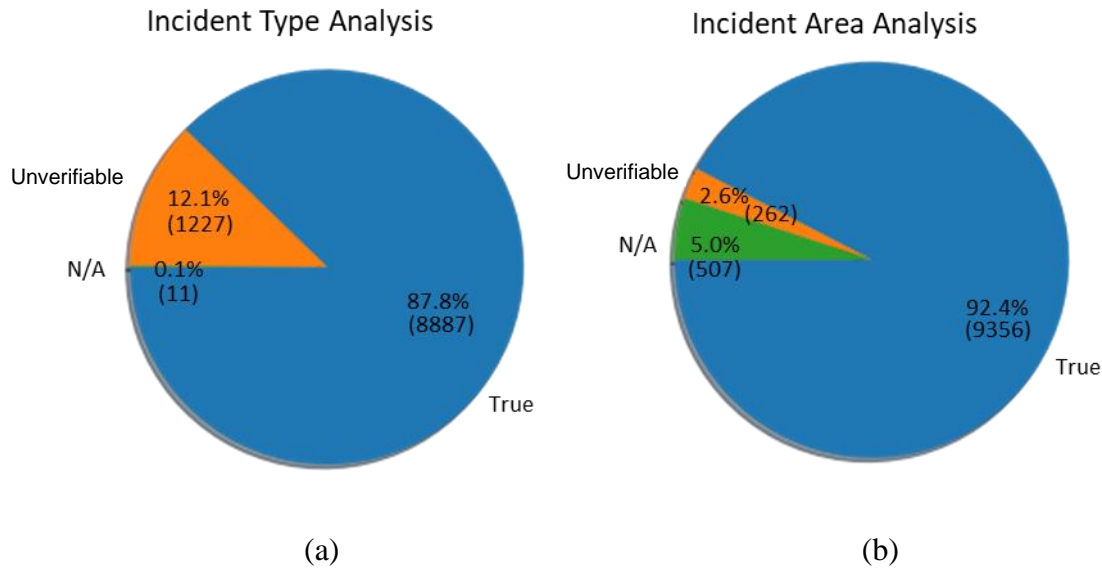
Figure 39. Charts. Results of: (a) incident presence and type, (b) incident area analysis.

Figure 39 (a) displays the proportion of true and unverifiable alarms detected by incident presence and type. Of the 10,125 incident alarms generated during the 3-month period, 8,887 incident type alarms were true (87.8 percent); 1,227 incident alarms (12.1 percent) were unverifiable; and 11 alarms (0.1 percent) were undetermined, as the images and video clips were unavailable. Figure 39 (b) provides the location accuracy findings, with 92.4 percent of alarms correctly located and 2.6 percent incorrectly located, and 507 alarms (5 percent) did not provide a detection lane or the video and images were unavailable. The majority of the location errors result from heavy-duty vehicles in the adjacent zones on the video, typically locations with flatter viewing angles, or when a vehicle stopped on the shoulder was identified as being in an active lane due to camera view angle.

A critical issue in this evaluation, and the AID system, is that "crashes" are not one of the incident types. Rather, the AID system reports incident types of stop, congested, slow, and wrong-way. While crashes will most likely be identified as "stop" incidents, the reverse is

not true, in that most stop incidents (92.7 percent of all incidents) are not related to crashes. Thus, identifying crashes is a primary need for which the AID system does not provide an efficient solution in the current context, with the vast majority of incident alarms unrelated to crashes.

AID Evaluation with Comparison Analysis with Existing Incident Logs

To further explore AID in relation to crashes, the AID data are next compared to crashes identified though the Georgia 511 (NaviGAtor) incident logs. Georgia 511 is provided through the Georgia Department of Transportation and is housed at the GDOT TMC. The Georgia 511 logged incidents are detected by a variety of methods, such as operator-detected (i.e., TMC operator observing highway video feeds), mobile operator (HERO, i.e. GDOT roadside assistance vehicles), motorist calls, police department / 911, etc.

During the same 3-month period as the AID I-475 data collection, Georgia 511 generated 104 incident logs (see figure 40). The detected incident logs include 12 crashes, 24 debris in roadway, 1 fire, 1 infrastructure issue, 65 stalled vehicles, and 1 unplanned (i.e., live animal presence) case. Of the 104 incidents, most incidents required temporary active lane or shoulder closures, as indicated in the TMC incident log system. It is recognized that crashes not captured by the existing Georgia 511 system or seen in the AID video clips will not be reflected in the subsequent analysis within this report; thus, a complete measure of missed crashes is not possible.
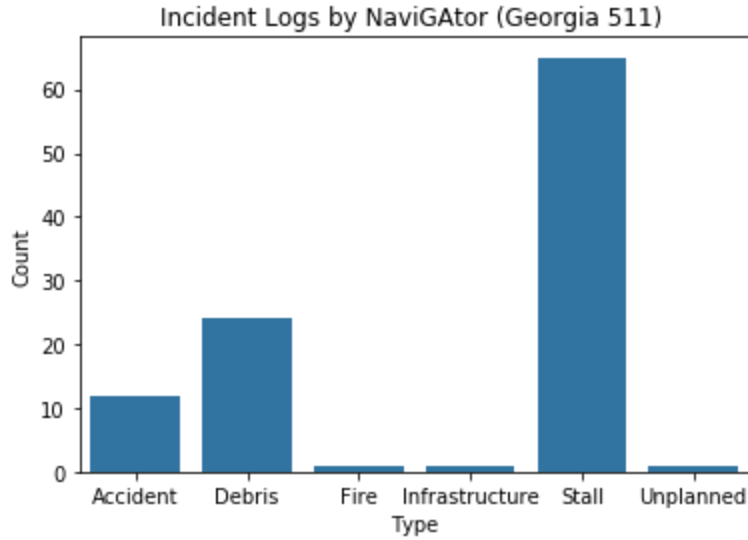
Figure 40. Graph. Count per incident type on NaviGAtor (n=104).

The current comparison focuses on the 12 crashes identified through the Georgia 511 system. To visually represent the comparison, a separate plot is developed for each of the 12 crashes. The time (time incident was first logged); location (latitude and longitude, based on location reports in the incident log); and direction of travel for each incident was plotted. Next, all AID incidents (stopped, congestion, slowed, wrong-way, and pedestrian) within 5 miles and with a reported start time within 1 hour (before or after) of the given crash were plotted. The utilized AID time is the time at which the alarm was triggered, and the location plotted is of the camera reporting the alarm; thus, the location is commonly within 1,000 ft of the plotted point. Each plot is centered on the NaviGAtor incident, with ±1 hour on the x-axis and ±5 miles (+ indicating upstream and – indicating downstream) on the y-axis, for a total 2-hour timeframe and 10-mile distance coverage. Figure 41 shows two example plots for crashes: (a) one recorded by Georgia 511 at 9:17 AM on August 2, 2018, at approximate mile marker 9; and (b) one recorded at 6:16 PM on September 2, 2018, at approximate mile marker 10.
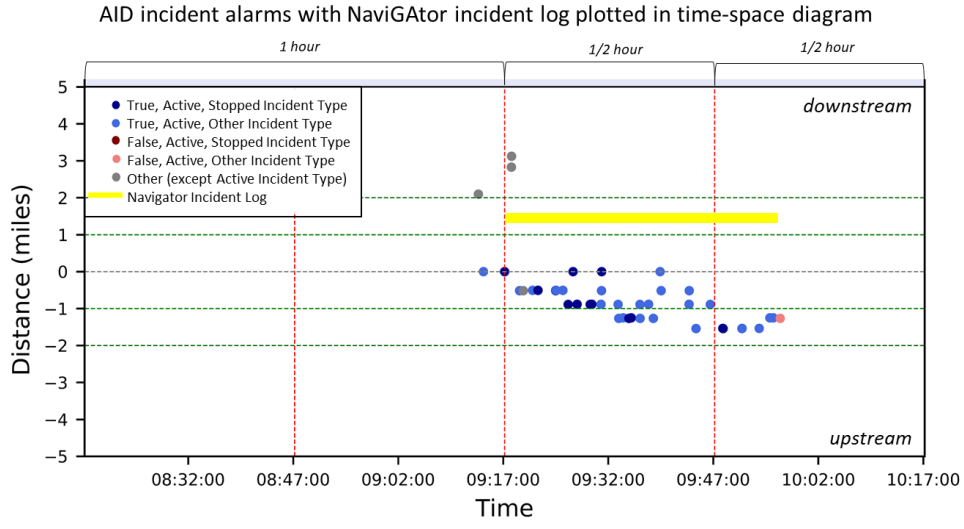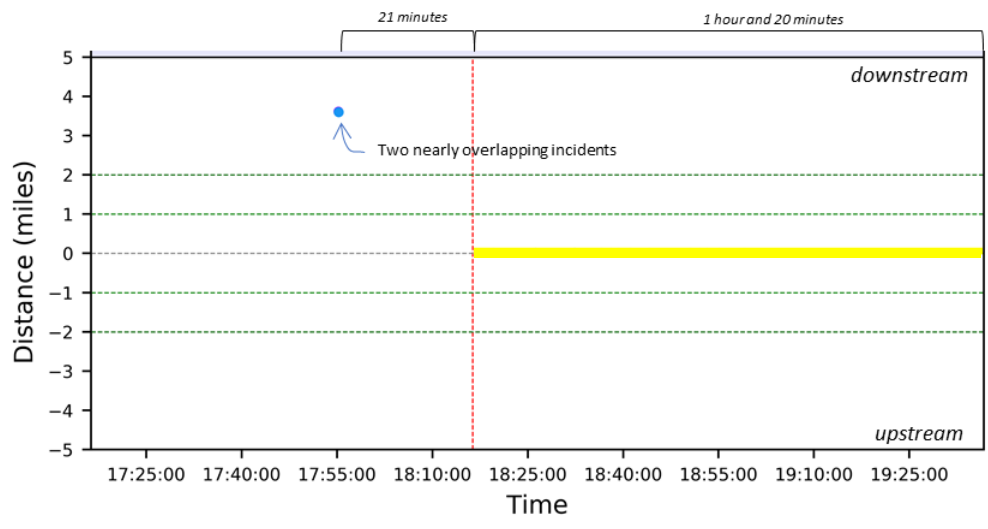
Figure 41. Plots. Georgia 511 and AID incident alarm time–space plots: (a) with significant AID cluster, (b) without AID cluster.

It is seen in figure 41 (a) that a number of AID alarms are triggered at or near the location of the crash, as well as within 45 minutes of the crash. A trend of the alarms moves upstream from the crash as time progresses. This cluster is likely associated with the crash and such a cluster provides a means to potentially identify crash incidents in AID data.

74

However, as seen in figure 41 (b), AID alarms do not always cluster around recorded crashes. Whether the lack of associated alarm clusters at some crashes is related to an AID algorithm issue or to incomplete camera coverage over the 16-mile section is unknown. Future efforts will seek to confirm coverage in these areas. (This is not readily done as this task is nontrivial, involving police records, site visits to confirm camera latitude/long and camera coverage, etc.) However, while not consistent, approximately half of the Georgia 511 crashes did seem to have an associated AID cluster, so this potential is further explored in the remainder of the chapter.

METHODOLOGY (MACHINE LEARNING FRAMEWORK)

The developed methodology seeks to identify alarms for high-impact incidents that need immediate attention from emergency and TMC responders. The methodology uses the co-occurrence of multiple alarms in the time–space vicinity, to determine whether an incident is significant or nonsignificant in terms of the impact of the incident on traffic. The framework is built on top of an unsupervised clustering machine learning algorithm. Consolidation strategies and filters are developed as additional layers over the ML algorithm to further tune the algorithm and eliminate the majority of false, unverifiable, and noncritical alarms. Confidence values are assigned to alarms to further assist in prioritizing which alarms to confirm before dispatching a response unit and are particularly useful during busy periods.

Cluster Analysis

Density-based spatial clustering of applications with noise (DBSCAN) (Pedregosa et al. 2011), a well-known unsupervised ML density-based clustering algorithm, was selected to

identify contiguous occurrences of multiple AID alarms. Unlike other clustering

algorithms, such as the K-means clustering, DBSCAN does not require the number of

clusters to be specified as an input parameter but rather infers the number of clusters as

output depending on the data structure. DBSCAN requires two input parameters for

execution: an epsilon value that defines the radius of the neighborhood around a point p,

and the minimum number of points in the epsilon neighborhood (including point p). If

point p satisfies the minimum number of points within the epsilon radius, a 'cluster' is

formed, and point p is considered a 'core point'. If other points in the cluster have also been

identified as core points, then the respective clusters are joined to create a single joint

cluster of all points. Points within a cluster that are not core points are labeled 'boundary

points', and points not clustered are labeled 'outliers'. The DBSCAN algorithm continues

until all points have been checked for being a core point.

The DBSCAN algorithm typically works on spatial clustering for a single factor, e.g.,

distance between points. The AID dataset, however, is multidimensional, including time

and space factors. Therefore, by adding a temporal feature in DBSCAN, a modified

version, DBSTCAN (density-based spatial and temporal clustering applications with

noise), has been developed whereby a temporal epsilon parameter is added to the existing

set of parameters in the DBSCAN algorithm. This is similar to the approach conducted by

Birant and Kut (2007); however, instead of four parameters required for their algorithm, we

simplified our algorithm with three parameters: *epsilon_distance* (from longitude and

latitude information), *epsilon_time*, and *min_points*. This is well suited to the traffic-

monitoring camera infrastructure, which is an approximately fixed distance between traffic

poles. The three parameters also offer increased computational efficiency, which is critical

in real-time operations. The DBSTCAN algorithm development is based on the AID

clusters visually identified as discussed above. The analysis focuses on the northbound

direction of travel, which contains more of the identified crashes.

DBSTCAN Calibration

As discussed, the DBSTCAN algorithm requires three parameters for cluster detection: the

time dimension radius (*epsilon_time*), the space dimension radius (*epsilon_distance*), and

the minimum number of points required in a cluster (*min_points*). For the space dimension

radius, longitude and latitude values of the points are utilized to determine a Euclidian

distance measurement. The objective of the DBSTCAN calibration was to find the best set

of parameters to accurately identify clusters. To rate the performance of a parameter set, the

performance metrics utilized are false alarm rate, detection rate, and mean time to detect

(MTTD), which are in line with performance measures used in previous AID studies. For

this analysis, the term false alarm is expanded to refer to an incident alarm where none

exists (or is unverifiable) or the incident is low-impact and is therefore more appropriately

referred to as the NCAR. The metric definitions are fine-tuned specifically to this study, as

in equations 1–3:

$$NCAR = \frac{number\ of\ low\text{-}impact\ clusters\ detected}{number\ of\ available\ high\text{-}impact\ clusters} \tag{1}$$

$$DR = \frac{number\ of\ detected\ high\text{-}impact\ clusters}{number\ of\ available\ high\text{-}impact\ clusters} \tag{2}$$

$$MTTD = \sum_{k=0}^{n}(t_d - t_o)/n \tag{3}$$

where,

    n = number of defined clusters

    t_d = time when a cluster is detected by the algorithm

    t_o = time of occurrence of the first alarm in the cluster

In terms of the performance metrics, the high-performance models are those that have the lowest NCAR with the highest DR while keeping MTTD within acceptable bounds.

To apply these performance measures, it is necessary to know the number of "high-impact" clusters in the data. To determine this, all 3 months of the AID were plotted on time–space diagrams, as in figure 41, and clusters were manually identified. A total of 15 clusters were identified. Of these, 12 were in the northbound direction and 3 were in the southbound direction. For efficiency, the described analysis will focus on the northbound direction of travel. Of the 12 northbound clusters, 4 were associated with crashes, 1 was a stall, and 1 was an animal in the roadway, as identified in Georgia 511. The remaining 6 were not in the Georgia 511 logs but will be retained in this analysis as the cluster of AID alarms is likely a significant traffic event worthy of review by a TMC operator. Thus, the set of 12 clusters is considered the high-impact clusters. Any other clusters identified by DBSTCAN are considered low-impact clusters.

A set of the 1,000-combination parameters was used to define a set of 1,000 DBSTCAN models that were applied to the AID dataset. The range and increments of the parameters were as follows:

- *epsilon_time*: 1 to 10 minutes in 1-minute increments.

- *epsilon_distance*: 0.2 to 2.0 miles in 0.2-mile increments.

- *min_points*: 1 to 10 points in increments of 1 point.

The ranges were selected to achieve a reasonable balance between the applicability of the results and the computational requirements to run the experiment. A future effort may seek to further fine-tune the stated ranges. The increment sizes for *epsilon_time* and *epsilon_distance* were chosen based on findings in Taylor et al. (2017).

DBSTCAN Parameter Selection Strategy (Model Selection Strategy)

DBSTCAN results for the 1,000 models were analyzed to identify the optimal parameter set, i.e., those that resulted in the lowest NCAR and highest DR values, while maintaining an acceptable MTTD. Figure 42 shows the relationships between these three parameters in a scatter matrix plot, which is a grid of scatter plots used to visualize bivariate relations between combinations of variables. Each point in the scatter plot represents the results of the DBSTCAN run of a parameter set, with 1,000 points in each plot. (Plots that seem to have fewer points have multiple points in a single location). The histograms show the distribution of each numeric variable.
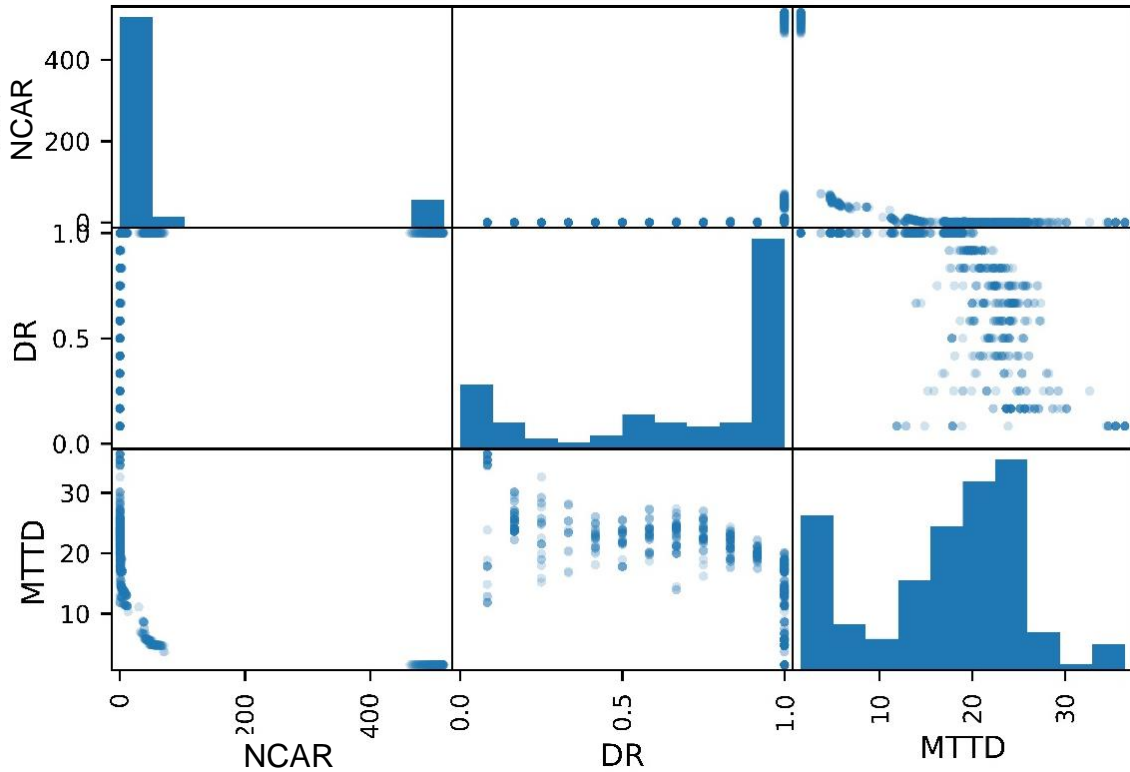
Figure 42. Scatter matrix plot. NCAR, DR, and MTTD for 1,000 DBSTCAN models.

