USDOT Region V Regional University Transportation Center Final Report

NEXTRANS Project No 003OY01

Length based vehicle classification on freeways from single loop detectors

By

Benjamin Coifman, Principal Investigator
Associate Professor of Civil and Environmental Engineering and Geodetic Sciences
Associate Professor of Electrical and Computer Engineering
The Ohio State University
Coifman.1@OSU.edu

Report Submission Date: October 15, 2009

DISCLAIMER

Funding for this research was provided by the NEXTRANS Center, Purdue University under Grant No. DTRT07-G-005 of the U.S. Department of Transportation, Research and Innovative Technology Administration (RITA), University Transportation Centers Program. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

USDOT Region V Regional University Transportation Center Final Report

TECHNICAL SUMMARY

NEXTRANS Project No 003OY01 Final Report, October 2009

Length based vehicle classification on freeways from single loop detectors

Introduction

Roadway usage, particularly by large vehicles, is one of the fundamental factors determining the lifespan of highway infrastructure, e.g., as evidenced by the federally mandated Highway Performance Monitoring System (HPMS). But the complexity of Weigh in Motion (WIM) and other classification stations makes them difficult and costly to maintain. Some of the classification stations employ axle counters, but the least expensive of these stations use dual loop detectors to measure vehicle length and classify vehicles based on this measurement. To date, collecting reliable length data from single loop detectors has been considered impossible due to the noisy speed estimates. Single loop detectors promise to be an inexpensive alternative to spread classification coverage to the existing count stations and existing traffic operations detector stations. By extending classification to the relatively high density of real time traffic monitoring stations in urban areas, the classification work could allow these urban traffic management systems to better monitor freight traffic within the metropolitan areas.

The research seeks to develop a means to reliably classify vehicles using estimated vehicle length from single loop detectors. Single loop detectors are the most common vehicle detector, yet they are not used for vehicle classification due to the inherent noise in the individual vehicle length estimates. This work has developed a means to extract more reliable vehicle speed estimates from single loop detectors, and thus, vehicle length estimates as well. This new, reliable, single loop detector methodology for classifying vehicles based on estimated vehicle length is significant because it will provide a low cost means of collecting vehicle classification data by extending the capabilities of existing single loop detectors. There are thousands of single loop detectors on the freeways within the region served by NEXTRANS.

This approach is meant to supplement the network of traditional vehicle classification stations rather than supplant them. However, this work also promises to improve the performance of the traditional classification stations by providing a viable means of estimating speed and length when one loop fails in a dual loop detector; thus, allowing such a station to remain on line while awaiting repairs. The research should yield results in the short term, being applied to existing single loop detector stations and will continue to yield benefits into the long term as long as these classifications are employed by the various

state DOTs. The research is advanced, because we are able to extract high precision individual vehicle data from a sensor that had previously been thought of as only being capable of providing aggregate measurements, it is also exploratory because we sought out new locations and challenging traffic conditions. In the process, we uncovered several chronic detector errors that degrade classification performance as well as aggregate measures of flow, speed and occupancy.

This work represents an innovative use of the existing traffic monitoring infrastructure to provide data that will feed in to larger systems perspective (e.g., augmenting existing HPMS data collection). The work feeds two of the center's three "research pillars," it explicitly focuses on collecting vehicle information (e.g., the number of trucks in the passing fleet), which in turn is important for modeling and forecasting the performance of the infrastructure (i.e., vehicle infrastructure interactions). Needless to say, a better measure of truck (and thus freight) movements will help maintain the infrastructure (in terms of financing, e.g., forecasting when a facility will need rehabilitation), and improve mobility to all classes of freeway travelers (highway passenger, highway freight, and highway transit). The new information on truck flows will help provide more efficient travel to freeway users.

Findings

Speed estimation, length estimation, and vehicle classification algorithms were developed and improved in the course of this work. Approximately 21 hours of directional traffic data were ground truthed from 34 different data sets collected at 22 different locations and an average of 3.3 lanes per set. A total of 78,774 detector actuations were manually ground truthed (in the absence of a detector error, there should be exactly one actuation per vehicle). Roughly a quarter of these data come from congestion. Three different, chronic detector errors were observed at several of the detector stations: splash-over (SO), pulse break up (PBU), and detector dropout without return (DOwoR). These errors degrade classification performance as well as conventional speed, flow and occupancy; at single loops and dual loops alike. Preliminary diagnostic algorithms for identifying SO and PBU errors were developed and should be transferable to most loop detector stations (single loops and dual loops alike). The SO algorithm only detects the presence of the problem. The PBU algorithm is able to go further, it can repair most of the observed errors. Working with ODOT, we adjusted the detector settings at four detector stations and we were successful in eliminating the chronic detector errors at most of these stations. If these results are typical, the improved detector calibration enabled by our research could lead to a very inexpensive means to improve the quality of loop detector data at existing stations.

During free flow: at stations without PBU and without SO we had a correct classification rate of 96%, of the errors (72% of the errors were due to non-vehicle pulses (NVP), in this case due to vehicles changing lanes over the detector). The correct classification rate drops to 92% from raw data at stations with PBU (47% of errors due to NVP, including extra pulses from PBU), but improves to 98% when using our diagnostic algorithms to eliminate PBU (78% of errors due to NVP). The correct classification rate drops to 89% at stations with SO (76% of errors due NVP). Note that this analysis was conducted on a per vehicle basis, so in error with one vehicle is not allowed to cancel an error with another vehicle. During congestion: all stations used for classification evaluation exhibited PBU or SO, we had a correct

classification rate of 85% (17% due to NVP), but improves to 88% when using our diagnostic algorithms to eliminate PBU (12% of errors due to NVP).

Performance from the single loop detectors is comparable to dual loop detectors when traffic is free flowing. The length based classification performance degrades by about 10% during congestion because the individual speed estimates are still based on a sample of vehicles and in heavy congestion it is possible for a given vehicle's true speed to be far from the center of the sample. These congested conditions can be identified based on the speed estimates, so if the degradation is unacceptable the classification results can be discounted or subsequent research can develop adjustment factors.

After excluding the chronic detector errors (PBU, SO, and DOwoR), most classification errors were due to a true vehicle length being close to the boundary between two bins and the estimated length falling just on the other side of the boundary. Using thresholds of 22 and 40 ft between vehicle classes, class 2 (the middle class) had a significantly higher error rate than the other two classes. The higher rate of class 2 errors arose for several reasons, first, class 2 has two boundaries, so unlike the other two classes, by definition, all class 2 vehicle lengths are within 9 ft of one boundary or the other and thus, more susceptible to the boundary issue noted above. Roughly 40% of the class 2 vehicles were within 4 ft of a boundary while only 15% of class 1 (the short vehicles) and under 10% of class 3 (the long vehicles) were within 4 ft of their respective boundaries. Such boundary errors also impact class 2 vehicles when using dual loop detectors to measure vehicle length.

Recommendations

Discovering the extent of the chronic detector errors was an unanticipated byproduct of this research, but it may also prove to be one of the most significant findings since it potentially impacts most loop detector deployments. With conventional detector aggregation, e.g., 30 sec or 5 min averaging, the chronic errors often go unnoticed unless they are severe. Our diagnostic algorithms show great promise for detecting PBU and SO. After further refinement, in the short term these algorithms could be incorporated into a field diagnostic tool to assess the performance of a given station, either by tapping into the data upstream of the controller, e.g., via the InfoTek Wizard, or running an alternate controller program for a day or two, e.g., Caltrans Log_170. In the longer run, such tests should be incorporated into the regular controller software so that the controller can continually assess the health of the detectors. More research is necessary for catching DOwoR since the resulting time series from these errors are usually indistinguishable from the passage of a shorter vehicle. We have made some progress in catching DOwoR by comparing vehicle actuations between successive stations, but more work is needed. In the mean time, as one might expect, all of the stations that we observed having DOwoR also exhibited PBU. So in these cases, it is still possible to identify that the station has a problem. Operating agencies and freeway vehicle detector manufacturers (loop detector and non-invasive detectors) should evaluate these tools for potential adoption.

Operating agencies with single loop detectors should consider deploying the vehicle classification scheme developed in this research as a means to extract more information from their existing detector

infrastructure. Similarly, manufacturers of non-invasive detectors that emulate single loop detectors (e.g., Image Sensing Systems-RTMS) should consider employing these ideas in their classification scheme.

Finally, a practical length based vehicle classification scheme needs to be robust to the large discrete steps between classes (whether from single or dual loop). Further work is needed to develop strategies for mitigating these boundary errors. One example is the simple strategy of using buffer regions, e.g., vehicles with lengths from 19 ft to 25 ft are considered "class 1 or class 2" vehicles and treated accordingly. Since these vehicles are definitely at the extreme end of their class, they might be treated differently than vehicles closer to the center of the class (e.g., borrowing ideas from fuzzy logic, instead of counting a 24 ft vehicle as 100% class 2, it might be counted as 0.8 class 2 and 0.2 class 1). Like the chronic detector errors, this discovery was an unanticipated byproduct of the detailed ground truthing and analysis. Determining the optimal correction was beyond the scope of the present work, but should be examined in future research.

Contacts

For more information:

Benjamin Coifman
Principal Investigator
Civil and Environmental Engineering and Geodetic Sciences
Ohio State University
Coifman.1@OSU.edu

NEXTRANS Center
Purdue University - Discovery Park
2700 Kent B-100
West Lafayette, IN 47906

nextrans@purdue.edu
(765) 496-9729
(765) 807-3123 Fax

www.purdue.edu/dp/nextrans

USDOT Region V Regional University Transportation Center Final Report

NEXTRANS Project No 003OY01

Length based vehicle classification on freeways from single loop detectors

By

Benjamin Coifman, Principal Investigator
Associate Professor of Civil and Environmental Engineering and Geodetic Sciences
Associate Professor of Electrical and Computer Engineering
The Ohio State University
Coifman.1@OSU.edu

Report Submission Date: October 15, 2009

ACKNOWLEDGMENTS

The author would like to acknowledge the help and input of many contributors to this research effort. Including the graduate research assistants: Ho Lee and Seoungbum Kim who lead the day to day research and made significant contributions to this report. Without assistance from ODOT to collect the raw data this work would never have been feasible. Matt Graf and Nick Hegemier stand out in their assistance from ODOT.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. **INTRODUCTION**

1.1 Background and motivation

Roadway usage, particularly by large vehicles, is one of the fundamental factors determining the lifespan of highway infrastructure, e.g., as evidenced by the federally mandated Highway Performance Monitoring System (HPMS). But the complexity of Weigh in Motion (WIM) and other classification stations makes them difficult and costly to maintain. As a result, there is interest both at the state and federal level for a lower cost vehicle classification system. Each state typically has several dozen WIM stations, supplemented with many more vehicle classification stations. Some of the classification stations employ axle counters, but the least expensive of these stations use dual loop detectors to measure vehicle length and classify vehicles based on this measurement (e.g., the state of Ohio currently has 216 permanent count stations, roughly half of which provide WIM or axle based classification, 50 provide length based classification from dual loop detectors and 49 only provide volume data from single loop detectors).

Dual loop detectors can measure speed and vehicle on-time directly, allowing for direct length measurement. At single loop detectors speed can only be estimated. To date, collecting reliable length data from single loop detectors has been considered impossible due to the noisy speed estimates. Single loop detectors promise to be an inexpensive alternative to spread classification coverage through the existing count stations and the existing traffic operations detector stations. The research seeks to enable such an extension to these existing detector stations. By extending classification to the relatively high density of real time traffic monitoring stations in urban areas, the classification work could allow these urban traffic management systems to better monitor freight traffic

within the metropolitan areas (e.g., within the region served by NEXTRANS, Chicago has 2,400 single loop detectors, and Minneapolis/St. Paul have 3,500 single loop detectors). New, out-of-pavement detectors seek to replace loop detectors using wayside mounted sensors, e.g., the Remote Traffic Microwave Sensor (RTMS), but most of these detectors emulate the operation of single loop detectors and this research would largely be applicable to those detectors as well.

Prior to undertaking the present effort, we had already overcome most of the speed and length estimation problems at single loop detectors, and demonstrated very good performance at two locations (Coifman, 2007, Coifman and Kim, 2009). The single loop based vehicle classification is within 95% agreement with concurrent measured length classification from dual loops. That work included manual verification with roughly six hours of concurrent video data (24 lane hours) and the single loop classification performance is comparable to dual loop classification performance. For reference, an in depth review of the state of the art in loop based vehicle classification is provided in Appendix A.

The present research focused on two issues that remained: (1) collect additional ground truth at more than the two locations and under different traffic conditions for further validation and development. (2) Address conditions that still challenge length based classification from loop detectors, specifically: (a) pulse breakup (PBU), a problem impacting single and dual loops alike when poorly tuned loops is to drop out in the middle of semi-trailer trucks, yielding data that would suggest two short vehicles passed when actually one long vehicle did; and (b) improving speed and length estimation performance under heavily congested conditions.

Item (1) is conceptually straightforward, we used video cameras to film traffic passing various detector stations under different conditions (e.g., congestion). But the task remains labor intensive. We have developed software to semi-automate the process and provide ground truth class and length, but it still takes a few seconds per vehicle for a human user to enter the data. The additional data was used for further validation and development of the speed and length estimation schemes and feed directly into Item (2).

This second item received most of the resources in this study. While studying pulse breakup, we uncovered two other chronic detector errors: splashover (SO) and dropout without return (DOwoR). So the scope of our study was broadened to include these errors as well. Of course catching these detector errors will have ripple effects, potentially benefiting most applications that use loop detectors to monitor traffic. The ground truth data was used to investigate if there are ways to further improve performance during congestion. Classification accuracy is around 97% during free flow, but drops to about 85% during congestion due to the fact that speed continuously changes in queued traffic.[1] Performance from the single loop detectors is comparable to dual loop detectors when traffic is free flowing. The length based classification performance degrades by about 10% during congestion because the individual speed estimates are still based on a sample of vehicles and in heavy congestion it is possible for a given vehicle's true speed to be far from the center of the sample. These congested conditions can be identified based on the speed estimates, so if the degradation is unacceptable the classification results can be discounted, they can be completely discarded, or subsequent research can develop adjustment factors. After excluding the chronic detector errors (PBU, SO, and DOwoR), most classification errors were due to a true vehicle length being close to the boundary between two bins and the estimated length falling just on the other side of the boundary from the true length.

While the main focus of this research is single loop detector based vehicle classification, the research has to ensure accurate on-time measurements, an outcome that will benefit both single loop and dual loop detector stations alike, whether or not they are deployed for vehicle classification. As such, many of the advances at single loop detectors will carry over to dual loop detectors. In fact, many of the advances are likely to carry over to non-invasive sensors like the RTMS as well.

---

[1] Unlike the earlier studies into single loop, length-based vehicle classification, these statistics do not allow over-counting errors to cancel under-counting errors.

1.2 Study objectives

The problem approach consisted of the following tasks:

1) Collect additional data: we have approximately 40 stations from CMFMS phase I available for this study. ODOT has roughly doubled this number with phase II, which just came on line at the start of the present study. These new stations provide data from several different freeways, including: I70, I270, I670, I71, and SR315.
2) We collected many hours of concurrent ground truth video data from the detector stations, and then manually extracted the vehicle lengths. The ground truth was then split into separate development sets and validation sets.
3) Using the development data sets, we found the detector errors, characterized the nature of the detector errors empirically, and conducted further theoretical development. Similarly, we continued to refine the speed estimation techniques using the expanded development data sets.
4) Used the validation data sets to evaluate the performance of the advances from item 3.
5) Documented the results, presenting them in a final report, as well as various conferences and peer reviewed journals.

1.3 Organization of the report

Accurate vehicle classification depends on accurate vehicle measurements and estimates. Most of this work is devoted to addressing detector errors. Chapter 2 presents the background on identifying chronic detector errors and reviews our data sources. Chapter 3 presents an algorithm to identify splashover. Chapter 4 presents an algorithm to identify pulse breakup. Chapter 5 presents a pilot study for identifying detector dropout without return. At which point we are then ready to proceed into chapter 6, vehicle classification from single loop detectors. The work closes with conclusions in Chapter 7.

CHAPTER 2. **IDENTIFYING CHRONIC DETECTOR ERRORS- BACKGROUND AND DATA SOURCES**

Loop detectors are effectively metal detectors embedded in the pavement. They are the most commonly used vehicle detector for automated surveillance. A typical loop detector station will either have one or two loops per lane (single or dual loop detectors, respectively). Data obtained from loop detectors can be used for applications such as ramp metering, incident detection, travel time prediction, and vehicle classification. The performance of such applications greatly depends on the quality of detector data. Data collected from loop detectors are prone to detector errors caused by hardware and software problems. Detector errors degrade the quality of detector data, and the impact of these errors will propagate to subsequent measurements such as flow, occupancy, and speed from the loop detectors. In the end, unreliable data incorporating detector errors could affect the control decisions and traveler information provided to drivers based on the detector's data.

There has been considerable research effort to screen the quality of loop detector data. Data screening methods have been developed at the macroscopic and microscopic levels. Macroscopic tests embody the formalization of heuristics to check average measurements from a given sample period against statistical tolerance, while microscopic tests examine the individual vehicle actuations, when the detector turns "on" and "off" for each vehicle that passes over a loop detector. The macroscopic tests are more common, because conventional practice discards the microscopic data at the controller cabinet after aggregation to macroscopic flow, occupancy and average speed.

As examples of the macroscopic level of approach, Jacobson et al. (1990) introduced a test for setting limits for acceptable values of flow for any given occupancy on the basis of plausible ratios between flow and occupancy within specific occupancy ranges. In particular, the algorithm was useful for detecting intermittent failures and short pulses (hanging–off) malfunctions of loop detectors. A later study by Cleghorn et al. (1991) presented several screening methods using macroscopic measurements. They claimed to have obtained a tighter upper boundary from feasible flow-occupancy pairs to be used to screen data from a single loop detector. They also presented additional screening for dual loop detectors that includes a comparison of the received speed-flow-occupancy points, a calibrated three-dimensional speed-flow-occupancy "acceptable region" as well as comparison of measurements between upstream and downstream loops. Chen et al. (2003) developed a macroscopic error detection test using the time series of flow and occupancy measurements. Statistics computed over a whole day at each detector are used to differentiate between a "bad" or "good" detector with respect to various specific loop detector malfunctions, e.g., stuck and hanging. The algorithm did not detect chattering or pulse breakup (i.e., a single pulse expected from a vehicle is separated into two or more pulses if the detector "drop out" in the middle), but they suggested that additional constraints, such as consistently high flow, should be useful to detect loops with these errors. Turochy and Smith (2000) developed an integrated data-screening procedure including a critical threshold value of measurements such as occupancy and flow, and tests utilizing the relationships between speed, flow, and occupancy. Among those tests included in the procedure, a maximum hourly flow threshold test (e.g., 3,100 vehicles/lane/hr) was used to catch detector errors causing higher flow.

At the microscopic level, Chen and May (1987) may have been the first to use individual vehicle actuations, rather than the macroscopic measurements, to verify detector data. They examined the ratio of a detector's average on-time to average on-time of all detectors at the loop station. This on-time ratio test provided a reliable indication of detector status, e.g., the change of sensitivity. In particular, their experiment found pulse

breakups and they surmised that the breakups might be caused by low loop sensitivity. In addition, they found unexpected detector actuations in their data due to: lane change maneuvers over the loop detectors, splashover (the erroneous detection in one lane of a vehicle from an adjacent lane), and phantom actuations that are not due to vehicles. Coifman (1999) presented a microscopic method utilizing the redundancy of a pair of loops to assess the performance of dual loop detector and to identify detector errors; namely, that during free flow conditions the on-times for a given vehicle from the pair of loops should be virtually identical regardless of vehicle length. The method detected a longer on-time problem due to delayed falling edge and cross-talk problems. Coifman and Lee (2006) presented the mode on-time test as a measure of performance for single loop detectors. The test calculates the most common on-time over a day. Assuming most vehicles are free flowing passenger cars, this mode on-time should fall within a small range. The test indirectly detects inappropriate level of loop sensitivity. Additionally, minimum on-time and maximum on-time tests were applied to catch extreme errors due to pulse breakup and detector sticking-on. Coifman and Dhoorjaty (2004) developed eight detector validation tests using microscopic data to identify various errors both at single and dual loop detectors. That work specifically classified errors into seven groups: either the rising or the falling edge being premature or delayed, pulse breakup, missed vehicle, and wrong detection. Cheevarunothai et al (2007) developed an algorithm to improve the quality of dual-loop truck data so as to identify and correct detector problems such as pulse breakups, cross-talk, and the difference of sensitivity in two loops.

Despite the previous research effort in this area, some significant detector errors have not received much attention due to the difficulty of identifying their occurrence. Splashover and pulse breakup are such cases. No study has explicitly attempted to offer any means of detecting the presence or absence of splashover error. Pulse breakup detection has been explored previously (Chen and May, 1987; Cheevarunothai et al., 2007), but the previous methods were not sufficient to resolve the problem under all conditions.

The overall objective of our research is to develop algorithms to identify splashover (SO) and pulse breakup (PBU). The algorithms are developed using loop detector data with concurrent video-recorded ground truth data from the in Columbus, Ohio. The splashover detection algorithm is designed to find detector stations exhibiting chronic splashover problem, while the pulse breakup detection algorithm is designed to identify pulse breakup from individual vehicle actuation data. Finally, the developed algorithms are evaluated from several detector stations with the ground truth data.

2.1 Data Sources

2.1.1 **Loop Detector Data**

Individual vehicle actuation data were collected from the 69 detector stations in the Columbus Metropolitan Freeway Management System (CMFMS), sampled at 240 Hz (Coifman, 2006a). These stations include 330 loop detectors on the northbound / eastbound freeway mainline lanes and 328 loop detectors on the southbound / westbound freeway mainline lanes. In detail, the 46 detector stations on I-70/I-71 were installed during the first phase of CMFMS, completed in 2001. These stations include 196 loop detectors on the northbound / eastbound freeway mainline lanes and 194 loop detectors on the southbound / westbound freeway mainline lanes. Another 23 detector stations were installed on SR 315 / I-270 / I-70 / I-670 during the second phase of the CMFMS, completed in 2006. These stations include 134 loop detectors on each of direction freeway mainline lanes and six detector stations have RTMS. Figure 2.1 shows a schematic of the study corridor. Roughly 90% of the Phase II detector stations have dual loop detectors, while only 35% of the Phase I detector stations have dual loop detectors. For most of the Phase I corridor there is one dual loop detector station every mile, with two single loop detector stations between dual loop stations. As noted previously, only one loop in a given dual loop detector is used to emulate a single loop detector. In many cases, however, we can than compare the single loop detector results against the dual loop detector results.

2.1.2 **Ground Truth Data**

Ground truth data is used to develop and validate the various algorithms. The task to extract the ground truth data consists of recording video of the loop detectors and the concurrent detector actuations, digitizing the video, extracting individual frames, time synchronization between loop data and video data, stepping through all of the loop detector actuations individually, loading the specific frame corresponding to a given loop detector actuation, and manually classify the detector actuation.

Most of the video comes from the existing traffic surveillance cameras in the CMFMS. There are currently 99 traffic surveillance video cameras are operated in the CMFMS, with 74 of them mounted near freeways. After reviewing camera views, we found the set of cameras that provide good views across all lanes for one or both directions at one or more detector stations. The traffic surveillance video cameras were recorded with a VCR in the Columbus Traffic Management Center (TMC). For the other detector stations that are not readily viewed from the surveillance cameras, a video camcorder was set up on overhead bridges or the side of the road. Video was collected at a total of 15 detector stations. Figure 2.2 typical views (A) of station 38 from a surveillance camera and (B) of station 41 eastbound from a camcorder set up on an overhead bridge (Woodcrest Rd in this case).

Both sources of video data were recorded in analog, thus an additional step of digitizing the video was necessary. The digital video was stored in AVI format and the frames were extracted in JPEG format at the rate of 30 frames per second (fps). A purpose built software ground truthing tool with graphical user interface (GUI) was developed in MATLAB to semi-automate the process of generating ground truth data. The GUI interface can step through the detector data in a given lane and display both the time series detector data for a few seconds before and after the actuation along with the frame corresponding to the actuation time (this GUI was inspired by VideoSync, Caltrans, 2007). The tool allows the user to manually record types of vehicles and errors from the direct comparison between concurrent detector and video data. A user can mark actuations as detector errors such as pulse breakup, splashover or a vehicle changing

lanes. The GUI also allows the users to classify the vehicle as: motorcycle, short vehicle (SV), medium vehicle (MV), or long vehicle (LV). In the research, a LV refers to a long truck or a semi-trailer truck, while a SV refers to car, van, and pick-up truck. Vehicles not included in SV, LV, and motorcycles are classified into MV, e.g., large vans, buses, most single unit trucks, or most SV pulling long trailers. Once an actuation has been classified, the user clicks a button and the GUI jumps to the next detector actuation in that lane. This process was repeated for each visible lane during the entire time period of video data. As discussed in the next chapter, for about half of the data a second pass was then made to actually measure the vehicle's length.

In the end, approximately 21 hours of directional traffic data were ground truthed from 34 different data sets collected at 22 different locations and an average of 3.3 lanes per set. A total of 78,774 detector actuations were manually ground truthed (in the absence of a detector error, there should be exactly one actuation per vehicle). Out of these data, 9 sets include congestion, spanning 4.5 hrs and 20,576 detector actuations.

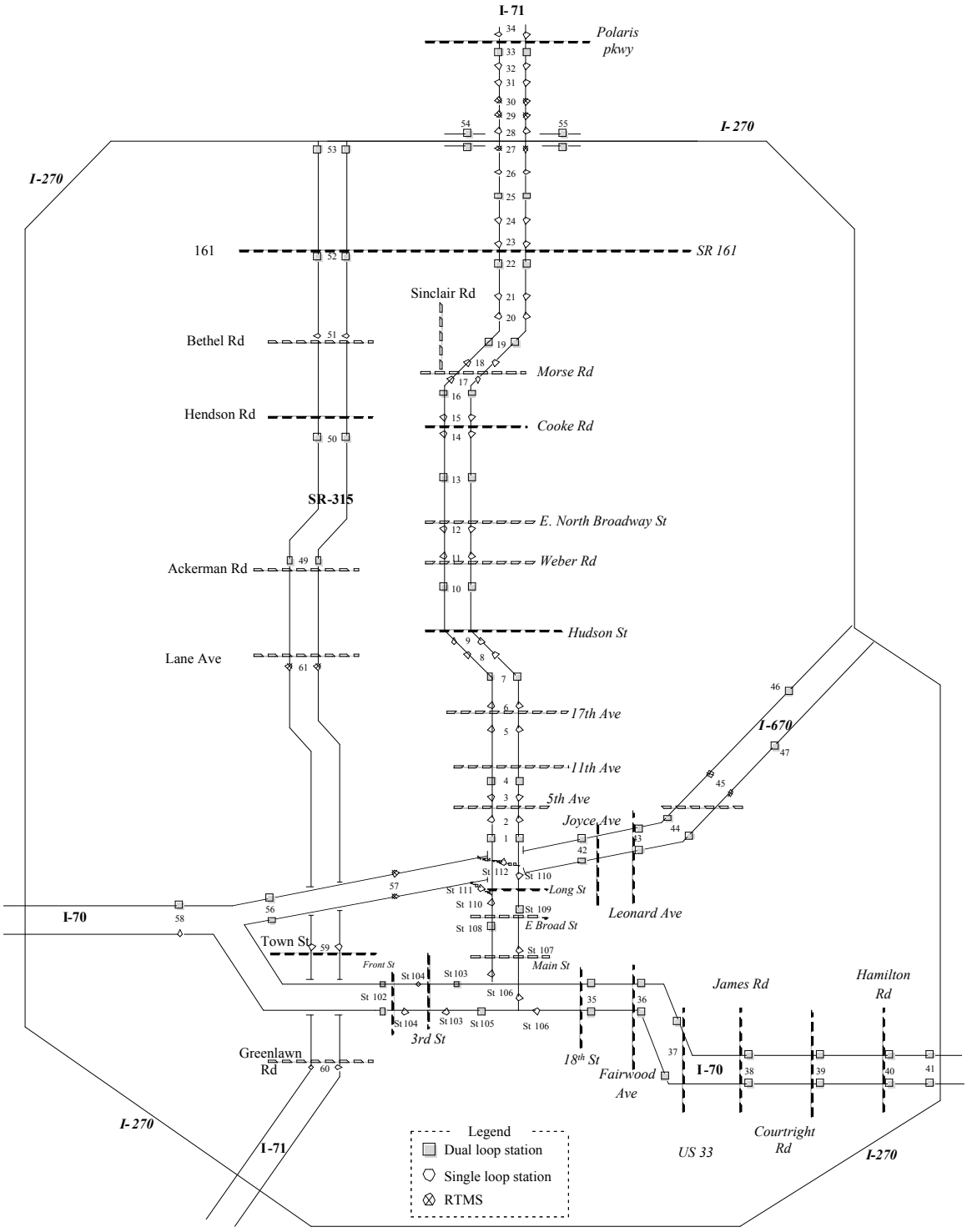


Figure 2.1. A schematic of the detector stations in the CMFMS

(a) (b)

Figure 2.2. (a) camera view from the traffic camera at station 56 and (b) camera view from a video camera mounted on overhead bridge near station 41

CHAPTER 3. **AN ALGORITHM TO IDENTIFY SPLASHOVER**

Splashover is the erroneous detection of a vehicle that is outside the desired detection zone, and commonly this error arises when vehicles from adjacent lanes erroneously actuate the detector. Typically when this error occurs the loop detectors in two neighboring lanes record the same vehicle at roughly the same time. If not caught, the splashover of one vehicle will likely be recorded as two distinct vehicles that almost simultaneously passed over two adjacent loops. The resulting double count can lead to inaccurate measurements, e.g., higher flow and occupancy, degrading the quality of loop detector data. Splashover has not received much attention in the earlier research. According to the Traffic Detector Handbook (Klein et al., 2006) splashover usually occurs when the sensitivity level of a loop detector is set too high or a loop detector is too close to the lane line. While the handbook offers some advice for fixing this problem, it does not offer any characteristics of splashover or any means of detecting the presence or absence of splashover. This section presents an algorithm to identify detectors exhibiting chronic splashover problems. The splashover detection algorithm exploits the characteristics of splashover revealed from the ground truth data.

3.1 Hypothetical Example of Splashover

The relation of an actual detection in one loop and a false detection from an adjacent loop is illustrated in a time-space diagram of a single vehicle in Figure 3.1. A schematic of the roadway is shown coincident with the distance axis. This hypothetical example shows that a vehicle passing in lane 2 is falsely detected in loop 1, where on_1 and on_2 denote the actuation times of rising transition in loop 1 and loop 2, off_1 and off_2 denote the actuation times of falling transition in both loops. OT_2 indicates the on-time

from actual detection of a vehicle while OT_1 indicates the on-time from the erroneous splashover detection from the same vehicle. Since two on-times (OT_1 and OT_2) are produced from a single vehicle, they just depend on the size of detection zones from loops (denoted by DZ_1 and DZ_2). The size of the detection zones will likely differ from the two detectors due to different sensitivity and distance to the vehicle. Technically, a loop's sensitivity is in inversely by proportional to the square of the distance to the vehicle (Klein et al., 2006), and typically (but not always) the detection zone (DZ_1) for splashover should be smaller than the actual lane (DZ_2). As seen in this figure, DZ_2 is larger than DZ_1, thus OT_2 from actual detection is longer than OT_1 from splashover. The difference of on-times (OT_1 - OT_2) is proportional to the difference of detection zone and could be separated into the difference in rising transition times (DRTT = on_1 - on_2) and the difference in falling transition times (DFTT = off_1 – off_2). When DZ_1 falls within DZ_2, actuation times (on_1 and off_1) from splashover should fall into interval of the actuation times (on_2 and off_2) from the actual detection of the vehicle.