As expected, no single combination provided an optimal solution across parameters and clusters. DR and NCAR have a weak positive relationship, while NCAR and MTTD have a weak negative relationship. Hence, a solution based on an ensemble algorithm approach was envisioned. Different strategies for using multiple parameter sets, rather than a single parameter set, were investigated. A distributed decision model was thereby created to assign confidence values to alarms.

Top 20 NCAR–TTD per Defined Cluster

To reduce the number of dimensions of the optimization problem, a multistep process was utilized. First, from the 1,000 DBSTCAN runs, only the models that successfully detected at least one of the 12 visually identified northbound clusters, i.e., high-impact clusters, were

selected to be carried forward. From these remaining models, the top 2 values of the lowest

TTD for each of the 12 high-impact clusters were identified.[2] Next, for models with the

identified TTDs for each high-impact cluster, those with the lowest NCAR were selected,

10 for each TTD value. The selection process is illustrated in figure 43 (a). The number of

selected "dots" within the yellow rectangles appears to be less than 20 because in several

cases different models (i.e., DBSTCAN runs with different parameter sets) have identical

NCAR and MTTD values and the points overlap in the plot. Figure 43 (b) shows the

NCAR vs. TTD plot of the 240 models selected from the original pool of 1,000,

corresponding to the 12 high-impact clusters. It was found that some parameter sets were

selected for multiple high-impact clusters. After the elimination of duplicate selections, the

number of unique parameter sets was 159.

---

[2] As a clarification, each model has an associated TTD for each cluster, with MTTD being the average TTD across clusters.
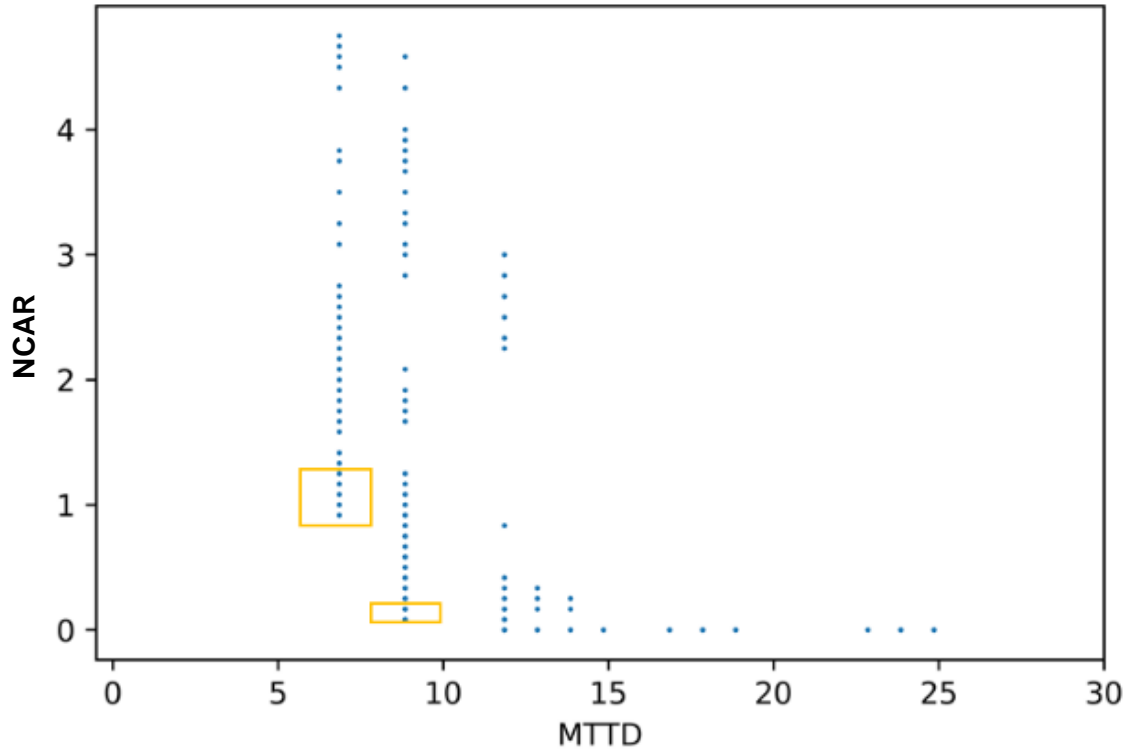
(a)
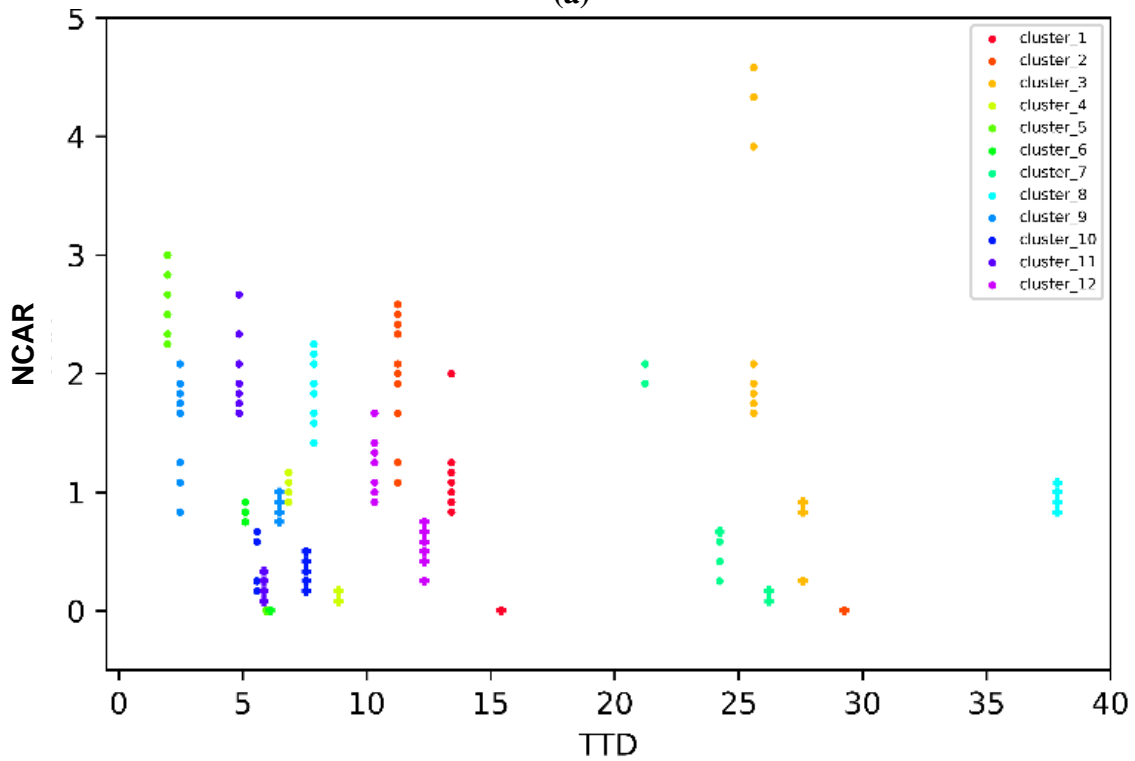
(b)

Figure 43. Plots. (a) Selection process example (4th high-impact cluster), (b) NCAR vs. TTD for the 240 selected models.

Pseudo–Real-time Analysis

While post hoc analysis is unconcerned with model runtime (within reason) and may utilize multiple days or months of data, real-time processing of the alarms to identify high-impact incidents must be computationally efficient. Thus, data processed within a given run is limited to ensure real-time processing speed. Since timely detection is critical, it is assumed that the algorithm would run once every minute and would process the last 1 hour of data. Real-time detection data older than an hour is considered to stagnate and would result in a TTD significantly longer than useful in practice. To replicate a real-time process, a sliding 1-hour window that shifts in 1-minute increments is used to create the input data for the DBSTCAN models. This process is repeated for every 1-minute increment over the 90 days of the dataset for a total of 20,606,400 runs. The processing by 159 models generated 4,093 clusters. However, a majority of these clusters had overlapping points. When clusters with overlapping points were consolidated, this yielded a total of 14 distinct clusters that overlapped with the identified 12 northbound high-impact clusters and 77 distinct clusters that did not overlap with high-impact clusters. Further investigation was made to understand why the model resulted in 14 clusters to capture the 12 visually identified clusters. For two of the visually identified clusters (second and eighth in the order of appearance), it was seen that the DBTSCAN model split these into two closely spaced clusters, as in figure 44.

2nd Cluster of the 12 Significant Incidents

(a)

8th Cluster of the 12 Significant Incidents
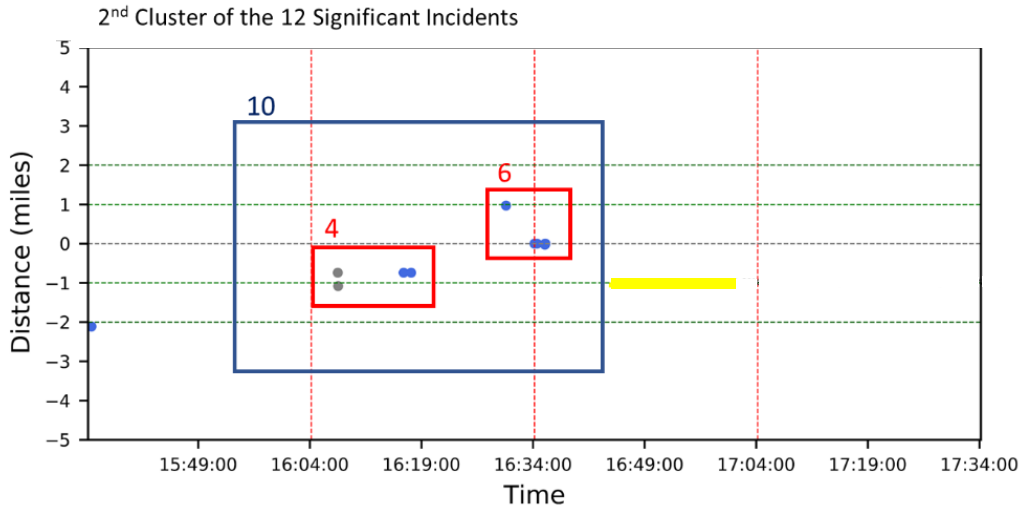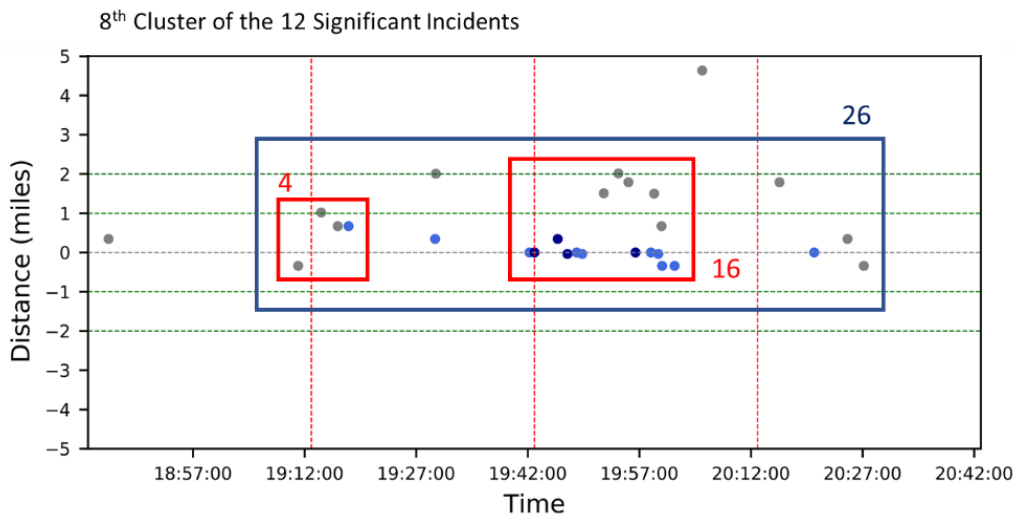
(b)

Figure 44. Plots. Overlap of model-identified clusters with visually identified clusters: (a) second cluster, (b) eighth cluster.

Figure 45 shows the number of models that detected each of the high-impact and nonrelevant clusters by chronological order for the study period. In most of the cases for the high-impact incidents, the number of models that detect the incident is quite high. However, in two of the high-impact incidents, where the original clusters are split into two,

the number of models that detect the first cluster in the pair is relatively low, at 10 and 20 out of 159. For the remaining high-impact incidents, the number of identifying models is above 90 (out of 159), representing a detection rate above 0.6. In comparison, only 6 of the 77 nonrelevant clusters have a detection ratio over 0.5.
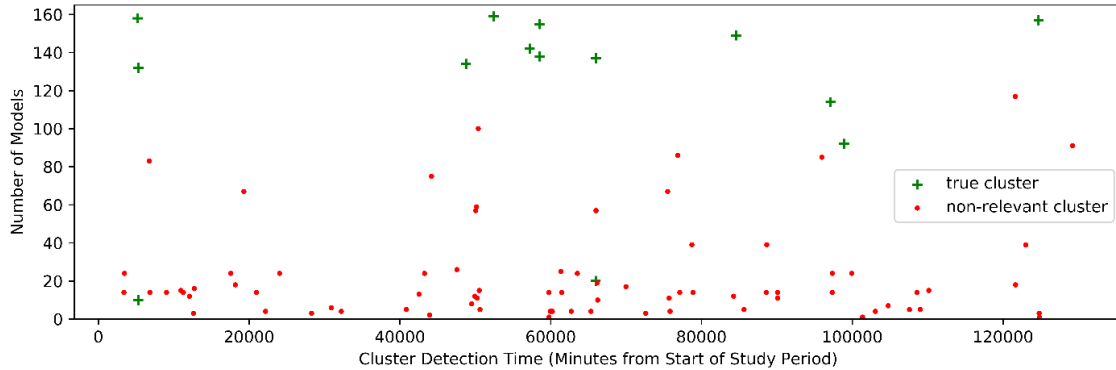
Figure 45. Plot. Number of combinations that detected clusters over the study period.

Figure 46 shows the box-and-whiskers plot of the time required to detect each incident based on the estimated incident occurrence time and the cluster detection time. It is seen that while most of the incidents are characterized by a mean TTD below 15 minutes, incidents 1 and 8 for the high-impact incidents have a larger mean TTD.
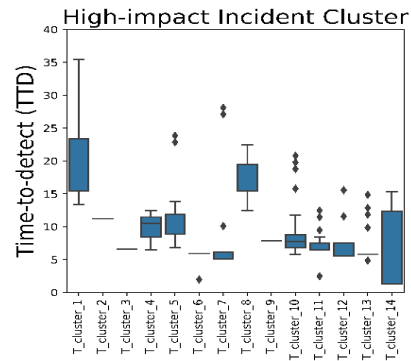
(a)
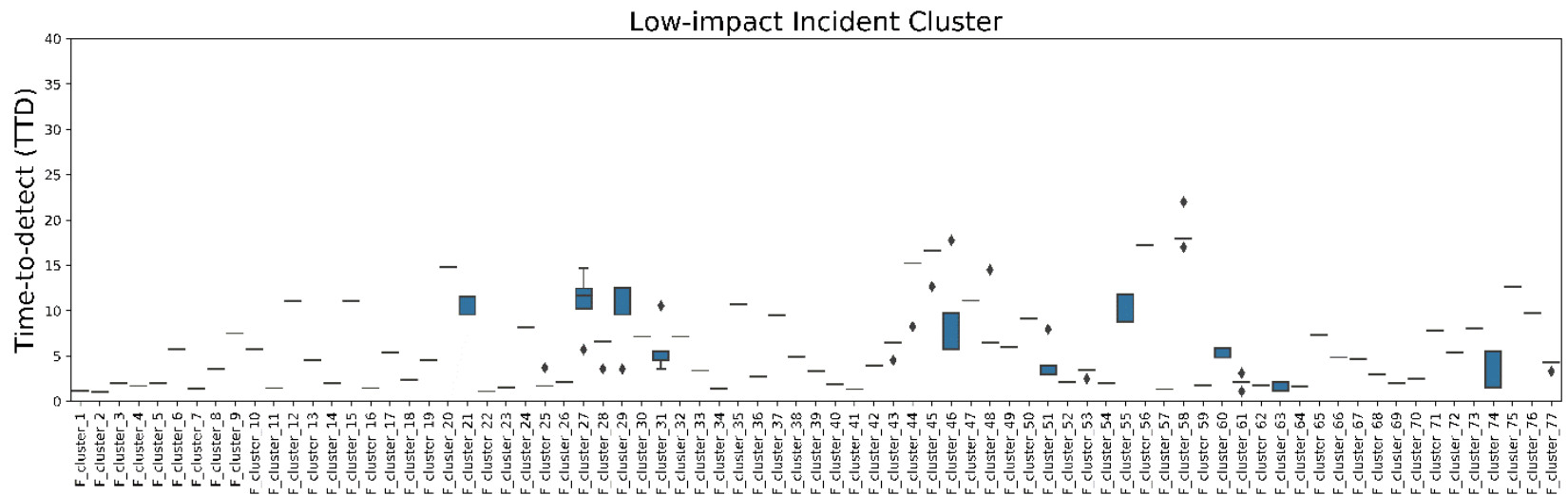
(b)

Figure 46. Plots. TTD boxplot for: (a) high-impact clusters, (b) low-impact clusters.

Cluster Performance Index

The combined DR, NCAR, and MTTD performance of the 159 models is reflected in figure 47. The figure confirms that the models with low NCARs (desirable) have high MTTDs (undesirable) and supports the hypothesis about the need for an ensemble approach whereby models with observed low MTTD can be used to identify a high-impact incident early, but would need to be confirmed either by the co-occurrence of multiple alarms or an alarm from a model with observed low NCAR.
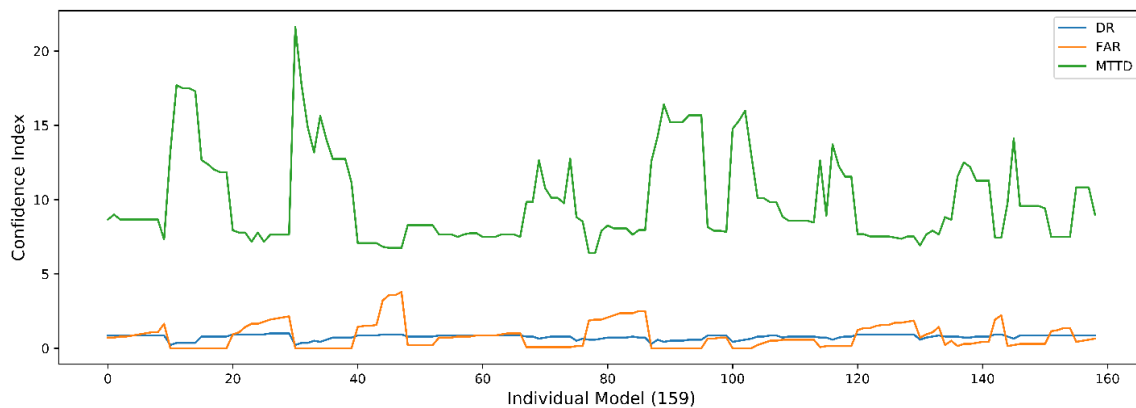
Figure 47. Plot. Performance metrics (DR, NCAR, and MTTD) by each combination.

For this purpose, a framework is developed for strategically merging the output of the individual models in the ensemble to produce a single signal that can be used to trigger a timely alarm for a high-impact incident. A *confidence index* is first assigned to each model in the ensemble. The *confidence index* is calculated as a function of the model's observed NCAR and DR values, as shown in equation 4. A *weight* value (a fraction ranging between 0 and 1) is used in the formula, to control the balance of importance assigned to NCAR vs. DR in computing the confidence, with a higher *weight* value assigning more importance to NCAR and less importance to DR.

$$Confidence\ Index\ (i) = weight \times \frac{max(NCAR) - NCAR(i)}{max(NCAR)} + (1 - weight) \times DR(i)$$

$$(4)$$

where,

i = 1 to 159 (ID of 159 combinations)

$$NCAR(i) = \frac{Number\ of\ nonrelevant\ clusters\ per\ model\ (i)}{Total\ number\ of\ relevant\ clusters}$$

$$DR(i) = \frac{Number\ of\ true\ clusters\ per\ model\ (i)}{Total\ number\ of\ relevant\ clusters}$$

As defined earlier, the NCAR was defined with the number of relevant clusters as the denominator instead of the total clusters to ensure that the NCAR value truly reflects the ratio of noncritical to true signals rather than becoming a function of the total number of alarms. NCAR values could, therefore, exceed 1. In the *confidence index*, the value is normalized so the *confidence index* values can be bounded between 0 and 1. An *impact index* is then computed at each time step for the entire spatial area of coverage as shown in equation 5. The *impact index* combines the cluster-detection output of all models in the ensemble by aggregating the weighted outputs of the model with the *confidence index* of each model as the *weight* for that model's output.

$$Impact\ Index\ (i,t) = \frac{\sum_{i=1}^{159} Confidence\ Index\ (i) \times X(i,t)}{\sum_{i=1}^{159} Confidence\ Index\ (i)}$$

$$(5)$$

where,

$X(i,t)$ = 1 or 0 as a Boolean data type for representing a cluster detection at time *t* by model *i*. If a cluster is detected by model *i* at time *t*, $X(i,t)$ = 1. Otherwise, $X(i,t)$ = 0.

Figure 48 shows the *impact index* signals (with *weight* = 0.2) along the 90-day timeline, with blue columns indicating signals matching the manually identified high-impact incidents and red columns corresponding to signals matching the timestamps of the low-impact incidents.

For a closer look at the variation of the *impact index* over time (with *weight* = 0.2), an overlapping 120-minute timeline for each cluster identification is plotted in figure 49, where the x-axis represents the time, for each cluster identification, since the first AID alarm corresponding to the cluster. A clear pattern emerges. The majority of the high-impact incident clusters indicated in blue (except the two cases that were created as a split of clusters 2 and 8) achieve a high *impact index* value within the first 20 minutes, most exceeding 0.6 within the first 10 minutes. On the other hand, the majority of the low-impact incidents never exceed an *impact index* of 0.6, with the highest concentration in the zone below an *impact index* of 0.2.
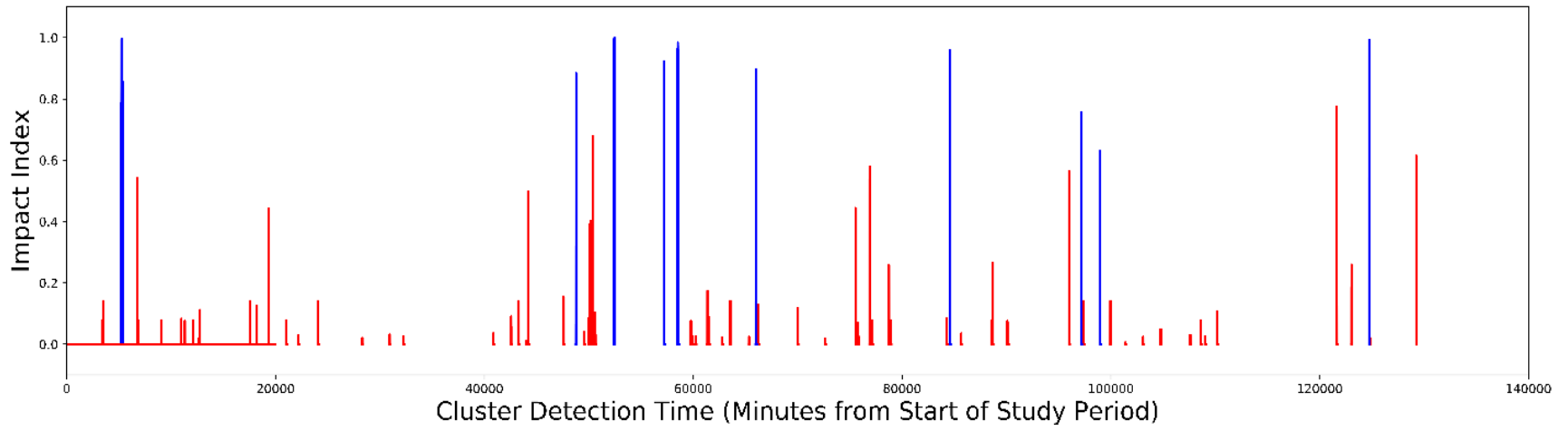
Figure 48. Plot. *Impact index* **of the derived clusters.**
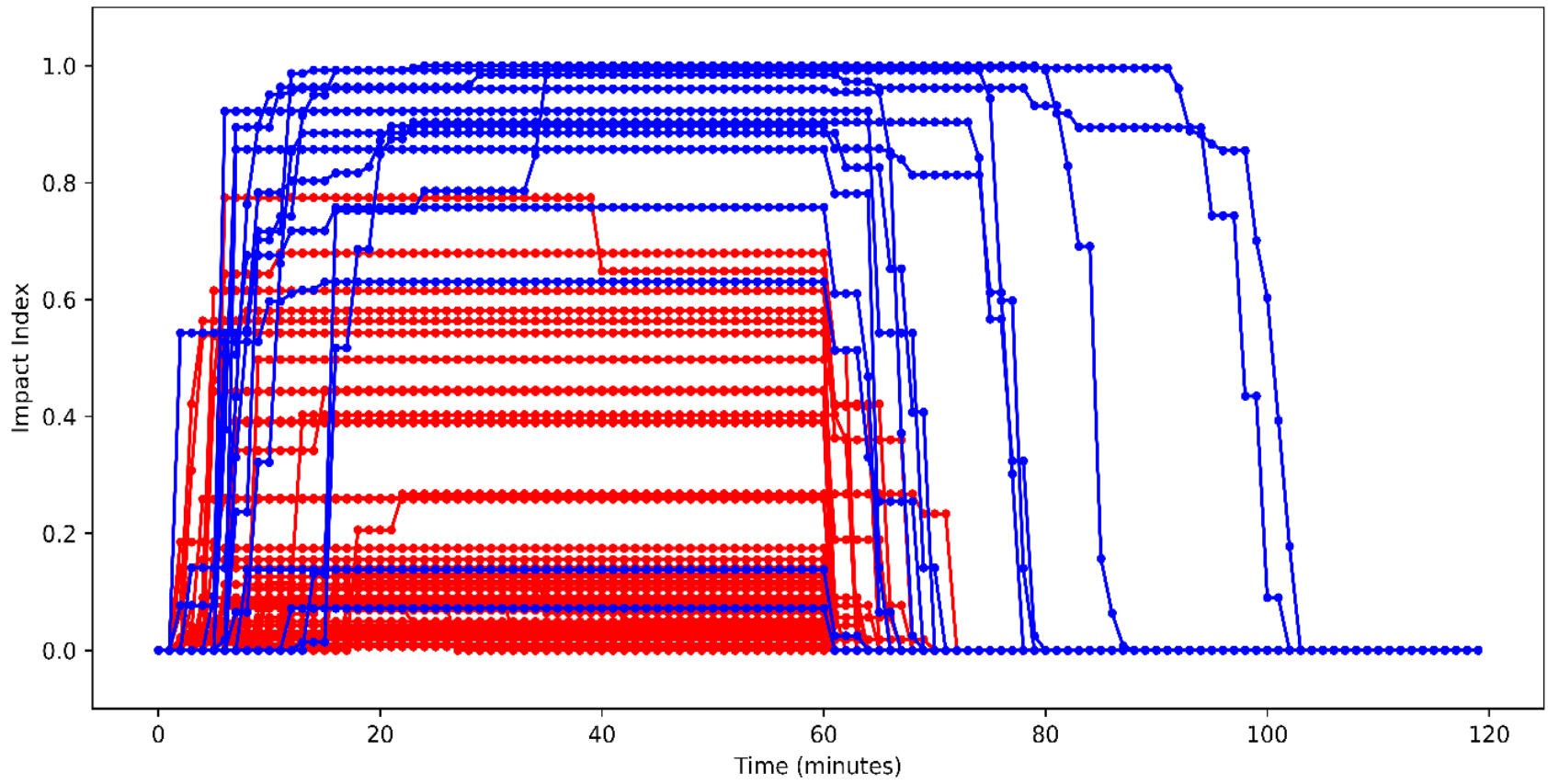
Figure 49. Plot. *Impact index* **for high-impact (blue) and low-impact (red) incident clusters.**

At this point, the effect of NCAR and DR is being captured successfully by the *confidence index*. It was, therefore, postulated that the initial filter applied during the development of the methodology, which reduced the number of models in the ensemble from 1,000 to 159, may not be necessary. The *confidence index* would automatically account for the variation of NCAR and DR and reduce the impact of the models that do not contribute to meaningful detection. However, the lowest value of each parameter, i.e., *epsilon_time* of 1, *epsilon_distance* of 0.2, and *min_points* of 1, were eliminated to improve the efficiency of the algorithm.

The only constraint used was that a single point was not allowed to define a cluster; the models with *min_points* of 1 were eliminated, resulting in 766 models in the ensemble rather than 1,000. Figure 50 shows the inflation effect on the range of NCAR due to the expansion of the ensemble from 159 in figure 50 (a) to 766 in figure 50 (b).

(a)   (b)
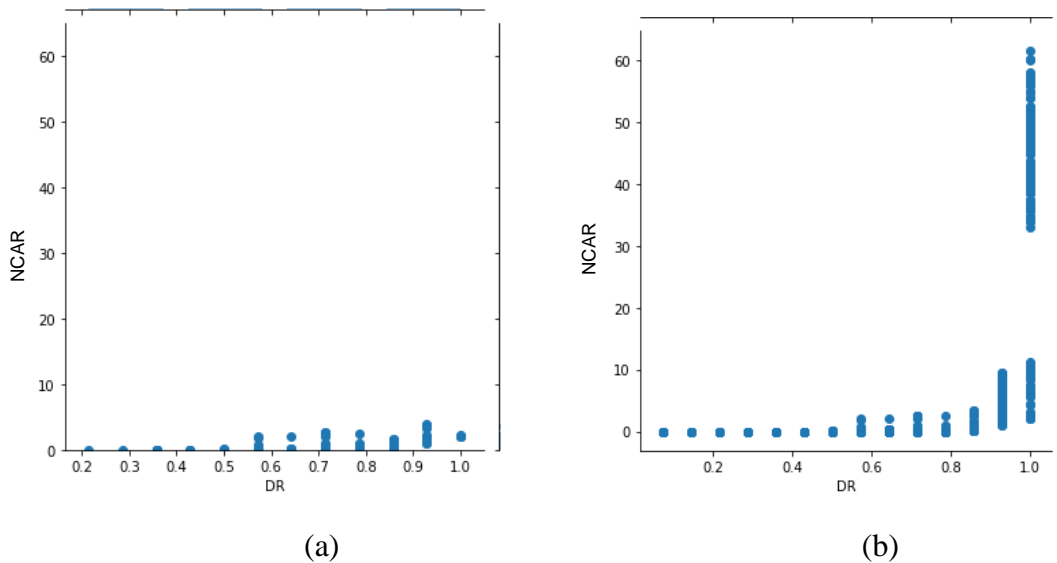
Figure 50. Scatter plots. NCAR and DR for: (a) 159-parameter combinations, (b) 766-combination set.

Figure 51 shows the plots of *impact index* values for the $14 + 77$ clusters, obtained with different variations of *weight* in the *confidence index* formula in equation 4. The left column of plots is for the 159-model ensemble, and the right column is for the 766-model ensemble. It was observed that with the use of the 766-model ensemble, the effect of the *weight* in separating the high-impact and low-impact incidents was significantly reduced. The high- and low-impact clusters were separated successfully by the *impact index* value, regardless of the *weight*. The position of the cutoff point of the separation, however, depended on the *weight*. For example, for *weight* = 0, the separation is between 0.57 and 0.65, whereas at *weight* = 1, the separation line can be drawn somewhere between 0.37 and 0.42. To strike a balance, and improve the stability of the algorithm, a *weight* of 0.5 with the cutoff threshold value of 0.5 for the *impact index* would be preferred. But this would be a parameter that can be tuned as necessary based on experience and feedback by the operators using the system for incident management operations. Additionally, the system could be set to ignore alarms where the first point in an identified cluster exceeds some time limit (e.g., 10 or 15 minutes), as the information is becoming stagnant.

The benefit of utilizing this ML approach to filter the AID alarms is significant. With this approach (with a *weight* of 0.5 and threshold of 0.5), for the given example, 7 out of the 14 high-impact clusters would be identified with the first 10 minutes of the initial AID alarm. Additionally, only one noncritical alarm would occur. Thus, a TMC could reduce to a total of eight alarms out of the original 10,125, which is a decrease of 99.9 percent. While it is readily acknowledged this is a single example based on a limited set of data, the potential benefit in the practical application of an AID system seems significant. Certain alarms, such as wrong-way, may be treated differently by a TMC.

Weight	159 Combinations	766 Combinations
0.0		
0.3		
0.5		
0.8		
1.0		
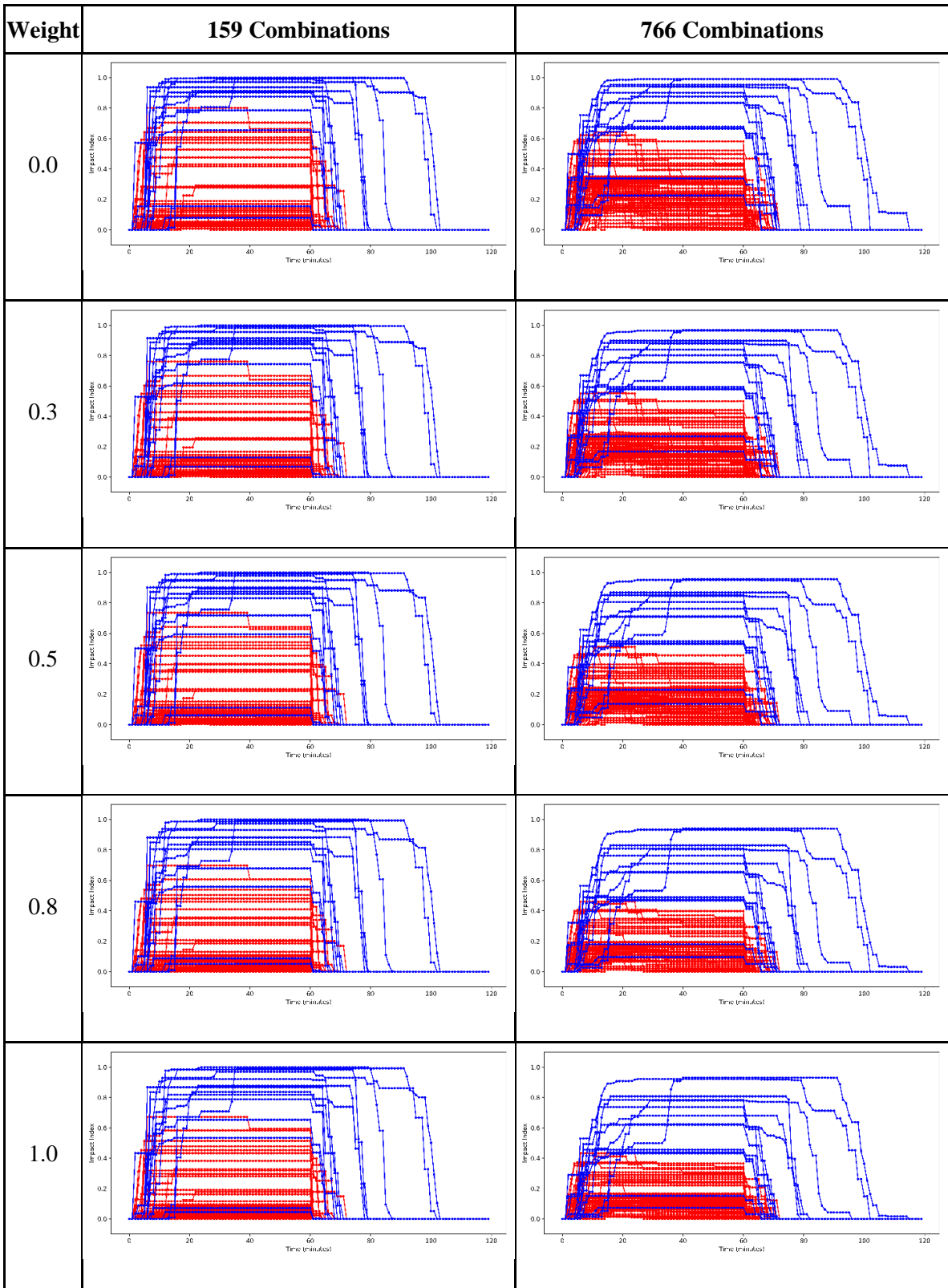
Figure 51. Plots. Confidence level at different weights (159-combination and 766-combination sets).

CONCLUSIONS

Overall, this study demonstrates the feasibility of using machine learning techniques to improve the efficiency of AIDs, minimizing noncritical alarms. The development of such a methodology is presented, using data from a corridor near Macon, Georgia, to demonstrate the methodology with a case study. During the development of the methodology, an extension of the conventional DBSCAN algorithm, which is typically used for cluster analysis, was proposed where the algorithm was extended to scan in a two-dimensional space (i.e., temporal, spatial) instead of a single dimension (i.e., spatial). The new algorithm developed was named DBTSCAN in recognition of the addition of the temporal aspect of the search.

The study developed an ensemble algorithm that takes advantage of the strides of advances in computing power that have been made since the development of the first AID algorithms. In particular, this algorithm is developed with the target of scalability whereby multiple models can run in parallel on distributed resources with a final combination of the outputs from the model ensemble into a single number for triggering an alarm for a high-impact incident that needs the operator's attention. The study also develops the concepts of a *confidence index* to capture the historical NCAR and DR performance measures of the ensemble models into a single parameter, and the concept of an *impact index* to facilitate the separation of AID alarms of high-impact incidents from low-impact incidents.

The advantage of this approach is that the methodology is agnostic to the underlying AID and can be used by incident management programs to enhance and manage the outputs from other AID algorithms. However, that also means that certain limitations of the

95

underlying AID carry through, as well, and impede overall accuracy. For example, if the underlying AID has some blind spots, which is quite possible in a video-based AID, the incidents might not be detected; the ML methodology developed in this study will not improve the probability of detecting such an incident. In addition, if the underlying AID generates a high volume of noncritical alarms with a certain spatial or temporal bias such that they trigger cluster-detection in a majority of the ensemble algorithms, then the current methodology will have to be enhanced with a feedback loop via an additional ML layer that will help identify the noncritical instances and provide a mechanism for the algorithm to continually retrain itself to eventually grow the capacity to filter out such instances by itself.