The example in Figure 3.1 shows a single vehicle actuating loops in adjacent lanes, but no other vehicles around. We term this situation "unique splashover". In contrast, things get a little more complicated when splashover combines with a true vehicle actuation in the given lane, the union of the splashover pulse and true pulse may be longer than either of the two pulses individually. Figure 3.2 shows four hypothetical examples where a lane 2 vehicle splashes over to lane 1, but a lane 1 vehicle also passes. In theses cases, the splashover may appear as if the lane 1 detector is sticking on (Figure 3.2A-B), which we term "combined splashover", it may obscure the lane 1 vehicle actuation (Figure 3.2C) or it may be completely obscured by the lane 1 vehicle actuation (Figure 3.2D). In all of these cases the splashover does not result in a count error, however, in Figure 3.2A-C lane 1 would measure an on-time longer than it should for the passing vehicle. The combined splashover pulse might not be bounded by the original pulse in lane 2.

3.2 The Nature of Splashover

Typically, when splashover occurs, loop detectors in adjacent lanes actuate in response to the same vehicle at roughly the same time, e.g., the hypothetical example in Figure 3.1. Figure 3.3 shows three examples of actual splashover in lane 1 at a single loop detector station (station 104 eastbound). Figure 3.3A shows the pulses from all three lanes as the detectors respond to vehicles (throughout this example lane 1 is at the bottom and lane 3 is at the top). In the absence of detector errors, a passage of a vehicle is manifested as a single pulse, but according to the concurrent video, all three of the pulses in lane 1 were erroneous. First, in Figure 3.3B, shows a unique splashover when a vehicle passing in lane 2 is also recorded in lane 1 in the absence of any vehicles in lane 1. Next, Figure 3.3C-D both show examples of combined splashover, where the valid on-time of a short vehicle in lane 1 is extended due to splashover from lane 2. In both cases the on-times are roughly twice as long as they should be for the given vehicle and speed. After ground truthing 10 min of data at this station, total vehicles observed from video in lane 1 to lane 3 are 164, 346, and 347 vehicles, respectively. Reviewing the detector data, 318 out of 347 vehicles in lane 2 are erroneously detected in lane 1, i.e., 92% of vehicles in lane 2 caused splashover in lane 1. Of these splashover events, 273 are unique splashover and 45 are combined splashover. In terms of macroscopic measurement errors, flow in lane 1 would be over-counted by the number of unique splashover events, i.e., more than 2/3 of the actuations in lane 1 are not due to vehicles in lane 1. While the combined splashover events do not impact flow, both types of splashover will lead to an erroneous increase in occupancy because the detector reports that it is on ($\sum OT_1$) for a longer cumulative time than it is actually occupied by vehicles.

Figure 3.4A plots the on-time for the splashovers seen in lane 1 versus the actual on-time as seen in lane 2, the vehicle's lane of travel (i.e., OT_1 versus OT_2). Each point represents the actual detection of a vehicle that traveled in lane 2 and its erroneous detection in lane 1. The dashed line shows the set of points where the two on-times are equal. In general, unique splashover errors tend to fall below the reference line, they have

a shorter on-time than the actual detection. While most of the combined splashovers fall above the reference line, having a longer on-time than the actual detection. Figure 3.4B-C tabulates the cumulative distribution function (CDF) of the difference of on-times between the two lanes for unique splashover and combined splashover, respectively. For unique splashover the on-time difference ranges between -13/60 and 3/60 seconds (a negative value indicating that the splashover is shorter than the actual pulse). About 90% of unique splashover pulses have shorter on-times than the actual pulses. According to the concurrent video data, the difference of on-times appear to be related to the location of vehicles in lane 2 relative to the detection zone of lane 1: generally the closer the vehicle is to the lane line, the more positive the difference becomes. For combined splashover the on-time difference ranges between -6/60 and 24/60 seconds. About 90% of combined splashover pulses in lane 1 have longer on-times than the actual pulses in lane 2. The negative values arise when the combined splashover in lane 1 is still shorter than the original pulse in lane 2, e.g., Figure 3.2C.

Next, the difference in rising transition times (DRTT) and the difference in falling transition times (DFTT) between a splashover pulse in lane 1 and the valid pulse in the vehicle's lane of travel, lane 2, are used to understand characteristics of splashover. Figure 3.5 illustrates the various relationships between DRTT and DFTT. The plane is divided into four regions (I to IV) with boundaries at zero seconds on both axes. The boundary of each region is defined as follows:

Region I: $DRTT > 0$ and $DFTT > 0$,

Region II: $DRTT \leq 0$ and $DFTT > 0$,

Region III: $DRTT < 0$ and $DFTT \leq 0$,

Region IV: $DRTT \geq 0$ and $DFTT \leq 0$.

Hypothetical examples of the relationship are shown between the splashover pulse in lane 1 and the valid pulse in lane 2 for each region. One would expect most unique splashovers from lane 2 to lane 1 to be like Figure 3.1 and fall in region IV since the lane 2 pulse starts before and ends after the lane 1 pulse, while the combined splashovers

should fall in regions I and III. Using the data from Figure 3.4, Figure 3.6 shows the relation of DFTT and DRTT between splashover and actual detection at station 104 eastbound for the splashover pulse in lane 1 (L1) and the valid pulse in lane 2 (L2). Within each region the brackets tally the total number of observations of [unique splashover, combined splashover] seen. As expected, 86% of unique splashovers (236 out of 273) fall in region IV. The concurrent video data reveals that the unique splashover observed in Region II occurred when the lane 2 vehicle traveled much closer to the lane boundary with lane 1 than the vehicles observed in Region IV. A loop detector's sensitivity is in inversely proportional to the square of the distance to the vehicle (Klein et al., 2006). So, typically DZ_1 for splashover should become larger and DZ_2 for valid vehicle detection should become smaller as a lane 2 vehicle gets closer to the boundary with lane 1. Unique splashovers falling in region I and region III might be due in part to lateral motion by the vehicles, though we did not test this hypothesis. Meanwhile, 82% of combined splashovers are observed in Region I (18 out of 45) and III (19 out of 45). Not surprisingly, the concurrent video data reveal that in for the combined splashovers in region I a vehicle in lane 1 is followed by a vehicle in lane 2 causing splashover (e.g., Figure 3.3C), while in region III a vehicle in lane 2 causing splashover is followed by a vehicle in lane 1 (e.g., Figure 3.3D). Combined splashovers in region II and region IV usually occurred when two vehicles passed over loop detectors at roughly same time. In free flow conditions unique splashovers are more frequently observed than combined splashovers. In this case about 76% of all splashovers in lane 1 are bounded by the valid pulses from the actual lane of travel, lane 2 (e.g., Figure 3.1). These results are typical of the ground truthed stations with splashover for free flow conditions.

3.3 Development of An Algorithm to Identify Loop Detectors with Splashover

The task of detecting splashover is complicated by the fact that other events yield similar trends in the detector data. On the one hand, all four of the relationships shown in Figure 3.5 can arise from valid actuations in neighboring lanes if two vehicles pass side by side, i.e., a non-error. The frequency of these events depends on the demand in each lane. As speeds decrease due to congestion, the rate of these non-error events will go up

simply because vehicles will reside over the loop detectors for a longer time. As a result, the present work is intended for (predominantly) free flow conditions. Since splashover usually arises due to a hardware fault, it should not depend on traffic conditions. So one can use the macroscopic measures from the detectors to ascertain when conditions are free flowing and then apply this work. On the other hand, it is not uncommon for a vehicle changing lanes over a detector station to actuate the loop detectors in both lanes. These lane change maneuver errors are similar to splashover errors, but are slightly different since the vehicle is actually residing between the two lanes. In any event, occasional splashover or lane change maneuver errors (under 1% of the actuations) are within normal tolerance of conventional loop detector stations.

This research seeks to identify detectors exhibiting chronic splashover problems. As seen in Figure 3.6, a splashover pulse is usually bounded by the pulse from the actual lane of travel, e.g., in this case 86% of unique splashover and 76% of all splashovers. This feature of splashover is used as the starting point for the splashover detection algorithm. A pair of loop detectors in adjacent lanes is selected for testing, one loop is arbitrarily taken as the "source lane", and for the test is assumed to provide an accurate record of vehicle actuations in that lane. The other loop is taken as the "target lane," and is evaluated as to whether or not it may be recording splashover pulses in response to vehicles in the source lane. The algorithm checks each pulse in the source lane to see if it spans a pulse in the target lane (i.e., region IV in Figure 3.5). Any time the check is true it is considered to be a suspected splashover event. Figure 3.7 shows a hypothetical example where indeed the pulse in the source lane, s1, spans a pulse in the target lane, t1. The algorithm repeats the process over all pulses in the source lane for the data set (in this study the duration of the ground truth video sequence). The roles of the lanes are exchanged and the process is repeated, then it is repeated for every other pair of adjacent lanes at the station.

The first two rows of Table 3.1 summarize the number of true splashovers (from ground truth) and suspected splashovers (from the algorithm) for each pair of adjacent lanes at station 104 eastbound. With lane 2 as source lane and lane 1 as target lane 318

actual splashover are observed at lane 1 from vehicles in lane 2 and 240 pulses in lane 1 are labeled suspected splashover from lane 2 by the algorithm. In this case, all 240 of the suspected splashovers turn out to be actual splashover events. But a suspected splashover does not always correspond to an actual splashover, e.g., if two vehicles pass simultaneously in adjacent lanes. Taking lane 1 as source lane and lane 2 as target lane there are no actual splashover events but the algorithm found 32 suspected splashovers. In fact all of the lane pairs had some suspected splashover events but only one lane pair had actual splashovers.

Again, the non-splashover events that are labeled as suspected splashover arise due to two vehicles passing in adjacent lanes. The frequency of these events depends on the demand in the two lanes. To estimate the rate of these non-splashover events, we make a second pass through the data. We take all of the arrivals in the source lane, shift them by ε (set to five seconds), and tally how many pulses in the target lane are spanned by the time-shifted pulses from the source lane. These intersections cannot arise from the given pulse splashing over. The result is called the "background non-splashover." Figure 3.7 shows that target lane pulse t3 begins within the time-shifted window from the source lane pulse, but it is not spanned by the time-shifted pulse and so it would not contribute to this background non-splashover. In any event, if vehicle arrivals are independent in the two lanes and there were no splashover events, the expected number suspected splashover and background non-splashover should be the same. The background non-splashover is listed in the third row of Table 3.1 and the difference between suspected splashover and the background is tabulated in the fourth row. While the difference is closer to zero, all lane pairs are still positive. It is possible that the arrivals in adjacent lanes are not completely independent (e.g., drivers may momentarily slow down as they are overtaking a vehicle in an adjacent lane), but in any case, both the suspected splashovers and the background non-splashover are random variables that will not always cancel each other out in the absence of actual splashover events.

Since the present work seeks to identify chronic splashover, we adopt a more liberal definition for the background non-splashover rate. Instead of requiring the entire

pulse in the target lane to fall within the time-shifted source lane pulse, a target lane pulse will be counted in the background if the target pulse's rising edge is bounded by the shifted source lane pulse (i.e., regions I and IV in Figure 3.5). So now pulse t3 in Figure 3.7 would contribute to the background non-splashover. The results for the on-going example at station 104 are tabulated in the final two rows of Table 3.1. The liberal definition far outnumbers the suspected splashover that arose in the three lane pairs due to non-splashover events (negative numbers in the final row), but it does not outnumber the true splashovers from L2 to L1. Any lane pairs that have a positive value after subtracting the liberal background non-splashover rate are considered to be from a loop detector with splashover.

This process is formalized in Equation (3.1). A positive R_T^S during free flow traffic is an indicator of chronic splashover. Since the rate of two vehicles passing in adjacent lanes simultaneously depends on demand in both lanes, the magnitude of R_T^S is not in itself a fair measure of comparison between lanes.

$$R_T^S = \max\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(P_{ij} - Q_{ij}\right)}{n}, \quad 0 \right) \tag{3.1}$$

where,

$$P_{ij} = \begin{cases} 1, \; if \; RT_i^S \le RT_j^T \le FT_i^S \;\; \& \;\; RT_i^S \le FT_j^T \le FT_i^S \\ 0, \; otherwise \end{cases}$$

$$Q_{ij} = \begin{cases} 1, \; if \; RT_i^S + \varepsilon \le RT_j^T \le FT_i^S + \varepsilon \\ 0, \; otherwise \end{cases}$$

and

R_T^S = adjusted ratio of suspected splashover between source lane (S) and target lane (T),

P_{ij} = suspected splashover of pulse j in the target lane matched to pulse i in the source lane,

Q_{ij} = background non-splashover of pulse j in the target lane matched to pulse i shifted by the constant delay in the source lane,

ε = constant delay for shifting a pulse in the source lane, currently set to five seconds,

n = total number of pulses in the source lane,

m = total number of pulses in the target lane,

i = i-th pulse in the source lane (i=1, 2, ... , n),

j = j-th pulse in the target lane (j=1, 2, ... , m),

RT_i^S = i-th pulse rising transition time in source lane,

FT_i^S = i-th pulse falling transition time in source lane,

RT_j^T = j-th pulse rising transition time in target lane,

FT_j^T = j-th pulse falling transition time in target lane

3.4 Correction by Daily Median On-Time

The preceding analysis assumes that a given target and source lane have roughly the same sensitivity level. This sensitivity depends on the hardware instillation and the settings of the loop sensor. If the sensitivity in the target lane is significantly greater than the source lane, it is possible to see an inversion, e.g., most unique splashovers in lane 1 from lane 2 falling in region II of Figure 3.5A. In which case the algorithm will detect the splashover in the lane pair, but attribute it to the wrong lane. Fortunately, such extreme cases can be detected via the median on-time test presented in Coifman (2006a) (a variant of the mode on-time test presented in Coifman and Lee, 2006). Each on-time measurement depends on vehicle length, vehicle speed, and the detector sensitivity. Although the speed and length vary from vehicle to vehicle, over a 24 hour period at a

typical detector, most of the vehicles should be free flowing and the majority should be passenger vehicles. As a direct result, the daily median on-time should usually fall within a small range that corresponds to the effective length of a passenger vehicle (i.e., physical length plus detection zone) at free flow speed. The detection zone length is a function of the detector sensitivity, so the daily median on-time will increase with the detector sensitivity.

For instance, the on-time from a 20 ft effective length vehicle traveling at 65 mph is 13/60 seconds and the daily median on-time is expected to be around that value. Assuming that most effective vehicle lengths indeed fall in the 18 to 22 ft range and drivers usually obey the posted speed limit in free flow conditions, the daily median on-time should fall between 11/60 to 14/60 seconds at 65 mph and 13/60 to 16/60 seconds at 55 mph. If the daily median on-time falls much outside the expected range, it is indicative of a transient event (e.g., a snowstorm) or improper loop sensitivity. Transient events can be addressed by looking at the results from several days or avoiding results on days with known incidents. If the location is known to have many hours of recurring congestion, the test can be modified to exclude congested traffic (either by time of day, day of week, or via the macroscopic data).

To correct these speed estimation errors at single loop detectors we calculate a multiplicative correction factor individually for each loop, as follows. First, the daily median speed is taken from speed in off-peak time periods (9a am-3 pm), and this process is repeated for all weekdays in a month. Next, the median of the daily speeds is found. The correction factor is then defined as the posted speed limit divided by this daily median speed. For example, in Coifman (2006a), the median speed in lane 2 at station 26 northbound is 81 mph, so the correction factor is 0.8 (the quotient of 65 mph and 81 mph). Naturally one could use a radar gun or other measurement device to validate the speeds more accurately, which is exactly what we did. To verify the process of generating correction factors is valid, the corrected single loop speeds are compared against the corresponding GPS velocity measurements from the probe vehicle runs. After applying the correction factors, most of the single loop detectors report speeds close to the GPS

velocity. Ultimately the correction factors simply reflect the fact that given the detector's sensitivity, the true effective vehicle length differs from 20 ft. Based on the correction factors for single loop detectors, the average effective vehicle length can be estimated. These estimated average effective vehicle lengths can improve the single loop detector speed estimates.

At the moment we are merely concerned about assigning splashover to the correct lane. When a detector is suspected of splashover, one should find the daily median on-time from the source and target lanes to ensure that the apparent source lane is not considerably more sensitive than the apparent target lane. If there is a large discrepancy, then the splashover may be in the opposite direction than indicated by Equation 3.1. In this study we did not observe any such extreme cases. Consider each loop at station 104 eastbound over 24 hrs. The posted speed limit is 55 mph. Daily median on-times are 14/60 seconds for lane 1, 16/60 seconds for lane 2, and 16/60 seconds for lane 3. The slightly lower median on-time in lane 1 is due in part to the large number of shorter pulses from the lane 2 vehicles splashing over.

3.5 Application and Results

A total of 19 directional ground truth data sets were generated in free flow conditions for this evaluation. Four of these directional sets exhibited some degree of actual splashover, as enumerated in Table 3.2. The remaining 15 data sets did not have any observed splashover. The total pulses listed in Table 3.3 tally the number of pulses recorded by the loop detector during the video data collection at the stations that exhibited splashover, while the total vehicles tally the corresponding number of vehicles that traveled in the lane as seen in the video. The total number of splashover events are reported for the given lane where the vehicle was incorrectly detected (i.e., the target lane) as well as the subtotals for unique splashover and combined splashover. The total pulses do not always correspond to the sum of total vehicles and unique splashovers, the remaining 106 extra pulses are due to vehicles changing lanes and being counted in both lanes (note that all recorded pulses, including those due to lane change maneuvers, are

included in Table 3.3). The source and target lanes giving rise to the splashover are shown in the second to the last column. The final column shows the percentage of splashovers in the target lane relative to total pulses in the source lane, i.e., the splashover rate. For example, lane 3 at station 38 westbound has 115 unique splashovers and 2 combined splashovers, all of which are caused by vehicles traveling in lane 2. So this lane has a splashover rate of 117/242.

Note that the splashover rate is relative to the total pulses rather than the total vehicles. As noted earlier, the source lane is assumed by the algorithm to provide an accurate record of vehicle actuations in that lane and in practice, one could not exclude all of the extra pulses due to splashover without an independent ground truthing process. However, if this point proves to be one of concern, one could reduce the denominator in the splashover rate by the number of suspected splashovers seen in the source lane. In any event, lane 1 at station 104 eastbound has the highest splashover rate (91%), and lane 2 at station 38 westbound has the lowest non-zero splashover rate (1.2%). There was a total of 537 splashovers over the four detector stations, 473 of which (88%), are unique splashover. So 473 out of 3,756 total pulses resulted from vehicles being counted a second time across these four stations, i.e., 12.6% over counting. As previously mentioned, while the combined splashover events do not impact flow, both types of splashover will lead to an erroneous increase in occupancy.

Table 3.4 presents the results of the splashover detection algorithm (i.e., R_T^S from Equation 3.1) applied to the loop detector data for the periods with ground truth. Of the 82 adjacent lane pairs from 56 loop detectors at 13 loop detector stations, a total of five lane pairs returned a positive ratio of adjusted suspected splashover, R_T^S. Although not shown, the comparison of daily median on-times at 82 adjacent lane pairs verified that there is not an extreme difference of loop sensitivity. Now employing the ground truth data from Table 3.2, the seven lane pairs that actually exhibited splashover are shaded in Table 3.4. The algorithm failed to identify two of the lane pairs with actual splashover as loops with splashover, specifically, splashover in lane 2 from lane 1 at station 38 westbound and at station 41 eastbound. The algorithm correctly classified all loop

detectors without splashover. The splashover rate from Table 3.3 is an upper bound on the total contribution of P_{ij} to Equation 3.1. The two lane pairs with splashover that were missed by the algorithm had a relatively small splashover rate. On the other hand, the flow in the target lane was significantly greater than the flow in the source lane, thereby increasing the chance of finding a background non-splashover event. In the end, the splashover rate was exceeded by the contribution of the liberal background non-splashover rate, Q_{ij}. These two cases represent the threshold of "chronic splashover" that the algorithm can detect. In contrast, lane 3 at station 56 westbound also has a low rate of splashover but the algorithm correctly identifies this lane because the flow in the target lane is about 20% of source lane and thus, Q_{ij} is small in this case.

As a result of this analysis, our team asked the operating agency (the Ohio Department of Transportation) to reduce the sensitivity setting on the detectors at station 56 westbound and station 104 eastbound. A second round of ground truth data was collected for each station after the change: 30 min at station 56 and 15 min at station 104 in free flow conditions. As discussed in Section, 4.7, no splashover events were found in the ground truth data at either station after the change and the algorithm labeled all of the lanes at those stations as being non-splashover.

3.6 Comparison of The Performance of Splashover Detection Algorithms

We compare the performance of three of the earlier error detection methodologies against our algorithm using the ground truth data from Table 3 (the process was repeated using a data from a 24 hr period selected at random and the results were similar). At each detector: Chen and May (1987) (C&M) tabulated the percent of individual actuations with an off-time under 15/60 seconds; Jacobson et al. (1990) (JNB) tabulated the percent of macroscopic data (20 sec samples) outside of the acceptable thresholds; Turochy and Smith (2000) (T&S) tabulated the percent of macroscopic data (30 sec samples) with flow greater than 3,100 vehicles/hr; and our method (L&C) as described above. The detectors were segregated into two groups, the seven with splashover, and the 53 without (non-splashover). Within each group, the min, max, mean, and median values were found

for each test. The results are shown on Figure 3.8 and are tabulated in Table 3.5. One should not compare absolute values between methodologies since they measure different features; rather, consider the relative values between splashover and non-splashover for a given methodology. Only T&S and L&C have a zero mean or median for the non-splashover detectors. But the difference between the splashover and non-splashover detectors is small for T&S. While L&C is the only test to have a zero maximum for the non-splashover tests. Our test exhibits the largest difference between the splashover and non-splashover conditions using the mean. Using the median JNB exhibits a slightly larger difference than L&C does between the two conditions, but JNB incorrectly catches several non-splashover loop detectors. JNB also exhibits an inversion from the median value to maximum value. These results are not surprising since, as noted previously, the other tests were not specifically designed to identify splashover.

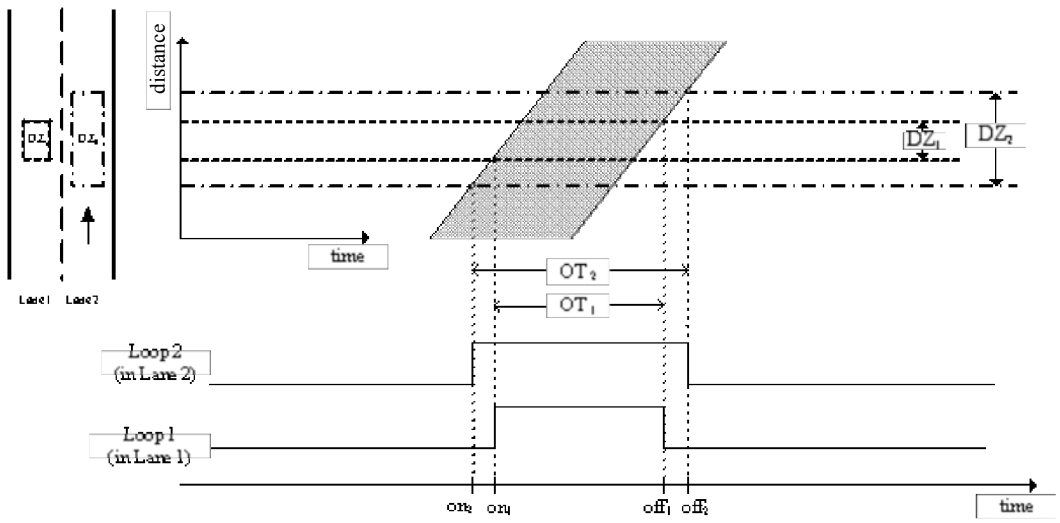


Figure 3.1, A hypothetical example of splashover from lane 2 to lane 1

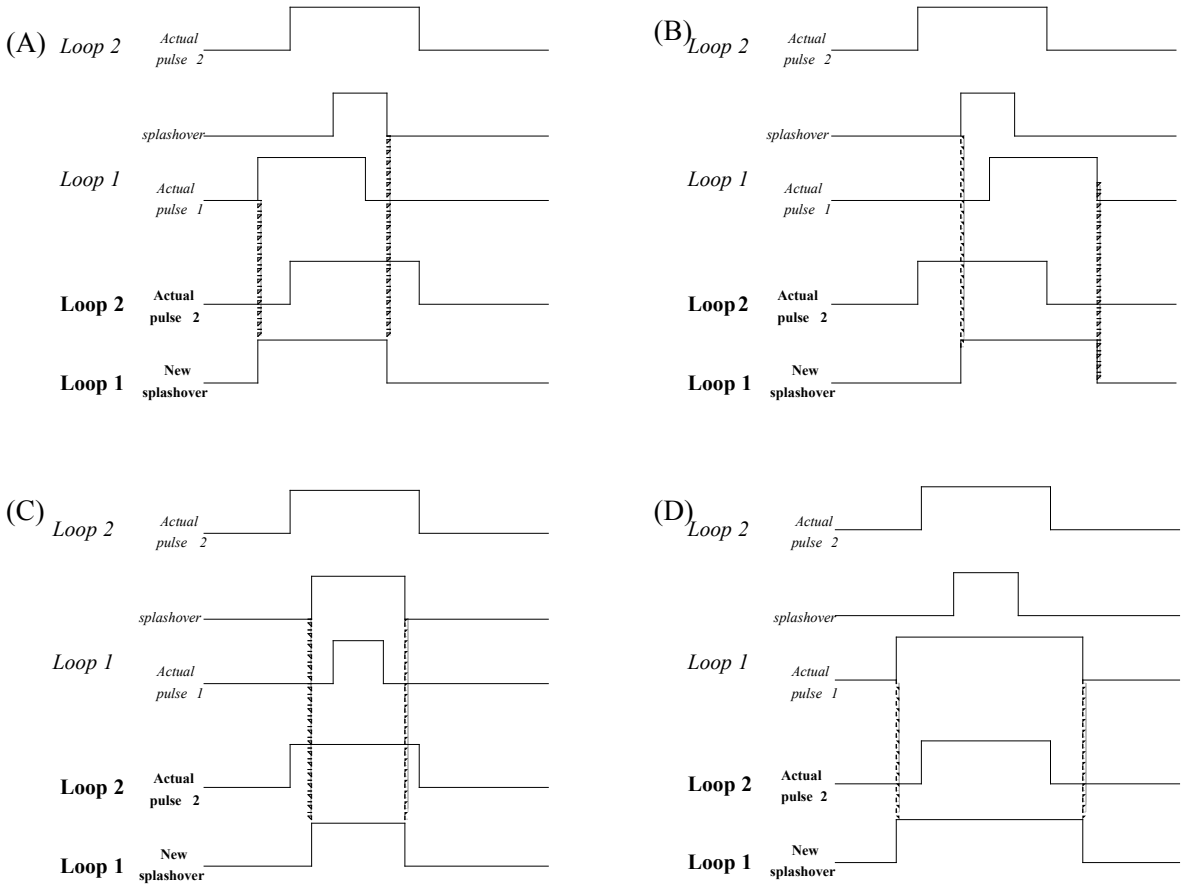


Figure 3.2, Hypothetical examples of the coupling effect of actual detection and splashover.