Finally, the authors note a specific limitation of the dataset that was used in the case study. The data for this study came from a portion of the freeway that has very little recurrent congestion, which made it feasible for the study to clearly attribute instances of vehicle stoppage and queuing to incidents rather than recurrent bottlenecks. The transferability of the methodology needs to be established more firmly with further testing on more congested roadways.

CHAPTER 5.   INCIDENT IMPACT ANALYSIS WITH INCIDENT DELAY ESTIMATION

INTRODUCTION

According to a study by the National Highway Traffic Safety Administration (NHTSA) in 2010, the costs of crashes resulting from delays, increased fuel consumption, and environmental pollution "amount to $28 billion" (Blincoe et al. 2015). Presumably, the economic and social cost of all traffic incidents combined is much more than this figure. However, efficient incident management can help mitigate some of these impacts and, thus, warrants comprehensive analysis. This study explores the delay estimation aspect of incident analysis; specifically, it reviews incident delay estimation methodologies based on queueing and shockwave theory. Recognizing that there have been a number of theoretical examinations of these approaches, the current study aims to analyze the practicability and accuracy of several of these approaches for field use. For this evaluation, a microscopic simulation model of an incident, calibrated to field data, is used as the baseline for comparison, given that it is impractical to directly measure delay in the field to obtain the ground truth. The evaluation is followed by a regression-based predictive delay model that aims to identify important traffic and incident characteristics and aid in resource allocation for incident management.

LITERATURE REVIEW

The topic of incident-induced delay estimation has been studied extensively in the past. The literature shows different methodologies that have been used to estimate delay, one of which is based on the deterministic queueing theory (DQT) approach, which was first

discussed by Moskowitz and Newman (1962). Models based on queueing theory assume a linear arrival and departure curve and estimate the delay based on reduced capacity, demand flow, and incident duration. A typical queueing diagram used to estimate incident-related delay is given in figure 52.
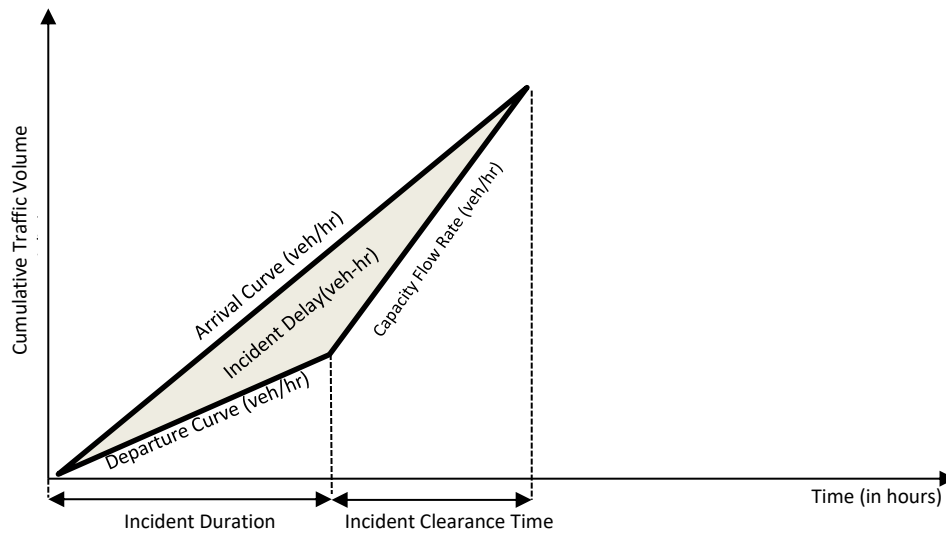
Figure 52. Graph. Typical deterministic queueing diagram.

(Moskowitz and Newman 1962)

This simple deterministic approach was used by Morales (M. Morales 1986) and was also part of a submodel developed by Sullivan (Sullivan Edward 1997) to predict the delay due to an incident for different time distributions of traffic demand. This approach was employed to evaluate the benefits of the incident management program of Oregon's Corridor Management Team (Bertini et al. 2004). A modified approach based on queueing theory was used by Guin et al. for evaluation of the NaviGAtor system in Georgia (Guin et al. 2007). For that evaluation, against the conventional assumption of constant reduced capacity, the study considered dynamic changes in reduced capacity during incident

clearance, and estimated delay for individual incidents rather than for an average incident; formulae used are given below in equations 6 to 9.

$$Delay = \frac{1}{2} * (\mu - \lambda) * \left(t_Q - (t_{R1} + t_{R2})\right)^2 + \frac{1}{2} * (\lambda - \mu_{R1}) * (t_{R1} + t_{R2})^2 +$$
$$\frac{1}{2} * (\mu_{R1} - \mu_{R2}) * (t_{R2}^2) \tag{6}$$

$$t_Q = t_{Q1} + t_{Q2} \tag{7}$$

$$t_{Q1} = t_{R1} * \frac{\mu - \mu_{R1}}{\mu - \lambda} \tag{8}$$

$$t_{Q2} = t_{R2} * \frac{\mu - \mu_{R2}}{\mu - \lambda} \tag{9}$$

where,

t_{R1} = incident duration from start of the incident until time of partial incident recovery (hr)

t_{R2} = incident duration from partial incident recovery until roadway clearance (hr)

t_Q = total time duration in queue (hr)

t_{Q1}= time in queue before partial incident recovery (hr)

t_{Q2}= time in queue after partial incident recovery (hr)

λ = demand (vph)

μ = capacity (vph)

μ_{R1} = reduced capacity until partial incident recovery (vph)

μ_{R2} = reduced capacity from partial incident recovery until roadway clearance (vph)

However, as discussed by Olmstead (1999), since the approach typically involves usage of an *average* incident duration, instead of an actual field-measured value, to estimate the delay, the final delay estimate from the DQT approach can be an underestimate. On the

other side, Skabardonis et al. (1996) noted that usage of an average incident duration value can lead to an overestimation because of the presence of a "few lengthy incidents." Studies like Fu and Rilett (1997) and Sheu (2003) have addressed these concerns of the DQT-based approach by assuming the incident duration as a stochastic variable. Additionally, Fu and Rilett (1997) gave a method to update the probability distribution of incident duration depending on updated real-time traffic information. Still, stochastic modeling of incident characteristics in real time remains a major challenge. Other major concerns of a DQT-based approach include:

- The impractical assumption of linear arrival and departure curves.
- The lack of consideration for effects of traffic diversion (Skabardonis et al. 1996).
- Sensitivity to the choice of the location, as the delay estimates from this approach are measured at a specific location rather than the entire area under the incident's influence.

This last concern has been addressed in previous studies by using a shockwave-based approach to estimate the spatiotemporal extent of an incident and estimate the delay in the form of vehicle-hours lost due to reduction in average speed compared to a predetermined reference speed (Al-Deek et al. 1995; Snelder et al. 2013; Z. Chen, Liu, et al. 2016; Skabardonis et al. 1996). Skabardonis et al. (1996) used this approach to evaluate the benefits of the Freeway Service Patrol (FSP) in Los Angeles, California; formulae used in their study are given in equations 10 and 11. For delay estimation using these equations, the freeway section under the incident's influence was divided into multiple short segments, and delay was estimated for each segment and then summed to obtain the total delay (Skabardonis et al. 1996; Al-Deek et al. 1995). Algorithms based on shockwave theory

100

were first discussed in Messer et al. (1973) and Wirasinghe (1978), where the authors use estimates of shockwave speeds and time–space diagrams to estimate the delay. Al-Deek et al. (1995) and Chen et al. (2016) also sought to segregate the delay due to a primary and a secondary incident, under the assumption of linear shockwaves. Al-Deek et al. (1995) justified the assumption of linear shockwaves by concluding that the assumption can only lead to an overestimation of congestion boundaries, which would increase the computational effort required to estimate delay in real time but would not change the final delay estimate. Chow (1976) compared these two approaches, shockwave-based and queueing theory, and found that under the assumption of time-independent density, both approaches give the same result. Chow (1976) also stated that for time-dependent traffic density, "shockwave analysis has more physical meaning." A glance at the formula (see equation 10) used in Skabardonis et al. (1996) suggests that the methodology is not applicable for an all-lanes-blocked incident, where the current average speed can go to zero. Also, the formula suggests usage of spot speed measurements, which implies the assumption of uniform distribution for instantaneous speeds between two adjacent detector stations. This assumption can be unrealistic for traffic in transition from a stable to an unstable state, such as in the case of an incident.

$$D_{ki} = L_k \frac{t}{60} Q_{ki} \left(\frac{1}{v_{ki}} - \frac{1}{v_{kif}} \right) \text{ for } 0 < v_{ki} < v_{kif} \tag{10}$$

$$D = \sum_{k=1}^{n} \sum_{i=1}^{m} D_{ki} \tag{11}$$

where,

k = freeway segment, k = 1,2,…n

i = time interval, i = 1,2,...m

Q_{ki} = traffic volume (vph) on segment k and time interval i

L_k = length (miles) of freeway segment k

v_{ki} = average travel speed (mph) of segment k during time interval i

v_{kif} = average travel speed under prevailing incident-free conditions of segment k during

   time interval i

t = time (minutes) interval length

D_{ki} = delay (vehicle-hour) perceived on segment k during time interval i

D = total incident delay (vehicle-hour)

Alternatively, Lawson et al. (1997) used the difference in cumulative counts at the bottleneck and an upstream location to estimate the delay. A schematic diagram of this approach is given in figure 53, where A(t) and D(t), respectively, are typical arrival and departure curves during an active bottleneck, and V(t) represents a "virtual arrival curve" constructed by shifting A(t) along the time axis by the magnitude of free-flow travel time. In that study, the authors have also explained the difference in "total time spent in queue" and "total delay," which they say is often "confused in the literature" (Lawson et al. 1997). Delay using this approach can be calculated using equations 12–15. A modified form of this approach was used in Wang et al. (2010) where, instead of a virtual arrival curve as in Lawson et al. (1997), a downstream curve for an incident-free scenario was constructed via a regression technique and then used to estimate the delay. While this approach is easy to implement and is promising for real-time implementation, the viability of this approach is

contingent on available count data quality. Previous studies by Guensler et al. (2013) and Suh et al. (2015) on count data quality, report that data inconsistency rates between video detection systems can be more than 18 percent (Guensler et al. 2013). Such high levels of inconsistency in count data can render the approach infeasible for application.
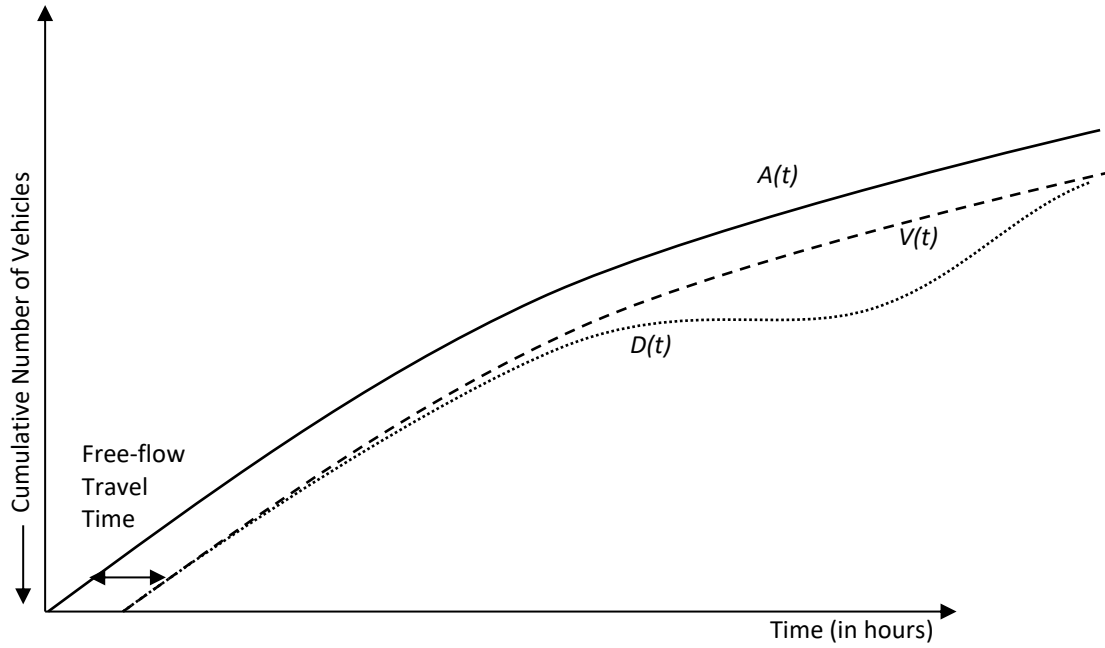
Figure 53. Graph. Typical arrival and departure curves during an active bottleneck.

$$A_{Ti} = \sum_{t=0}^{t=T} A_{ti} \tag{12}$$

$$D_{Ti} = \sum_{t=0}^{t=T} D_{ti}$$

$$TT_{Ti} = (A_{Ti} - D_{Ti}) * \Delta t \tag{13}$$

$$TFT_{Ti} = (D_{Ti} - D_{(T-1)i}) * (FT_i) \tag{14}$$

$$DE_{Ti} = TT_{Ti} - TFT_{Ti} \tag{15}$$

$$TDE_i = \sum_{t=0}^{t=T} DE_{ti}$$

where,

T = current time (or time at which travel time calculation is being made)

t = time interval, $t = 0,1,\ldots,T-1,T$

A_{Ti} = cumulative upstream arrivals on segment i from $t = 0$ to $t = T$

D_{Ti} = cumulative downstream arrivals on segment i from $t = 0$ to $t = T$

TT_{Ti} = total travel time on segment i during time interval T

Δt = length of a time interval

TFT_{Ti} = aggregate total free-flow travel time for segment i over interval $T-1$ to T

FT_i = estimated free-flow travel time for a vehicle over segment i

DE_{Ti} = delay experienced on segment i over interval $T-1$ to T

TDE_i = total delay experience over segment i from time $t = 0$ to time $t = T$

In addition to the above-described macroscopic methods to estimate delay, many studies have used microscopic traffic simulation for incident-related delay estimation. Examples of such studies are:

- Evaluation of The Hoosier Helper Freeway Service Patrol (Latoski Steven et al. 1999).

- Evaluation of incident response team by Carson et al. (Carson et al. 1999).

- Other studies, such as Birst and Smadi (1999), Gillen (2001), and Khattak and Rouphail (2004).

Approaches based on traffic simulation provide valuable insights for an individual incident but are impractical when it comes to analyzing a large number of incidents with a wide

variation of ambient conditions. Nevertheless, due to lack of ground truth, microscopic

simulation has been used in this study for baseline comparison.

DETECTOR DATA INCONSISTENCY

As part of the current study, the data inconsistency as noted by Guenseler et al. (2013) and

Suh et al. (2015) was further investigated. For example, traffic count and speed data on a

1-mile section of the I-285 northbound freeway corridor (at approximately mile marker 43)

in Atlanta, Georgia, was examined. Three adjacent VDS stations within this roadway

section, without any intermediate corridor access or egress points, are shown in figure 54,

namely: GDOT-STN-2851970, GDOT-STN-2851971, and GDOT-STN-2851972, listed in

order of direction of travel. For this site, a general pattern has been observed where GDOT-

STN-2851971 and GDOT-STN-2851970, respectively, give an over- and undercount

estimate compared to GDOT-STN-2851972. The pattern can also be observed for a typical

day (04/30/2018) in figure 55. This data inconsistency is seen to be further exacerbated

during incident and inclement-weather conditions. This issue is illustrated in the cumulative

count curves for the three stations on one such incident day (04/23/2018) in figure 56.

There were four incidents on that day, two of which occurred during the morning peak

hour, and the others were detected at approximately 4:30 PM and 5:00 PM, respectively.

The evening peak hour incidents were detected at locations immediately downstream of

station GDOT-STN-2851972. From figure 56, it can be observed that the order of count

differences between the three stations over the duration of 24 hours may be conservatively

estimated as approximately 8,000 vehicles. Furthermore, the cumulative curves are not

sequential according to expectations; GDOT-STN-2851970's curve is lying below both

GDOT-STN-2851971 and GDOT-STN-2851972. With such differences it begins to become apparent how these errors can significantly impact any delay estimations.

As discussed in previous studies (Suh et al. 2015; Bonneson and Abbas 2002; Martin et al. 2004; Rhodes et al. 2006). Such variability in data quality across locations and over time makes it difficult to rectify these inconsistencies using a "uniform correction factor" (Suh et al. 2015). Usage of a calibration factor, even when one is created specific to each site, makes the calculated delay highly sensitive to even small calibration errors and limits the ability to estimate delay in real time. More importantly, with the error level depending on the occlusion, which is a function of the density, a single calibration factor is not valid throughout the day; this makes the development of the calibration factors quite arduous, as well.
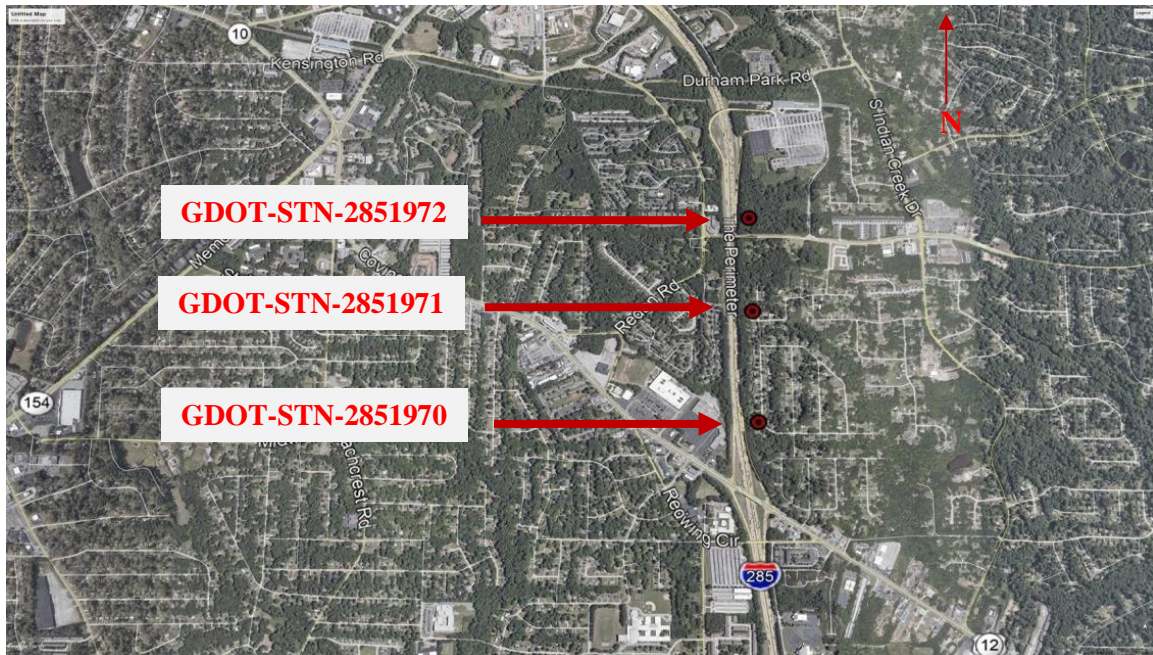
Figure 54. Map. GDOT detection station camera locations on one-mile section of the I-285 northbound freeway corridor (at approximately mile marker 43), Atlanta, Georgia. Source: Google Earth
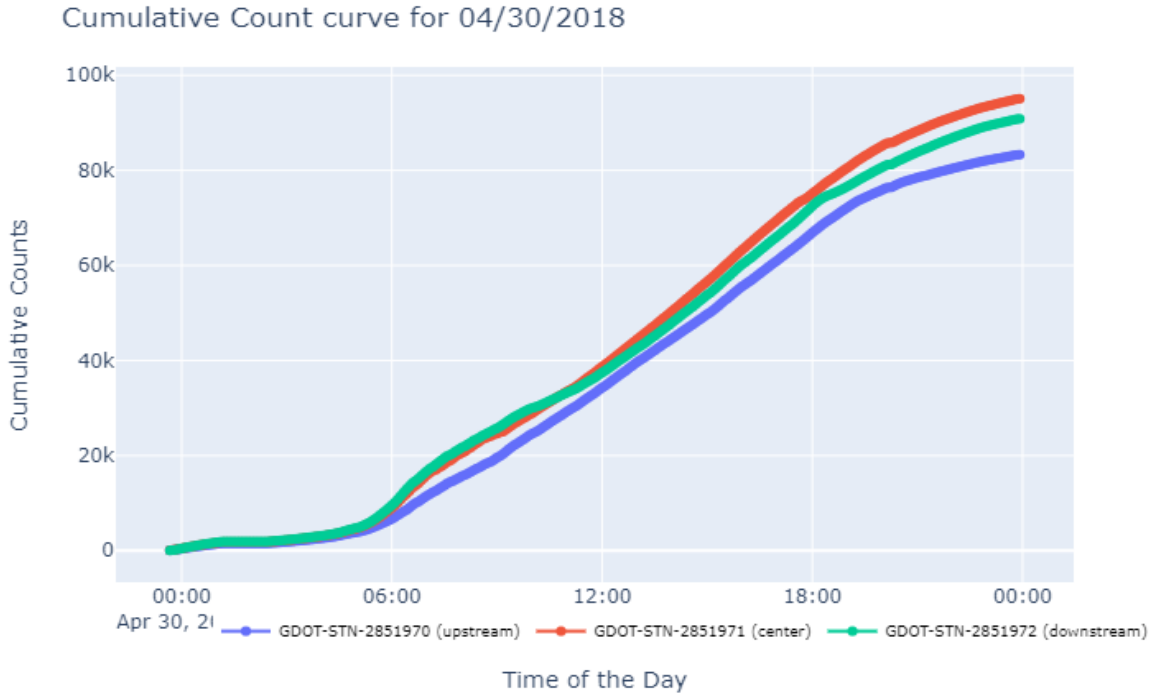
Figure 55. Graph. Cumulative count curve for a typical day (04/30/2018) at the site.
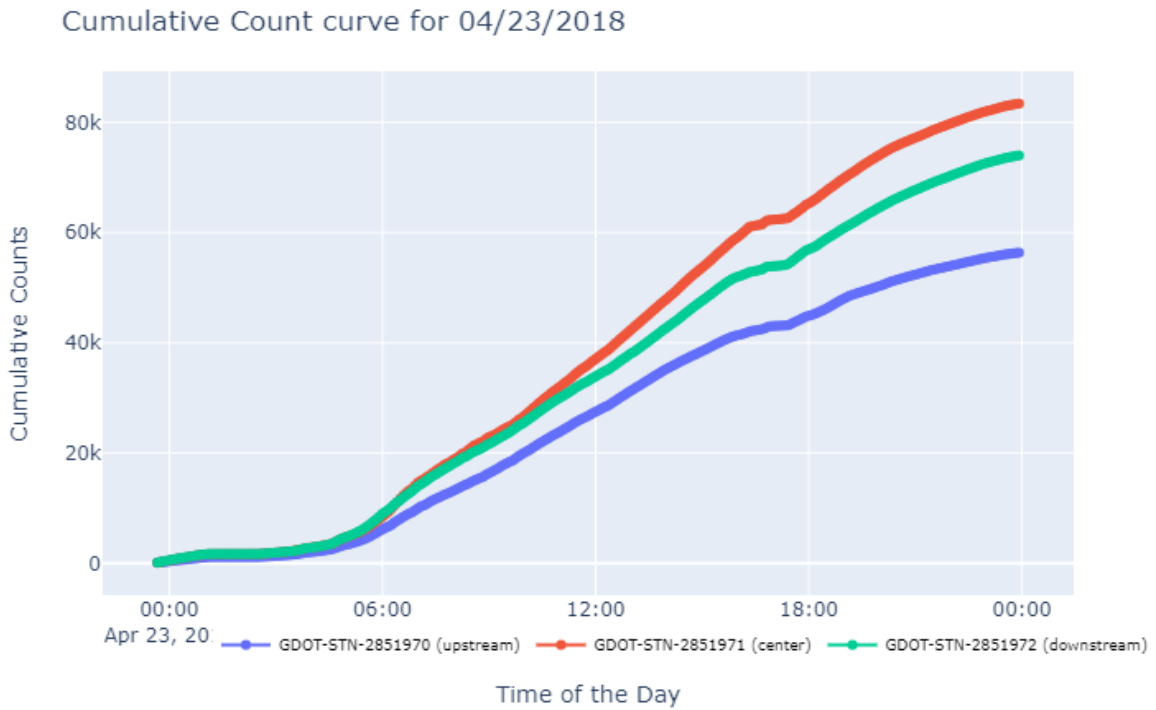
Figure 56. Graph. Cumulative count curve for an incident day (04/23/2018) at the site.

COMPARISON

For comparison of delay estimation methodologies, a Vissim® model was calibrated to simulate a real-life incident. The calibration was more notional than rigid in the sense that the throughput and capacity drops during the incident were reflected closely, but the demand used was based on flows from a typical day, instead of the demand on the incident day, which was reduced due to diversions from upstream ramps. Once calibrated, a comparative analysis between the queuing models and the simulated results for the range of demands and incident lengths was performed.

Incident data were obtained from NaviGAtor, the advanced traffic management system owned and operated by the Georgia Department of Transportation (Wells 2016). The incident occurred at I-285 and SR-6 Camp Creek Parkway, as shown in figure 57 (a), in the southbound direction, on April 19, 2018. According to the NaviGAtor incident logs, the incident was detected at 4:15 PM; however, the data show a continuous drop in average speed after 3:55 PM, as can be observed from figure 57 (b). This suggests that the incident occurred around 3:55 PM and was not reported (or recorded) until 4:15 PM. According to the incident logs, the roadway was cleared at 7:33 PM and the shoulder was cleared in another 32 minutes, which makes the incident duration equal to 250 minutes, using the start time as 3:55 PM. At the start of the incident, all lanes were blocked and after approximately 119 minutes one lane was opened (closest to the median). The developed model simulated an approximately 5-mile section along I-285. A 50-mile-long upstream freeway corridor consisting of four lanes without any on/off ramps was appended to the 5-mile section to ensure vehicles could queue within the model network, allowing for their reflection in the delay estimation. It is recognized that such queuing would not occur in the field due to

dynamic rerouting of vehicles to avoid the incident-related delay; however, the intent of this analysis to compare delay estimation approaches requires simplifying out confounding factors such as rerouting. Similarly, the corridor Vissim® empirical distribution for desired speed was set with a lower and upper bound equal to 69.90 and 70 mph, respectively, reflecting the observed freeway free-flow speed. While this is likely a narrower desired speed band than that in the field, this was chosen to allow for a more direct comparison to the queuing model, eliminating desired speed variability as a confounding variable, as Vissim® defines delay as the difference between desired and actual travel time (PTV 2018a). The travel time in Vissim® is recorded through vehicle travel-time segments and data collection points, which were set up every one-third mile. The length of a travel-time measurement segment represents the typical distance between adjacent VDS stations in the NaviGAtor system (Guensler et al. 2013). Data collection frequency was set at 1 minute for this model.

(a)
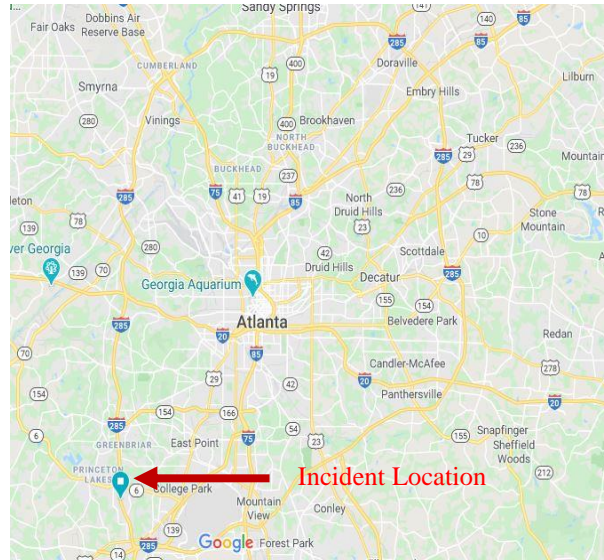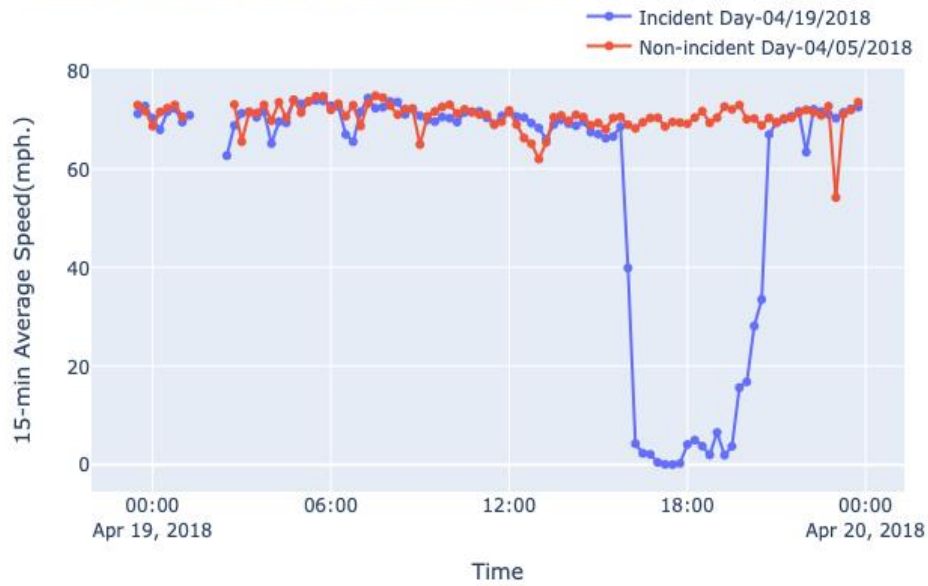
(b)

Figure 57. Map and graph. (a) Incident location for event ID 1164269 on 04/19/2018 detected at 4:15 PM (Source: Google Earth), (b) average speed (in mph) during incident day vs. non-incident day.

The corridor was first calibrated for a saturation flow equal to observed throughput at the incident location, following the procedure described in Hunter et al. (Hunter et al. 2017). The vehicle input for the corridor was set according to the throughput curve observed at the incident location for an incident-free day, April 5, 2018, which was the same day of the week but 2 weeks prior. The incident was simulated by creating a bottleneck using parking lots and parking routes, with parking duration assigned to each parking lot according to the block duration of that particular lane (PTV 2018b).

The timeline of the simulation can be traced through figure 58 and figure 59, which show the field observed and simulated aggregated vehicle counts and average speed (mph) for the incident and incident-free day, respectively, at the incident location. Additionally, for comparison, Figure 58 and figure 59 also show count and average speed from a station near the input location of the Vissim® network. The incident was initiated in Vissim® after a warmup period of 1.67 hr and then closely followed the incident timeline obtained from the NaviGAtor incident logs. Field data at the incident location showed lower throughput than the expected capacity when only a single lane was open during the interval from partial incident recovery to complete clearance, which indicates the occurrence of some rubbernecking, lane shifting, partial blockage of the "open" lane, or other effects during that interval. To simulate this reduced capacity, the headway distribution of the upstream link was changed for this interval. However, reviewing figure 58 and figure 59, at approximately 6:00 PM, it is seen that the model still somewhat overestimates queue discharge, although speed is well matched. Assuming the same arriving traffic as on the non-incident day (i.e., no rerouting), the experiment showed that free flow was restored 6.25 hr after the incident had been cleared and the back of the queue reached about

32.67 miles upstream of the incident location before dissipating completely. As is clear

from figure 59, free-flow speed at the site itself was restored at around 8:45 PM. From the

lack of difference between the throughput curves beyond 8:45 PM for the incident day and

the non-incident day in figure 58, it is clear that significant diversion did occur. Thus, the

Vissim® delay is much higher than the delay experienced on the freeway. However, during

peak periods, most roadways are already operating near their capacity and the traffic

diversion process effectively distributes the delay over more vehicles and alternate

facilities. Residual capacity in the network used by the diversion paths would mitigate

some part of the delay; however, in many instances the rerouting serves to move the delay

from the freeway facility to alternate facilities, rather than eliminate the delay.
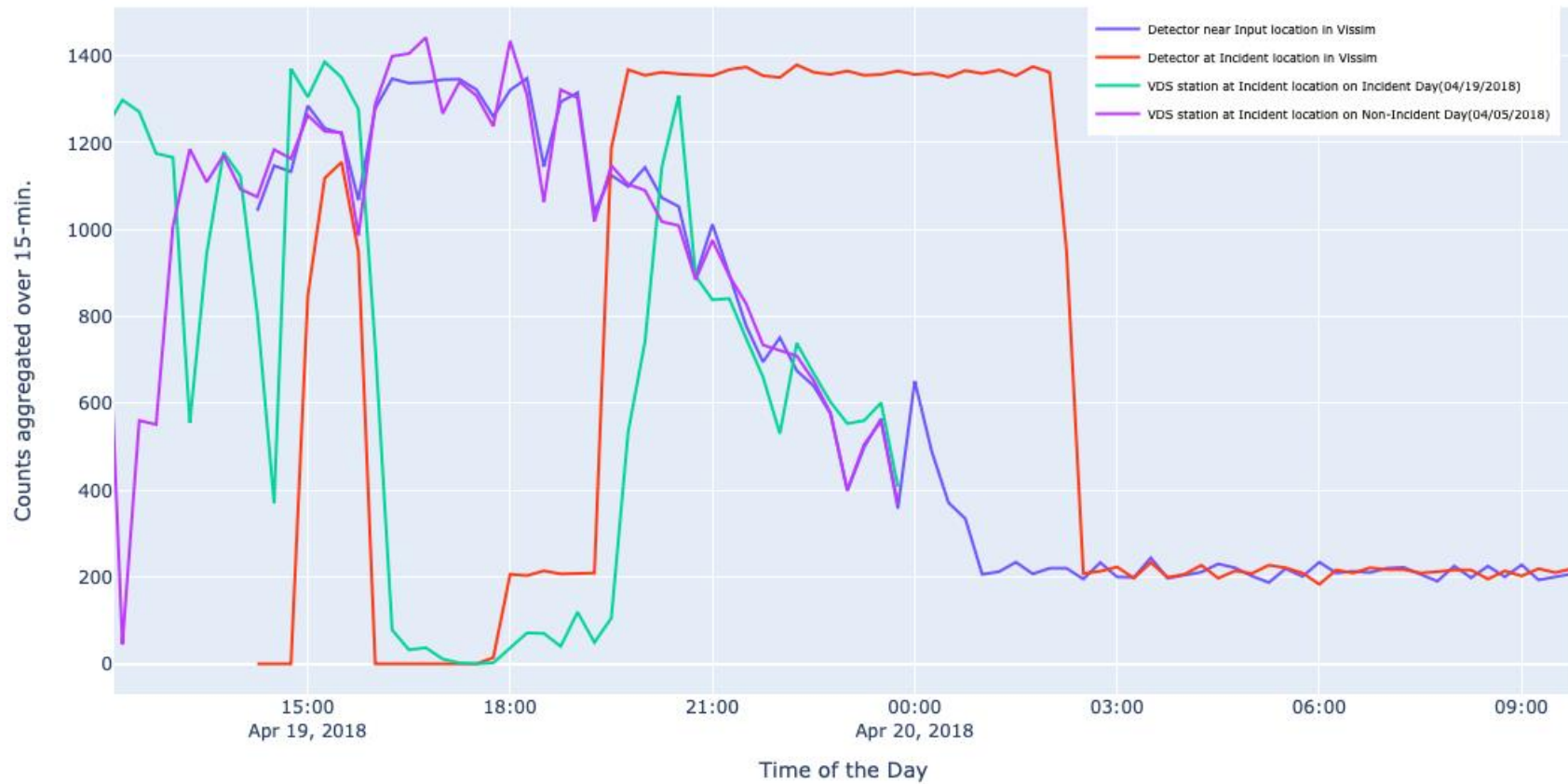
Figure 58. Graph. Vehicle count (aggregated over 15 min) of the Vissim® model and the VDS station at incident location for the incident day and incident-free day.

Figure 59. Graph. Average per lane speed (mph) of the Vissim® model and the VDS station at incident location for the incident day and incident-free day.

Delay Estimation Results

To compare the delay estimates from Vissim® to that from queueing theory, formulae described in Guin et al. were used, as they account for partial incident recovery (Guin et al. 2007) These formulae are provided in equations 6-9 for ease of reference. For this equation, t_{R1}, t_{R2}, were assigned the values of 1.98 hr, 1.6 hr, and 0 vph, respectively, and μ and μ_{R2} were estimated using queue discharge flow from the station in the simulation located immediately downstream to the incident location. Demand was estimated using average throughput observed at the station where the back of the queue was traced in the model. Typically, there are challenges with using the throughput as a reflection of demand, as observed throughput is generally constrained by recurrent congestion for implementation of the approach with field data. However, that is not necessarily a major concern for this particular application case where the removal of the recurrent congestion-related delay from the measured delay would be desirable to help isolate the incident-related delay. Nonetheless, at this site, the non-incident day experienced free-flow speeds over the period of interest, and it would be safe to assume that the throughput was an accurate reflection of the total demand, assuming zero diversion.

As discussed in the Literature Review section, two variants of the incident delay estimation methodologies based on shockwave theory—one based on current average speed and another on cumulative counts—were also tested. These variants, respectively, will henceforth be referred to as the approaches based on 'difference-in-speed' and 'difference-in-cumulative count' for the remainder of this chapter. For testing difference-in-speed and difference-in-cumulative counts, equations 10–11 and equations 12–14 were used,

115

respectively. Figure 60 shows a time series of the calculated delays using the difference-in-speed and difference-in-cumulative count approaches for one of the ⅓-mile-long segments. The horizontal line observed for delay estimated using the difference-in-cumulative count approach corresponds to the delay experienced during complete blockage of the roadway; the difference-in-speed approach fails to capture any delay during the same period since no vehicles were crossing the detection points. The results of this comparison are shown in Figure 61 and Figure 62, which, respectively, show the average of total delay estimate from three replications in vehicle-hours and the spatial distribution of this delay for each segment for one of the replications.