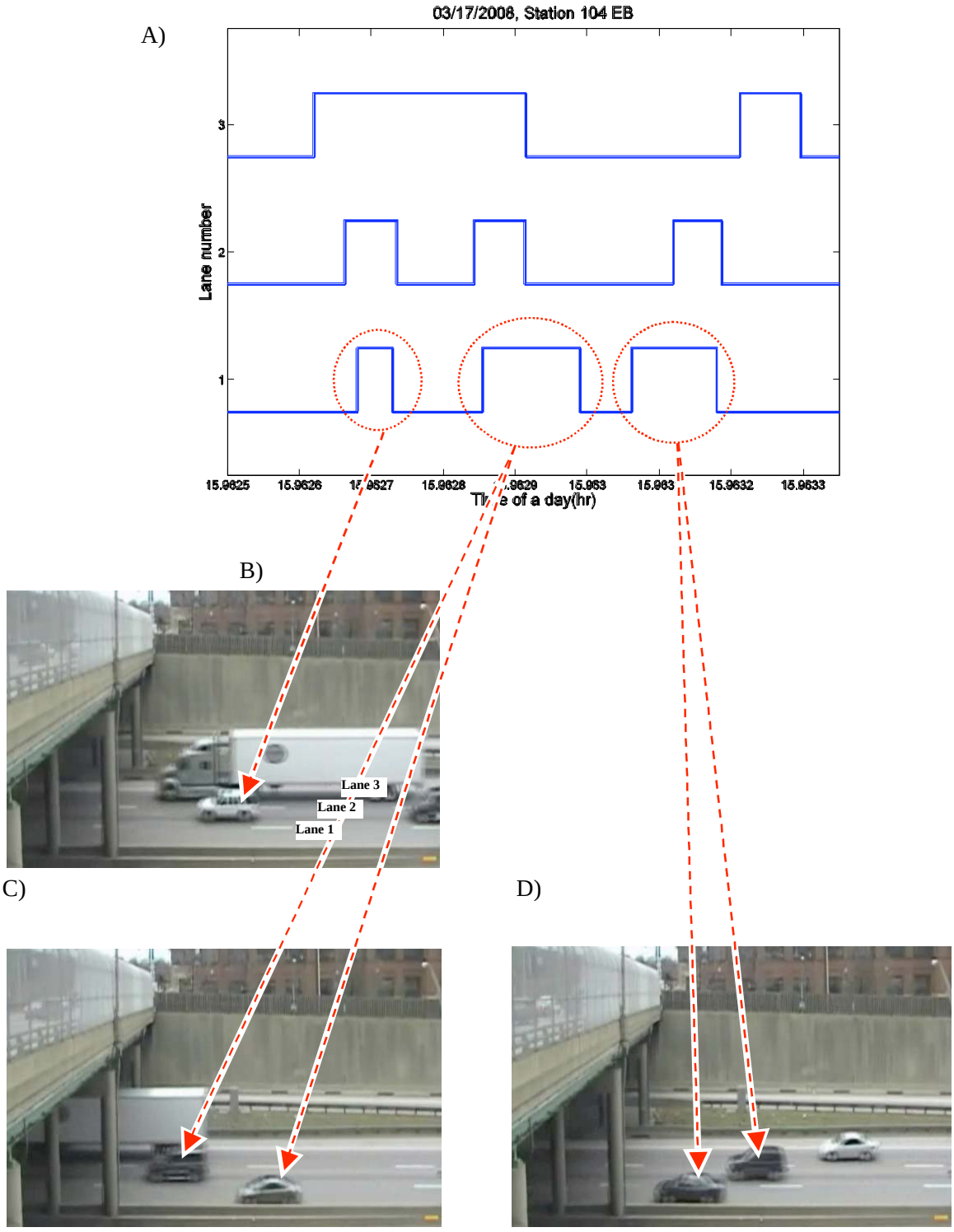


Figure 3.3, (A) A plot of detector actuations, and (B-D) the corresponding video image at station 104 eastbound; (B) unique splashover and (C and D) combined splashover

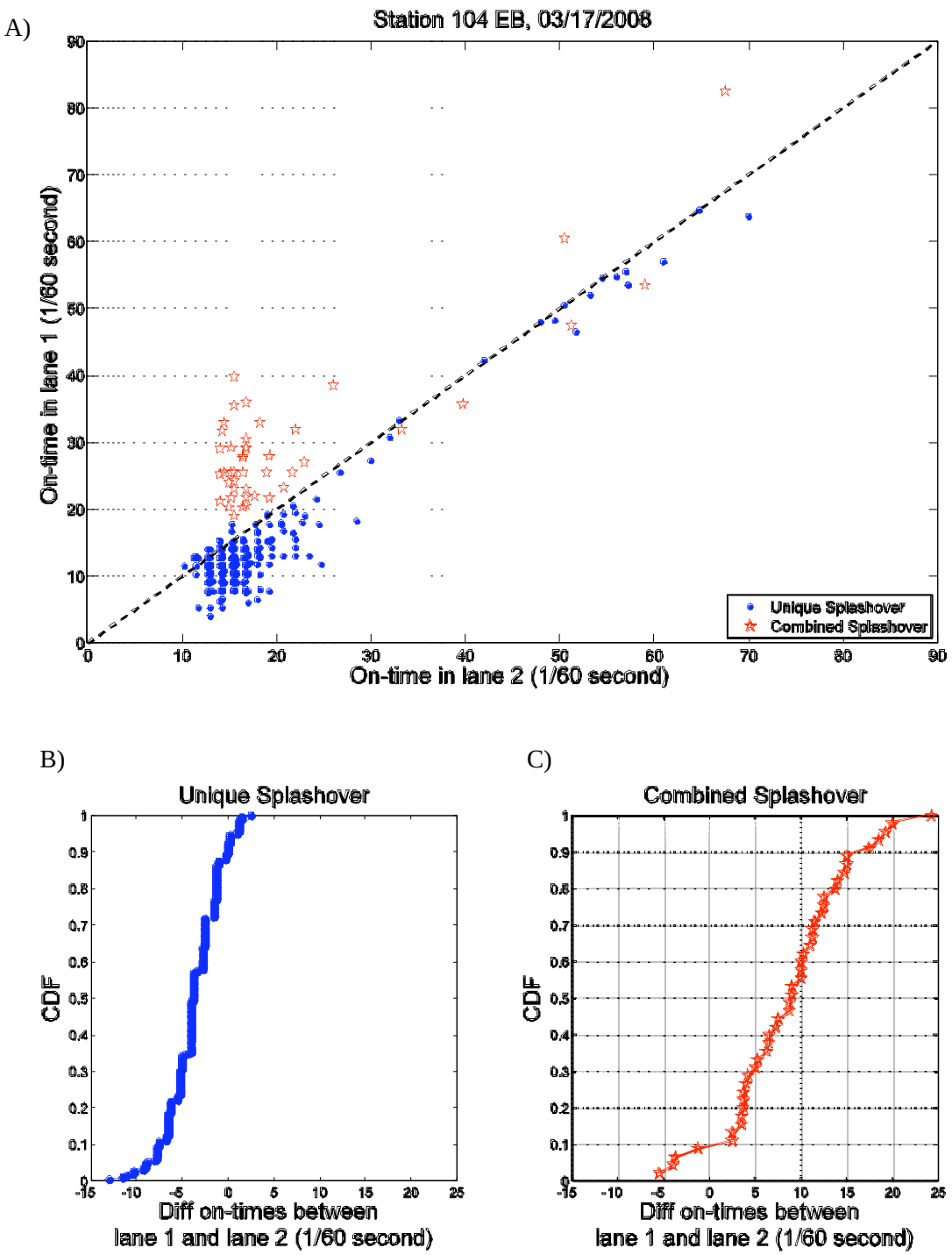


Figure 3.4, (a) A scatter plot of on-times in unique splashover and combined splashover, (b) CDF of the difference of on-time in unique splashover and (c) combined splashover:

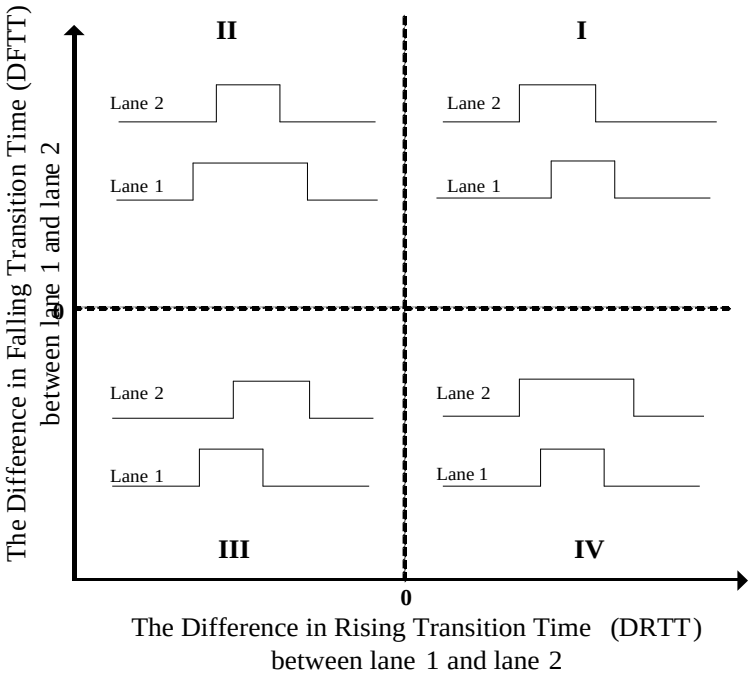


Figure 3.5, The expected relationship between the difference in falling transition times and the difference in rising transition times: (lane 1 - lane 2)

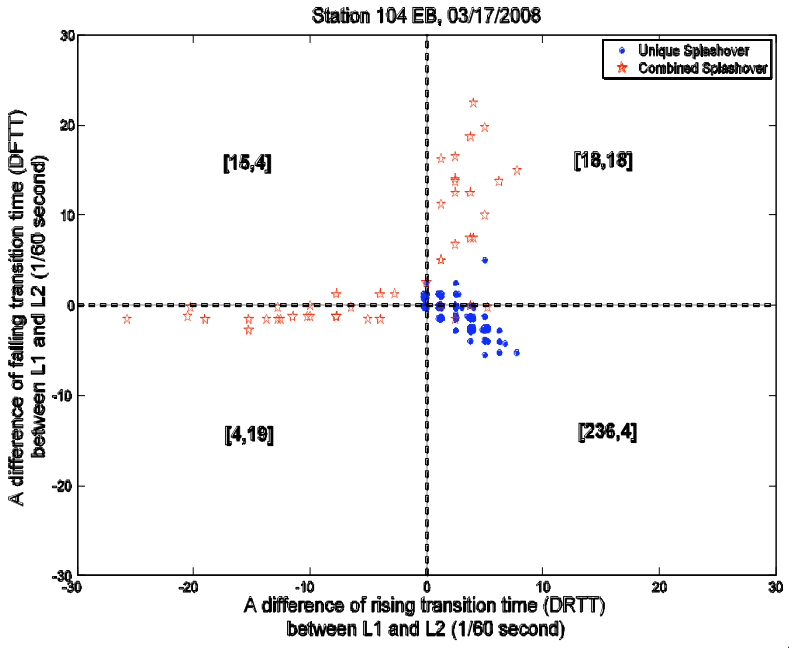


Figure 3.6, Scatter plot of a different time in rising and falling time in splashover from GTD; Within each region the brackets tally the total number of observations of [unique splashover, combined splashover]

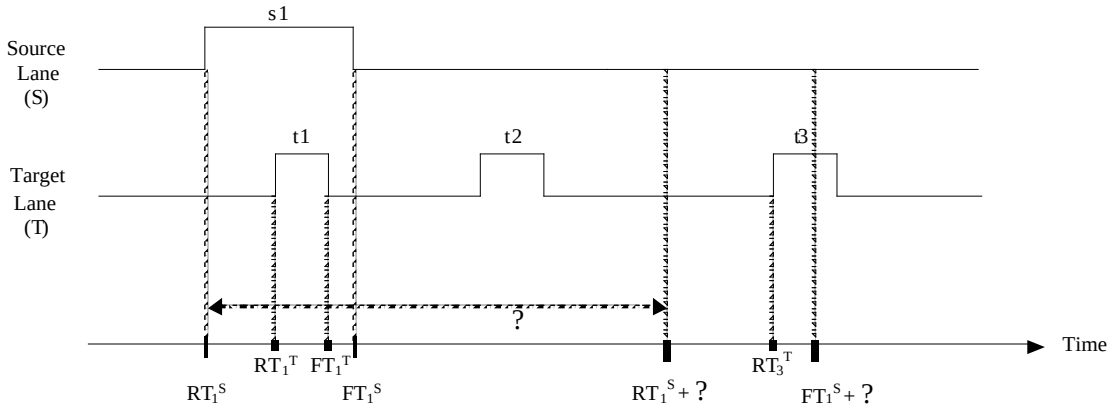


Figure 3.7, The splashover detection algorithm to select suspected splashover and background non-splashover.

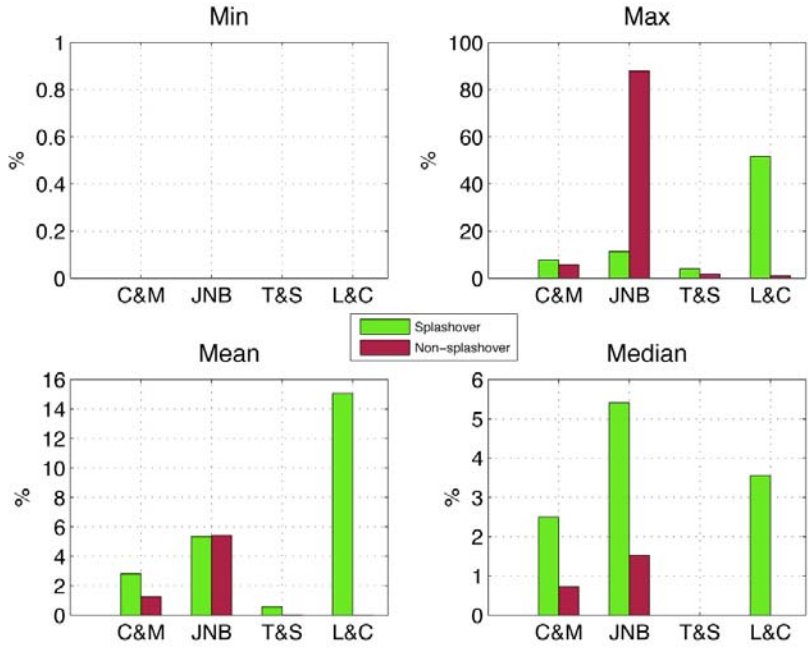


Figure 3.8, Bar chart comparing the max, min, mean, and median results for detectors with splashover and non-splashover from the four error detection methods. Note that vertical scales are differ between the plots.

Criteria of background non-splashover	St 104 EB	[Source lane (i) → Target lane (j)]			
		L1 → L2	L2 → L1	L2 → L3	L3 → L2
(A): Pulse	Actual splashover	0	318	0	0
	Suspected splashover (I)	32	240	7	35
	Background non-splashover (II)	10	14	6	24
	(I) - (II)	22	226	1	11
(B): Rising transition	Background non-splashover (III)	61	60	47	74
	(I) - (III)	-29	180	-40	-39

Table 3.1, Application of the splashover detection algorithm to station 104 eastbound

the presence or the absence of splashover	Station number	Dire-ction	Number of lanes	Date	Start Time (hh:min)	End Time (hh:min)	Duration of time (hh:min)
With splashover	38	WB	4	09/09/2008	12:05	12:25	0:20
	41	EB	2	09/09/2008	11:00	11:35	0:35
	56	WB	3	11/21/2008	09:00	09:40	0:40
	104	EB	3	03/17/2008	16:00	16:10	0:10
Without splashover	2	NB	4	03/09/2009	17:21	17:50	0:29
	3	NB	4	03/17/2008	10:57	11:20	0:23
	3	SB	4	04/18/2008	15:55	16:55	1:00
	4	SB	4	03/17/2008	10:15	10:35	0:20
	6	NB	3	04/18/2008	15:55	16:55	1:00
	9	NB	3	06/05/2006	12:20	14:20	2:00
	9	SB	3	06/05/2006	12:20	14:20	2:00
	15	NB	3	03/10/2009	17:18	17:47	0:29
	18	NB	3	03/09/2009	08:24	08:57	0:33
	19	NB	3	03/17/2008	09:25	09:40	0:15
	31	NB	4	11/21/2008	10:35	11:05	0:30
	38	EB	3	08/29/2008	15:05	15:25	0:20
	43	EB	3	09/02/2008	08:50	09:15	0:25
	43	WB	3	09/02/2008	08:50	09:15	0:25
	56	EB	3	09/03/2008	16:40	17:25	0:45
	102	EB	3	03/10/2009	17:05	17:20	0:15
	104	WB	3	03/12/2009	17:00	17:18	0:18

Table 3.2, Information of the ground truth data used in this experiment

Condi-tion	Station # (Dire-ction)	Lane	Total Pulses	Total vehicles	Number of splashover			Mechanism of splashover [Source lane (i) → Target lane (j)]	% splashover rate
					Total	Unique splash-over	Combined splash-over		
Free flow	38 (WB)	1	172	172	0	0	0	-	-
		2	242	235	2	2	0	L1 → L2	1.2%
		3	206	90	117	115	2	L2 → L3	48.3%
		4	56	39	17	17	0	L3 → L4	8.3%
	41 (EB)	1	336	274	53	39	14	L2 → L1	10.5%
		2	506	475	11	8	3	L1 → L2	3.3%
	56 (WB)	1	345	340	0	0	0	-	-
		2	632	610	0	0	0	-	-
		3	121	84	19	19	0	L2 → L3	3.0%
	104 (EB)	1	441	164	318	273	45	L2 → L1	91.1%
		2	349	347	0	0	0	-	-
		3	350	347	0	0	0	-	-
Total			3,756	3,177	537	473	64		

Table 3.3, Summary of the ground truth data with splashover in free flow

Condition	Station #	Dire-ction	A ratio of adjusted suspected splashover: R_T^S [**Source lane (S)** → Target lane (T)]					
			L1→ L2	L2→ L1	L2→ L3	L3→ L2	L3→ L4	L4→ L3
Splashover	38	WB	0%	0%	41.3%	0%	6.8%	0%
	41	EB	0%	3.6%	-	-	-	-
	56	WB	0%	0%	2.2%	0%	-	-
	104	EB	0%	51.6%	0%	0%	-	-
Non-splashover	2	NB	0%	0%	0%	0%	0.6%	0%
	3	NB	0%	0%	0%	0%	0%	0%
	3	SB	0%	0%	0%	0%	0%	0%
	4	SB	0%	0%	0%	0%	0%	0%
	6	NB	0%	0%	0%	0%	-	-
	9	NB	0%	0%	0%	0%	-	-
	9	SB	0%	0%	0%	0%	-	-
	15	NB	0%	0%	0%	0%	-	-
	18	NB	0%	0%	0%	0%	-	-
	19	NB	0%	0%	0%	0%	-	-
	31	NB	0%	0%	0%	0%	0.9%	0%
	38	EB	0%	0%	0%	0%	-	-
	43	EB	0%	0%	0%	0%	-	-
	43	WB	0%	0%	0%	0%	-	-
	56	EB	0%	0%	0%	0%	-	-
	102	EB	0%	0%	0%	0%	-	-
	104	WB	0%	0%	0%	0%	-	-

Table 3.4, Percentage of adjusted suspected splashover relative to source lane. Light shaded cells indicate a loop detector with splashover verified from the ground truth data, dark shaded cells are those with unexpected results. All of the non-shaded cells represent detectors that did not exhibit splashover in the ground truth data.

Methods	Data	Min	Max	Mean	Median
C&M	Splashover	0.0%	7.7%	2.8%	2.5%
	Non-splashover	0.0%	5.7%	1.2%	0.7%
	Difference	0.0%	2.0%	1.6%	1.8%
JNB	Splashover	0.0%	11.4%	5.3%	5.4%
	Non-splashover	0.0%	87.9%	5.4%	1.5%
	Difference	0.0%	-76.6%	-0.1%	3.9%
T&S	Splashover	0.0%	4.0%	0.6%	0.0%
	Non-splashover	0.0%	1.6%	0.0%	0.0%
	Difference	0.0%	2.4%	0.5%	0.0%
L&C	Splashover	0.0%	51.6%	15.1%	3.6%
	Non-splashover	0.0%	0.9%	0.0%	0.0%
	Difference	0.0%	50.7%	15.1%	3.6%

Table 3.5, Comparison of the max, min, mean, and median results for detectors with splashover and non-splashover

CHAPTER 4. **AN ALGORITHM TO IDENTIFY PULSE BREAKUP**

Before addressing pulse breakups, it is important to correct several other detector errors that might be present. If the loop detector sensitivity is significantly too high or too low, the assumed effective vehicle length (mean or median) will yield inaccurate speed estimates. So the daily median on-time should be tracked and be adjusted by setting the monthly median speed to the expected value (see Section 3.4, Correction by Daily Median On-Time). Of course splashover or other stray pulses could also look like pulse breakup, so it is important to eliminate those errors (via the previous chapter) before proceeding to address pulse breakup.

4.1 Problems of Pulse Breakup

In the absence of detector errors, a single vehicle is recorded as a single pulse with a rising transition and a falling transition. Sometimes however, what should be a single pulse from a vehicle breaks up into two or more pulses. Pulse breakup most often occurs when multi-unit vehicles, e.g., trucks or vehicles with trailers, pass over a loop detector (Cheevarunothai et al., 2007). An example of this error is evident in the comparison between loop detector data and concurrent video shown in Figure 4.1, in lane 2 at a single loop detector station (station 9 northbound). Figure 4.1A shows the pulses from all three lanes as detectors respond to vehicles. Figure 4.1B shows that the two pulses in lane 2 result from a single truck passing over the loop detector. As illustrated in Figure 4.1C, the on-time (OnT) denotes the period when the loop detector should have been occupied by the truck in the absence of pulse breakup. But in the recorded data OnT is divided into two distinct on-times (OnT_1 and OnT_2) and one off-time ($OffT_1$). It is clear from this figure that pulse breakup causes flow to be high and occupancy to be low. In

addition, the inaccurate on-time will cause inaccurate speed estimates. Even if we were able to estimate speed correctly (e.g., via a dual loop detector), the pulse breakup will lead to an inaccurate estimated vehicle length.

4.2 Limitation of Previous Research

Pulse breakups in previous studies were detected using a threshold of the time gap (Chen and May, 1987) or the time headway (Cheevarunothai et al, 2007) between two consecutive pulses. Both methods ultimately use short off-time as the indicator of pulse breakups. Figure 4.2, shows a histogram of off-times corresponding to pulse breakups by lane at station 9 northbound on 6/05/2006 during 2 hrs in free flow condition. In these data, all of the off-times arising from pulse breakup are less than 20/60 seconds. The largest such off-times are only 5/60 second longer than the threshold of pulse breakup used in Chen and May (1987). While short off-time might be a good indicator to find pulse breakups, a short off-time does not always correspond to pulse breakup. A short off-time can also arise due to tailgating and other maneuvers. Meanwhile, when traffic is congested, the resulting off-time in a pulse breakup could easily exceed a static boundary used to find pulse breakups, as mentioned in Chen and May (1987).

Figure 4.3 shows the CDF of off-times from the ground truth data after excluding pulse breakups, i.e., manually verified to be true and valid off-times. If an off-time threshold of 20/60 seconds were used to detect pulse breakups in these data, some real off-times would be considered erroneous and attributed to pulse breakups. For example, 3% of the valid data in lane 1 will erroneously be marked as pulse breakups in this case. The concurrent video data reveal that two consecutive pulses with short off-time are often related to two actual vehicles.[2]

The off-time arising from pulse breakup during congestion should be larger than that during free flow. Figure 4.4 shows a CDF of off-time corresponding to pulse

---

[2] Of course there may be other detector errors that are causing the short off-times, e.g., sticking on. Whenever possible, those external sources of error should be corrected. But they are not pulse breakup.

breakups by a lane at station 3 northbound during 1 hr on 4/18/2008 in congested conditions. Off-times in pulse breakups during congestion span a much larger range than during free flow conditions. For instance, off-time in pulse breakup at lane 2 is distributed in a range of 4/60 to 135/60 seconds, with only 30% of pulse breakup off-times falling below 20/60 seconds. Obviously there is a trade-off, a larger off-time threshold should catch more pulse breakups, but more non-pulse breakups can be mistakenly considered as pulse breakups. Simply using short off-time as an indicator of pulse breakup is not sufficient to identify pulse breakups in free flow and congested conditions.

4.3 Development of Algorithm to Identify Pulse Breakup for a Single Loop Detector

The pulse breakup detection algorithm is designed to identify pulse breakup from individual vehicle actuation data. The method is based on the nature of pulse breakups revealed from video recorded ground-truth data. Ground truth data at station 9 northbound during a 2 hr sample period is the primary data set used for this development. After ground truthing data at this station, the following totals were observed from video in lane 1 to lane 3: 2,372, 2,689, and 2,182 vehicles, respectively. Reviewing the detector data, totals 306 out of 7,243 vehicles cause pulse breakups. Of these pulse breakup events, 298 vehicles, i.e., 97% of the pulse breakups arise from LV (semi-trailer trucks), while 8 arise from a MV (single unit truck) pulling a trailer, with a combined length below 42 ft. All observed pulse breakups at this location consist of two pulses, and our algorithm focuses on identifying a pulse breakup consisting of two pulses. Like the earlier works, we begin with a simple threshold on the off-time, but then include several comparisons of the adjacent on-times with respect to traffic condition, as follows.

4.3.1 **Dynamic Off-Time**

As seen above, short off-time should be a good indicator to identify pulse breakup during free flow conditions, but the same threshold will do a poor job during congestion, and any static threshold will have a high error rate during congestion. Pulse breakup often arises from a semi-trailer truck, e.g., Figure 4.1. The ground clearance of a typical semi-

trailer truck's undercarriage changes significantly over the length of the vehicle, it is relatively close to the ground under the tractor, but then rises significantly under the trailer, only to come close to the ground once more with the trailer's axles. Since a loop detector's sensitivity is inversely proportional to the square of the distance of the vehicle's undercarriage, loop detectors are more likely to drop out right after the tractor passes and the ground clearance jumps up to the bottom of the trailer.

Assuming the speed of a vehicle passed over a loop detector is almost constant, two on-times and off-time from the vehicle in free speed (V_f) and in lower speed (V_c) could be expressed as follows:

[Free flow conditions]

$$OnT_1 = \frac{D_1 + DZ}{V_f} = \frac{EVL_1}{V_f}$$

$$OnT_2 = \frac{D_2 + DZ}{V_f} = \frac{EVL_2}{V_f} \qquad (4.1)$$

$$OffT = \frac{D_3 - DZ}{V_f}$$

[Congested conditions]

$$OnT_1' = \frac{D_1 + DZ}{V_c} = \frac{EVL_1}{V_c}$$

$$OnT_2' = \frac{D_2 + DZ}{V_c} = \frac{EVL_2}{V_c} \qquad (4.2)$$

$$OffT' = \frac{D_3 - DZ}{V_c}$$

Where,

EVL_i = Effective vehicle length associated with pulse i,

D_1 = the physical length of the object associated with pulse 1

D_2 = the physical length of the object associated with pulse 2

D_3 = the physical length of the object associated with the gap

DZ = a size of detection zone,

V_f = free flow speed,

V_c = congested speed

From Equation 4.1 and Equation 4.2,

$$\frac{V_c}{V_f} = \frac{OffT}{OffT'} = \frac{OnT_1}{OnT_1'} = \frac{OnT_2}{OnT_2'}$$

In other words,

$$OffT' = \frac{V_f}{V_c} \times OffT$$
$$= \frac{OnT_1'}{OnT_1} \times OffT \; = \frac{OnT_2'}{OnT_2} \times OffT \qquad (4.3)$$

Equation 4.3 shows the relationship of off-time in a pulse breakup between congestion and free flow conditions. As one would expect, the off-time is greater in congestion because speed is lower than free speed ($V_f / V_c > 1$). Since speeds can not be measured at single loop detectors, the ratio of speeds can be replaced by the ratio of on-times between both traffic conditions.

To obtain the feasible off-time of pulse breakup in congestion from Equation 4.3, it is necessary to know free speed and off-time of a pulse breakup in free flow conditions. First, speed from a single loop detector is estimated from effective vehicle length divided by median on-time in the given sample period (Coifman et al., 2003). As the length and speed cannot be measured directly at a single loop detector, effective vehicle length is usually assumed to be some constant value, e.g., 20 ft. Of course one must first make sure

the assumed effective vehicle length is accurate, via the correction factors in Section 3.4. We expect median speed in off-peak time periods to usually correspond to follow free-flow speed, for this work we use 9 hr to 15 hr. Assuming free speed is nearly constant across vehicles, the off-time of pulse breakup in free flow just depends on the length of the portion of the vehicle that is undetected. As shown in Figure 4.2, off-times in pulse breakups at station 9 northbound are distributed in the range of 4/60 second to 20/60 second. So for this study we set the maximum off-time of suspected pulse breakup in free flow to be 20/60 second (as per Figure 4.3, any larger and the threshold would start selecting a large number of non-pulse breakup events).

Returning to Equation 4.3, the threshold off-time of a pulse breakup in congestion is scaled up by the factor of V_f/V_c, where V_c comes from the assumed effective vehicle length divided by the median on-time over a window of a fixed number of pulses (41 pulses in this study), centered on the current vehicle. The off-time of a pulse breakup in congested condition is expressed by Equation 4.4. As a result, the off-time threshold of suspected pulse breakup in congested condition just depends on the median on-time in 41 consecutive pulses, and we call it the "dynamic off-time".

$$
\begin{aligned}
OffT' &= \frac{V_f}{V_c} \times OffT \\
&= \frac{\dfrac{20\,ft}{median \;\; on-time^{(off-peak\ time\ period)}}}{\dfrac{20\,ft}{median \;\; on-time^{41 vehicles}}} \times OffT \\
&= \frac{median \;\; on-time^{41 vehicles}}{median \;\; on-time^{(off-peak\ time\ period)}} \times OffT
\end{aligned}
\tag{4.4}
$$

4.3.2 **The Ratio of On-Times**

Since many of the pulse breakups arise in the middle of semi-trailer trucks, OnT_1 and OnT_2 in Figure 4.1C should be proportional to the length of the tractor and the trailer axles, respectively. After including DZ, the tractor is typically about twice as long as the

trailer axles. Figure 4.5A shows the relation of on-times between the following pulse and the preceding pulse for the station 9 northbound pulse breakups (recall that these are in free flow conditions). The dashed line shows the set of points where the two on-times are equal. In general, the preceding pulses have a longer on-time than the following pulses. The ratio of on-times can be used to highlight the difference of on-times. The ratio is used rather than the difference because the ratio of on-times is less constrained by traffic conditions. From Equation 4.1, the ratio of two on-times in a pulse breakup can be expressed via Equation 4.5. The ratio of on-times corresponds to the ratio of two effective vehicle lengths. Assuming that the composition of the vehicle fleet does not change significantly between free flow and congestion, the feasible boundary of on-time ratio during free flow time periods should also be applicable during congested time periods.

$$\frac{OnT_2}{OnT_1} = \frac{\frac{EVL_2}{V}}{\frac{EVL_1}{V}} = \frac{EVL_2}{EVL_1} \tag{4.5}$$

Figure 4.5B shows the CDF of the ratio of on-times between two pulses in a pulse breakup, i.e., OnT_2/OnT_1. This on-time ratio ranges between 0.12 and 1.74, for about 99% of pulse breakups, $OnT_2 < OnT_1$. But this pattern also arises when a short vehicle follows a long vehicle. Figure 4.6A shows CDFs of the ratio from Equation 4.5 for pulse breakups and separately for successive non-pulse breakup on-times in the ground truth data at station 9 northbound. Choosing a threshold on the ratio at 1, 99% of pulse breakups are correctly selected, but 50% of non-pulse breakups are erroneously marked as pulse breakup as well. Taking the difference of the two CDFs, Figure 4.6B, we select the ratio corresponding to the maximum difference between the two functions, i.e., 0.72.

Figure 4.7 shows a scatter plot of off-time versus on-time ratio for the pulse breakups at station 9 northbound, showing the 8 MV separate from the LV results. The observations to the left of the vertical delineation (at an on-time ratio of 0.72) satisfy both the on-time ratio and off-time threshold. The on-time ratios from most of LV with pulse breakups are less than 0.72, while on-time ratio of most MV are greater than 0.72. But the

MVs with pulse breakup have a relatively short off-time, less than 10/60 second,[3] in fact the maximum off-time of a pulse breakup from a MV when on-time ratio is greater than 0.72 is 6/60 second. If a second, more restrictive threshold of 6/60 seconds off-time is used independent of the on-time ratio, then 50% of the pulse breakups undetected by the first two criteria (i.e., on-time ratio < 0.72 and off-time < 20/60 seconds) are caught. Since minimum off-time from non-pulse breakups is 8/60 second, the additional condition, on-time ratio > 0.72 and off-time <6/60 second, does not increase any false detection of pulse breakup.

4.3.3 **Ratio of Off-Time and the Preceding On-Time**

Figure 4.8A shows a scatter plot of $OffT_1$ versus OnT_1. The dashed dot line shows the set of points where the off-time and on-time are equal and the number of observations in either side is shown on the plot. For the observed pulse breakups, the on-time of the preceding pulse is greater than the off-time. As with the on-time ratio, the ratio of $OffT_1/OnT_1$ just depends on the physical characteristics of the vehicle, not the traffic condition. Figure 4.9 shows CDFs of the ratio of off-time and the preceding on-time in pulse breakups and separately for non-pulse breakups in the ground truth data at station 9 northbound. The ratio of off-time and on-time from pulse breakups ranges between 0.12 and 1.65, while the ratio from non-pulse breakups ranges between 0.35 and 62. Only 10% of non-pulse breakup events fall in the range of the pulse breakup events. Taking the difference of the two CDFs, Figure 4.9B, we select the ratio corresponding to the maximum difference between the two functions, i.e., 1.2.

4.3.4 **20th Percentile Off-Time**

As shown in Equation 4.4, the dynamic off-time in congestion depends on the median on-time over 41 consecutive pulses, centered on the current vehicle. Usually

[3] From the concurrent video, the pulse breakups from a MV pulling a trailer occurr at the trailer hitch, i.e., the smallest cross-section of the vehicle; however, the pulse breakups from LVs occur at the end of the tractor, when the ground clearance suddenly increased.

speeds are stable enough for this constraint to hold, but under heavy congestion, the median on-time over 41 consecutive pulses is sometimes larger than the local traffic conditions would dictate (e.g., if one could measure speed or sample reliably over just 5 or 7 pulses). The larger threshold is more likely to erroneously select non-pulse breakup events and mark them as suspected pulse breakups. To accommodate these errors, we exploit the fact that the off-time in a pulse breakup is usually shorter than the off-time between two consecutive vehicles. Or formalizing it in terms of a rule, the off-time in a suspected pulse breakup should fall within the lowest 20% of off times observed in the 41 consecutive pulses. For example, Figure 4.10A shows a CDF of off-time in 41 consecutive pulses, highlighting the off-time in a pulse breakup at 19/60 seconds. This pulse breakup falls just below the 10th percentile of the distribution. According to the concurrent video data, the three observations with off-time shorter than 19/60 seconds are due to tailgating. Repeating this procedure for each pulse breakup at station 9 northbound, Figure 4.10B shows the CDF of the percentile of off-time of the pulse breakups. We can see that the off-time associated with pulse breakup is usually under the 20th percentile in the sample of 41 successive pulses. Consequently, any suspected pulse breakup falling above the 20th percentile off-time for the given 41 pulses is discarded.