Using delay estimates from Vissim® as the baseline, the approaches based on difference-in-speed, queueing theory, and difference-in-cumulative counts gave average differences of about 87, 45, and, 0.04 percent, respectively, for three replications. The severe underestimation by the difference-in-speed approach can be attributed to the application not meeting the underlying assumption of this approach. This approach assumes the speeds throughout the segment between consecutive detection points are the same, and equal to the speed at the detection point; i.e., the traffic is homogeneous. Additionally, the average travel speed must be sufficient for a vehicle to traverse a time segment within a single time interval. However, under severe congestion, as in the given case study, with speeds at or approaching zero for extended periods, this condition is not met. The underestimation for the delay estimated using the queueing theory approach is predominately a result of the assumption of a single constant arrival rate. This assumption fails to capture the variability in demand or, in this case, throughput. As is known, the magnitude of underestimation could be reduced (potentially significantly) by utilizing a series of arrival rates to better fit

116

observations. As for the underestimation associated with the difference-in-cumulative counts approach, it can partially be explained by the difference in discharge rates before and after the incident. Presumably, this error could be reduced by collecting traffic counts more frequently, but this assertion needs further study.
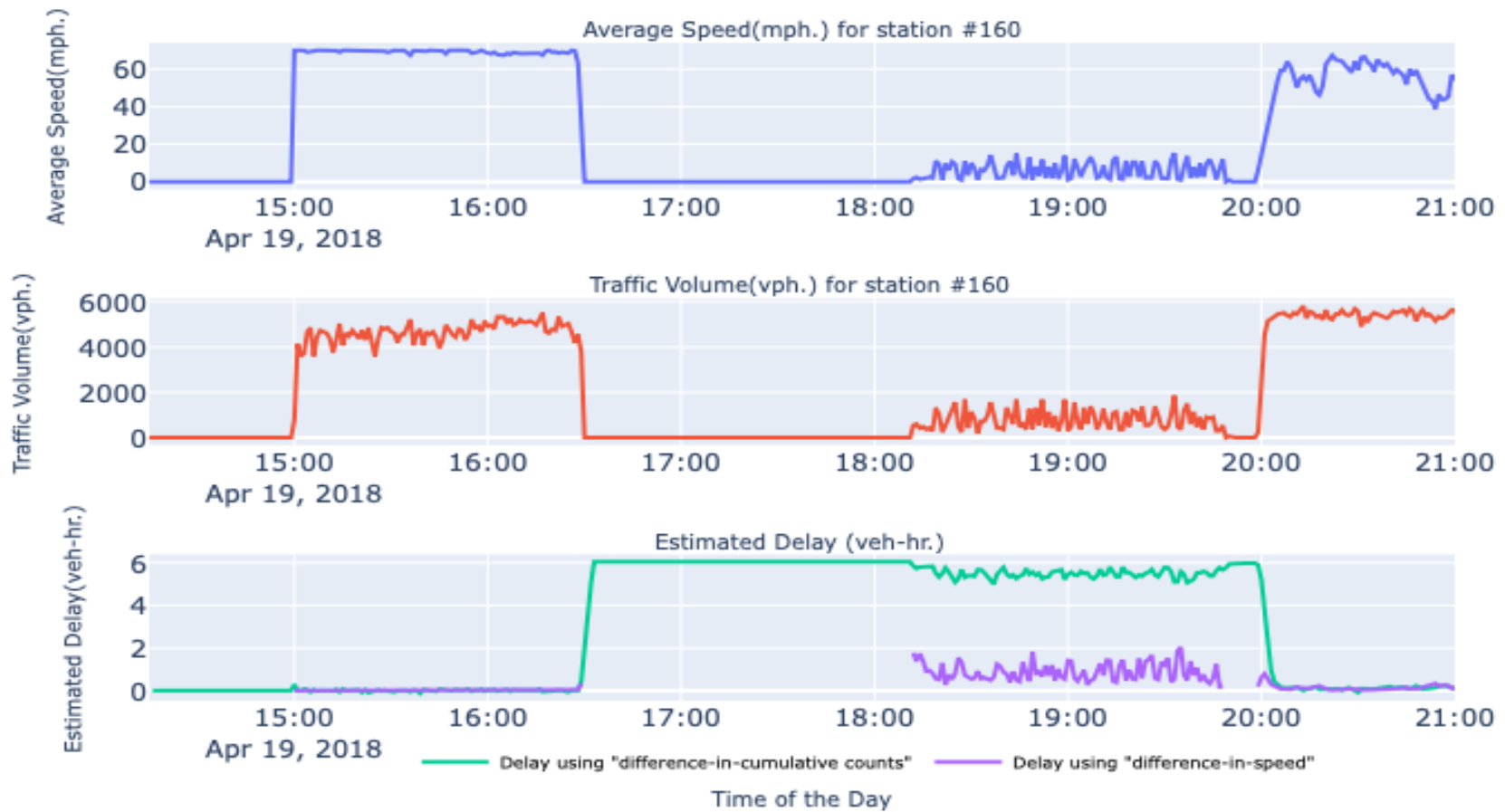
Figure 60. Graphs. Delay (veh-hr) estimation using 'difference-in-cumulative counts' and 'difference-in-speed' approaches for station #160.
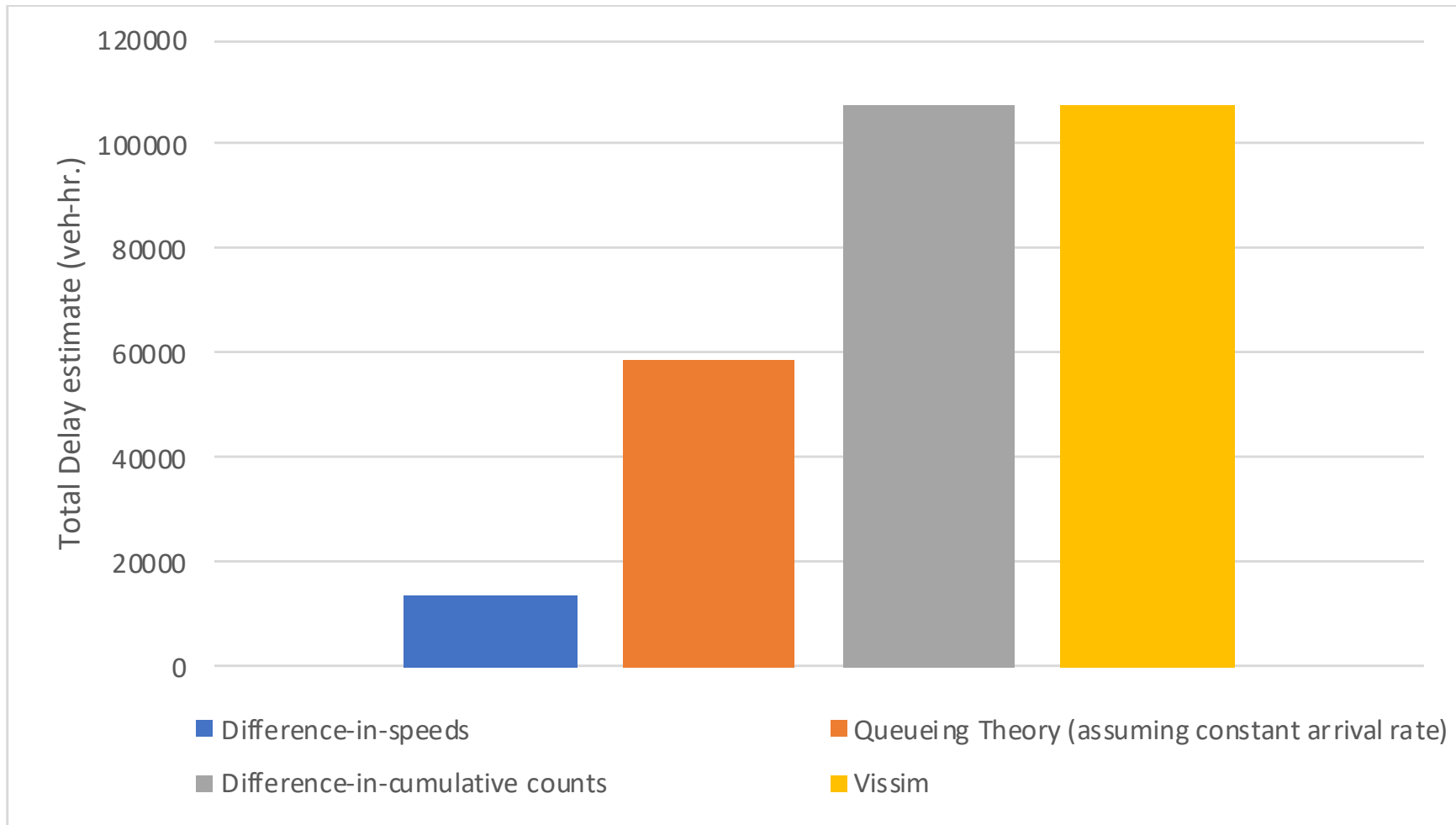
Figure 61. Graph. Total delay (veh-hr) estimated using different estimation methods.
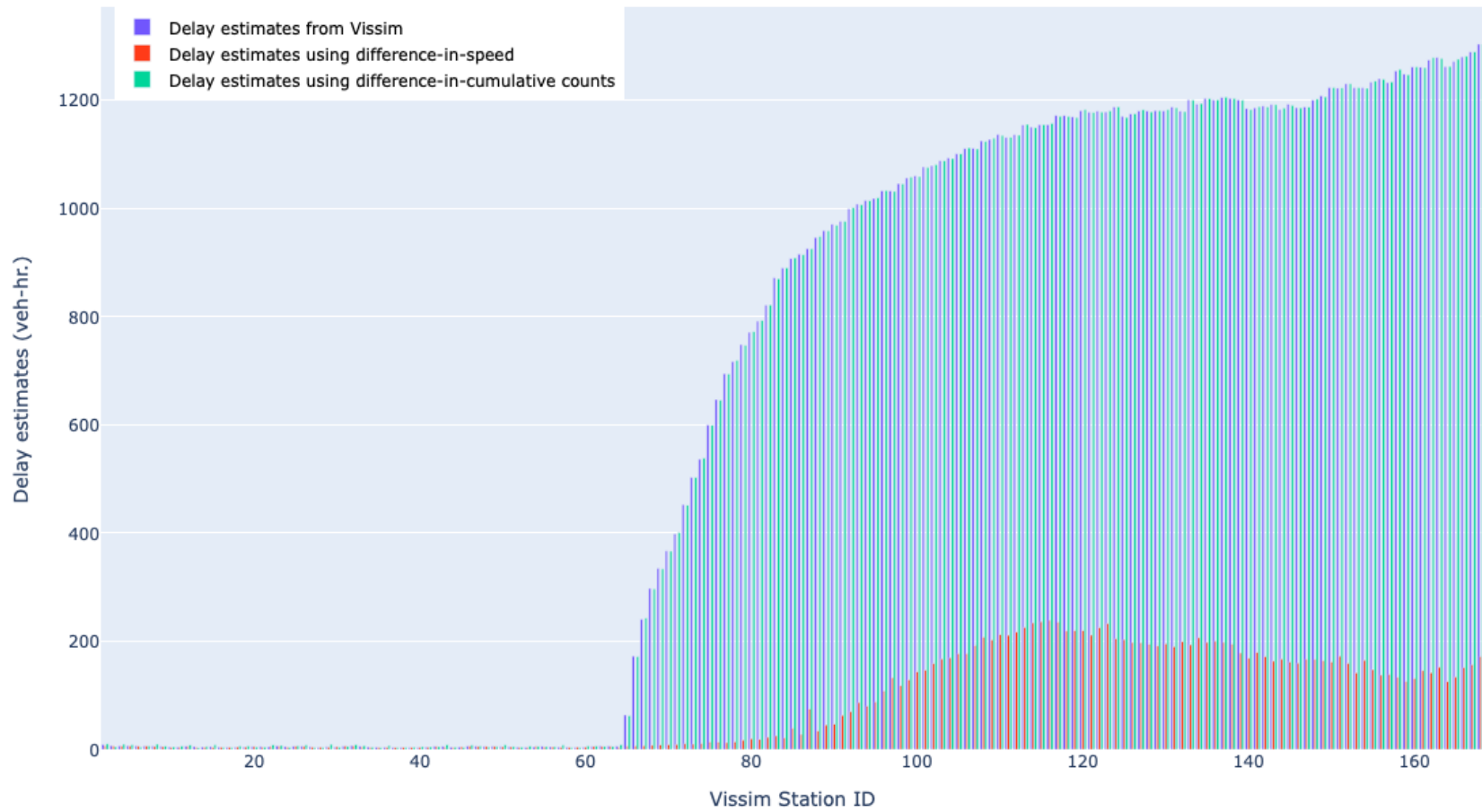
Figure 62. Graph. Spatial distribution of delay (veh-hr) estimated from different methods.

REGRESSION MODEL

The above validation experiment exhibited that delay estimates from commonly used approaches are likely to be erroneous even with fairly well controlled variability in a simulation environment. The most easy-to-use method based on spot speeds yields the largest errors. The best estimates are generated by the cumulative-count method; however, the weakness of this method is that it relies on the counts to be very precise, to the extent that conservation of vehicles can be validated. However, results presented in figure 5 show that such an expectation is very likely unrealistic under field conditions, especially where the resources available for maintaining a large system of detectors are understandably constrained.

This study has, therefore, developed a regression-based predictive model for incident delay to rapidly obtain delay estimates for incidents with varying characteristics occurring under different base conditions. The model can be helpful for resource allocation during incident management, especially in the case of poor data quality. Previously, some studies have developed regression models to estimate incident delay. For example, Garib et al. (1997) developed two regression models to predict incident delay. These models were developed based on relevant incident, traffic, weather, and geometric characteristics, and delay estimates calculated based on equations 10–11. However, we have seen that the incident delay estimates generated by that approach could have a high degree of error due to heterogeneity in traffic. For the current study, therefore, data produced from simulation of different incident scenarios have been used. While the simulation might not be completely reflective of field conditions, such as drivers using shoulder lanes to bypass queues, or other geometric conditions specific to an incident site, the results do not suffer from the

field data measurement errors and can be demonstrated to be fairly closely reflective of "typical" conditions. The simulation scenarios were produced using a combination of independent variables, namely, throughput demand (in vehicle/hr/lane), total number of lanes available to travel in a direction, number of lanes blocked, and incident duration. An earlier study by Hadi et al. (2007) had found incident location to be an insignificant variable for "reduction in capacity" for Vissim® simulations. Therefore, incident location was not used as a variable in this model. The analysis considered all independent variables as deterministic.

Since simulation of an exhaustive set of incident scenarios considering all possible values for the independent variables was impractical, a limited set of values was considered. The selected values were based on experience and can be found in table 18. Different combinations of these variables gave a total of 108 scenarios. For statistical testing, 10 replications for each scenario were generated, making the total number of simulations equal to 1,080. Incident scenarios were automatically generated by enumerating all scenarios and iterating over them using the COM interface with Python®. The underlying model used for these scenarios had a straight corridor that was approximately 30-mile-long without any on/off ramps for traffic diversion and was assigned a default driver behavior. With these constraints and the given set of traffic characteristics, the simulation gave an upper bound for real-life total incident-induced delay.

Further elaborating on calculation of the dependent variable, total incident-induced delay can be understood as the product of two variables, average delay per vehicle (in hours) and total number of vehicles that experienced delay. Out of these variables, the total number of vehicles that experience the incident delay largely is a function of demand or, in this case,

throughput volume of the road. This makes it difficult to isolate the effect of other independent variables on total number of vehicles affected by an incident. So, for this study, a relationship between average delay per vehicle, summed over spatial extent of the incident, and the above-described independent variables is studied. The dependent variable used in this study was calculated using equation 16.

$$Average\ Delay\ per\ vehicle\ (hr) = \sum_{j=1}^{k} \frac{\sum_{i=1}^{n_j} c_{ij} d_{ij}}{\sum_{i=1}^{n_j} c_{ij}} \tag{16}$$

where,

i = time interval (1 min)

j = travel segment (⅓-mile long)

k = total number of segments upstream of the incident

n_j = total number of time intervals for which incident's impact was experienced for every j^{th} segment

c_{ij} = number of vehicles that have crossed the jth segment during i^{th} time interval

d_{ij} = average delay per vehicle (in hr) reported by Vissim® in i^{th} time interval for j^{th} segment

Table 18. Variables for different incident scenarios.

Variable Type	Values
Demand (veh/hr/lane)	High (1700), Medium (1100), and Low (500)
Total number of lanes	4, 6, and 8
Number of lanes blocked	1, 2, 3, and 4
Incident duration (in minutes)	15, 30, and 60

Before carrying out the modeling exercise, the dependent variable was examined for validity of the normality assumption and was transformed using a logarithmic function. A plot of dependent versus independent variables is given in figure 63. The figure shows the total average delay per vehicle for different demand conditions, incident durations, and number of lanes blocked, with a total of four lanes for travel in a direction. The exploratory analysis of the dependent and independent variables revealed different underlying behaviors for different degrees of saturation. The exploratory analysis indicated that the ratio of number of lanes blocked to total number of lanes would be useful as an independent variable in the regression model. This ratio is hereafter referred to as 'severity'. For the regression model, severity was treated as a continuous predictor. To represent the difference in behavior, two separate regression models were developed, divided based on a predetermined level of degree of saturation equal to 0.95. The model was further divided based on residual capacity level. A representation of the model is given in figure 64, and the final model parameters are given in table 19.
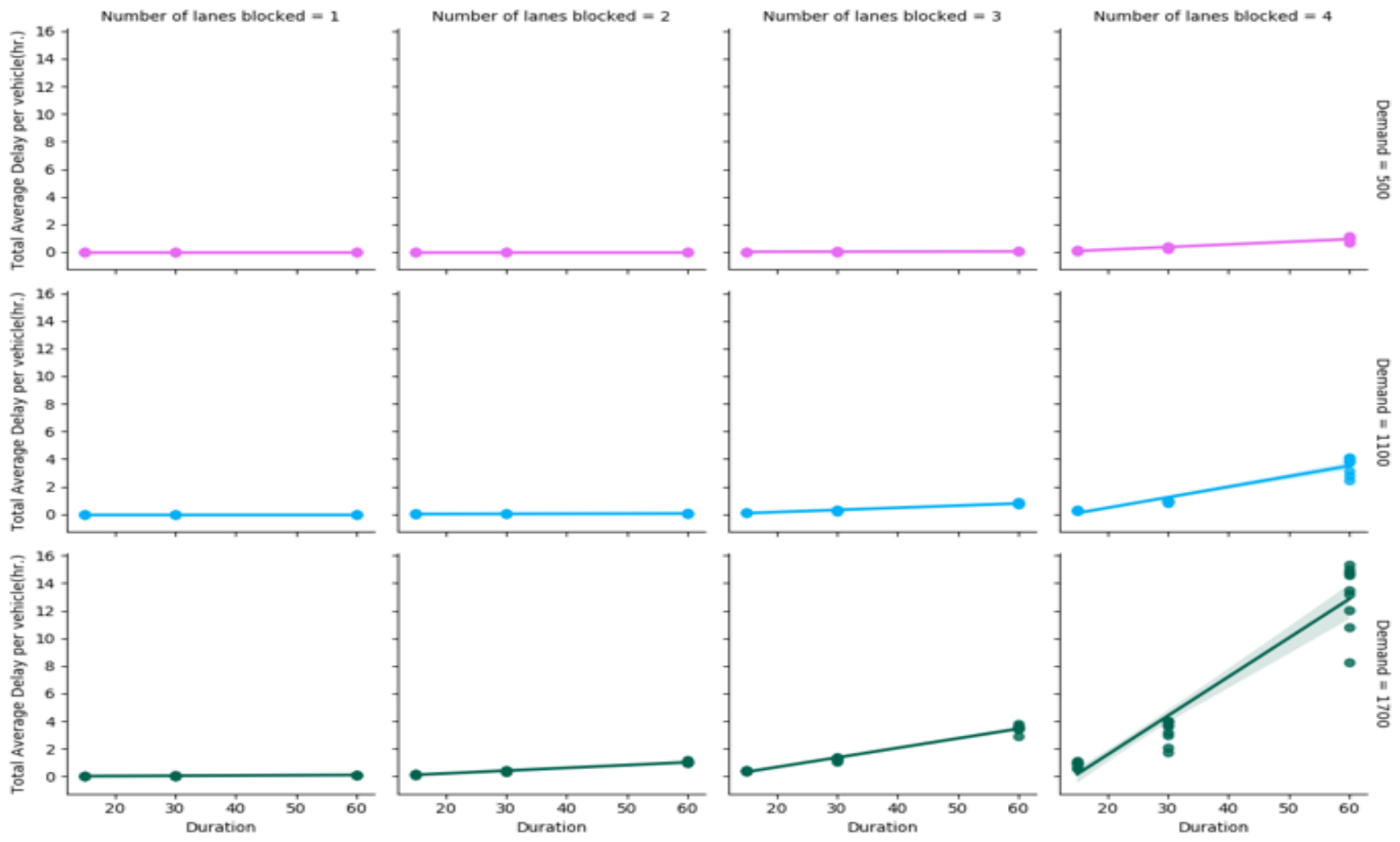
Figure 63. Graphs. Total average delay per vehicle (in hr) vs. demand (veh/hr/lane) and incident duration (in min) for total of four lanes.