4.3.5 **Maximum Vehicle Length**

When a pulse breakup occurs, OnT is just the sum of OnT_1, OnT_2 and $OffT_1$. Given the estimated speed, this OnT can be converted to an estimated vehicle length in the absence of a pulse breakup. So when we suspect a pulse breakup we check that the estimated vehicle length from OnT is shorter than the maximum possible vehicle length. As mentioned previously, speed from a single loop detector is estimated from effective vehicle length divided by median on-time in the given sample period. Using dual loop detectors, we established that few vehicles should have true effective lengths over 80 ft. However, the estimated vehicle length could be longer than actual vehicle length because a LVs speed in free flow condition may systematically be lower than SV free speed, e.g., some locations have the different speed limit for passenger cars and trucks. The median on-time over 41 consecutive vehicles should usually be representative of the SVs, and

faster than the actual speed of a LV. For that reason the maximum allowable estimated vehicle length is set to 100 ft. If an estimated vehicle length in a merged on-time from a suspected pulse breakup is greater than 100 ft, the suspicion is dropped.

4.3.6 **The Pulse Breakup Detection Algorithm for a Single Loop Detector**

Combining all of these tests, the flowchart of the algorithm to identify pulse breakup from a single loop detector is shown in Figure 4.11. The process consists of six steps. If two consecutive pulses satisfy all of the checks, these pulses are considered a suspected pulse breakup. Otherwise, these pulses are considered to come from non-pulse breakup. The process is repeated over all pulses in each lane.

4.4 Evaluating the Pulse Breakup Detection Algorithm

First, the algorithm from Figure 4.11 is applied to the 2 hr long development data set from station 9 northbound and Table 4.1 summarizes the performance. The total pulses listed in the table tallies the number of pulses recorded by the detector during the video data collection. Actual pulse breakup tallies the pulse breakup verified by the ground truth data, while suspected pulse breakup the events that the algorithm suspects as being pulse breakups. The final three columns are generated by comparing the individual suspected pulse breakups against the actual pulse breakups. Where "Success" counts the number of times that the algorithm correctly caught an actual pulse breakup, while "False" counts the number of times that the algorithm erroneously labeled a non-pulse breakup as a suspected pulse breakup. Any actual pulse breakups that were not included in the "Success" column are counted in the "Failure" column, i.e., the algorithm failed to catch the given pulse breakup. The algorithm correctly identifies 295 out of 306 actual pulse breakups (96%) and it missed 11 actual pulse breakups (4%). Three suspected pulse breakups are false errors since they do not correspond to actual pulse breakups. Two of the three false errors were due to tailgating. The other false error is due to a lane changing maneuver.

4.4.1 **Free Flow Condition**

Moving now to the test data, Table 4.2 shows the ground truthed data sets recorded during free flow conditions. The data are sorted by those sets with and without pulse breakup (based on the ground truth data reduction). The data include 8 hr 20 min from 10 directional locations with pulse breakup (including the one development set) and 5 hr and 12 min from 10 locations without pulse breakup. None of the locations with pulse breakup suffered from splashover, but four of the locations without pulse breakup did, as shown with an asterisk in the station number column. Much as was done for the development set, in Table 4.1 all of the data sets from Table 4.2 were used to evaluate the performance of the pulse breakup algorithm for a single loop detector. The performance of the algorithm during free flow conditions is summarized in Table 4.3. The non-pulse breakup data are shown combined, and then repeated a second time, split between splashover and non-splashover stations. Detailed results from the stations with pulse breakup are presented in Table 4.4 and for stations without pulse breakup in Table 4.5. From Table 4.3, the algorithm correctly catches 683 out of 722 pulse breakups (94.6%), thus 39 pulse breakups are not caught by our algorithm. From all of the data (i.e., both with and without pulse breakup), 76 out of 45,197 pulses (0.17%) are erroneously marked as suspected pulse breakup.

The last three columns of Tables 4.3-4.5 tally the underlying reason whenever two valid consecutive pulses were erroneously marked as a suspected pulse breakup, i.e., a false error. As one might expect, "Tailgating" indicates two vehicles pass with a very small headway, while lane change maneuver ("LCM") indicates that at least one of the two pulses is generated from a vehicle changing lanes over the given loop detector. Finally, "Splashover" indicates that one of the two pulses was due to a splashover error from an adjacent lane (as per the preceding chapter). Overall the false errors due to splashover account for 29 out of 76 (38%). At the four locations with splashover, splashover is a dominant cause of false error, about 67% (29 out of 43). All of the stations and lanes exhibiting splashover in this set were previously labeled as stations

with splashover in Table 3.4. If a loop detector with splashover can be fixed, obviously the false errors due to splashover will be reduced.

4.4.2 **Congested Conditions**

Now moving to the more challenging congested conditions, Table 4.6 shows the ground truthed data sets recorded during congested conditions. Like the free flow conditions, the data are sorted by those sets with and without pulse breakup. The data include 2 hr 15 min from 4 directional locations with pulse breakup and 2 hr and 21 min from 5 locations without pulse breakup. None of the locations with pulse breakup suffered from splashover, but three of the locations without pulse breakup did, as shown with an asterisk in the station number column. The time series of speed from these locations is presented in Appendix B. The performance of the algorithm during congested conditions is summarized in Table 4.7. Once more the non-pulse breakup data are shown combined, and then repeated a second time, split between splashover and non-splashover stations. Detailed results from the stations with pulse breakup are presented in Table 4.8 and for stations without pulse breakup in Table 4.9. From Table 4.7, the algorithm correctly catches 157 out of 169 pulse breakups (92.8%), thus 12 pulse breakups are not caught by our algorithm. From all of the data (i.e., both with and without pulse breakup), 180 out of 20,576 pulses (0.87%) are erroneously marked as suspected pulse breakup. Overall the false errors due to splashover now account for only 29 out of 180 (16%). Compared to the performance of the algorithm in free flow condition, the success rate has dropped by almost 2%, but remains above 92% and the false alarm rate has increased by a factor of 5, but remains below 1%.

4.5 Sensitivity of the Parameters of Variables of the Algorithm

There are several parameters in the algorithm to identify pulse breakup in single loop detector data that were derived from one detector station using only a 2 hr long sample. The preceding results are based on the assumption that the nature of pulse breakups observed at the one location is similar to all of the detector stations. While it is not possible to test stations for which we do not have data, the assumption will be

examined using the data from the evaluation set. This section examines the optimal threshold for the ratio of on-times between two pulses and the optimal threshold for the ratio of off-time and the preceding on-time in the algorithm, repeating the analysis from Figures 4.6 and 4.9, but now using the entire free flow evaluation data set from Table 4.2 (excluding station 9 northbound, which was used for development).

First the on-time ratio threshold is evaluated, holing the other parameters constant. The on-time ratio is stepped from 0.1 to 1.1 at increments of 0.01 in this evaluation. The ratio is also set to infinity, i.e., the results of the algorithm without the process of on-time ratio, and the result is plotted on the abscissa of 5. Figure 4.12 shows the evolution of the false error, failure error and sum of two errors. In general, as on-time ratio increases, false error increases, but failure error decreases. The sum of the two is minimized when the on-time ratio threshold is between 0.71 and 0.76 (all of the tested values except 0.74). The original on-time ratio, 0.72, falls within this range.

Second the ratio of off-time and preceding on-time is evaluated, holing the other parameters constant. The ratio is stepped from 0 to 1.5 at increments of 0.1 in this evaluation. The ratio is also set to infinity, i.e., the results of the algorithm without the process of off-time ratio, and the result is plotted on the abscissa of 36,000. Figure 4.13 shows the evolution of the false error, failure error and sum of two errors. In general, as off-time ratio increases, false error increases, but failure error decreases. The sum of the two is minimized when the off-time ratio threshold is between 1.2 and 1.4. The original off-time ratio, 1.2, falls within this range.

Finally, varying both the on-time and off-time ratio thresholds, Figure 4.14 shows the resulting performance. The off-time ratio is varied from 0.7 to 1.5 at increments of 0.1, separated by the bold vertical dashed lines. Between each pair of dashed lines, the on-time ratio is varied between 0.67 and 0.76 at increments of 0.01. In total 150 combinations are tested, 15 values of the on-time ratio threshold and 10 values of the off-time ratio threshold. The sum of the two errors is minimized when the off-time ratio threshold is between 1.3 and 1.4 when the on-time ratio is 0.71. However, the original thresholds of 1.2 and 0.72 yield a performance that has just one more error than the

optimal values. These results indicate that the calibration from one location transferable to the other locations in this study. If such microscopic event data become available from other metropolitan areas, it would be advisable to test the calibration on those facilities as well.

4.6 Comparison of the Performance of Pulse Breakup Detection Algorithm

We compared the performance of two earlier pulse breakup detection methodologies against our algorithm using the data underlying Table 4.2 and 4.6. In previous studies, Chen and May (1987) (C&M) used a threshold of the time gap, while Cheevarunothai et al. (2007) (CYN) used a threshold of the time headway. Next, our pulse breakup detection algorithm (L&C) is applied to the loop detectors. The performance of each test is evaluated by a number of success, false, and failure, as shown in Table 4.10. Overall, our algorithm exhibits the lowest rate of false alarms and failures than the two previous methods, and it catches more of the actual pulse break up events.

4.7 Field Testing the Results

Working with ODOT, we adjusted the detector settings at four detector stations and we were successful in eliminating the chronic detector errors at most of these stations. If these results are typical, the improved detector calibration enabled by our research could lead to a very inexpensive means to improve the quality of loop detector data at existing stations. We selected two stations with significant splashover events (based on the previous chapter) and two stations with significant pulse breakup problems (based on this chapter). ODOT engineers went to the field and turned the detector sensitivity down at the stations with splashover and up at the stations with pulse breakup. While most loop detector sensors have at least five sensitivity levels, often there is no clear guidance in which level is correct for the particular location. This work seeks to provide such guidance.

Table 4.11 shows the sensitivity levels before and after the change, as well as the date of the change. At which point another round of concurrent video and loop detector data were collected, as enumerated in Table 4.12. Table 4.13 shows the detector

performance by lane and Table 4.14 shows the performance by station, before and after, based on the ground truth.[4] Note that all of the existing problems were solved, i.e., we resolved splashover at the stations with splashover and pulse breakup at stations with pulse breakup. However, at one detector we overcompensated and went from suffering from pulse breakup to suffering from splashover, the only errors seen in the after set.

Next, applying our detection algorithms, Table 4.15 shows that the splashover detection algorithm correctly found one and only one detector with splashover in the set. the one lane Table 4.16 shows that the pulse breakup algorithm had no failures (though it was infeasible to do so since there were no errors to miss) and had a false alarm rate under 0.5%, comparable to our results in Table 4.10 for free flow, non-pulse breakup case.

[4] Note that we did not use the same sample size before and after. The number of vehicles are enumerated here, see Table 3.2, 4.2 and 4.12 for the filming durations.

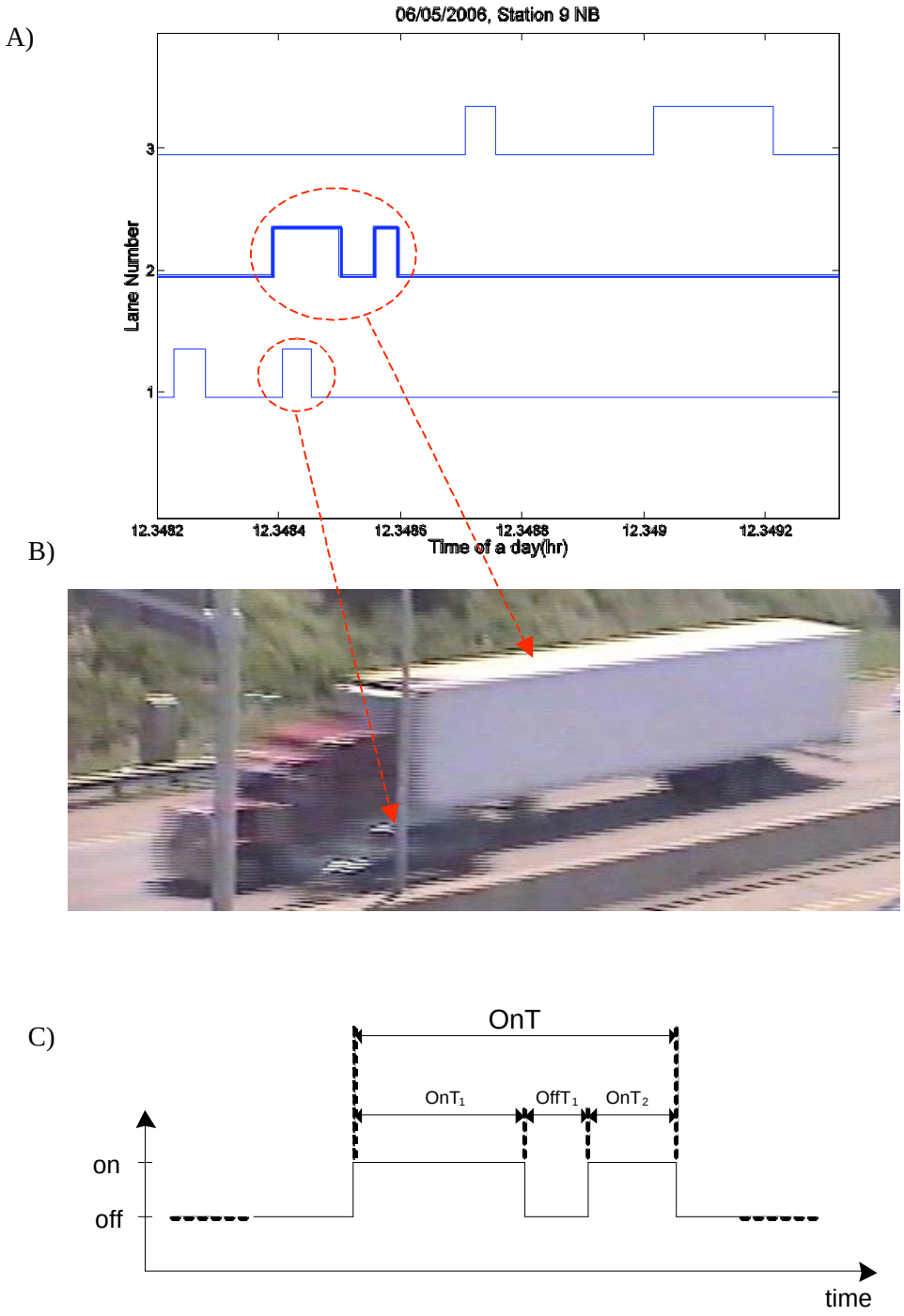


Figure 4.1, (a) A plot of detector actuations with pulse break-up over a short time period at station 9 northbound, (b) the corresponding video image at station 9 northbound and (c) anatomy of the pulse breakup

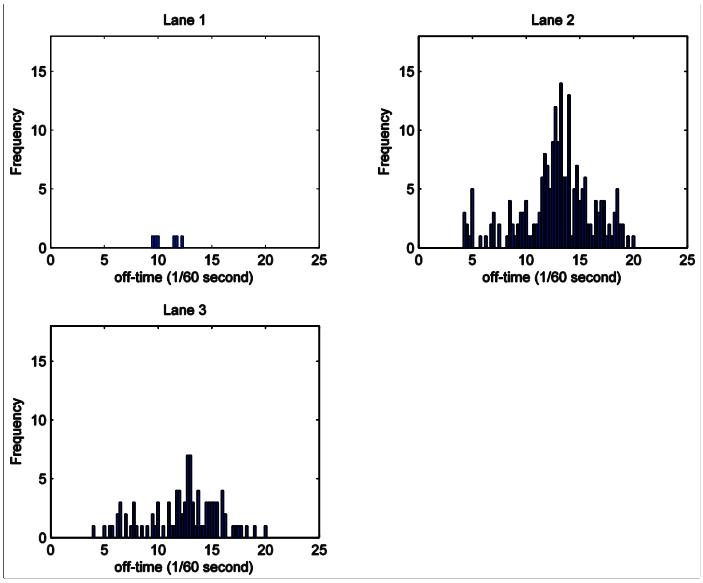


Figure 4.2, Frequency plot for off-time corresponding to pulse breakup in each lane at station 9 northbound on 6/05/2006 during free flow conditions. Maximum observed off-time from pulse breakup is 20/60 seconds.

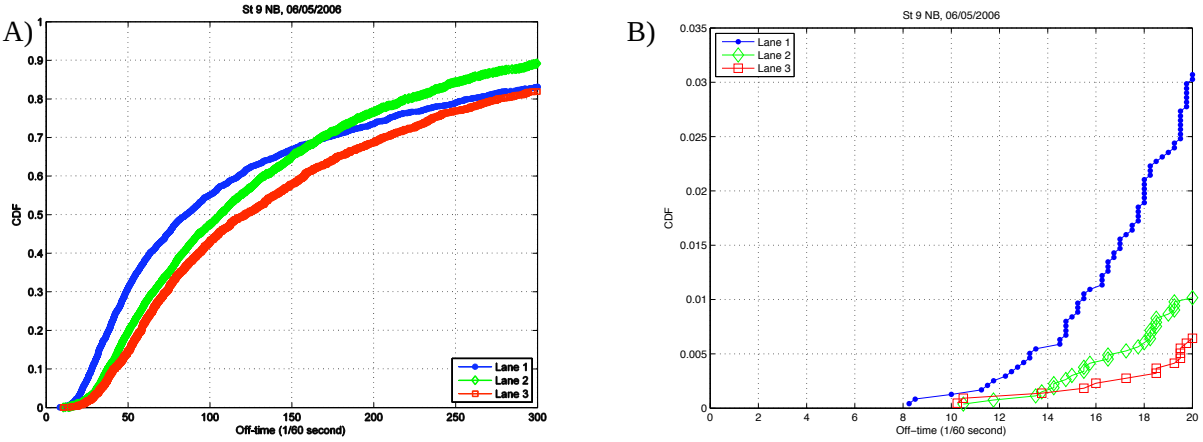
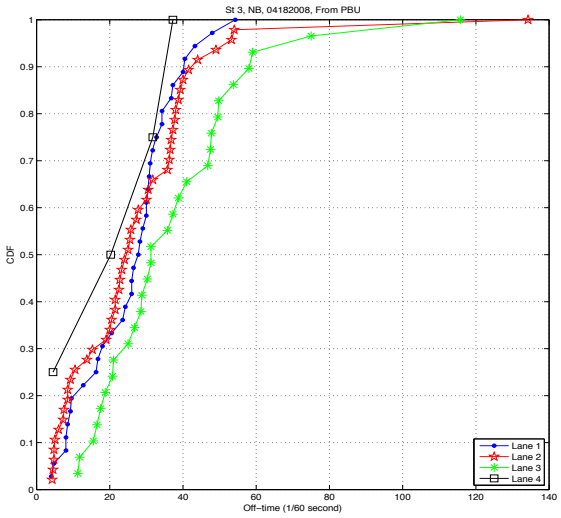


Figure 4.3, Cumulative density function (CDF) of off-times from ground truth data excluding pulse breakups. (a) shows off-time distribution on a large vertical scale while (b) repeats the data on a smaller scale.

A)



B)

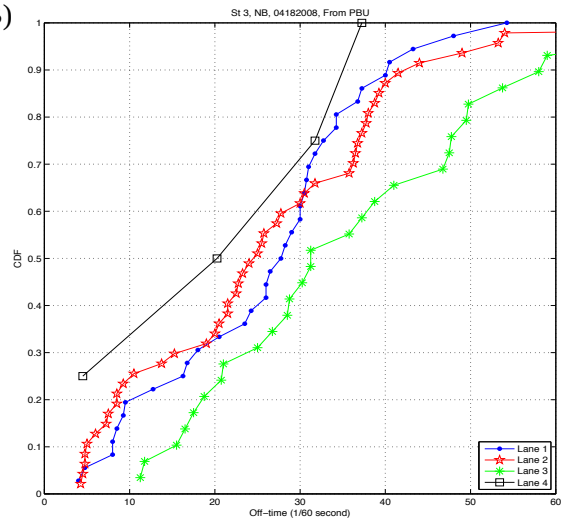
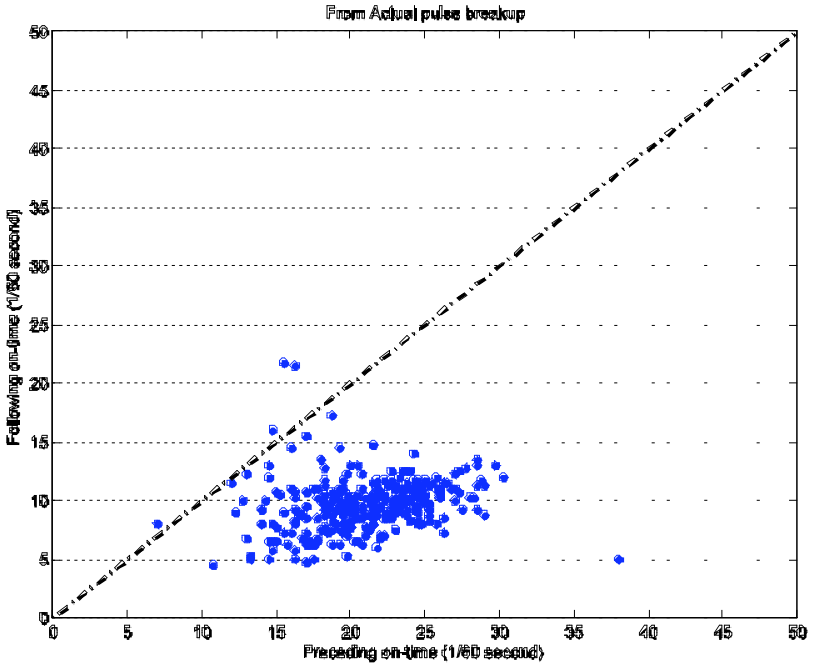


Figure 4.4, CDF of off-time from pulse breakups by lane at station 3 northbound in congestion on (a) a large horizontal scale (b) repeats the data on a smaller scale

A)



B)

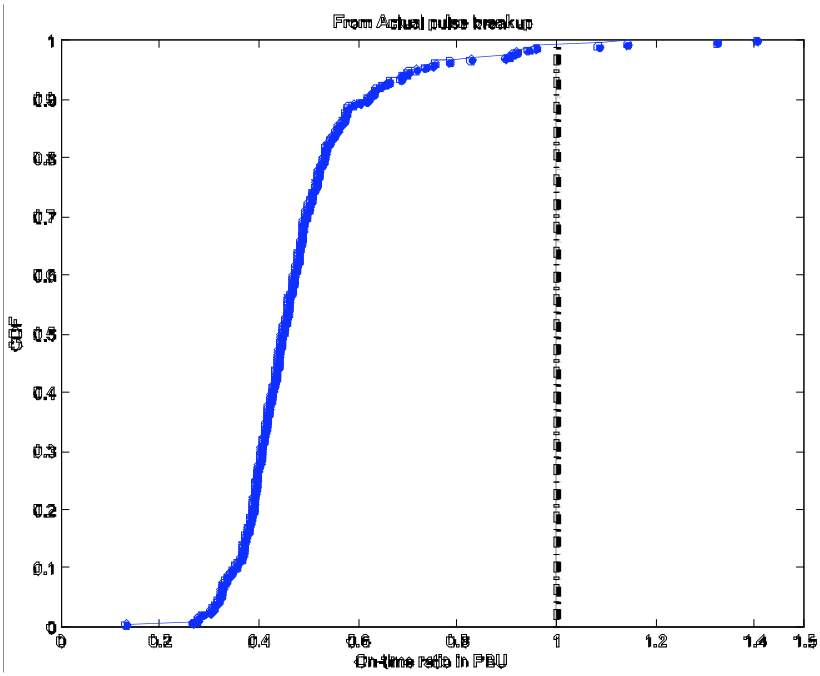
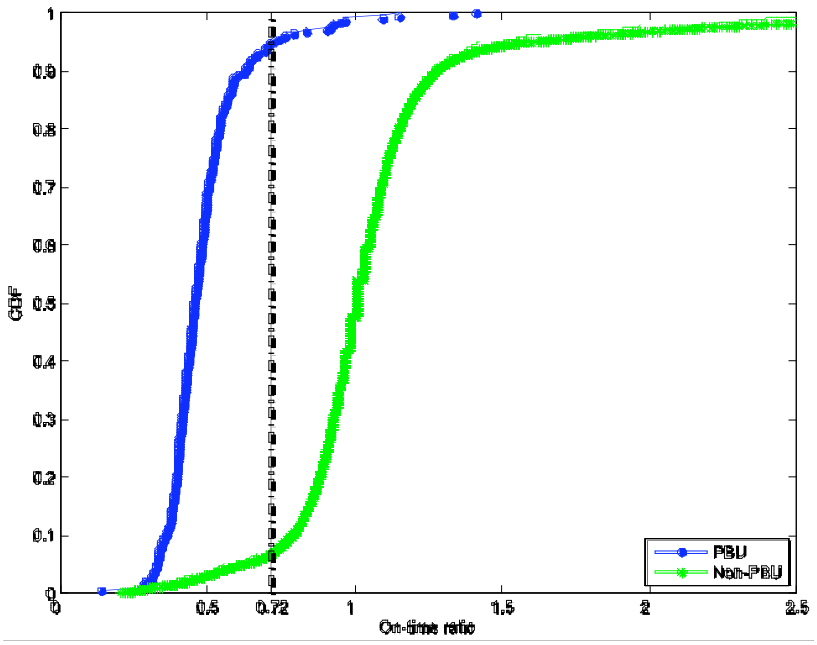


Figure 4.5, (a) A scatter plot of on-times between two pulses in pulse breakup, (b) Cumulative distribution function of on-time ratio in pulse breakups

A)



B)

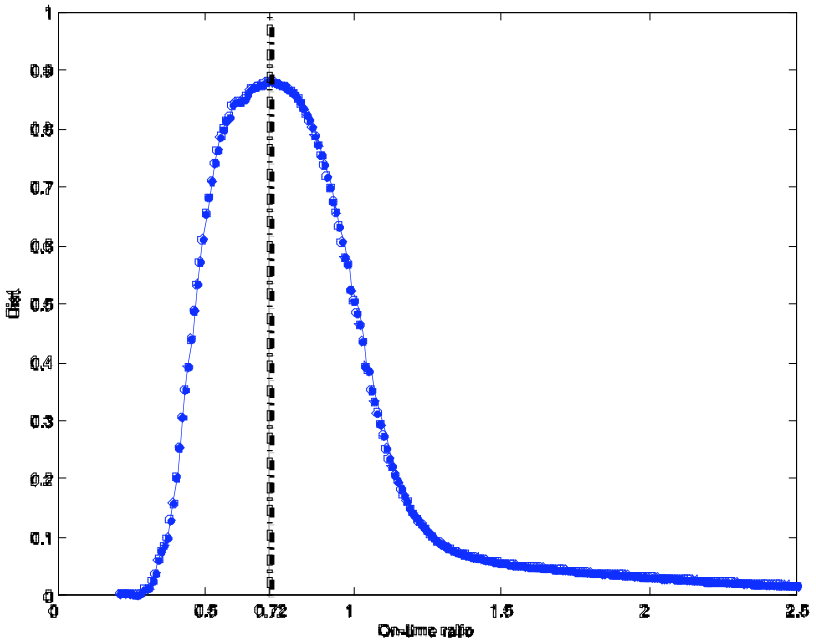


Figure 4.6, (a) CDF of on-time ratio from pulse breakup and from non-pulse breakup (b) the difference of the two functions in a range of on-time ratio from 0 to 2.5

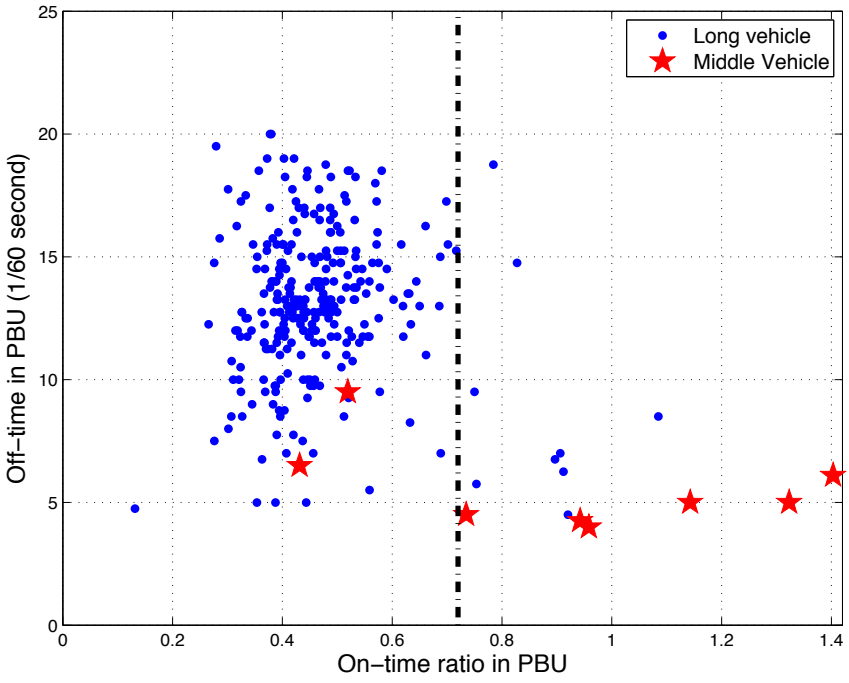


Figure 4.7, Scatter plot of off-time ratio versus on-time ratio in pulse breakup

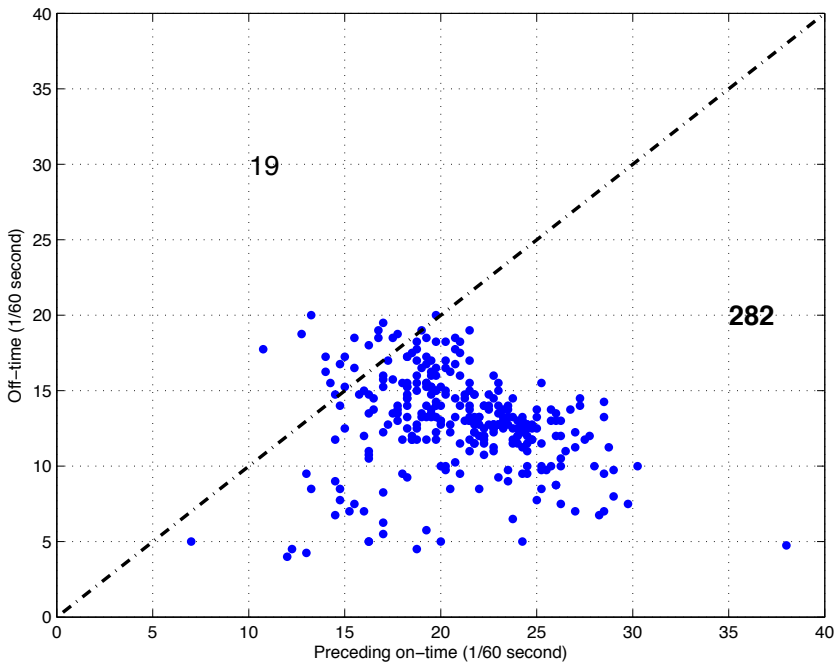
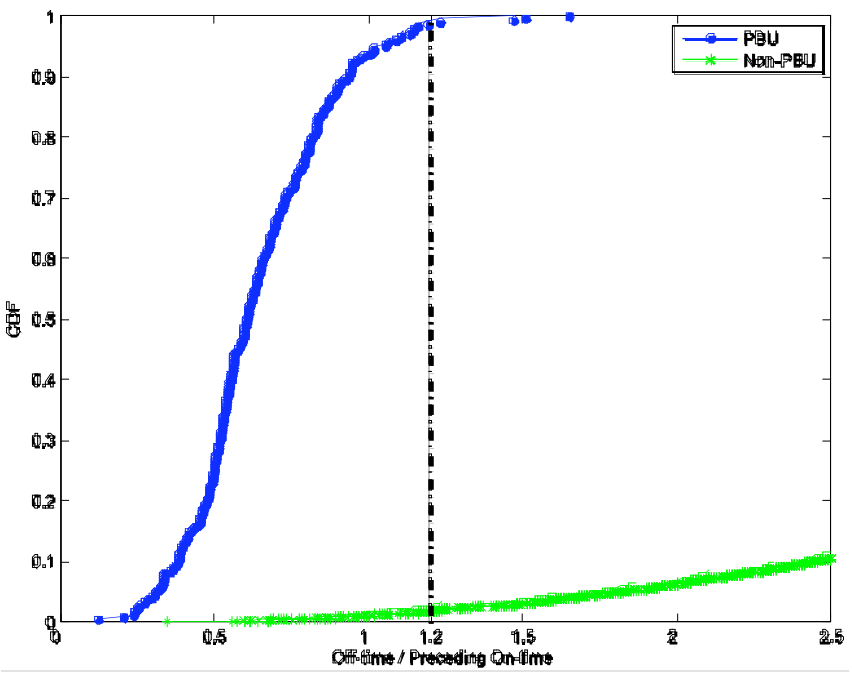


Figure 4.8, A scatter plot of off-time and preceding on-time of pulse breakups at station 9 northbound

A)



B)

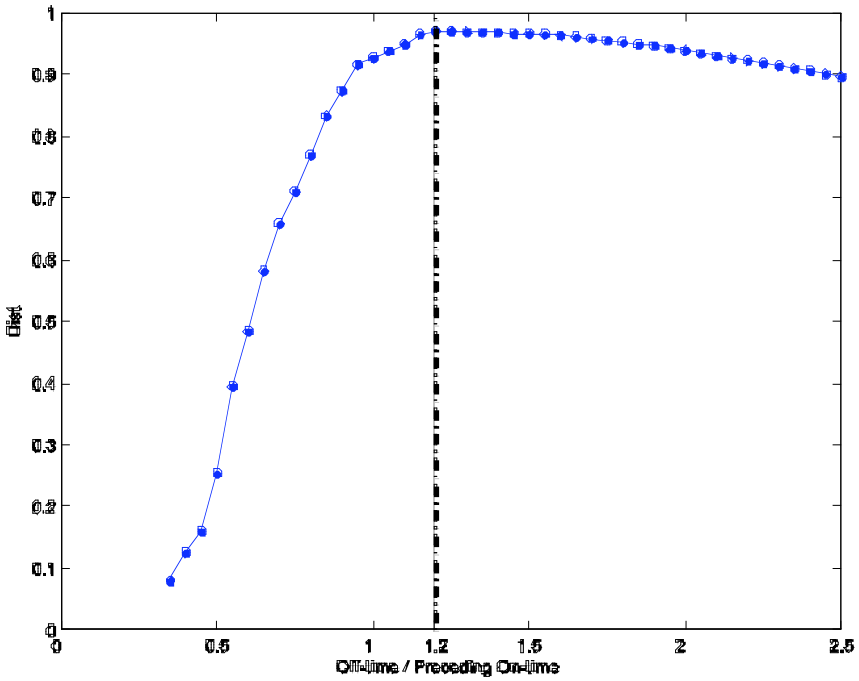
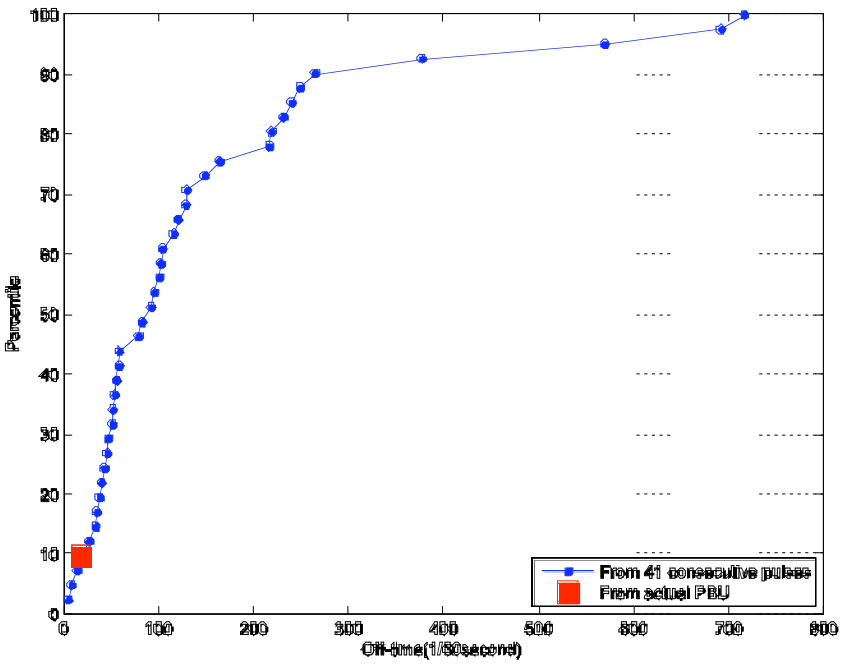


Figure 4.9, (a) CDF of off-time ratio from pulse breakup and from non-pulse breakup (b) the difference of the two functions.

A)



B)

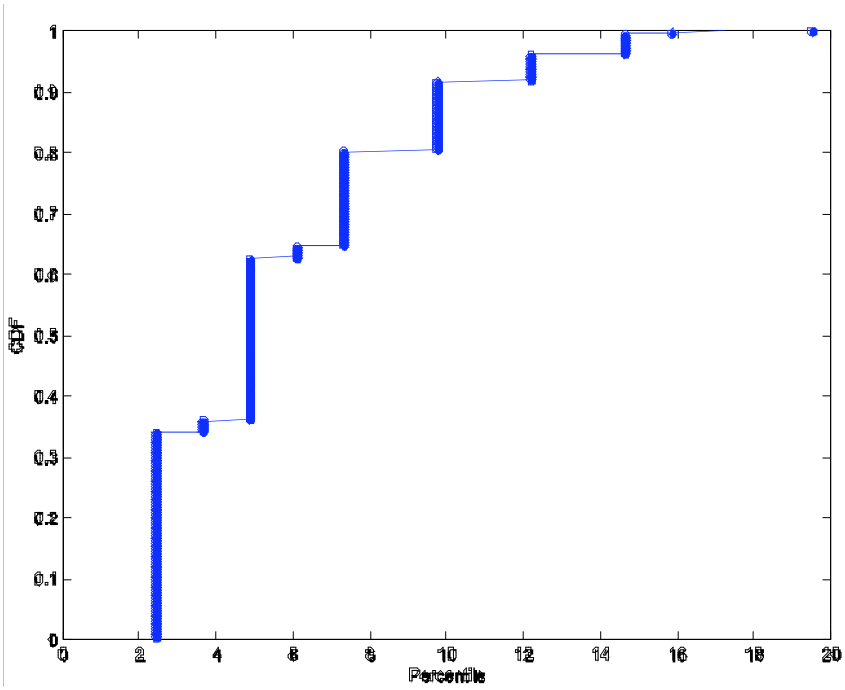


Figure 4.10, (a) CDF of off-time in 41 consecutive pulses in lane 2, including an actual pulse breakup with off of 18.5/60 seconds, falling at the 10th percentile. (b) A plot of the corresponding off-time percentiles from each of the actual pulse breakups at station 9 northbound

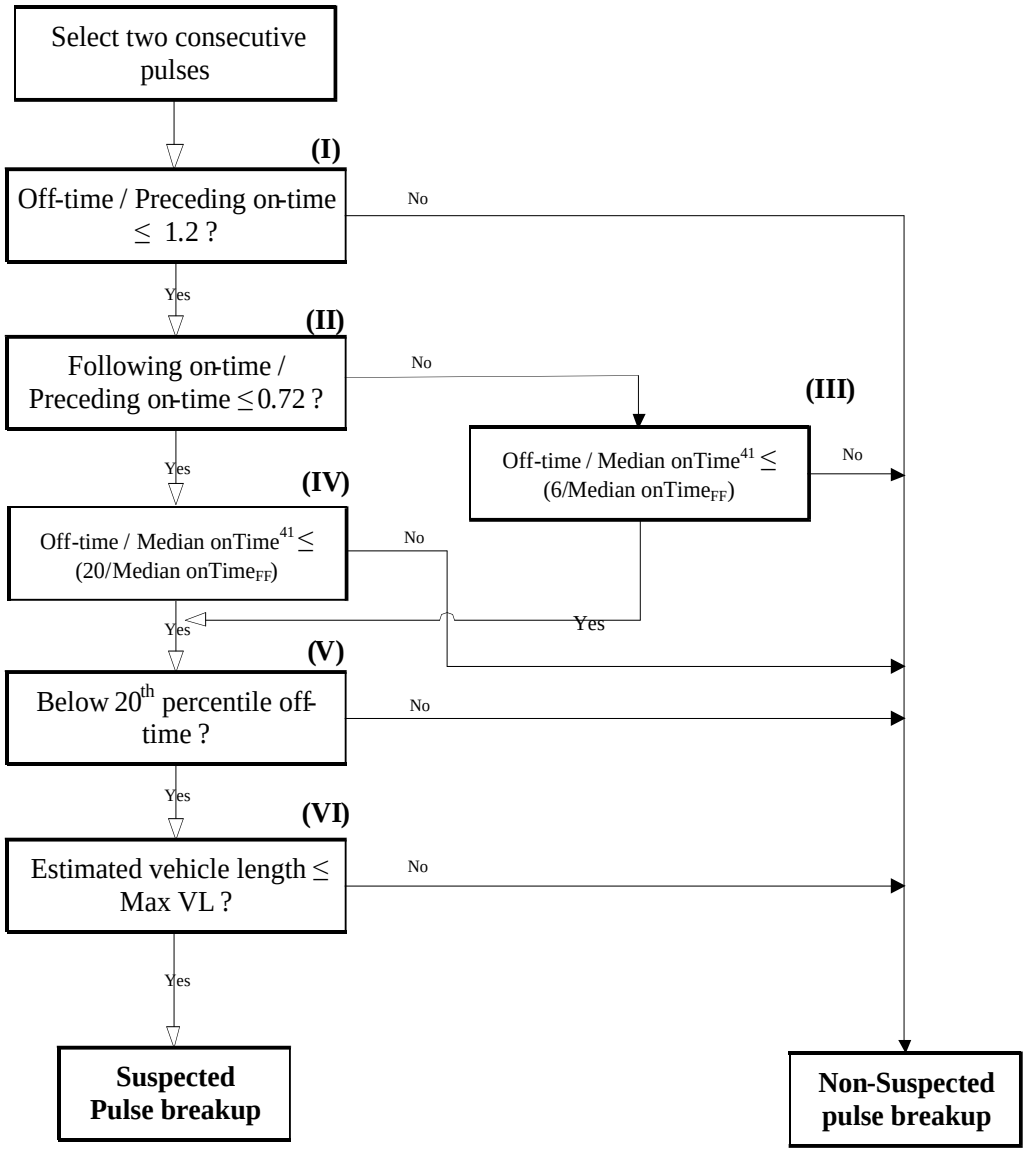


Figure 4.11, A flowchart of the algorithm to identify pulse breakup from a single loop detector

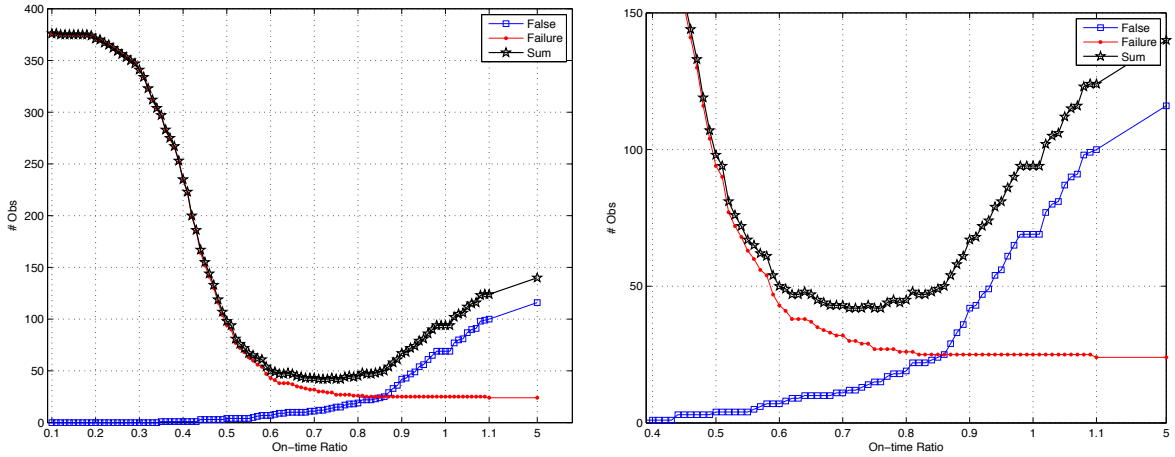


Figure 4.12, Sensitivity analysis of the algorithm performance relative to the on-time ratio threshold

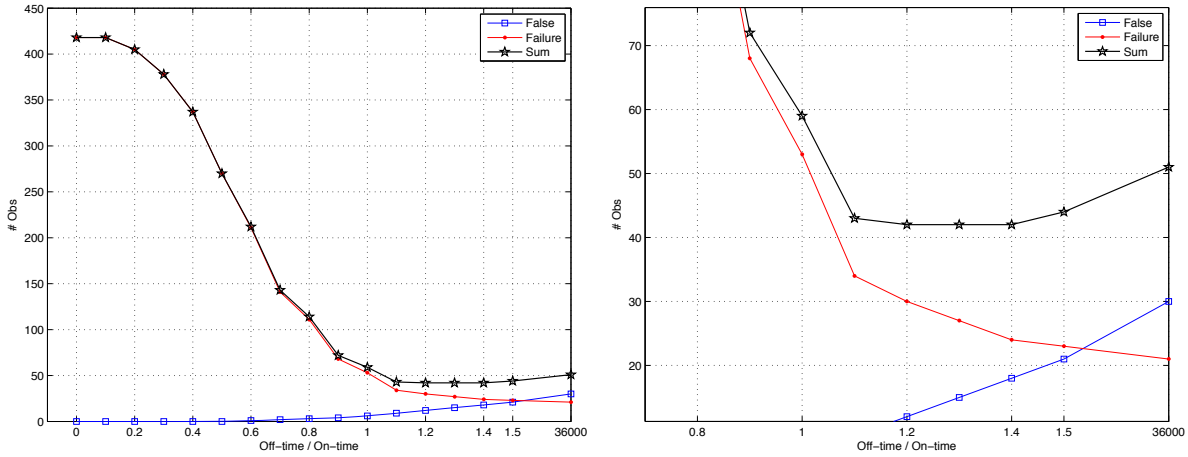


Figure 4.13, Sensitivity analysis of the algorithm performance relative to the off-time ratio threshold

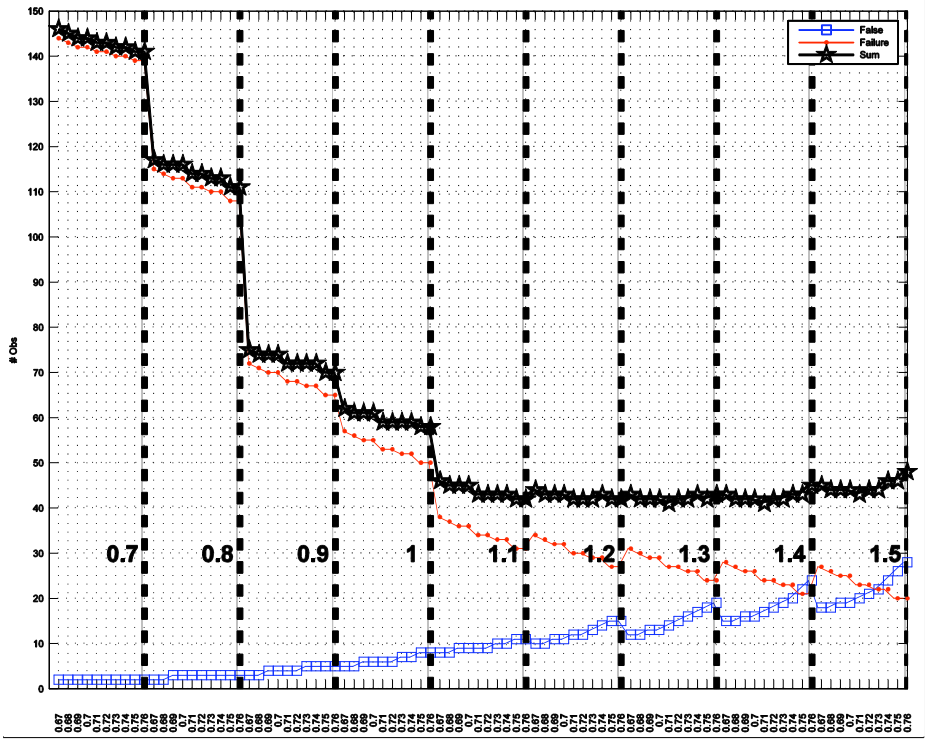


Figure 4.14, Sensitivity analysis of the algorithm performance relative to the on-time ratio threshold and off-time ratio threshold combined

Station Num (Direction)	Lane number	Total pulses	Actual pulse breakup	Suspected pulse breakup	Performance		
					Success	False	Failure
9 (NB)	1	2,386	6	8	6	2	0
	2	2,900	208	200	200	0	8
	3	2,277	92	90	89	1	3
Total		7,563	306	298	295	3	11

Table 4.1, The performance of the proposed algorithm to identify pulse breakup in free flow condition at station 9 northbound in 2hr sample data

the presence or the absence of pulse breakup	Station number	Dire-ction	Number of lanes	Date	Start Time (hh:min)	End Time (hh:min)	Duration of time (hh:min)
With pulse breakup	3	NB	4	03/17/2008	10:57	11:20	0:23
	3	SB	4	04/18/2008	15:55	16:55	1:00
	4	NB	4	09/09/2008	16:55	17:15	0:20
	4	SB	4	03/17/2008	10:15	10:35	0:20
	6	NB	3	04/18/2008	15:55	16:55	1:00
	9	NB	3	06/05/2006	12:20	14:20	2:00
	9	SB	3	06/05/2006	12:20	14:20	2:00
	15	NB	3	03/10/2009	17:18	17:47	0:29
	18	NB	3	03/09/2009	08:24	08:57	0:33
	19	NB	3	03/17/2008	09:25	09:40	0:15
Without pulse breakup	2	NB	4	03/09/2009	17:21	17:50	0:29
	31	NB	4	11/21/2008	10:35	11:05	0:30
	38	EB	3	08/29/2008	15:05	15:25	0:20
	38*	WB	4	09/09/2008	12:05	12:25	0:20
	41*	EB	2	09/09/2008	11:00	11:35	0:35
	43	EB	3	09/02/2008	8:50	9:15	0:25
	43	WB	3	09/02/2008	8:50	9:15	0:25
	56	EB	3	09/03/2008	16:40	17:25	0:45
	56*	WB	3	11/21/2008	9:00	9:40	0:40
	102	EB	3	03/10/2009	17:05	17:20	0:15
	104*	EB	3	03/17/2008	16:00	16:10	0:10
	104	WB	3	03/12/2009	17:00	17:18	0:18

Table 4.2, Data information of the ground truth data with free flow conditions, total recorded time of video data from locations with pulse breakup is 500 min and 312 min for the locations without pulse breakup. Stations with splashover indicated with "*"

During free flow condition	Total pulses	Actual PBU	Suspected PBU	Performance			Reason of False		
Data				Success	False	Failure	Tail-gating	LCM	Splash-over
with pulse breakup	34,401	722	699	683	16	39	12	4	0
without pulse breakup	13,304	-	67	-	67	-	15	23	29
with splashover	3,758	-	43	-	43	-	7	7	29
without splashover	9,546	-	24	-	24	-	8	16	0
Total	47,705	722	766	683	83	39	27	27	29

Table 4.3, Summary of the performance of pulse breakup's algorithm for a single loop detector during free flow conditions.

Date	St # (Direc-tion)	Lane	Total vehicles	Actual PBU	Suspected PBU	Performance			Reason of False		
						Success	False	Failure	Tail-gating	LCM	Splash-over
03/17/08	3 (NB)	1	310	6	6	6	0	0	-	-	-
		2	420	23	21	21	0	2	-	-	-
		3	359	16	15	14	1	2	1	0	0
		4	168	2	2	2	0	0	-	-	-
04/18/08	3 (SB)	1	995	1	2	1	1	0	1	0	0
		2	1,806	9	9	9	0	0	-	-	-
		3	1,537	4	4	4	0	0	-	-	-
		4	1,139	2	3	2	1	0	1	0	0
09/09/08	4 (NB)	3	619	10	10	9	1	1	0	1	0
		4	124	1	0	0	0	1	-	-	-
03/17/08	4 (SB)	1	225	0	0	0	0	0	-	-	-
		2	523	19	18	18	0	1	-	-	-
		3	533	21	15	15	0	6	-	-	-
		4	83	1	0	0	0	1	-	-	-
04/18/08	6 (NB)	1	2,249	26	25	25	0	1	-	-	-
		2	1,962	58	55	55	0	3	-	-	-
		3	1,747	8	9	8	1	0	0	1	0
06/05/06	9 (NB)	1	2,386	6	8	6	2	0	1	1	0
		2	2,900	208	200	200	0	8	-	-	-
		3	2,277	92	90	89	1	3	1	0	0
06/05/06	9 (SB)	1	2,434	2	7	2	5	0	4	1	0
		2	2,964	21	22	20	2	1	2	0	0
		3	2,288	107	106	105	1	2	1	0	0
03/10/09	15 (NB)	1	1,173	5	3	3	0	2	-	-	-
		2	940	24	23	23	0	1	-	-	-
		3	899	11	8	8	0	3	-	-	-
03/09/09	18 (NB)	1	197	0	0	0	0	0	-	-	-
		2	227	13	13	13	0	0	-	-	-
		3	140	3	3	3	0	0	-	-	-
03/17/08	19 (NB)	1	186	0	0	0	0	0	-	-	-
		2	297	13	13	13	0	0	-	-	-
		3	294	10	9	9	0	1	-	-	-
Total			34,401	722	699	683	16	39	12 (75%)	4 (25%)	0 (0%)

Table 4.4, Summary of the performance of pulse breakup's algorithm for a single loop detector from the data with pulse breakup during free flow conditions.

Date	St # (Direc-tion)	Lane	Total pulses	Actual PBU	Suspected PBU	Performance			Reason of False		
						Success	False	Failure	Tail-gating	LCM	Splash-over
03/09/2009	2 (NB)	1	628	0	2	0	2	0	1	1	0
		2	642	0	1	0	1	0	0	1	0
		3	526	0	1	0	1	0	0	1	0
		4	45	0	0	0	0	0	-	-	-
11/21/2008	31 (NB)	1	124	0	0	0	0	0	-	-	-
		2	296	0	0	0	0	0	-	-	-
		3	220	0	3	0	3	0	1	2	0
		4	27	0	0	0	0	0	-	-	-
8/29/2008	38 (EB)	1	355	0	1	0	1	0	1	0	0
		2	331	0	1	0	1	0	1	0	0
		3	164	0	1	0	1	0	1	0	0
9/9/2008	38 (WB)	1	172	0	0	0	0	0	-	-	-
		2	242	0	1	0	1	0	1	0	0
		3	206	0	9	0	9	0	0	0	9
		4	56	0	1	0	1	0	0	0	1
9/9/2008	41 (EB)	1	337	0	7	0	7	0	1	0	6
		2	507	0	3	0	3	0	3	0	0
9/2/2008	43 (EB)	1	262	0	1	0	1	0	0	1	0
		2	419	0	0	0	0	0	-	-	-
		3	384	0	1	0	1	0	0	1	0
9/2/2008	43 (WB)	1	500	0	0	0	0	0	-	-	-
		2	590	0	3	0	3	0	0	3	0
		3	444	0	2	0	2	0	0	2	0
9/3/2008	56 (EB)	1	596	0	1	0	1	0	0	1	0
		2	771	0	4	0	4	0	3	1	0
		3	322	0	0	0	0	0	-	-	-
11/21/2008	56 (WB)	1	345	0	0	0	0	0	-	-	-
		2	632	0	5	0	5	0	0	5	0
		3	121	0	1	0	1	0	0	0	1
3/10/2009	102 (EB)	1	189	0	1	0	1	0	0	1	0
		2	320	0	1	0	1	0	0	1	0
		3	322	0	0	0	0	0	-	-	-
3/17/2008	104 (EB)	1	441	0	12	0	12	0	0	0	12
		2	349	0	1	0	1	0	0	1	0
		3	350	0	3	0	3	0	2	1	0
3/12/2009	104 (WB)	1	359	0	0	0	0	0	-	-	-
		2	415	0	0	0	0	0	-	-	-
		3	295	0	0	0	0	0	-	-	-
Total			13,304	0	67	0	67	0	15 (22%)	23 (34%)	29 (43%)

Table 4.5, Summary of the performance of pulse breakup's algorithm for a single loop detector from the data without pulse breakup during free flow conditions.

the presence or the absence of pulse breakup	Date	Station number	Direction	Start Time (hh:min)	End Time (hh:min)	Duration of time (hh:min)
With pulse breakup	03/21/2008	3	NB	16:35	16:50	0:15
	04/18/2008	3	NB	15:55	16:55	1:00
	09/09/2008	4	NB	17:15	17:55	0:40
	04/07/2008	9	SB	07:50	08:10	0:20
Without pulse breakup	03/12/2009	41*	EB	16:40	17:06	0:26
	03/12/2009	43	EB	17:07	17:48	0:41
	09/03/2008	56*	WB	16:40	17:25	0:45
	03/10/2009	102	EB	16:46	17:05	0:19
	03/17/2008	104*	EB	16:10	16:20	0:10

Table 4.6, Data information of the ground truth data with congestion. Stations with splashover indicated with "*"

During congestion Data	Total pulses	Actual PBU	Suspected PBU	Performance			Reason of False		
				Success	False	Failure	Tail-gating	LCM	Splash-over
with pulse breakup	10,721	169	181	157	24	12	24	0	0
without pulse breakup	9,855	-	156	-	156	-	105	22	29
with splashover	5,177	-	98	-	98	-	51	18	29
without splashover	4,678	-	58	-	58	-	54	4	0
Total	20,576	169	337	157	180	12	129	22	29

Table 4.7, Summary of the performance of pulse breakup's algorithm for a single loop detector during congestion.

Date	St # (Direc-tion)	Lane	Total vehicles	Actual PBU	Suspected PBU	Performance			Reason of False		
						Success	False	Failure	Tail-gating	LCM	Splash-over
03/21/08	3 (NB)	1	404	4	2	2	0	2	-	-	-
		2	369	6	6	6	0	0	-	-	-
		3	387	10	10	10	0	0	-	-	-
		4	208	3	3	3	0	0	-	-	-
04/18/08	3 (NB)	1	1789	36	43	36	7	0	7	0	0
		2	1619	47	48	44	4	3	4	0	0
		3	1482	29	30	25	5	4	5	0	0
		4	920	4	5	3	2	1	2	0	0
09/09/08	4 (NB)	3	1201	12	11	10	1	2	1	0	0
		4	298	1	1	1	0	0	-	-	-
04/07/08	9 (SB)	1	758	0	2	0	2	0	2	0	0
		2	704	3	6	3	3	0	3	0	0
		3	582	14	14	14	0	0	-	-	-
Total			10,721	169	181	157	24	12	24 (100%)	0 (0%)	0 (0%)

Table 4.8, Summary of the performance of the pulse breakup algorithm to the congested ground truth data at stations with pulse breakup. During congestion, the performance of our algorithm degrades, the rates of false and failure errors increased. All false errors are observed from the interaction of two actual vehicles' movement.

Date	St # (Direc-tion)	Lane	Total vehicles	Actual PBU	Suspected PBU	Performance			Reason of False		
						Success	False	Failure	Tail-gating	LCM	Splash-over
03/12/09	41 (EB)	1	935	0	32	0	32	0	11	6	15
		2	733	0	24	0	24	0	23	1	0
03/12/09	43 (EB)	1	1,417	0	15	0	15	0	13	2	0
		2	1,322	0	27	0	27	0	25	2	0
		3	1,174	0	14	0	14	0	14	0	0
09/03/08	56 (WB)	1	1,350	0	7	0	7	0	5	2	0
		2	1,171	0	16	0	16	0	7	9	0
		3	329	0	3	0	3	0	1	0	2
03/10/09	102 (EB)	1	171	0	0	0	0	0	-	-	-
		2	286	0	0	0	0	0	-	-	-
		3	308	0	2	0	2	0	2	0	0
03/17/08	104 (EB)	1	308	0	12	0	12	0	0	0	12
		2	204	0	0	0	0	0	-	-	-
		3	147	0	4	0	4	0	4	0	0
Total			9,855	0	156	0	156	0	105 (67%)	22 (14%)	29 (19%)

Table 4.9, Summary of the performance of the pulse breakup algorithm to the congested ground truth data at stations without pulse breakup. In this case, we can see relatively high number of tailgating causing false error.

Condition	Method	Total pulses	Actual PBU	Suspected PBU	Performance		
					Success	False	Failure
Free flow & Pulse breakup	C&M	34,401	722	636	521	115	201
	CYN			1,769	546	1,223	176
	L&C			699	683	16	39
Congestion & Pulse breakup	C&M	10,721	169	55	49	6	120
	CYN			54	14	40	155
	L&C			181	157	24	12
Free flow & Non-pulse breakup	C&M	13,304	0	159	-	159	-
	CYN			509	-	509	-
	L&C			67	-	67	-
Congestion & Non-pulse breakup	C&M	9,855	0	127	-	127	-
	CYN			104	-	104	-
	L&C			156	-	156	-
Overall	C&M	68,281	891	977	570	407	321
	CYN			2,436	560	1,876	331
	L&C			1,103	840	263	51

Table 4.10, Comparison of our proposed methodsagainst previous methods for detecting pulse breakup. Our method has the smallest false error and failure error.

Date adjusting sensitivity	Station number	Direction	Lane number	Old Sensitivity Level	New Sensitivity Level
6/9/2009	3	NB	1	Normal	High
			2	Normal	High
			3	Normal	High
			4	Normal	High
		SB	1	Normal	High
			2	Normal	High
			3	Normal	High
			4	Normal	High
	9	NB	1	Low	High
			2	Low	High
			3	Low	High
		SB	1	Low	High
			2	Low	High
			3	Low	High
6/10/2009	56	WB	3	Normal	Low
	104	EB	1	High	Low

Table 4.11, Detector sensitivity of 16 loop detectors at four detector stations.