$$Ln(Average\ delay\ per\ vehicle(hr))$$

Residual Capacity=0

Residual Capacity>0

$$\frac{Volume}{Capacity} \leq 0.95$$

$$\frac{Volume}{Capacity} > 0.95$$

$$\alpha_1 + \beta_1 * Demand + \varepsilon_1 * Duration$$

$$\alpha_2 + \beta_2 * Demand + \gamma_2 * \frac{Number\ of\ lanes\ blocked}{Total\ number\ of\ lanes} * + \varepsilon_2 * Duration$$

$$\alpha_3 + \beta_3 * Demand + \gamma_3 * \frac{Number\ of\ lanes\ blocked}{Total\ number\ of\ lanes} * + \varepsilon_3 * Duration$$
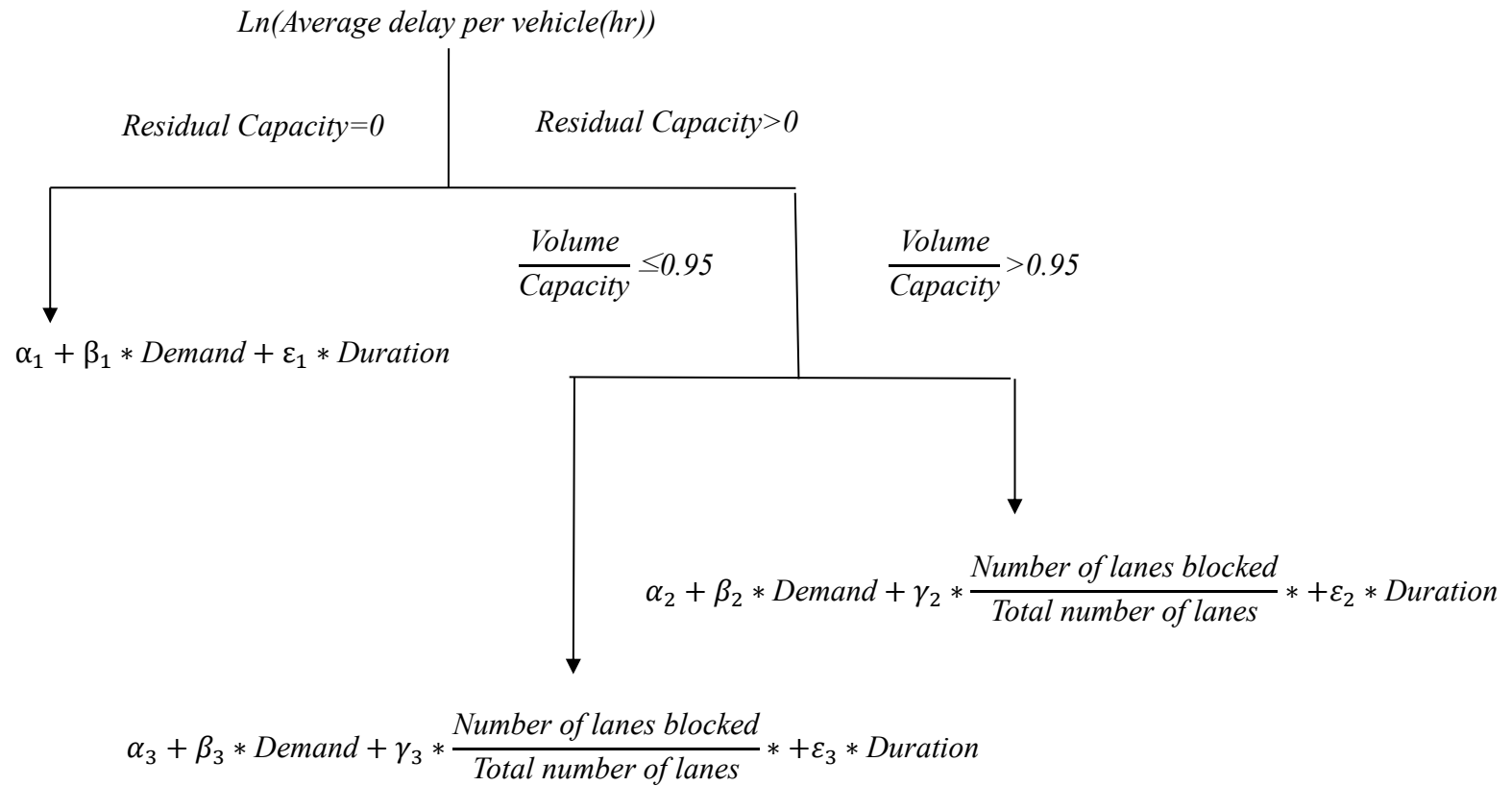
Figure 64. Model. Tree regression model for the dependent variable.

Table 19. Model coefficients for tree-based regression.

Variable	Coefficient	Standard Error	t-statistic	*p*-value
Residual Capacity = 0	Adjusted R^2: 0.95			Sample Size: 180
Constant	−4.101	0.073	−56.284	0.000
Demand	0.002	0.000	41.246	0.000
Duration	0.053	0.001	42.215	0.000
($\frac{Volume}{Capacity}$)>0.95, Residual Capacity ≠ 0	Adjusted R^2: 0.90			Sample Size: 360
Constant	−12.593	0.205	−61.564	0.000
Severity	8.097	0.164	49.427	0.000
Demand	0.003	0.000	34.990	0.000
Duration	0.036	0.001	26.433	0.000
($\frac{Volume}{Capacity}$)≤0.95, Residual Capacity≠ 0	Adjusted R^2: 0.77			Sample Size: 540
Constant	−9.389	0.098	−96.190	0.000
Severity	6.055	0.144	41.920	0.000
Demand	0.001	0.000	23.227	0.000
Duration*	0.003	0.001	2.140	0.033*

*Insignificant at 1 percent level

The adjusted R^2 for the three models, operating in the three residual capacity regimes, are 0.95, 0.90, and 0.77, respectively. Coefficients for all independent variables, including the constant term, are significant at a 5 percent significance level. However, at a 1 percent significance level, incident duration becomes an insignificant variable in the model for low degrees of saturation, indicating that it is not an important factor in prediction of delay per vehicle under such conditions. The signs of coefficients for severity, demand, and duration

are positive, implying that these variables contribute to increasing the value of average

delay per vehicle, which is in line with the formulae used under the queueing and

shockwave theories. These signs intuitively make sense, but the reader should be cognizant

of the log transformation while quantifying the impact of change in an independent variable

on change in average delay. The overall model structure implies that at a high degree of

flow saturation, a higher percentage of the estimated delay can be explained in terms of

incident severity, traffic demand at the time of the incident, and incident duration than at a

low degree of saturation. A likely explanation of the observation is that at a low degree of

saturation, vehicles mainly experience delay due to the weaving that is required in the

proximity of the incident location, while at a high degree of saturation the main source is

capacity constraint.

The residual analysis in the form of residual vs. fitted values for the current model structure

is given in figure 65. The figure justifies the underlying assumption of homoscedasticity for

regression-based modeling. It is to be noted that while the residuals seem to be large, with a

sizeable deviation from zero, on the left half of the plots, these residuals correspond to very

small values of the delays. For example, the delay value corresponding to a fitted value of

-2 would be $^1/_{100}$ of a vehicle-hour, i.e., 36 seconds of delay per vehicle. Since the impact

of such small values of delay on incident management decisions is minor, the lack of

accuracy in this region is not a critical failure for the model. The negative coefficients for

the intercept of the model indicate that the unobserved influences have a negative impact

on the dependent variable. These unobserved influences could be due to cooperative lane

changing, simulated braking behavior, or other aspects of a simulated merging behavior.

Future efforts will further explore this issue. Additionally, one can also test the influence of

interaction terms of the independent variables, but the model is kept relatively simple for field use.

(a)

(b)

Residual versus Fitted Values for Queue growth
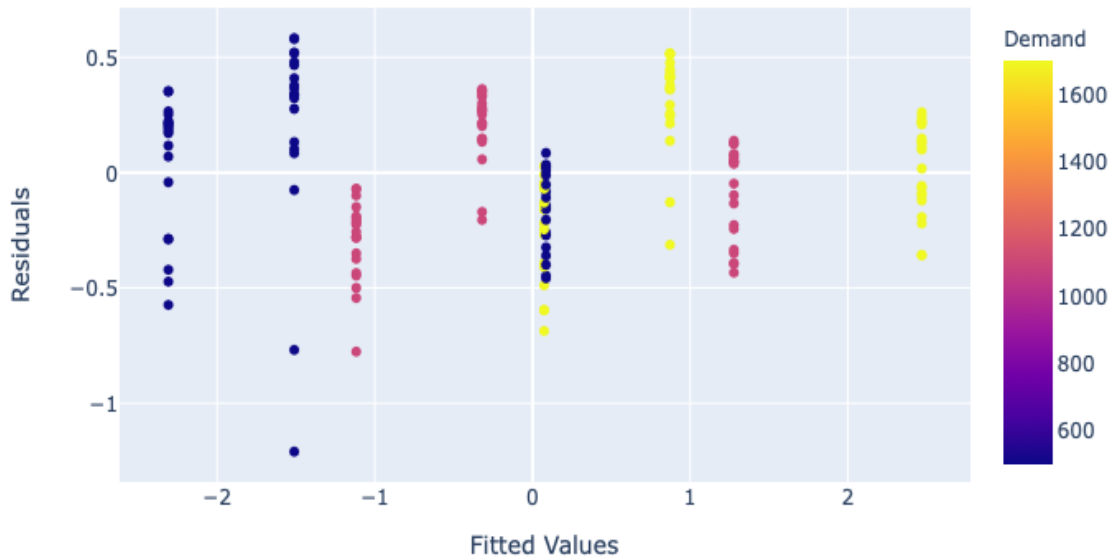
(c)

**Figure 65. Plots. Residual vs. fitted values for models with: (a) residual capacity ≠ 0
and *volume ÷ capacity* ≤ 0.95, (b) residual capacity ≠ 0 and *volume ÷ capacity* > 0.95,
(c) residual capacity = 0.**

DISCUSSION

This study develops a simulation-based regression model approach for rapid estimation of

incident delay. The need for a simulation-based approach is validated by a two-pronged

approach:

1. An analysis of the field data that serve as the inputs to the conventional models is

    performed to demonstrate the data quality issues that would prevent a theoretically

    valid method, such as the cumulative count approach, from producing accurate

    delay estimates. The persistent errors in the field data often create directional biases

in the cumulative counts and violate the conservation of vehicles assumption, which naturally leads to errors in the incident delay estimate.

2. A comparative analysis of the conventional delay-estimation methods within the framework of a simulated case study demonstrates how the violation of the homogeneity assumption inherent to the speed-based methods causes a severe underestimation of the incident delays. While the method is theoretically valid, the limitations on the application using spot speeds from fixed-point detectors make the application too simplistic to represent the dynamics of unstable traffic.

The analysis also highlighted the need for considering alternative data sources for incident delay estimation such as probe-based travel time data or more emerging sources like data from connected vehicle (CV) technology. The advancement of CV technology suggests a trend of moving away from aggregate volume counts and speed estimates to individual travel time and speed estimates for incident analysis. However, this supposition is somewhat premature in the sense that currently, most of these technologies lack sufficient coverage to be able to analyze incidents with varying characteristics. In the meantime, the simulation-based model developed in this study would be a valuable addition to the practitioners' toolbox for rapid, yet conservative delay estimation.

While the models gave excellent results in terms of the model's goodness of fit, there are some limitations of the current approach. Firstly, the demand variable here was treated as a deterministic time invariant variable, which limits the ability to study the effect of incident occurrence on vehicle delay with respect to current and future demand from a typical time variant demand curve. Secondly, the simulation scenarios used in the model development did not model the rubbernecking effect that is typically observed in the field, which could

131

lead to an underestimate. Lastly, a microscopic simulation-based approach makes it difficult to take traffic diversion into account; large-scale dynamic traffic assignment (DTA) models are better suited to study the effect of diversion on delay. Although this results in a possible overestimate of the delays, especially if the remaining roadways had excess capacity to absorb the diverted traffic, from an incident management perspective, a conservative delay estimate might be desirable. While traffic diversion may reduce delay per vehicle on the subject corridor, it might not necessarily help in overall network delay mitigation, especially during peak traffic conditions.

Future work for this research would include comparison of delay estimation methods against ground truth, using travel time measurements. Also, demand-estimation methods currently used in delay estimation can be improved further and account for recurrent congestion. The case study presented in this research only analyzed the influence of a single incident, and traffic dynamics and delay estimation resulting from secondary or multiple closely located incidents still remain to be explored.

CHAPTER 6.   CONCLUSIONS AND RECOMMENDATIONS

DISCUSSION

This project essentially performed four closely related studies. Chapter 2 presented an accuracy evaluation of a vehicle detection technology. Chapter 3 presented an evaluation of the feasibility of using crowdsourced smartphone application–based incident detection for reducing incident detection times. Chapter 4 presented the evaluation of an AID technology and the use of that AID technology in improving incident management. Chapter 4 also presented the machine learning–based methodology that was developed for use on top of a base AID algorithm to enable automated identification of potential high-impact incidents. Chapter 5 presented the development of a method to quantify the impact of incidents in terms of vehicle delay to lay the foundations of automated decision support for real-time management of emergency response resources.

Results of the accuracy evaluation of the vehicle detection technology revealed that the count and speed measurements are highly accurate, with less than 2 percent error under normal circumstances. The errors in vehicle classification were in the range of 6–7 percent under these conditions, which is typically considered acceptable for most applications. The count errors, however, increase significantly with a downward bias, i.e., the detector fails to detect vehicles, under inclement weather conditions such as heavy rain or snow. The speed measurements had a consistent upward bias when tested at average roadway operating speeds between 40 and 70 mph. However, the errors were typically less than 5 mph (about 10 percent of the average speed).

The evaluation of the feasibility of using crowdsourced smartphone application–based incident detection for reducing incident detection times was performed by comparing detections from the Waze® logs with detections in NaviGAtor's incident logs. With the data fusion methodology developed, about 46 percent of the NaviGAtor incidents could be re-identified in the Waze logs in the Atlanta area and about 39 percent in the Macon area. A correlation analysis with the Waze incident attributes confirmed that incidents with lower *report rating* of 0 or 1 in Waze have a slightly lower match rate with NaviGAtor logs. Incidents with a higher *confidence* number in Waze have a higher match rate, and incidents with a *reliability* of 10 in Waze have a higher match rate than the average.

Among the incidents that matched between the two logs, it was observed that in about 57 percent of the cases, the incident appeared in the Waze log before it appeared in the NaviGAtor incident log. In these 57 percent of the cases, the gain in the *time to detect* was largely in the 5–15-minute range. However, in the other 43 percent of the cases, Waze took longer to detect and log the incident than NaviGAtor, with most delays in the range of 0–30 minutes.

The evaluation of the accuracy of an AID technology involved an intensive effort of manual review of videos and images associated with 10,125 incident alarms generated by the AID over a period of 91 days. About 12 percent of the alarms could not be verified to be true because of the lack of evidence based on the videos and images available. About 2.6 percent of the alarms were misplaced in terms of lane assignment. However, there was not enough information available to verify whether the AID missed any incident. An incident detection rate for the AID technology, quantifying the ratio of the incidents detected to the total number of actual incidents, therefore, could not be established.

Neither of the unverifiable or misplaced alarm cases, in itself, would likely be a reason for not using AID. The sheer bulk of the "true" alarms generated by the AID, however, consists of very minor incidents that have very little impact on traffic operations or traffic safety. Therefore, a methodology for reducing the high number of noncritical alarms, such as shoulder stalls, is proposed. The study uses a clustering machine learning framework for developing consolidation strategies and filters that will eliminate most false, unverifiable, and noncritical alarms and associate confidence values with the alerts, thereby allowing for a focus on higher confidence alerts during busy periods. Clustering evolution patterns of the appearance of multiple alarms, where the basic alarms are generated by the AID system based on traffic anomalies, are used to train the ML algorithm to separate potential high-impact incidents from normal congestion or noncritical-related stops and slowdowns. The results indicated a significant potential of the framework in consolidating the AID-generated alarms to a small number of high-confidence clusters that can be used real-time incident management operations. This methodology might be particularly useful in controlling the number of alarms if AID is deployed over a large coverage area.

In regard to the feasibility of use of AID, it is important to recognize a limitation of the evaluation. The I-475 testbed provided a stretch of freeway with very little recurrent congestion. This helped in the ability to easily confirm the validity of the alerts during the manual review process. However, this also means that the test scenarios did not include recurrent congestion conditions. The performance of the AID technology, when used on freeways with recurrent congestion, has not been evaluated in this study.

Lastly, the project involved a study to develop a method to quantify the impact of incidents in terms of vehicle delay. Spot speed and vehicle count measurement have been the most

widely accepted performance monitoring methods for traffic operations data collection by transportation agencies. Delay estimation methods based on spot speed and cumulative count are typically deployed by practitioners and researchers alike for rapid estimation of delays as a precursor to congestion mitigation. In this chapter, these commonly used incident-induced delay estimation methodologies, which are based on queuing theory or shockwave analysis models, are reviewed and validated against microscopic simulation of a real-life incident. For the simulation model, NaviGAtor speed–volume data were used. The incident timeline was constructed using NaviGAtor incident logs. The comparison revealed challenges related to noisy data and the failure of spot-speed measurements to adequately capture heterogeneity in congested traffic, which rendered the methodologies impractical for field use. In the absence of any alternative method to accurately quantify delay within the constraints of field observational data, a regression model was developed using data from a non-exhaustive set of incident scenarios simulated using Vissim®, to help obtain rapid estimates of delays for incidents with varying characteristics occurring under varying base conditions. This regression model can aid in resource allocation for efficient incident management and identification of influence factors.