Station number	Direction	Number of lanes	Date	Start Time (hh:min)	End Time (hh:min)	Duration of time (hh:min)
3	NB	4	06/17/2009	10:21	10:51	0:30
	SB	4	06/17/2009	10:21	10:51	0:30
9	NB	3	06/17/2009	10:05	10:41	0:36
	SB	3	06/17/2009	10:05	10:41	0:36
56	WB	3	06/17/2009	09:33	10:03	0:30
104	EB	3	06/26/2009	13:14	13:29	0:15

Table 4.12, Detail information of video data recorded for examination of detector sensitivity. All six directional locations were recorded during free flow conditions.

Station number	Dire-ction	Lane number	Before			After			Date ["Before" data, "After" data]
			Total pulses	Splash-over	Pulse breakup	Total pulses	Splash-over	Pulse breakup	
3	NB	1	310	0	6	479	0	0	[03/17/2008, 06/17/2009]
		2	420	0	23	687	0	0	
		3	359	0	16	595	0	0	
		4	168	0	2	324	0	0	
3	SB	1	995	0	1	409	0	0	[04/18/2008, 06/17/2009]
		2	1,806	0	9	831	0	0	
		3	1,537	0	4	611	0	0	
		4	1,139	0	2	397	0	0	
9	NB	1	2,386	0	6	657	0	0	[06/05/2006, 06/17/2008]
		2	2,900	0	208	784	0	0	
		3	2,277	0	92	624	0	0	
9	SB	1	2,434	0	2	803	68	0	[06/05/2006, 06/17/2009]
		2	2,964	0	21	908	0	0	
		3	2,288	0	107	673	0	0	
56	WB	1*	345	0	0	234	0	0	[11/21/2008, 06/17/2009]
		2*	632	0	0	446	0	0	
		3	121	19	0	70	0	0	
104	EB	1	441	318	0	136	0	0	[03/17/2008, 06/26/2009]
		2*	349	0	0	435	0	0	
		3*	350	0	0	401	0	0	
Total			24,221	337	499	10,504	68	0	

Table 4.13, Performance during free flow conditions before and after the detector sensitivity change. The four detectors that were not changed are shown with *.

Station number	Dire-ction	Before			After		
		Total pulses	Splash-over	Pulse breakup	Total pulses	Splash-over	Pulse breakup
3	NB	1,257	0	47	2,085	0	0
	SB	5,477	0	16	2,248	0	0
9	NB	7,563	0	306	2,065	0	0
	SB	7,686	0	130	2,384	68	0
56	WB	1,098	19	0	750	0	0
104	EB	1,140	318	0	972	0	0
Total		24,221	337	499	10,504	68	0

Table 4.14, Comparison of before and after study across all lanes

Station number	Direction	A ratio of adjusted suspected splashover: R_T^S [**Source lane (S)** → Target lane (T)]					
		L1 → L2	L2 → L1	L2 → L3	L3 → L2	L3 → L4	L4 → L3
3	NB	0%	0%	0%	0%	0%	0%
3	SB	0%	0%	0%	0%	0%	0%
9	NB	0%	0%	0%	0%	-	-
9	SB	0%	0.2%	0%	0%	-	-
56	WB	0%	0%	0%	0%	-	-
104	EB	0%	0%	0%	0%	-	-

Table 4.15, Percentage of adjusted suspected splashover relative to source lane from stations where the detector sensitivity was changed.

St # (Direc-Tion)	Lane	Total Pulses	Actual PBU	Suspected PBU	Performance			Reason of False;			% False
					Success	False	Failure	Tail-gating	LCM	Splash-over	
3 (NB)	1	479	0	1	0	1	0	1	0	0	0.2%
	2	687	0	2	0	2	0	2	0	0	0.3%
	3	595	0	1	0	1	0	0	1	0	0.2%
	4	324	0	0	0	0	0	-	-	-	0.0%
3 (SB)	1	409	0	0	0	0	0	-	-	-	0.0%
	2	831	0	0	0	0	0	-	-	-	0.0%
	3	611	0	3	0	3	0	0	3	0	0.5%
	4	397	0	2	0	2	0	2	0	0	0.5%
9 (NB)	1	657	0	2	0	2	0	0	2	0	0.3%
	2	784	0	1	0	1	0	1	0	0	0.1%
	3	624	0	1	0	1	0	1	0	0	0.2%
9 (SB)	1	803	0	9	0	9	0	3	0	6	1.1%
	2	908	0	4	0	4	0	3	1	0	0.4%
	3	673	0	1	0	1	0	1	0	0	0.1%
56 (WB)	1	234	0	0	0	0	0	-	-	-	0.0%
	2	446	0	2	0	2	0	1	1	0	0.4%
	3	70	0	0	0	0	0	-	-	-	0.0%
104 (EB)	1	136	0	0	0	0	0	-	-	-	0.0%
	2	435	0	1	0	1	0	0	1	0	0.2%
	3	401	0	2	0	2	0	2	0	0	0.5%
Total		10,504	0	32	0	32	0	17	9	6	0.3%

Table 4.16, Summary of the pulse breakup detection algorithm performance on stations where the detector sensitivity was changed.

CHAPTER 5. **DROPOUT WITHOUT RETURN- A PILOT STUDY**

Each on-time measurement, i.e., the duration during which a vehicle occupies a loop detector, depends on: vehicle length, vehicle speed, and the detector sensitivity. However, detector errors can cause the on-time to be longer or shorter than expected. For example, pulse breakup splits a vehicle's pulse into several shorter on-times. An extreme case occurs when the detector without return, e.g., when a semi-trailer truck passes the detector turns off at the end of the tractor, but instead of turning back on for the trailer axles (as it would in pulse breakup), the detector stays off. Or more generally, dropout without return indicates that a loop detector recognizes a part of a vehicle passed over a loop detector but it did not recognize all of the vehicle. Figure 5.1 shows such an example, using a video frame and concurrent loop detector data when a long truck entered the detection zone at lane 2 at station 2 northbound on 3/09/2009. This example clearly shows that on-time of the long vehicle is too short, 4/60 seconds, which is even shorter on-time than the on-time from the passenger car immediately before, 8/60 seconds.

After ground truthing 29min of data at station 2 northbound on 3/09/2009, the following totals and bracketed subtotals, [short vehicles, middle vehicles, long vehicles], were observed from video from lane 1 to lane 4: 623 [608, 8, 7], 638 [604, 5, 29], 519 [505, 3, 11], and 38 vehicles [37, 1, 0], respectively. Figure 5.2 shows the on-time CDFs by vehicle class for the four lanes at station 2 northbound. Note that no long vehicles are observed at lane 4. For all three classes the distribution of on-times in lane 2 is to the left of the corresponding distribution from the other lanes. The biggest difference of on-times between lane 2 and the other lanes is observed from long vehicles. While all three classes of vehicles exhibited shorter on-times in lane 2 than the other lanes, the difference is

severe for the long vehicles. This lane has a chronic problem of dropping out without return. All long vehicles observed at lane 2 have on-times less than 18/60 seconds, the distribution appears to be similar to the distribution of on-times from short vehicles. This fact and the absence of short off times makes it particularly difficult to differentiate these long vehicle dropouts without return from the short vehicle measurements using data from a single station.

We bring up this problem here to make clear that there are other detector errors that our tests do not catch. Obviously extreme dropouts without return would bring the on-times of all vehicles down (e.g., if the detectors were inadvertently set to pulse mode). For the present work we focus on the next level up, and observe that long vehicles should be the ones that are most susceptible to this error, e.g., Figure 5.2. Due to the difficulties distinguishing between long vehicles that dropped out and short vehicles that were measured correctly, we only seek to catch chronic problems (e.g., at least 30% of the long vehicles are impacted at a given station). Broadly, the flow of long vehicles should be similar between successive stations. But that approach will only help if it is known that one of the stations does not suffer from dropout without return. So borrowing vehicle reidentification ideas from Coifman (1998) and reversing the process, a vehicle observed at the upstream (or downstream) detector station must be seen at the downstream (or upstream) detector station offset by the travel time unless the vehicle exits (or enters) in between the stations. A vehicle's travel time between the two locations depend on its speed. Limiting this work to free flow conditions, we assume a range link speeds between 55 mph and 75 mph, i.e., speed limit (65 mph) ±10 mph, the free flow travel times can be estimated from the two bounding values of speed over the link. Consequently, most vehicles detected in upstream (or downstream) station would have a match in one of the lanes of the downstream (or upstream) station within a time window of reasonable free flow travel times. In this respect long vehicles have two additional desirable properties: because they typically pass less frequently than short vehicles there are fewer possibilities to consider and their longer range makes them easier to differentiate from one another

Obviously if there is a queue between the detector stations that does not reach either station the travel time will be larger than the assumed free flow travel time. But such conditions will be transient. This work is intended to be used over a longer period, e.g., at least a day, in which case it can reasonably be assumed that most of the time that both of the detectors stations indicate locally free flowing traffic that the entire link spanning the two stations is freely flowing. In fact this approach is complementary with the vehicle reidentification techniques presented Coifman (2003), if long vehicles are normally seen between successive stations within the free flow travel time window, the sudden absence can be used to detect the presence of queuing before it reaches either detector station.

Only long vehicles and locally free flow conditions at the pair of detector stations are considered in the following analysis. We selected all of the long vehicles from the ground truth video data at station 3 northbound on 6/17/2009, and then used the station 2 detector data to try to find the corresponding vehicles. For station 2 (and whenever one does not have video ground truth) in this proof of concept study we sought to bypass the confounding interaction between vehicle length and speed have on on-time at a single loop detector, so we used a simple threshold: on-time in excess of 35/60 seconds to defines a long vehicle.

Figure 5.3 shows a schematic around station 2 and station 3 northbound. Since there is an on-ramp between station 2 and station 3 northbound, flows at station 3 northbound should be greater or equal to flows at station 2 northbound. The ground truth data include 30 min sample data, 10:21 to 10:51 and traffic was free flowing. At station 3 the following totals of long vehicles are observed from video from lane 1 to lane 4: 5, 57, 48, and 3 vehicles, respectively. The free flow travel times between station 2 and station 3 are between 18 and 25 seconds (if link speed is between 55 and 75 mph). The two free flow travel times are subtracted from the arrival time for each long vehicle at station 3, thereby yielding an arrival time window at station 2. First checking on a lane-by-lane basis, and then all-lanes-to-all-lanes, whenever an on-time in the window exceeds the 35/60 seconds threshold the station 3 vehicle is recorded as having a possible match. The

results between station 2 and 3 are summarized in the first few columns of Table 5.1. The second column indicates the number of long vehicles seen in the video data at station 3, sorted by lane. A total 113 long vehicles are observed at station 3 northbound and as noted above, these vehicles were used to define the search time windows for the detector data at station 2 northbound. Only 53 out of 113 vehicles (about 50% of long vehicle) have a matched vehicle at station 2 northbound. Looking at the lane-by-lane results, lane 2 has only one matched long vehicle between the stations, while the two other through lanes have at least 80% of the long vehicles matched between the stations.

The process is repeated substituting data from station 1 (about a half mile further upstream) in place of station 2. The lane geometry is the same at stations 1 and 2, only the travel time and travel time windows increase. Now 109 out of the 113 long vehicles at station 1 northbound have a match (about 96% of long vehicles). Most notably, the number of matched vehicles in lane 2 between station 1 and station 3 is much higher than the previous pair, now with 51 out of 57 vehicles (89%) having a match. The other through lanes each have a rate equal to or lower than the given lane when using station 2 data, as one would normally expect over the longer distance (at least until the time windows become too large). These results are consistent with the CDFs of on-times observed directly at station 2, in Figure 5.2.

Generally one would not have video ground truth, in which case this method can be applied to the loop detector data in both directions- matching from station 3 to station 2 (as per above) but also from station 2 to station 3. Now, however, all stations would use the 35/60 seconds threshold to define a long vehicle.

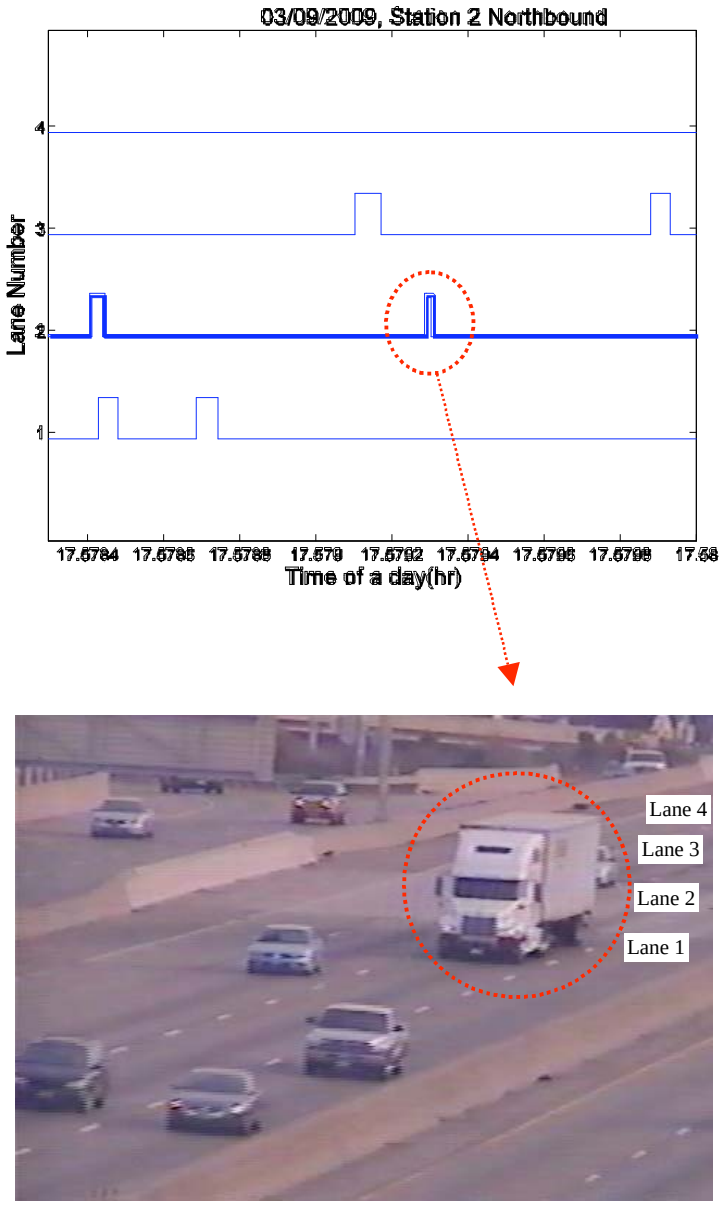


Figure 5.1, An example of drop-out without return (DOwoR) from a long vehicle in lane 2 during the 29 min of ground truth (17:21 to 17:50) on 3/09/2009.

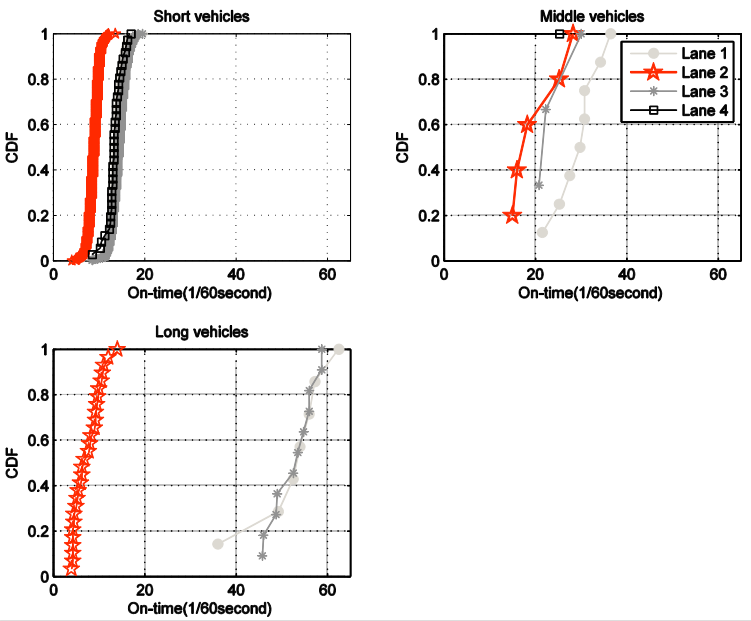


Figure 5.2, CDF of on-times at each loop detector by three classes of vehicles as measured from the concurrent video data

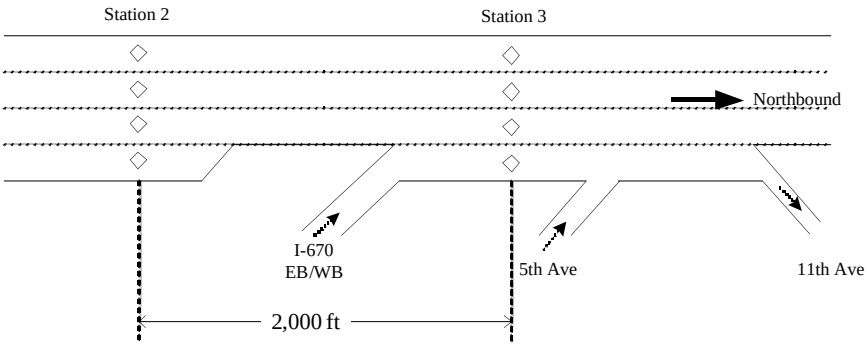


Figure 5.3, Study site used to find matched long vehicles between station 2 and station 3 northbound

Method	Number of long vehicles at station 3 [from GTD]	Station 3 vs. Station 2		Station 3 vs. Station 1	
		Number of matched vehicles	%	Number of matched vehicles	%
Lane 1 by Lane 1	5	4	80%	4	80%
Lane 2 by Lane 2	57	1	2%	51	89%
Lane 3 by Lane 3	48	39	81%	35	73%
Lane 4 by Lane 4	3	1	33%	0	0%
All lanes by All lanes	113	53	47%	109	96%

Table 5.1, Summary of the number of matched vehicles between station 3 and station 2 northbound and between station 3 and station 1 northbound

CHAPTER 6. **VEHICLE CLASSIFICATION FROM SINGLE LOOP DETECTORS**

6.1 Relationship Between the Standard 13 FHWA Vehicle Classes and Length Classes

As a starting point for the length based classification, we examine the relationships between the standard 13 FHWA vehicle classes and length classes. To remove any impact of detector errors, we use a data set where the vehicles were filmed from a 90 degree angle relative to the direction of travel and their lengths were extracted from video (see Coifman, 2007). These data were collected over four hours on I70 at Brice Rd. under free flow conditions. We developed a graphic user interface (GUI) to load each individual vehicle's image according to its detection on a vehicle classification station, manually measure its length, and classify it as per the 13 FHWA classes. A total of 9,746 vehicles were recorded at the location and we generated the ground truth data for 9,372 of them. The remaining 374 vehicles were sufficient obscured when they passed that we cannot identify number of axles and/or its length.

Using the ground truthed vehicle data, Figure 6.1 plots the distribution of vehicle lengths for each of the 13 FHWA vehicle classes. Looking at the distribution of each of the 13 FHWA vehicle classes, we group several neighboring FHWA vehicle classes together into three groups, as shown in Table 6.1. From the table, one might define or redefine X_{short} and X_{long}, the boundaries between the length based classes, by looking at measured vehicle length within those three groups. For this research, however, unless otherwise noted, we adopt the two divisions employed by ODOT: 22 ft and 40 ft. These divisions are superimposed on each plot in Figure 6.1 and are shown in the final column in Table 6.1. Figure 6.2 shows the histogram and CDF of vehicle length for the three clusters of FHWA classes. No matter what value of X_{short} and X_{long} are used, it should be

clear that there is not a pure one-to-one relationship between the clustered FHWA classes and the length classes, the overlapping tails in the length distributions ensure that vehicles from two different classes have similar lengths. More importantly, as enumerated in Table 6.2, over 80% of the vehicles fall into class 1. As we will see, attempts to minimize the average absolute error across all of the vehicles will tend to favor this group simply because it is the largest. So we also generated a set of synthetic data from the ground truth data, with the same number of vehicles, but now all three classes have the same number of vehicles. A given vehicle the synthetic data is sampled at random from the empirical ground truth data for the assigned class (thus ensuring that e.g., there will still be roughly the same percentage of class 1 vehicles that fall above Xshort). But since the total number of class 1 vehicles make up a smaller percent of the synthetic data, they will have a smaller impact on the optimal values for X_{short} and X_{long}.

Figure 6.3 shows the resulting histograms of length for each of the three clustered classes, and the histogram for the combined set on the bottom. The left column shows the empirical data, while the right column shows the synthetic data. Looking at class 2 vehicles, X_{short} and X_{long} seem to constraining (i.e., suboptimal) for the synthetic data. So Figure 6.4 shows all combinations of X_{short} and X_{long} and tallies the resulting number of incorrect classifications. As before the left hand column show the results for the empirical data and the right hand column show the results for the synthetic data. Here we assume a vehicle is incorrectly classified if its vehicle class from the clustered FHWA vehicle class does not match the class from the associated length based vehicle classification scheme, e.g., a vehicle falling into clustered FHWA vehicle class is 1 would be correctly classified if it also had a length class of 1, but it would be misclassified if the length class is 2. To understand how to read the plots, consider Figure 6.4A. Across the bottom axis, the X_{short} is stepped from 20 to 23 ft in 1 ft increments. Within each step, there are five clusters of bars that correspond to the sweep of X_{long} from 40 to 44 ft. Meanwhile, a cluster of bars show the number of incorrectly classified vehicles for given pair of X_{short} and X_{long}. In each cluster of bars the leftmost bar is for class 1 and the right most for class 3. The below, Figure 6.4C sums across the cluster of bars, showing the total number of

vehicles that would be misclassified for the given pair of X_{short} and X_{long}. Then Figure 6.4E & F present the same results in terms of percent of vehicles within the class or across all classes, respectively. The right hand side of the figure repeats the process for the synthetic data. From the empirical data, the optimal values are X_{short}=21 ft and X_{long}=40 ft, while from the synthetic data, the optimal values are X_{short}=20 ft and X_{long}=43 ft. The empirical data are very close to the 22 and 40 ft delineations that ODOT uses, but it also underscores the facts (1) the optimal boundary for the fleet is a function of the vehicle fleet, and (2) any underrepresented class will literally be squeezed by the length delineations.

6.2 Probability That a Given Length Based Vehicle Classification is True

One initially puzzling finding that we reported in Coifman (2007) and Coifman and Kim (2009) is the fact that our initial single loop detector, length based vehicle classification scheme performed much better on class 1 and class 3 vehicles (typically over 95% accuracy) than it did on class 2 vehicles (typically on the order of 75% accuracy). In this section we return to this question and answer it. Like the earlier study, we continue to use the distribution based method for estimating vehicle speed from single loop detectors. Basically this approach looks at the distribution of on-times, rather than just the mean or median, to deduce the speed. It provided consistent results from 10% vehicles to 90% long vehicles. Except during heavy congestion, the accuracy approaches that of dual loop detectors. Vehicle lengths are then estimated from the product of the estimated speed and measured on-time.

In the present study we construct a model to calculate the probability that a given estimated vehicle length based vehicle class corresponds to the vehicles true length class by employing the video based ground truth length data from I70 at Brice Rd. Since the distribution of the length estimation error can be constructed (i.e., the difference between the measured length from ground truth and estimated vehicle length from single loop detectors), we can calculate the probability that a given length estimate will fall in the same length class as that vehicle's true length given the boundaries between classes. For

example, suppose the estimated vehicle length of a vehicle is 18 ft and the distribution of estimation error followed a normal distribution with the mean 0 and variance σ^2, as shown in Figure 6.5. From the distribution we calculate the probability that the estimated vehicle length will exceed the 22 ft boundary between class 1 and class 2, i.e., the shaded area in the figure. Or conversely, from the distribution one can deduce the probability that a given vehicle class based on estimated vehicle length is consistent with that vehicle's true length class.

To test normality of our data, we took 100 samples randomly and generated normal probability plot as illustrated in Figure 6.6. So given that the distribution of estimation error follows normal distribution, we looked at how the sample variance changes as a function of vehicle length. Figure 6.7 shows the sample variance calculated from the data sampled every 2ft and the fitted line from linear regression. As can be seen, generally variance tends to increase as vehicle length increase. There are some sample variances off from the fitted line between 22 ft and 50 ft, potentially due simply to an insufficient number of samples in this range. Consequently we decided to use the fitted line to calculate the variance for a given vehicle length. The solid curves in Figure 6.8 show the resulting estimates for each of the three length ranges corresponding to the vehicle classes. In each case the probability drops to 50% at the boundary between classes. Next, we calculate the proportion of correct classification for vehicles actually observed, in 2ft bins, e.g., we counted total number of true class 1 vehicles out of all vehicles whose estimated length falls between 18ft and 20ft. The results are shown with points (which may also appear as a bold, piecewise horizontal line). The empirical results roughly follow the theoretical model, with the largest difference occurring for class 2 vehicles between 22ft and 26ft.

So the higher rate of class 2 classification errors is not surprising, it arose for several reasons. First, class 2 has two boundaries, so unlike the other two classes, by definition, all class 2 vehicle lengths are within 9 ft of one boundary or the other and thus, more susceptible to the boundary issue noted above. Roughly 40% of the class 2 vehicles were within 4 ft of a boundary while only 15% of class 1 (the short vehicles) and

under 10% of class 3 (the long vehicles) were within 4 ft of their respective boundaries. This problem is exasperated by the fact that the variance in the length estimation error increases with vehicle length, so it impacts class 2 more. The median length in class 3 is 28 ft away from the boundary, so even with a higher variance these vehicles are less likely to be misclassified. Such boundary errors also impact class 2 vehicles when using dual loop detectors to measure vehicle length since the measured speed also exhibits an error distribution while the on-time error distribution should be similar to that of single loop detectors.

6.3 Systematic Reasons Why a Single Loop Detector Length Based Classification might be Erroneous

6.3.1 **Distribution of Measured Speed for Each Class**

Since the speed estimation algorithm uses modes in the distribution of on-time to estimate individual vehicle speed, slower or faster speed from specific vehicle class might bias the speed estimation (e.g., slow moving heavy truck or fast moving passenger car). One might typically expect that class 1 vehicles are generally faster than class 2 and 3; and the majority of vehicles passing through the I71 corridor are class 1 vehicles. Thus speed and length of class 2 or 3 vehicles would be over-estimated due the fact that the speed estimation is more likely to favor the faster class 1 vehicles. Figure 6.9 shows the distribution of measured speed for each class from a dual loop detector station (station 1) along I71. Here we only take samples at free speed (between 45 and 80 mph) over three days of data and the vehicle class was derived from the measured speed.

As can be seen in Figure 6.9, the distribution for class 1 is distinct from the other two classes separately in the three through lanes[5] as well as the combined distribution across these lanes. In short, the speed of class 2 and 3 vehicles tends to be slower than that of class 1 vehicles. Table 6.3 tabulates the median speed in each of the lanes and

[5] Recall that this station actually has four lanes, but the right hand lane ends about one half mile downstream, it sees little traffic and even fewer long vehicles, to conserve space it is excluded from the presentation.

across the three. This table illustrates how different the speed of class 1 vehicles is from the other two classes, in each case the median speed for class 1 vehicles is about 5 mph faster than the other two classes. From this analysis, we recognize that aggregated speed and estimated vehicle length from it at a single loop detector might see some accuracy limitations not observed at dual loop detectors.

6.4 Examine the Mean and Median Speed for Correctly and Incorrectly Classified Vehicles Near the 22 ft Boundary

To investigate incorrect vehicle classification near the boundary between class 1 and 2 from single loop detectors, we looked at the median and mean of measured speed and estimated speed for all vehicles between 17 and 31 ft (i.e., class 1 and class 2 vehicles near the boundary between the two classes) from the ground truth data on I70 at Brice Rd. Table 6.4 shows the number of these near-boundary vehicles that fall into each of four combinations, first each vehicle is classified from the single loop detector data as being class 1 or class 2, and this determines which column it falls in ("EL" for 'estimated length'). Second, these vehicles are similarly classified from the video data, which determines which row ("GL" for 'ground truth length'). For each vehicle we find the measured (from dual loop detector) and estimated (from single loop detector) speed. Looking at the counts in misclassification cells (1^{st} row 2^{nd} column and 2^{nd} row and 1^{st} column), more GL class 2 vehicles are incorrectly classified than GL class 1. This result is consistent with Table 6.3, class 1 vehicles are more likely to travel faster than median speed. Therefore they are less likely to be overclassified. To verify this supposition, within each combination from Table 6.4, we then calculate the ratio and the difference between the measured and estimated mean speed for each cell from the dual and single loop detectors, then repeat the process for the median speed, as enumerated in Table 6.5. This table clearly shows that a difference in speed between incorrect and correct classification indeed exists. The difference is closer to zero in the two cells where both classifications agree compared to the two cells where the classifications differ. If a class 1

vehicle is moving slower than the estimated speed, it is more likely to be placed in class 2 based on the single loop detector data, similarly, if a class 2 vehicle is moving faster than the estimated speed, it is more likely to be placed in class 1. As noted previously, the speed estimation algorithm still seeks the center speed of a sample (unlike chapter 4, now the sample is only 31 vehicles long), so if a vehicle's true speed is far from the center, these errors can arise. While these errors degrade classification performance some, from Table 6.4, they impact less than 10% of the vehicles in the study set and much less than 10% of the entire population. As will be seen shortly, the classification algorithm still yields good results in spite of these systematic errors.

6.5 Test Performance Against Additional Ground Truth Data

Moving beyond the existing ground truth data from I70 at Brice Rd, we collected many hours of ground truth data throughout the CMFMS (as presented in Chapters 3 and 4). Since the views from the video were not ideal for measuring vehicle length, we manually classified each vehicle into small, median, long vehicle class based on the clustered FHWA classes in Table 6.1. The additional data sets are listed in Table 6.6. We evaluated the results with and without the pulse breakup data cleaning algorithms from Chapter 4 (henceforth, referred to as cleaned data). As can be seen in Table 6.6, in addition to several sets from free flow conditions, three sets of data have were collected during congested conditions and another three during mixed conditions where we can observe free flowing and congested traffic in one data set. As a by-product from ground truthing, we found several stations with splashover, pulse breakup and various stray detector errors (e.g., due to lane change maneuvers). These features are used to cluster the ground truth data sets in Table 6.6, we categorized each set of ground truth data into three groups: stations without actual pulse breakup and splash over; stations with pulse breakup but no splash over; and stations with splash over.

After clustering the data in this fashion, we evaluate the single loop detector, length estimation algorithm by comparing length based vehicle class with ground truthed vehicle class. Table 6.7 shows the individual station results from the raw data for those

stations without pulse breakup and without splashover. Each subtable is for one data set and combines the data from all lanes. Our classification determines the row in which a given sample will fall in (C1-C3 based on EL). If we had merged pulses suspected of arising from pulse breakup that were actually distinct cars (as we will in the next table), these would have been recorded in the final row. Whenever the pulse is indeed due to a real vehicle, the result will fall in one of the first three columns based on its clustered FHWA class (C1-C3 based on GL). In this case there were no pulse breakups, but had there been, the first pulse would have fallen in C1-C3 columns and the 2nd (or more) would be recorded in the pulse breakup (PBU) column. Finally, there were several pulses due to vehicles changing lanes over the detectors (LCM) or simply splashover (SO), these are recorded in the final column. The table itself does not allow overcounting errors to cancel undercounting errors. The successfully classified vehicles fall on the diagonal, (C1,C1), (C2,C2), and (C3,C3). Below each subtable we note the number of vehicles (if any) that were obscured in the ground truth data. Then subtract the row sum (EL) from the column sum (GL) for each class to find the net incorrect classification with and without the non-vehicle pulses (NVP).

Table 6.8 repeats this process for those stations with pulse breakup. Now, however, the tables are doubled, on the left are results from the raw data, and on right the results from the cleaned data. In the raw data most of the pulse breakup events wind up being erroneously classified as C1, but in the cleaned data, most of these events have been resolved (though some errors clearly remain).

Table 6.9 repeats the process for those stations with splashover. Note that combined splashover events in the main grid are now highlighted parenthetically, e.g., "9(8)" means eight of the nine observations in that cell are due to combined splashover events.

Finally, Table 6.10 sums across each of the sub-groups. The correct classifications are in shaded cells. Below, the "Correctly classified" row takes the sum of the shaded cells, as well as the total number of vehicles shown on the given subtable. The next row

enumerates the number of pulses that were incorrectly classified, as well as tallying how many of those were due to non-vehicle pulses and thus, could not be correctly classified.

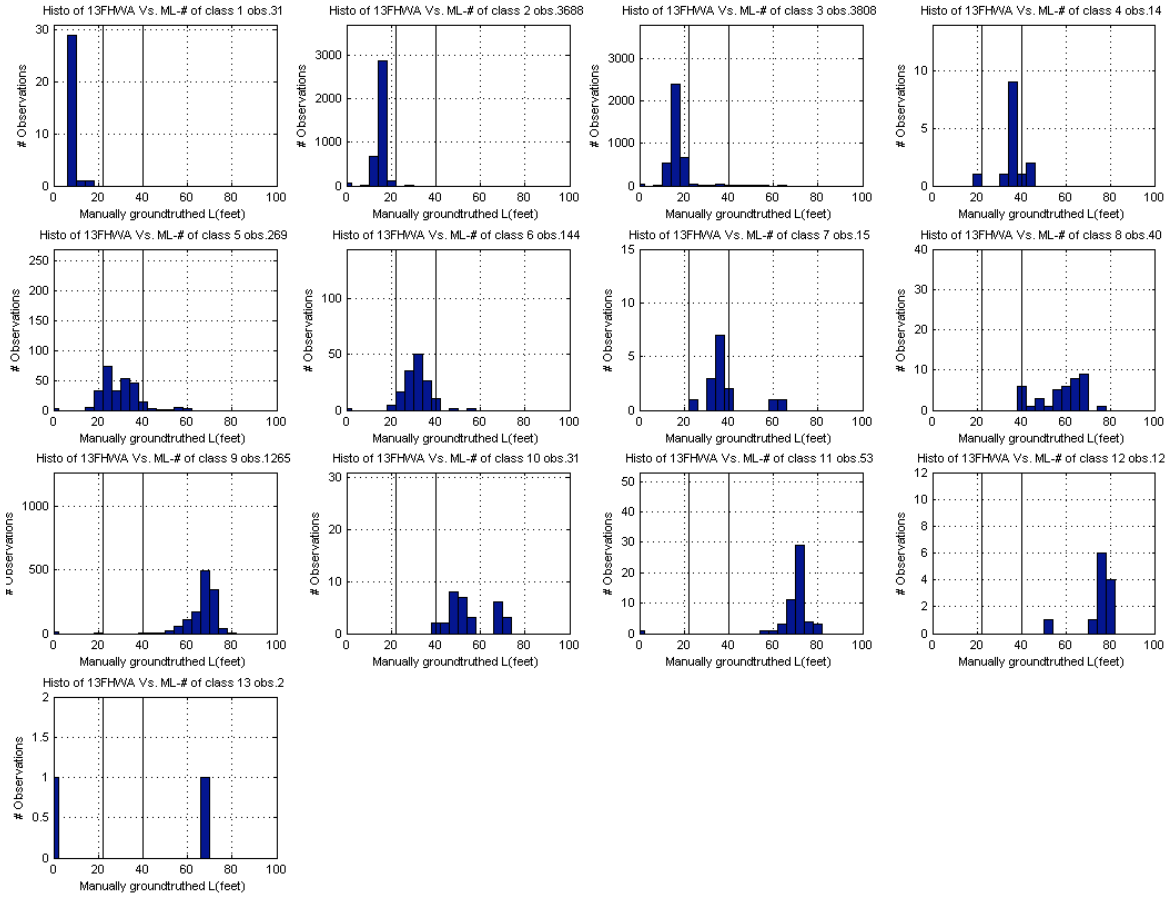
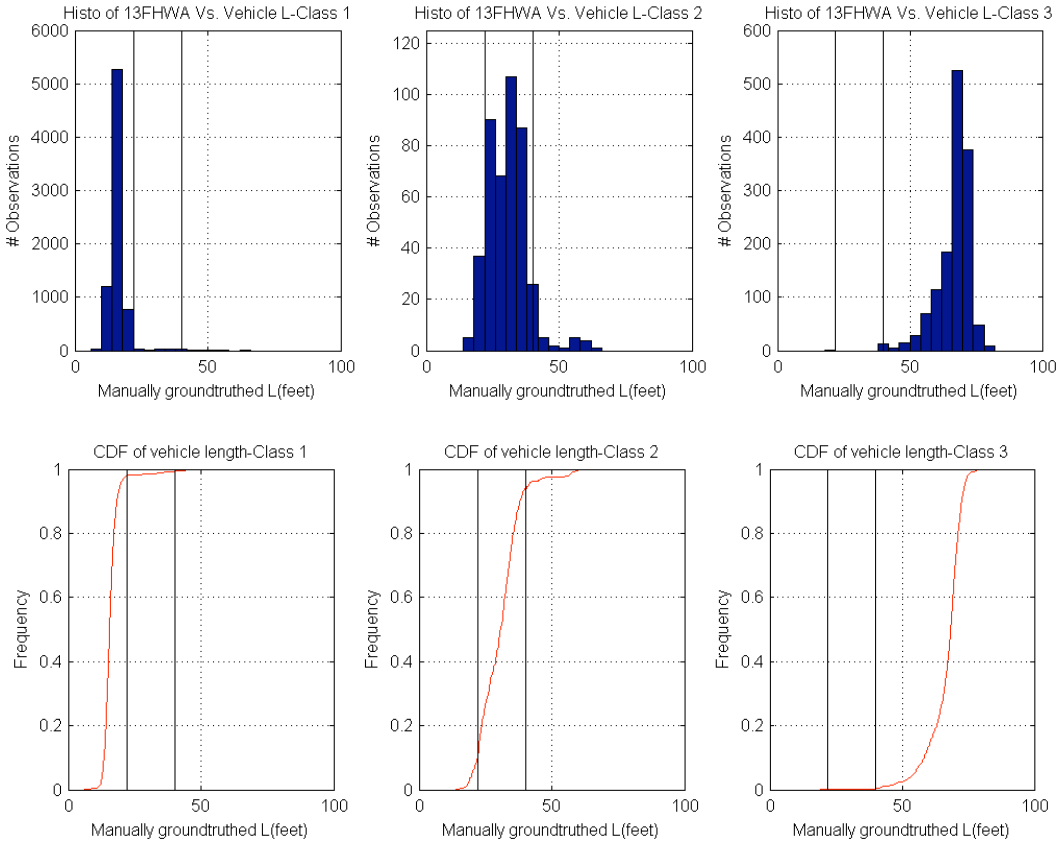


Figure 6.1, Histogram of ground truth, measured length for each of the 13 FHWA vehicle classes



# of vehs above 22ft: 154(2%)

# of vehs below 22ft: 42 (9.5%)
# of vehs above 40ft: 3 (0.06%)

# of vehs below 40ft: 3 (0.2%)

Figure 6.2, Distributions of measured vehicle length for the three clustered FHWA vehicle classes

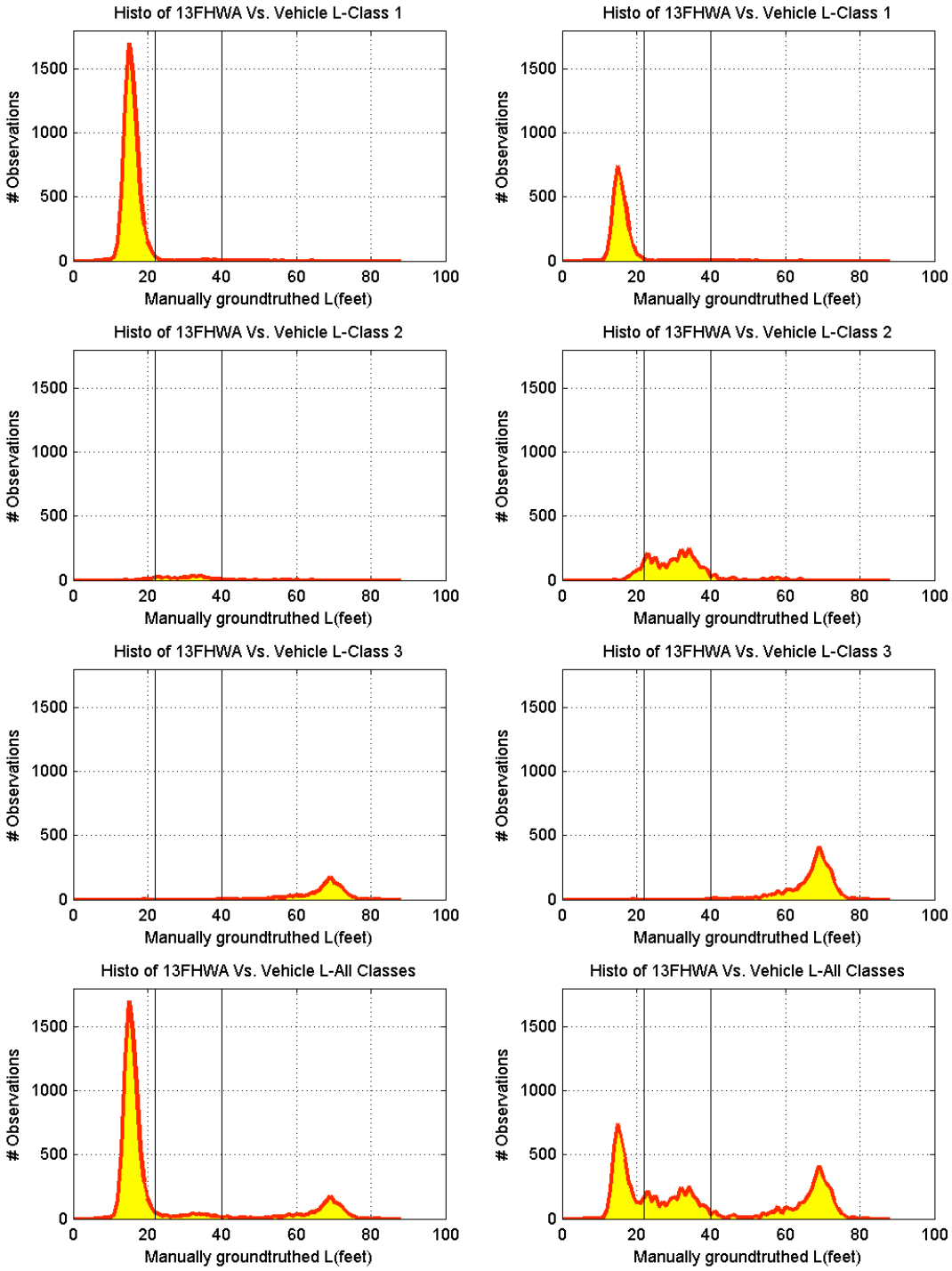


Figure 6.3, Histogram of measured vehicle length in ground truth data for each of the three clustered classes, and the histogram for the combined set on the bottom. The left column shows the empirical data, while the right column shows the synthetic data.

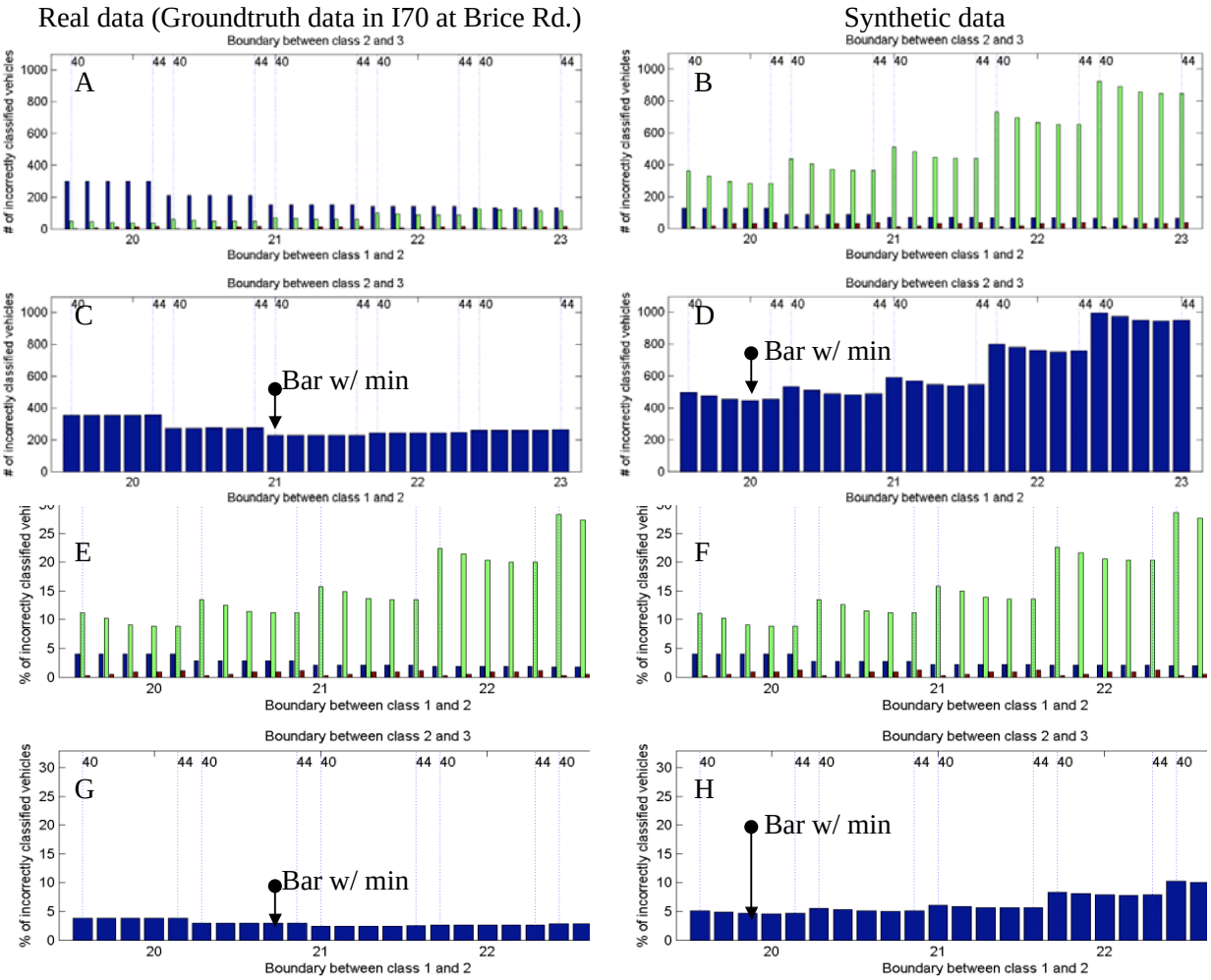


Figure 6.4, Incorrect vehicle classification (A&B: # of incorrectly classified vehicle for each class, C&D: # of incorrectly classified vehicle for all classes, E&F: % of incorrectly classified vehicle for each class, G&H: % of incorrectly classified vehicle for all classes)

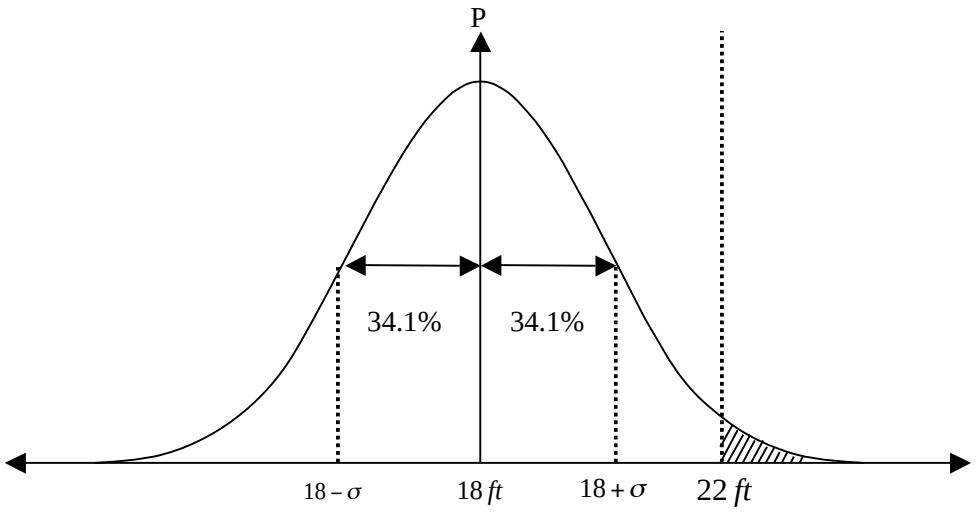


Figure 6.5, Distribution of measurements from an 18 ft vehicle

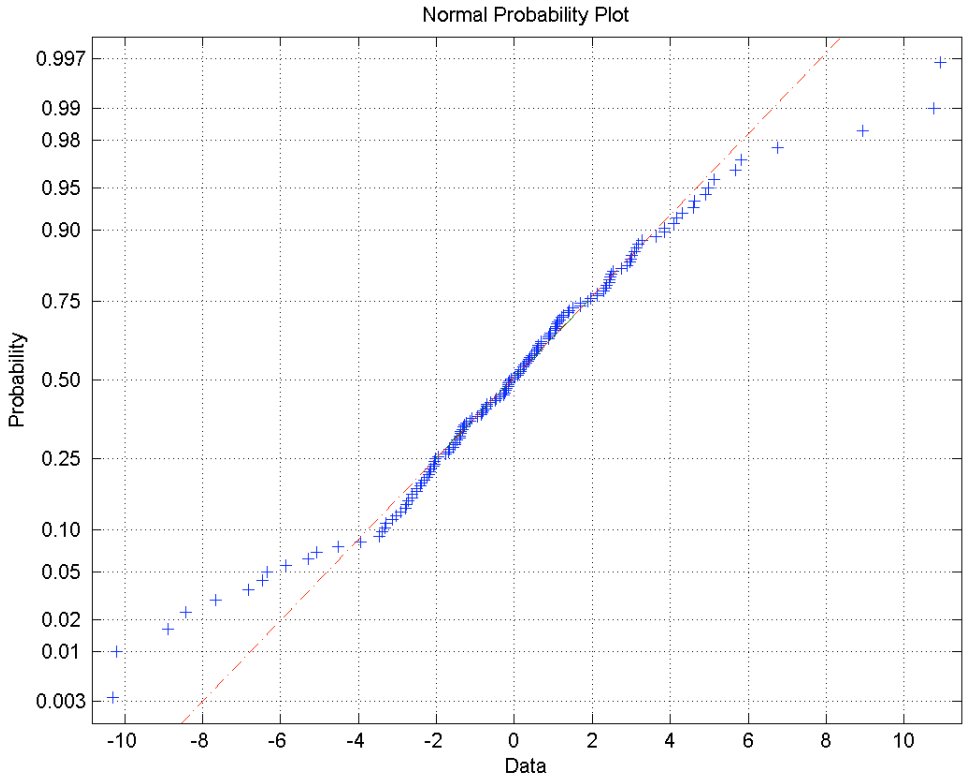


Figure 6.6, Normality probability plot

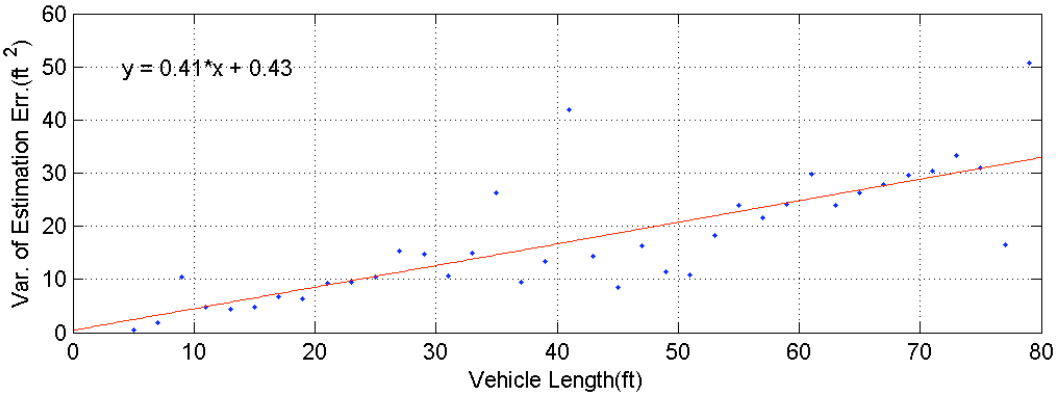


Figure 6.7, Variance of estimation error as a function of vehicle length

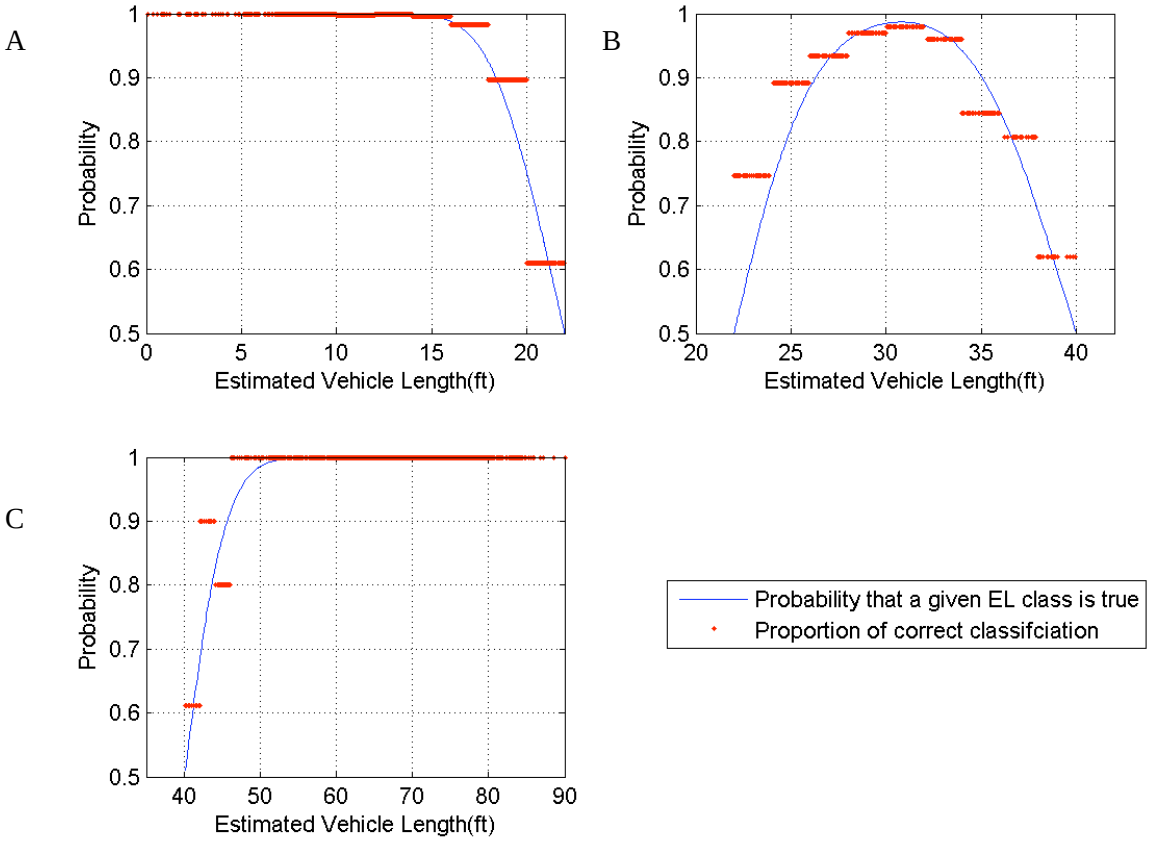


Figure 6.8, Probability that estimated length based vehicle class is true as a function of length (A: Class 1, B: Class 2, C: Class 3)

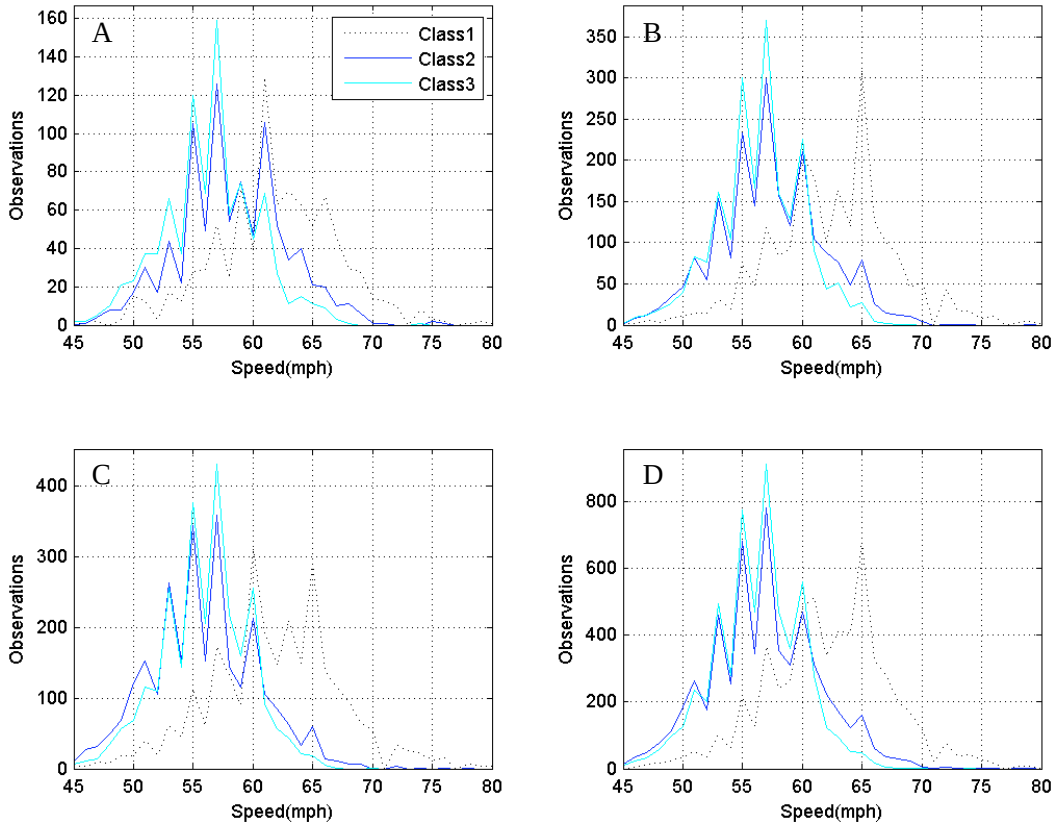


Figure 6.9, Distribution of speed (A: Lane 1, B: Lane 2, C: Lane 3, D: All lanes combined) using three days of typical data, excluding speeds beyond 45 and 80 mph.

	Clustered FHWA VCS.	Length Based VCS.
Class 1	FHWA class 1,2, and 3	$0<ML< X_{short}$
Class 2	FHWA class 4, 5, 6 and 7	$X_{short}<ML< X_{long}$
Class 3	FHWA class 8, 9,10,11,12 and 13	$X_{long}<ML$

Table 6.1, Clustered FHWA vehicle classification scheme (VCS) and length based vehicle classification scheme

Number of samples	Class 1	Class 2	Class 3
"Sort by FHWA"	7424	438	1385
"Sort by true vehicle len"	7324	478	1472

Table 6.2, Number in each class

	Class 1	Class 2	Class 3
Lane 1	62.22 mph	58.13 mph	56.64 mph
Lane 2	62.76 mph	57.27 mph	56.56 mph
Lane 3	61.91 mph	55.87 mph	55.87 mph
All lanes	62.22 mph	56.56 mph	56.56 mph

Table 6.3, Median free speed for each length class, by lane

Number of vehicles	Class 1 from EL	Class 2 from EL
Class 1 from GL	650	40
Class 2 from GL	48	301

Table 6.4, Classifying vehicles between 17 and 31 ft based on ground truth length (GL) and estimated length (EL), this table presents the distribution.

A1) Mean speed	Measured */ Estimated**	
	Class 1(EL)	Class 2(EL)
Class 1(GL)	60.75/62.97	50.51 /59.38
Class 2(GL)	63.49/59.94	62.21/61.98

B1) Median speed	Measured */ Estimated**	
	Class 1(EL)	Class 2(EL)
Class 1(GL)	61.60/62.50	50.60 /61.32
Class 2(GL)	62.70/59.10	61.60/60.80

A2)Difference	Measured *- Estimated**	
	Class 1(EL)	Class 2(EL)
Class 1(GL)	-2.22	-8.87
Class 2(GL)	3.55	0.23

B2)Difference	Measured *- Estimated**	
	Class 1(EL)	Class 2(EL)
Class 1(GL)	-0.9	-10.72
Class 2(GL)	3.6	0.80

*: from dual loop detector in I70 at Brice

**: from distribution method

Table 6.5, Mean and median of measured and estimated speed.