RECOMMENDATIONS

In determining AID zone and device placement location, specific attention should be given to merge or diverge points, weaving areas, or other zones with a higher potential for an incident. Should an incident occur outside of the detection zone of the AID device (such as the view of a camera in video-based AID), the AID will not provide feedback until the results of the incident (e.g., spillback) encroach into the detection zones (i.e., come into the view of the cameras). Thus, placement should consider the potential of such lag in

136

receiving information. In addition, for video-based AID, attention should be paid to items such as seasonal growth of vegetation and other potential temporary obstructions in the camera frame as they may be interpreted as an incident. If moveable cameras (pan-tilt-zoom cameras) are used for AID, automatic detection of change in background view and blocking of alarms during such movements would be a desirable feature for the AID system. Finally, for a video-based AID, the same cautions as required for video-based vehicle detection systems are recommended. For example, the camera angle should be as steep (overhead) as possible to limit occlusion-related errors (both vertical, i.e., within lane, and horizontal, i.e., across lanes). This can be a particular challenge where a camera angle precludes the AID from being able to distinguish between a vehicle on the shoulder and in the right travel lane, as vehicles on shoulders result in the majority of detections and it is often desirable to filter out or assign a lower priority to these alarms. During nighttime conditions, flat camera angles can produce views that generate false alarms of wrong-way detection from reflections of headlight on roadside objects such as barrier walls. A balance needs to be achieved between producing a larger area of detection by using flatter angles of the cameras, which will lead to lesser "blind spots," versus a higher quality of detection within a smaller area produced by steeper angles of cameras.

ACKNOWLEDGEMENTS

REFERENCES

Ahmed, Mohammed S., and Allen R. Cook. 1979. "Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques." *Transportation Research Record* (722): pp 1-9. Obtained from: http://onlinepubs.trb.org/Onlinepubs/trr/1979/722/722-001.pdf, and https://trid.trb.org/view/148123. Last accessed October 30, 2020.

Al-Deek, H., A. Garib, and E. Radwan. 1995. "New Method for Estimating Freeway Incident Congestion." *Transportation Research Record* (1494): 30-39. Obtained from: http://onlinepubs.trb.org/Onlinepubs/trr/1995/1494/1494-004.pdf, https://trid.trb.org/view/452634. Last accessed October 30, 2020.

Amin-Naseri, Mostafa, Pranamesh Chakraborty, Anuj Sharma, Stephen B. Gilbert, and Mingyi Hong. 2018. "Evaluating the Reliability, Coverage, and Added Value of Crowdsourced Traffic Incident Reports from Waze." *Transportation Research Record* 2672 (43): 34-43. Obtained from: https://doi.org/10.1177/0361198118790619. Last accessed October 30, 2020.

Bertini, Robert L. , Michael W. Rose, and Ahmed M. El-Geneidy. 2004. *Using Archived Data to Measure Operational Benefits of Its Investments, Volume 2: Region 1 Incident Response Program. Final Technical Report Tnw2004-01.2.* Oregon Department of Transportation, Salem. Obtained from: http://bertini.eng.usf.edu/pdf/Incident_Response.pdf. Last accessed October 30, 2020.

Birant, Derya, and A. Kut. 2007. "St-Dbscan: An Algorithm for Clustering Spatial-Temporal Data." *Data Knowl. Eng.* 60: 208-221.

Birst, Shawn; , and Ayman Smadi. 1999. *An Application of ITS for Incident Management in Second-Tier Cities.* (Mid-Continent Transportation Symposium 2000, Ames, Iowa).

Blincoe, Lawrence, Ted R. Miller, Eduard Zaloshnja, and Bruce A. Lawrence. May 2015 2015. *The Economic and Societal Impact of Motor Vehicle Crashes, 2010 (Revised).* National Highway Traffic Safety Administration, National Center for Statistics and Analysis, National Highway Traffic Safety Administration.

Bonneson, James , and Montasir Abbas. September 2002 2002. *Video Detection for Intersection and Interchange Control.* The Texas A&M University System Texas Transportation Institute, Texas Transportation Institute (Texas Department of Transportation).

Carson, Jodi L., Fred L. Mannering, Bill Legg, Jennifer Nee, and Doohee Nam. 1999. "Are Incident Management Programs Effective? Findings from Washington State." *Transportation Research Record* 1683 (1): 8-13. Obtained from: https://doi.org/10.3141/1683-02. Last accessed October 30, 2020.

CEOS. 2020. "TIRTL." CEOS. Accessed 02/15/2020. Obtained from: https://www.ceos.com.au/products/tirtl/.

Chen, Luyi, Ping Xu, Tiaojuan Ren, Yourong Chen, B. Zhou, and Hexin Lv. 2016. "A Svm-Based Approach for VANET-Based Automatic Incident Detection."

Chen, Shuyan, and Wei Wang. 2009. "Decision Tree Learning for Freeway Automatic Incident Detection." *Expert Systems with Applications* 36: 4101-4105. Obtained

from: https://doi.org/10.1016/j.eswa.2008.03.012. Last accessed October 30, 2020.

Chen, Zhuo, Xiaoyue Cathy Liu, and Guohui Zhang. 2016. "Non-Recurrent Congestion Analysis Using Data-Driven Spatiotemporal Approach for Information Construction." *Transportation Research Part C: Emerging Technologies* 71: 19-31. Obtained from: https://doi.org/https://doi.org/10.1016/j.trc.2016.07.002. and http://www.sciencedirect.com/science/article/pii/S0968090X16301000. Last accessed October 30, 2020.

Chow, W. 1976. "A Study of Traffic Performance Models under an Incident Condition." *Transportation Research Record* 567: 31-36.

Collins, J. F., C. M. Hopkins, J. A. Martin, Transport, and Laboratory Road Research. 1979. *Automatic Incident Detection - TRRL Algorithms HIOCC and PATREG.* Obtained from: https://trid.trb.org/view/156059. Last accessed October 30, 2020.

Dogru, N., and A. Subasi. 2018. "Traffic Accident Detection Using Random Forest Classifier." 2018 15th Learning and Technology Conference (L&T), 25-26 Feb. 2018.

Dudek, Conrad L, Carroll J Messer, and Nelson B Nuckles. 1974. "Incident Detection on Urban Freeways." *Transportation Research Record* 495: 12-24.

Fu, Liping, and Laurence R. Rilett. 1997. *Real-Time Estimation of Incident Delay in Dynamic and Stochastic Networks*. Vol. 1603.

Gall, Ana I., and Fred L. Hall. 1989. "Distinguishing between Incident Congestion and Recurrent Congestion: A Proposed Logic." *Transportation Research Record* (1232): p. 1-8. Obtained from:

http://onlinepubs.trb.org/Onlinepubs/trr/1989/1232/1232-001.pdf and

https://trid.trb.org/view/308664. Last accessed October 30, 2020.

Garib, A., A. E. Radwan, and H. Al-Deek. 1997. "Estimating Magnitude and Duration of

Incident Delays." *Journal of Transportation Engineering* 123 (6): 459-466.

Obtained from: https://doi.org/10.1061/(ASCE)0733-947X(1997)123:6(459). Last

accessed October 30, 2020.

Gillen, David. 2001. *Caltrans Tops Evaluation: Assessing the Net Benefits of Its*

Applications.

Guensler, Randall, Vetri Elango, Angshuman Guin, Michael Hunter, Jorge Laval, and et

al. 2013. "Atlanta I-85 Hov-to-Hot Conversion : Analysis of Vehicle and Person

Throughput." Obtained from: https://rosap.ntl.bts.gov/view/dot/23015. Last

accessed October 30, 2020.

Guin, Angshuman, Michael Hunter, Michael Rodgers, James Anderson, Scott Susten, and

Kiisa Wiegand. 2016. "Integrating Intersection Traffic Signal Data into a Traffic

Monitoring Program." *Transportation Research Record* 2593 (1): 74-84.

Obtained from: https://doi.org/10.3141/2593-08. and

https://journals.sagepub.com/doi/abs/10.3141/2593-08. Last accessed October 30,

2020.

Guin, Angshuman, Michael Hunter, Wonho Suh, James Anderson, Atlanta Georgia

Institute of Technology, Transportation Georgia Department of, and

Administration Federal Highway. 2013. *Feasibility Study for Using Video*

Detection System Data to Supplement Automatic Traffic Recorder Data. Obtained

from: http://g92018.eos-intl.net/eLibSQL14_G92018_Documents/11-13.pdf,
https://trid.trb.org/view/1396132. Last accessed October 30, 2020.

Guin, Angshuman, Christopher Porter, Bayne Smith, and Carla Holmes. 2007. "Benefits
Analysis for Incident Management Program Integrated with Intelligent
Transportation Systems Operations: Case Study." *Transportation Research
Record* 2000 (1): 78-87. Obtained from: https://doi.org/10.3141/2000-10. Last
accessed October 30, 2020.

Hadi, Mohammed, Prasoon Sinha, and Amy Wang. 2007. "Modeling Reductions in
Freeway Capacity Due to Incidents in Microscopic Simulation Models."
Transportation Research Record 1999: 62-68. Obtained from:
https://doi.org/10.3141/1999-07. Last accessed October 30, 2020.

Hunter, Michael P., Angshuman Guin, Michael O. Rodgers, Ziwei Huang, and Aaron T.
Greenwood. 2017. "Cooperative Vehicle–Highway Automation (CVHA)
Technology : Simulation of Benefits and Operational Issues." Obtained from:
https://rosap.ntl.bts.gov/view/dot/36001. Last accessed October 30, 2020.

Iqbal, Zafar, and Majid Iqbal Khan. 2018. "Automatic Incident Detection in Smart City
Using Multiple Traffic Flow Parameters Via V2X Communication." *International
Journal of Distributed Sensor Networks* 14 (11): 1550147718815845. Obtained
from: https://doi.org/10.1177/1550147718815845. Last accessed October 30,
2020.

Khattak, Asad J., and Nagui Rouphail. January 2004 2004. *Incident Management
Assistance Patrols: Assessment of Investment Benefits and Costs.* Center for
Urban & Regional Studies, Department of City and Regional Planning and

Institute for Transportation Research and Education, NC State University (North Carolina Department of Transportation: North Carolina Department of Transportation). Obtained from: https://connect.ncdot.gov/projects/research/pages/ProjDetails.aspx?ProjectID=2003-06. Last accessed October 30, 2020.

Latoski Steven, P., Raktim Pal, and C. Sinha Kumares. 1999. "Cost-Effectiveness Evaluation of Hoosier Helper Freeway Service Patrol." *Journal of Transportation Engineering* 125 (5): 429-438. Obtained from: https://doi.org/10.1061/(ASCE)0733-947X(1999)125:5(429). Last accessed October 30, 2020.

Lawson, Tim W., David J. Lovell, and Carlos F. Daganzo. 1997. "Using Input-Output Diagram to Determine Spatial and Temporal Extents of a Queue Upstream of a Bottleneck." *Transportation Research Record* 1572: 140-147.

Virtualdub 1.10.4.

Levin, Moshe, and Gerianne M. Krause. 1978. "Incident Detection: A Bayesian Approach."

Liu, Q., J. Lu, and S. Chen. 2014. "Design and Analysis of Traffic Incident Detection Based on Random Forest." *Journal of Southeast University (English Edition)* 30: 88-95. Obtained from: https://doi.org/10.3969/j.issn.1003-7985.2014.01.017. Last accessed October 30, 2020.

M. Morales, Juan. 1986. *Analytical Procedures for Estimating Freeway Traffic Congestion*. Vol. 50.

Martin, Peter T., Gayathri Dharmavaram, and Aleksandar Stevanovic. December 2004 2004. *Evaluation of Udot's Video Detection Systems* Department of Civil and Environmental Engineering, Department of Civil and Environmental Engineering, University of Utah Traffic Lab.

Messer, Carroll J., Conrad L. Dudek, and John D. Friebele. 1973. "Method for Predicting Travel Time and Other Operational Measures in Real-Time During Freeway Incident Conditions."

Moskowitz, Karl, and Leonard Newman. 1962. "Notes on Freeway Capacity." *Highway Research Record* 27: 44-68.

Office of Research, Development, and Technology, Office of Infrastructure, RDT. 2014. Verification, Refinement, and Applicability of Long-Term Pavement Performance Vehicle Classification Rules. edited by Federal Highway Administration US Department of Transportation: Federal Highway Administration.

Olmstead, Todd. 1999. "Pitfall to Avoid When Estimating Incident-Induced Delay by Using Deterministic Queuing Models." *Transportation Research Record* 1683 (1): 38-46. Obtained from: https://doi.org/10.3141/1683-06. and https://doi.org/10.3141/1683-06. Last accessed October 30, 2020.

Payne, Howard J., and Samuel C. Tignor. 1978. "Freeway Incident-Detection Algorithms Based on Decision Trees with States." *Transportation Research Record* (682): 30-37. Obtained from: http://onlinepubs.trb.org/Onlinepubs/trr/1978/682/682-005.pdf. Last accessed October 30, 2020.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. "Scikit-Learn: Machine Learning in Python." *J. Mach. Learn. Res.* 12 (null): 2825–2830.

PTV. 2018a. *PTV Vissim 10 User Manual.* PTV Group.

PTV. 2018b. "PTV Vissim Faqs: How Can I Model an Incident Where One or More Lanes Are Blocked?". Accessed 10 October 2019. Obtained from: http://vision-traffic.ptvgroup.com/en-us/training-support/support/ptv-vissim/faqs/visfaq/show/VIS19167/. Last accessed October 30, 2020.

Rhodes, Avery, Edward J. Smaglik, and Darcy M. Bullock. May 2006 2006. *Vendor Comparison of Video Detection Systems* Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana (Indiana Department of Transportation).

Saroj, Abhilasha, Nishu Choudhary, Han Gyol Kim, Angshuman Guin, Michael O Rodgers, and Michael Hunter. 2018. "Video Tool for Manually Extracting Complex Traffic Data." Transportation Research Board 97th Annual Meeting, Washington DC, United States, 01/08/2018.

Sheu, Jiuh-Biing. 2003. "A Stochastic Modeling Approach to Real-Time Prediction of Queue Overflows." *Transportation Science* 37 (1): 97-119. Obtained from: http://www.jstor.org/stable/25769136. Last accessed October 30, 2020.

Skabardonis, Alexander, Karl Petty, Hisham Noeimi, Daniel Rydzewski, and Pravin P. Varaiya. 1996. "I-880 Field Experiment: Data-Base Development and Incident

Delay Estimation Procedures." *Transportation Research Record* 1554 (1): 204-212. Obtained from: https://doi.org/10.1177/0361198196155400124. Last accessed October 30, 2020.

Snelder, M., T. Bakri, and B. van Arem. 2013. "Delays Caused by Incidents: Data-Driven Approach." *Transportation Research Record* 2333 (1): 1-8. Obtained from: https://doi.org/10.3141/2333-01. https://doi.org/10.3141/2333-01. Last accessed October 30, 2020.

Suh, Won Ho, James Anderson, Angshuman Guin, and Michael Hunter. 2015. "Evaluation of Traffic Data Collection Method." *Applied Mechanics and Materials* 764-765: 905-909. Obtained from: https://doi.org/10.4028/www.scientific.net/AMM.764-765.905.

Sullivan Edward, C. 1997. "New Model for Predicting Freeway Incidents and Incident Delays." *Journal of Transportation Engineering* 123 (4): 267-275. Obtained from: https://doi.org/10.1061/(ASCE)0733-947X(1997)123:4(267). Last accessed October 30, 2020.

Taylor, Nicholas, Johan Olstam, Viktor Bernhardsson, and Philippe Nitsche. 2017. "Modelling Delay Saving through Pro-Active Incident Management Techniques." *European Transport Research Review* 9 (4): 17p. Obtained from: https://doi.org/10.1007/s12544-017-0265-5. Last accessed October 30, 2020. https://trid.trb.org/view/1483838.

Toth, Christopher, Wonho Suh, Vetri Elango, Ramik Sadana, Angshuman Guin, Michael Hunter, and Randall Guensler. 2013. "Tablet-Based Traffic Counting Application Designed to Minimize Human Error." *Transportation Research Record* 2339 (1):

39-46. Obtained from: https://doi.org/10.3141/2339-05. and

https://journals.sagepub.com/doi/abs/10.3141/2339-05. Last accessed October 30,

2020.

Vallejos, Sebastián, Diego G. Alonso, Brian Caimmi, Luis Berdun, Marcelo G.

Armentano, and Álvaro Soria. 2020. "Mining Social Networks to Detect Traffic

Incidents." *Information Systems Frontiers*. Obtained from:

https://doi.org/10.1007/s10796-020-09994-3. Last accessed October 30, 2020.

Vermeulen, E. 2014. "Automatic Incident Detection (Aid) with Thermal Cameras." Road

Transport Information and Control Conference 2014 (RTIC 2014), 6-7 Oct. 2014.

Villela, Karina, Claudia Nass, Renato Novais, Paulo Simões Jr, Agma Juci Machado

Traina, José Fernando Rodrigues Junior, Jose Manuel Menendez, Jorge Kurano,

Tobias Franke, and Andreas Poxrucker. 2018. "Reliable and Smart Decision

Support System for Emergency Management Based on Crowdsourcing

Information." In *Exploring Intelligent Decision Support Systems : Current State

and New Treds*. Springer.

Wang, Jiawei, Xin Li, Stephen Shaoyi Liao, and Zhonsheng Hua. 2013. "A Hybrid

Approach for Automatic Incident Detection." *IEEE Transactions on Intelligent

Transportation Systems* 14 (3): 1176-1185. Obtained from:

https://doi.org/10.1109/TITS.2013.2255594 and

http://www.scopus.com/inward/record.url?scp=84883821651&partnerID=8YFLo

gxK. Last accessed October 30, 2020.

Wang, Yinhai, Runze Yu, Yunteng Lao, and Timothy Thomson. 2010. "Quantifying

Incident-Induced Travel Delays on Freeways Using Traffic Sensor Data : Phase

II." Obtained from: https://rosap.ntl.bts.gov/view/dot/6443. Last accessed October 30, 2020.

Wells, Bill. 2016. "NaviGAtor." Georgia Department of Transportation. Last Modified June 14, 2016. Accessed July 25, 2020. Obtained from: http://www.itsga.org/navigator/. Last accessed October 30, 2020.

Williams, B. M., and A. Guin. 2007. "Traffic Management Center Use of Incident Detection Algorithms: Findings of a Nationwide Survey." *IEEE Transactions on Intelligent Transportation Systems* 8 (2): 351-358. Obtained from: https://doi.org/10.1109/TITS.2007.894193.

Willsky, A. S., E. Y. Chow, S. B. Gershwin, C. S. Greene, A. L. Kurkjian, P. K. Houpt, Electrical Institute of, and Engineers Electronics. 1980. "Dynamic Model-Based Techniques for the Detection of Incidents on Freeways." *IEEE Transactions on Automatic Control* AC-2 (3): p. 347-360. Obtained from: https://trid.trb.org/view/161451. Last accessed October 30, 2020.

Wirasinghe, S. 1978. "Determination of Traffic Delays from Shock-Wave Analysis." *Transportation Research* 12: 343-348.

Xavier, Emerson M. A., Francisco J. Ariza-López, and Manuel A. Ureña-Cámara. 2016. "A Survey of Measures and Methods for Matching Geospatial Vector Datasets." *ACM Comput. Surv.* 49 (2): Article 39. Obtained from: https://doi.org/10.1145/2963147. Last accessed October 30, 2020.

Zhu, L., F. Guo, R. Krishnan, and J. W. Polak. 2018. "A Deep Learning Approach for Traffic Incident Detection in Urban Networks." 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 4-7 Nov. 2018.

Zuo, Fan, Abdullah Kurkcu, Kaan Ozbay, and Jingqin Gao. 2018. "Crowdsourcing

    Incident Information for Emergency Response Using Open Data Sources in Smart

    Cities." *Transportation Research Record: Journal of the Transportation Research

    Board*: 036119811879873. Obtained from:

    https://doi.org/10.1177/0361198118798736. Last accessed October 30, 2020.