Date	Station	[1] # of true PBU	[2] # of susp. PBU	[3] # of true PBU in [2]	[4] # of incor. PBU in [2]	Total # of pulse	[1]/[2]	Cond.
-Stations without actual pulse break up and splash over								
20090312	104WB	0	1	0	1	1519	NaN	FF
20080902	43EB	0	1	0	1	1058	NaN	FF
20080902	43WB	0	5	0	5	1518	NaN	FF
20080903	56EB	0	5	0	5	1688	NaN	FF
20090310	102EB	0	5	0	5	1596	NaN	Mix
20081121	31NB	0	3	0	3	638	NaN	FF
-Stations with pulse break up but no splash over								
20090309	18NB	16	15	15	0	549	0.94	FF
20090310	15NB	40	34	34	0	2978	0.85	Mix
20080317	19NB	24	24	24	0	753	1	FF
20080317	3NB	49	44	43	1	1214	0.88	FF
20080418	3NB	120	113	101	12	5709	0.84	CG
-Stations with splash over								
20080317	104EB	0	44	0	44	1799	NaN	Mix
20081121	56WB	0	5	0	5	1098	NaN	FF
20090312	41EB	0	53	0	53	1668	NaN	CG
20080903	56WB	0	30	0	30	2850	NaN	CG
20080909	41EB	0	14	0	14	842	NaN	FF

[1]: # of true PBU: number of pulse break ups observed via ground truthing video
[2]: number of suspected PBU: number of pulse break ups that the data cleaning algorithm considers as pulse break ups
[3]: # of true PBU in [2]: number of suspected pulse break ups that turns out to be true via ground truting video
[4]: number of incur PBU in [2]: number of suspected pulse break ups that turns out not to be true via ground truting video

Table 6.6, Summary of cleaned data

[1] 20090312/station104WB/ FF

W/ Raw Data		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	1362	6	0	0	13
	C 2	5	26	0	0	0
	C 3	0	4	103	0	0
Incorrect PBU		0	0	0	0	0

# of obscured vehicle not included in table: 0
Net incorrect classification w/o NVP (-1, 5, -4)
Net incorrect classification w/ NVP (-14, 5, -4)

[2] 20080902/station43EB/FF

W/ Raw Data		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	975	4	0	0	14
	C 2	1	43	0	0	0
	C 3	0	2	19	0	0
Incorrect PBU		0	0	0	0	0

# of obscured vehicle not included in table: 7
Net incorrect classification w/o NVP (-3, 5, -2)
Net incorrect classification w/ NVP (-17, 5, -2)

[3] 20080902/station43WB/FF

W/ Raw Data		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	1409	15	0	0	41
	C 2	1	36	0	0	0
	C 3	0	4	12	0	0
Incorrect PBU		0	0	0	0	0

# of obscured vehicle not included in table: 16
Net incorrect classification w/o NVP (-14, 18, -4)
Net incorrect classification w/ NVP (-55, 18, -4)

[4] 20080903/station56EB/FF

W/ Raw Data		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	1514	6	0	0	86
	C 2	3	46	0	0	2
	C 3	0	3	28	0	0
Incorrect PBU		0	0	0	0	0

# of obscured vehicle not included in table: 0
Net incorrect classification w/o NVP (-3, 6, -3)
Net incorrect classification w/ NVP (-89, 4, -3)

[5] 20090310/station102EB/Mix

W/ Raw Data		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	1428	8	0	0	16
	C 2	15	28	0	0	1
	C 3	0	5	95	0	0
Incorrect PBU		0	0	0	0	0

# of obscured vehicle not included in table: 0
Net incorrect classification w/o NVP (7, -2,-5)
Net incorrect classification w/ NVP (-9, -3,-5)

[6] 20081121/station31NB/FF

W/ Raw Data		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	457	6	0	0	17
	C 2	0	20	0	0	0
	C 3	0	7	129	0	2
Incorrect PBU		0	0	0	0	0

# of obscured vehicle not included in table: 0
Net incorrect classification w/o NVP (-6, 13,-7)
Net incorrect classification w/ NVP (-23, 13,-9)

Table 6.7, Vehicle classification at stations without actual pulse breakup or splash over

[1] 20090309/station18NB/FF

W/ Raw Data

		Groundtruth			Non veh. Pulse		W/ Cleaned Data Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO	C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	495	0	0	14	6	495	0	0	0	6
	C 2	3	18	12	1	0	3	18	1	0	0
	C 3	0	3	11	1	0	0	3	22	1	0
Incorrect PBU		0	0	0	0	0	0	0	0	0	0

# of obscured vehicle not included in table: 0

Net incorrect classification w/o NVP (3, -12, 9) (3, -1, -2)

Net incorrect classification w/ NVP (-17, -13, 8) (-3, -1, -3)

[2] 20090310/station15NB/Mix

W/ Raw Data

		Groundtruth			Non veh. Pulse		W/ Cleaned Data Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO	C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	2879	11	1	39	2	2879	9	1	6	2
	C 2	9	7	32	1	0	9	9	4	0	0
	C 3	0	13	18	0	0	0	13	46	0	0
Incorrect PBU		0	0	0	0	0	0	0	0	0	0

# of obscured vehicle not included in table: 0

Net incorrect classification w/o NVP (-3, -17, 20) (-1, 9, -8)

Net incorrect classification w/ NVP (-44, -18, 20) (-9, 9, -8)

[3] 20080317/station19NB/FF

W/ Raw Data

		Groundtruth			Non veh. Pulse		W/ Cleaned Data Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO	C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	677	6	0	24	0	677	6	0	0	0
	C 2	1	28	20	0	0	1	28	0	0	0
	C 3	0	2	19	0	0	0	2	39	0	0
Incorrect PBU		0	0	0	0	0	0	0	0	0	0

# of obscured vehicle not included in table: 0

Net incorrect classification w/o NVP (-5, -13, 18) (-5, 7, -2)

Net incorrect classification w/ NVP (-29, -13, 18) (-5, 7, -2)

Table 6.8, Vehicle classification at stations with pulse break up but no splash over- part 1 of 2

[4] 20080317/station3NB/FF

W/ Raw Data

		Groundtruth			Non veh. Pulse		W/ Cleaned Data Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO	C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	1078	4	5	49	0	1076	2	2	6	0
	C 2	12	37	32	0	0	12	39	4	0	0
	C 3	0	8	32	0	0	1	8	63	0	0
Incorrect PBU		0	0	0	0	0	1	0	0	0	0

# of obscured vehicle not included in table: 0

Net incorrect classification w/o NVP (3, -32, 29) (10, -6, -3)

Net incorrect classification w/ NVP (-46, -32, 29) (4, -6, -3)

[5] 20080418/station3NB/CG

W/ Raw Data

		Groundtruth			Non veh. Pulse		W/ Cleaned Data Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO	C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	4652	44	8	105	32	4742	40	1	14	32
	C 2	577	56	63	14	0	505	55	3	4	0
	C 3	164	35	59	1	0	134	40	126	1	0
Incorrect PBU		0	0	0	0	0	12	0	0	0	0

# of obscured vehicle not included in table: 0

Net incorrect classification w/o NVP (689, -561, -128) (610, -428, -170)

Net incorrect classification w/ NVP (552, -575, -129) (564, -432, -171)

Table 6.8, Vehicle classification at stations with pulse break up but no splash over- part 2 of 2

*: Table with combined splash over
(#): total # of combined splash over

[1] 20080317/station104EB/ Mix*

W/ Raw Data		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	1083(25)	5	0	0	401
	C 2	122(57)	32	1	0	10
	C 3	17(13)	4	99	0	25
Incorrect PBU		0	0	0	0	0

# of obscured vehicle not included in table: 0
Net incorrect classification w/o NVP (134, -114,-20)
Net incorrect classification w/ NVP (-267, -124,-45)

[5] 20080909/station41EB/FF*

W/ Raw Data		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	608	9	0	0	68
	C 2	9(8)	24	0	0	22
	C 3	9(9)	6	82	0	5
Incorrect PBU		0	0	0	0	0

# of obscured vehicle not included in table: 0
Net incorrect classification w/o NVP (9, 6,-15)
Net incorrect classification w/ NVP (-59, -16,-20)

[2] 20081121/station56WB/ FF

W/ Raw Data		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	921	10	0	0	58
	C 2	2	49	0	0	6
	C 3	0	4	48	0	0
Incorrect PBU		0	0	0	0	0

# of obscured vehicle not included in table: 0
Net incorrect classification w/o NVP (-8, 12,-4)
Net incorrect classification w/ NVP (-66, 6,-4)

[3] 20090312/station41EB/CG*

W/ Raw Data		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	1408	4	0	0	22
	C 2	131(9)	27	0	0	1
	C 3	20(13)	5	50	0	0
Incorrect PBU		0	0	0	0	0

# of obscured vehicle not included in table: 0
Net incorrect classification w/o NVP (147, -122,-25)
Net incorrect classification w/ NVP (125, -123,-25)

[4] 20080903/station56WB/CG

W/ Raw Data		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	2516	15	0	0	74
	C 2	128	34	0	0	5
	C 3	42	9	22	0	5
Incorrect PBU		0	0	0	0	0

# of obscured vehicle not included in table: 0
Net incorrect classification w/o NVP (155, -104,-51)
Net incorrect classification w/ NVP (81, -109,-56)

Table 6.9, Vehicle classification at stations with splash over

[1] Summary of stations w/o PBU or SO (All FF)

W/ Raw Data		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	5717	37	0	0	171
	C 2	10	171	0	0	2
	C 3	0	20	291	0	2
Incorrect PBU		0	0	0	0	0

Correctly classified: **6179/6421**
Incorrectly classified: **242 (67 w/o Non veh. Pulse)**
Net incorrect classification w/o NVP (-27, 47,-20)
Net incorrect classification w/ NVP (-198, 45,-22)

[2] Summary of stations w/ PBU

W/ Raw Data during FF

		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	2250	10	5	87	6
	C 2	16	83	64	1	0
	C 3	0	13	62	1	0
Incorrect PBU		0	0	0	0	0

Correctly classified: **2395/2598**
Incorrectly classified: **203 (108 w/o Non veh. Pulse)**
Net incorrect classification w/o NVP (1,-57, 56)
Net incorrect classification w/ NVP (-92,-58, 55)

W/ Cleaned Data during FF

		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	2248	8	2	6	6
	C 2	16	85	5	0	0
	C 3	1	13	124	1	0
Incorrect PBU		1	0	0	0	0

2457/2516
59 (46 w/o Non veh. Pulse)
(8, 0, -7)
(-4, 0,-8)

W/ Raw Data during CG (st. 3NB)

		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	4652	44	8	105	32
	C 2	577	56	63	14	0
	C 3	164	35	59	1	0
Incorrect PBU		0	0	0	0	0

Correctly classified: **4767/5810**
Incorrectly classified: **1043(891 w/o Non veh. Pulse)**
Net incorrect classification w/o NVP (689,-561,-128)
Net incorrect classification w/ NVP (552,-575,-129)

W/ Cleaned Data during CG (st. 3NB)

		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	4742	40	1	14	32
	C 2	505	55	3	4	0
	C 3	134	40	126	1	0
Incorrect PBU		12	0	0	0	0

4923/5709
786 (735 w/o Non veh. Pulse)
(610,-428,170)
(564,-432,-171)

[3] Summary of stations w/ SO

During FF

W/ Raw Data		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	1529	19	0	0	126
	C 2	11	73	0	0	28
	C 3	9	10	130	0	5
Incorrect PBU		0	0	0	0	0

Correctly classified: **1732/1940**
Incorrectly classified: **208 (49 w/o Non veh. Pulse)**
Net incorrect classification w/o NVP (1, 18,-19)
Net incorrect classification w/ NVP (-125,-10,-24)

During CG (st. 41EB and 56WB)

W/ Raw Data		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	3924	19	0	0	96
	C 2	259	61	0	0	6
	C 3	62	14	72	0	5
Incorrect PBU		0	0	0	0	0

4057/4518
461 (354 w/o Non veh. Pulse)
(302, -259,-76)
(206,-232,-81)

Table 6.10, Summary of FF and CG (does not include Mix samples)- part 1 of 2.

[4a] Summary during congestion

W/ Raw Data		Groundtruth			Non veh. Pulse	
		C 1	C 2	C 3	PBU	LCM/SO
EL	C 1	8576	63	8	105	128
	C 2	836	117	63	14	6
	C 3	226	49	131	1	5
Incorrect PBU		0	0	0	0	0

Correctly classified: **8824/10328**
Incorrectly classified: **1504 (1245 w/o Non veh. Pulse)**
Net incorrect classification w/o NVP (991,-787,-204)
Net incorrect classification w/ NVP (758,-807,-210)

[4b] Summary during congestion excluding known error (e.g., non veh pulse, combined so, and first pulse associated with PBU)

W/ Raw Data		Groundtruth		
		C 1	C 2	C 3
EL	C 1	8576	59	0
	C 2	827	113	1
	C 3	213	49	93
Incorrect PBU		0	0	0

Correctly classified: **8782/10069**
Incorrectly classified: **1149**

Table 6.10, Summary of FF and CG (does not include Mix samples)- part 2 of 2.

CHAPTER 7. **CONCLUSIONS**

This chapter summarizes the research, highlights its contributions, and proposes directions for future research.

7.1 Summary

Most of the states served by this regional UTC have deployed a large number of single loop detectors for real time traffic monitoring (e.g., Chicago has 2,400 single loop detectors, while Mineapolis/St. Paull has 3,500). Almost all of the states in this region also have single loop detector count stations used to measure AADT without classifying vehicles. The research seeks to refine and further develop a method to provide a reliable estimate of individual vehicle speed and length, initiated by the predecessor to NEXTRANS. This new, reliable, single loop detector methodology for classifying vehicles based on estimated vehicle length is significant because it will: (i) provide a low cost means of collecting vehicle classification data to supplement existing systems, leveraging the existing detector infrastructure, integrating new data collection with the existing surveillance system at the thousands of single loop detectors on the freeways within the region served by the regional UTC (it is meant to supplement rather than supplant other vehicle classification technologies), (ii) provide a viable means of estimating speed and length at a conventional classification station when one loop fails in a dual loop detector. The research has secondary benefits as well, in a draft research statement from the TRB Committee on Highway Traffic Monitoring, "Classification based solely on vehicle length is an alternative to axle-based classification but there has been no systematic study of how well it works -- or how it should work." The proposed

research will have to address many of these issues, and it should further the state of the art in dual loop detector based vehicle classification as well.

The resulting expansion of classification coverage and performance will in turn feed the various applications that rely on vehicle classification data. The new truck counts should benefit a wide range of applications, ranging from infrastructure maintenance and rehabilitation to better modeling of freight shipments in urban areas for the planning process.

Finally, tertiary benefits should be expected in the area of improved error corrections from both single and dual loop detectors and speed estimates from single loop detectors. The need to catch detector dropouts will benefit most means of conventional traffic monitoring, while the efforts to improve single loop speed estimation during congestion will benefit the many existing single loop detector based systems (and the emerging detectors that emulate single loop detectors).

7.2 Findings

Speed estimation, length estimation, and vehicle classification algorithms were developed and improved in the course of this work. Approximately 21 hours of directional traffic data were ground truthed from 34 different data sets collected at 22 different locations and an average of 3.3 lanes per set. A total of 78,774 detector actuations were manually ground truthed (in the absence of a detector error, there should be exactly one actuation per vehicle). Roughly a quarter of these data come from congestion. Three different, chronic detector errors were observed at several of the detector stations: splash-over (SO), pulse breakup (PBU), and detector dropout without return (DOwoR). These errors degrade classification performance as well as conventional speed, flow and occupancy; at single loops and dual loops alike. Preliminary diagnostic algorithms for identifying SO and PBU errors were developed and should be transferable to most loop detector stations (single loops and dual loops alike). The SO algorithm only detects the presence of the problem. The PBU algorithm is able to go further, it can repair most of the observed errors. Working with ODOT, we adjusted the detector settings at four detector stations and we were successful in eliminating the chronic detector errors at

these stations. If these results are typical, the improved detector calibration enabled by our research could lead to a very inexpensive means to improve the quality of loop detector data at existing stations.

During free flow: at stations without PBU and without SO we had a correct classification rate of 96%, of the errors (72% of the errors were due to non-vehicle pulses (NVP), in this case due to vehicles changing lanes over the detector). The correct classification rate drops to 92% from raw data at stations with PBU (47% of errors due to NVP, including extra pulses from PBU), but improves to 98% when using our diagnostic algorithms to eliminate PBU (78% of errors due to NVP). The correct classification rate drops to 89% at stations with SO (76% of errors due NVP). Note that this analysis was conducted on a per vehicle basis, so in error with one vehicle is not allowed to cancel an error with another vehicle. During congestion: all stations used for classification evaluation exhibited PBU or SO, we had a correct classification rate of 85% (17% due to NVP), but improves to 88% when using our diagnostic algorithms to eliminate PBU (12% of errors due to NVP).

Performance from the single loop detectors is comparable to dual loop detectors when traffic is free flowing. The length based classification performance degrades by about 10% during congestion because the individual speed estimates are still based on a sample of vehicles and in heavy congestion it is possible for a given vehicle's true speed to be far from the center of the sample. These congested conditions can be identified based on the speed estimates, so if the degradation is unacceptable the classification results can be discounted or subsequent research can develop adjustment factors.

After excluding the chronic detector errors (PBU, SO, and DOwoR), most classification errors were due to a true vehicle length being close to the boundary between two bins and the estimated length falling just on the other side of the boundary. Using thresholds of 22 and 40 ft between vehicle classes, class 2 (the middle class) had a significantly higher error rate than the other two classes. The higher rate of class 2 errors arose for several reasons, first, class 2 has two boundaries, so unlike the other two classes, by definition, all class 2 vehicle lengths are within 9 ft of one boundary or the

other and thus, more susceptible to the boundary issue noted above. Roughly 40% of the class 2 vehicles were within 4 ft of a boundary while only 15% of class 1 (the short vehicles) and under 10% of class 3 (the long vehicles) were within 4 ft of their respective boundaries. Such boundary errors also impact class 2 vehicles when using dual loop detectors to measure vehicle length.

7.3 Future directions

Discovering the extent of the chronic detector errors was an unanticipated byproduct of this research, but it may also prove to be one of the most significant findings since it potentially impacts most loop detector deployments. With conventional detector aggregation, e.g., 30 sec or 5 min averaging, the chronic errors often go unnoticed unless they are severe. Our diagnostic algorithms show great promise for detecting PBU and SO. After further refinement, in the short term these algorithms could be incorporated into a field diagnostic tool to assess the performance of a given station, either by tapping into the data upstream of the controller, e.g., via the InfoTek Wizard, or running an alternate controller program for a day or two, e.g., Caltrans Log_170. In the longer run, such tests should be incorporated into the regular controller software so that the controller can continually assess the health of the detectors. More research is necessary for catching DOwoR since the resulting time series from these errors are usually indistinguishable from the passage of a shorter vehicle. We have made some progress in catching DOwoR by comparing vehicle actuations between successive stations, but more work is needed. In the mean time, as one might expect, all of the stations that we observed having DOwoR also exhibited PBU. So in these cases, it is still possible to identify that the station has a problem. Operating agencies and freeway vehicle detector manufacturers (loop detector and non-invasive detectors) should evaluate these tools for potential adoption.

Operating agencies with single loop detectors should consider deploying the vehicle classification scheme developed in this research as a means to extract more information from their existing detector infrastructure. Similarly, manufacturers of non-

invasive detectors that emulate single loop detectors (e.g., Image Sensing Systems-RTMS) should consider employing these ideas in their classification scheme.

Finally, a practical length based vehicle classification scheme needs to be robust to the large discrete steps between classes (whether from single or dual loop). Further work is needed to develop strategies for mitigating these boundary errors. One example is the simple strategy of using buffer regions, e.g., vehicles with lengths from 19 ft to 25 ft are considered "class 1 or class 2" vehicles and treated accordingly. Since these vehicles are definitely at the extreme end of their class, they might be treated differently than vehicles closer to the center of the class (e.g., borrowing ideas from fuzzy logic, instead of counting a 24 ft vehicle as 100% class 2, it might be counted as 0.8 class 2 and 0.2 class 1). Like the chronic detector errors, this discovery was an unanticipated byproduct of the detailed ground truthing and analysis. Determining the optimal correction was beyond the scope of the present work, but should be examined in future research.

REFERENCES

Benekohal, R., Girianna, M. (2003) "Technologies for Truck Classification and Methodologies for Estimating Truck Vehicle Miles Traveled," *Transportation Research Record 1855*, pp. 1–13.

Caltrans (2007), VideoSync, http://www.dot.ca.gov/research/operations/videosync/index.htm, accessed on June 5, 2007.

Cheevarunothai, P., Wang, Y., Nihan, N. (2007) "Using Dual-Loop Event Data to Enhance Truck Data Accuracy", *Proc. of the Transportation Research Board Annual Meeting*.

Chen, C., Kwon, J., Rice, J., Skabardonis, A., and Varaiya, P. (2003). "Detecting Errors and Imputing Missing Data for Single Loop Surveillance Systems," *Transportation Research Record 1855*, National Research Council, Transportation Research Board, Washington, D.C., pp. 160-167

Chen, L., and May, A. (1987). "Traffic Detector Errors and Diagnostics," *Transportation Research Record 1132*, TRB, Washington, DC, pp 82-93.

Cheung, S., Coleri, S., Dundar, B., Ganesh, S., Tan, C., Varaiya, P., (2005) "Traffic Measurement and Vehicle Classification with Single Magnetic Sensor," *Transportation Research Record 1917*, pp. 173–181.

Cleghorn, D., Hall, F., and Garbuio, D. (1991). "Improved Data Screening Techniques for Freeway Traffic Management Systems," *Transportation Research Record 1320*, TRB, Washington, DC, pp 17-31.

Coifman, B. (1998). "Vehicle Reidentification and Travel Time Measurement in Real-Time on Freeways Using the Existing Loop Detector Infrastructure", *Transportation Research Record 1643*, Transportation Research Board, pp 181-191.

Coifman, B. (1999) "Using Dual Loop Speed Traps to Identify Detector Errors," *Transportation Research Record no. 1683*, Transportation Research Board, pp 47-58.

Coifman, B., (2001) "Improved Velocity Estimation Using Single Loop Detectors", *Transportation Research: Part A*, vol 35, no 10, pp. 863-880.

Coifman, B. (2003). "Identifying the Onset of Congestion Rapidly with Existing Traffic Detectors", *Transportation Research: Part A*, vol 37, no 3, pp. 277-291.

Coifman, B., Dhoorjaty, S., and Lee, Z., (2003) "Estimating Median Velocity Instead of Mean Velocity at Single Loop Detectors", *Transportation Research: Part C*, vol 11, no 3-4, pp 211-222.

Coifman, B., Dhoorjaty, S., (2004) "Event Data Based Traffic Detector Validation Tests", *ASCE Journal of Transportation Engineering*, Vol 130, No 3, pp 313-321.

Coifman, B. (2006a), *The Columbus Metropolitan Freeway Management System (CMFMS) Effectiveness Study: Part 2 - The After Study*, Ohio Department of Transportation.

Coifman, B. (2006b), "Vehicle Level Evaluation of Loop Detectors and the Remote Traffic Microwave Sensor," *ASCE Journal of Transportation Engineering*. Volume 132, No 3, pp 213-226.

Coifman, B., Lee, H. (2006), "A Single Loop Detector Diagnostic: Mode On-Time Test," *Proc. of Applications of Advanced Technology in Transportation*, ASCE, August 13-16, 2006, Chicago, IL. pp 623-628

Coifman, B., (2007) *Vehicle Classification from Single Loop Detectors*, Project 05-02, Midwest Regional University Transportation Center, University of Wisconsin, Madison.

Coifman, B., Kim, S., (2009) "Speed Estimation and Length Based Vehicle Classification from Freeway Single Loop Detectors" *Transportation Research Part-C*, Vol 17, No 4, pp 349-364.

Dailey, D. (1999) "A Statistical Algorithm for Estimating Speed from Single Loop Volume and Occupancy Measurements." *Transportation Research-B*, Vol 33B, No 5, June 1999, pp 313-322. Elsevier Science, London.

EIS, (2007) excerpts from: French Engineering, *Traffic Data Collection Methodologies Final Report*, prepared for Pennsylvania Department of Transportation, Bureau of Planning and Research, April 2006, (accessed on August 1, 2007) http://www.rtms-by-eis.com/pdf/Methodologies_PnDOT06EISresponse.pdf

Gajda, J., Sroka, R., Stencel, M., Wajda, A., Zeglen, T. (2001) "A Vehicle Classification Based on Inductive Loop Detectors" *Proc. of IEEE Instrumentation and Measurement Technology Conference*, Budapest, Hungary, May 21-23.

Jacobson, L, Nihan, N., and Bender, J. (1990). "Detecting Erroneous Loop Detector Data in a Freeway Traffic Management System," *Transportation Research Record 1287*, TRB, Washington, DC, pp 151-166.

Ki, Y., Baik, D. (2005) "Vehicle Classification Model for Loop Detectors Using Neural Networks" *Transportation Research Record 1917*, pp. 164–172.

Ki, Y., Baik, D. (2006) "Vehicle-Classification Algorithm for Single-Loop Detectors Using Neural Networks" *IEEE Transactions on Vehicular Technology*, Vol. 55, No. 6, pp. 1704-1711.

Klein, L.A., Mills, M.K., Gibson, D.R.P, (2006) *Traffic Detector Handbook, Third Edition*, FHWA, U.S., Department of Transportation.

Kwon, J., Varaiya, P., Skabardonis, A. (2003) "Estimation of Truck Traffic Volume from Single Loop Detectors with Lane-to-Lane Speed Correlation," *Transportation Research Record 1856*, pp. 106–117.

Mikhalkin, B., Payne, H., Isaksen, L. (1972) "Estimation of Speed from Presence Detectors." *Highway Research Record 388*, pp 73-83. Highway Research Board, Washington, DC.

Neelisetty, S., Coifman, B., (2004) "Improved Single Loop Velocity Estimation in the Presence of Heavy Truck Traffic" *Proc. of the 83rd Annual Meeting of the Transportation Research Board*.

Nihan, N. (1997). "Aid to Determining Freeway Metering Rates and Detecting Loop Errors," *Journal of Transportation Engineering*, Vol 123, No 6, ASCE, pp 454-458.

Oh, S. Ritchie, S., Oh, C. (2002) "Real-Time Traffic Measurement from Single Loop Inductive Signatures," *Transportation Research Record 1804*, pp 98-106

Pushkar, A., Hall, F., Acha-Daza, J. (1994) "Estimation of Speeds from Single-Loop Freeway Flow and Occupancy Data Using Cusp Catastrophe Theory Model." *Transportation Research Record 1457*, pp 149-157. Transportation Research Board, Washington, DC.

Reijmers, J., (1979) "On-Line Vehicle Classification," *Proc. of the International Symposium on Traffic Control Systems*, Vol 2B, Institute of Transportation Studies, University of California at Berkeley, pp 87-102.

Sun, C, Ritchie, S. (2000) "Heuristic Vehicle Classification Using Inductive Signatures on Freeways", *Transportation Research Record 1717*, pp130-136

Turochy, R.E., and Smith, B.L., (2000) "New Procedure for Detector Data Screening in Traffic Management Systems," *Transportation Research Record 1727*, TRB, National Research Council, Washington, D.C., pp. 127-131.

Wang, Y., Nihan, N. (2000) "Freeway Traffic Speed Estimation With Single Loop Outputs," *Transportation Research Record 1727*, pp120-126, Transportation Research Board, Washington, DC.

Wang, Y., Nihan, N. (2003) "Can Single-Loop Detectors Do the Work of Dual-Loop Detectors?" *Journal of Transportation Engineering*, Vol. 129, No. 2. ASCE, pp 169-176.

Wang, Y., Nihan, N. (2004) "Dynamic Estimation of Freeway Large-Truck Volumes Based on Single-Loop Measurements", *Journal of Intelligent Tansportation Systems*, Vol 8, No 3, pp 133-141.

Zhang, G., Wang, Y., Wei, H. (2006) "Artificial Neural Network Method for Length-Based Vehicle Classification Using Single-Loop Outputs," *Transportation Research Record 1945*, pp. 100–108.

Zhang, L., Rilett, L., Jones, E., Naik, B., Appiah, J. (2007) "An Analysis of Aggregated and Disaggregated Traffic Data Obtained from Non-Intrusive Traffic Detection Sensors," *Proc. of the Transportation Research Board Annual Meeting*.

Zhang, X., Nihan, N., Wang, Y. (2005) "Improved Dual-Loop Detection System for Collecting Real-Time Truck Data," *Transportation Research Record 1917*, pp. 108–115.

Zwahlen, H. T., Russ, A., Oner, E. and M. Parthasarathy. (2005) "Evaluation of Microwave Radar Trailers for Non-intrusive Traffic Measurements." *Transportation Research Record 1917*, pp127-140.

APPENDIX A

Roadway usage, particularly by large vehicles, is one of the fundamental factors determining the lifespan of highway infrastructure. Benekohal and Girianna (2003) note that "[i]t is necessary to encourage state DOTs to include classification counts in their annual traffic-monitoring program" While "some states lacked the necessary resources to adequately sample ADTs on the local road systems."

Meanwhile, single-loop detectors are the most common vehicle detector in use to monitor traffic, both for real-time operations and for collecting census data such as Annual Average Daily Travel (AADT). New, out-of-pavement detectors seek to replace loop detectors using non-invasive, wayside mounted sensors, but most of these detectors emulate the operation of single-loop detectors. In either case, collecting reliable length data from these detectors has been considered impossible due to the noisy speed estimates provided by conventional data aggregation at single-loop detectors (and in the case of non-invasive sensors, the noisy on-time measurements as well, Coifman, 2006b).

This research continue work originally funded by the Midwest Regional University Transportation Center (MRUTC, the predecessor to NEXTRANS) and which had very promising results to date (Coifman, 2007). Within this research, we refined unconventional techniques for estimating speed at a single-loop detector, yielding estimates that approach the accuracy of a dual-loop detector's measurements. Employing these speed estimation advances, this research enabled length based vehicle classification to single-loop detectors. The research promises to extend vehicle classification to single-loop detector count stations and the many single-loop detector stations already deployed for real-time traffic management. The work also offers a viable treatment in the event that one of the loops in a dual-loop detector classification station fails.

In short, this research seeks to mainstream advances in speed and length estimation from single-loop detectors and develop a vehicle classification methodology for these detectors. As noted in a draft research statement from the TRB Committee on Highway Traffic Monitoring, "Classification based solely on vehicle length is an alternative to axle-based classification but there has been no systematic study of how well it works -- or how it should work."

There has been considerable research on vehicle classification leading to the conventional technologies as well as on-going work in emerging technologies. Needless to say, the body of work is broad. For length-based classification from loop detectors, there are three interrelated parameters that can be measured or estimated for each passing vehicle, namely length (l), speed (v) and the amount of time the detector is "on", i.e., the *on-time* (on). These parameters are related by Equation A-1,

$$l = v \cdot on \tag{A-1}$$

The distinction between different detection technologies is important. Conventional dual-loop detectors can measure both on-time and speed directly, and so they are often employed to classify vehicles based on length (via Equation A-1). Conventional single-loop detectors can only measure on-time. In the absence of accurate speed estimation from single-loops, these detectors have not been used to estimate vehicle length or classify vehicles. As already noted, single-loop detectors are the most common vehicle detector in use to monitor traffic, and this research seeks to extend vehicle length estimation and vehicle classification to these detectors.

As a precursor, many researchers have sought better estimates of speed from single-loop detectors. The preceding research has emphasized techniques that use many samples of aggregate flow (q) and occupancy (occ) to reduce the estimation error, e.g., Mikhalkin et al (1972), Pushkar et al (1994), Dailey (1999), Wang and Nihan (2000), Coifman (2001). Although rarely noted, these techniques effectively seek to reduce the bias due to long vehicles in measured occupancy. Rather than manipulating aggregate

data, we developed new aggregation methods to reduce the estimation errors (Coifman et al, 2003; Neelisetty and Coifman, 2004; Coifman, 2007).

A thorough review of the literature on loop-detector based vehicle classification will reveal the following four main thrusts: inductive signature based classification; data cleaning at dual-loop detectors; estimates from conventional single-loop detectors; and classification from non-invasive detectors. The following reviews each one of these thrusts in turn. When reviewing the literature, it is stunning to find that all of the classification studies exhibit either a very small validation data set (under 1,000 vehicles, often under 100) or they were only compared against the results generated from dual loop detectors without any manual validation (and thus, any errors present in the dual-loop detectors would go unaccounted for). Presumably the problems of a small data set are obvious, but trusting that the dual-loop results are accurate can be equally problematic (e.g., we found loop detectors were "dropping out" in the middle of semi-trailer trucks, a problem that impacted both dual and single-loop detector classifications alike). In either case, the studies have all been limited to a small number of detector stations, often employing just a single station, but never more than 10 detector stations. Such a small sample size precludes capturing the impacts of different vehicle fleet compositions (long versus short vehicles), traffic conditions (free flow versus congested), and variable detection hardware performance. Except for our research, few have explicitly sought out the most challenging conditions: congestion, and high percentages of long vehicles.

A.1 Inductive Signature Based Classification

Conventional loop detectors only report a binary state- occupied or empty. In the process of making this determination, the detectors measure the loop's inductance several hundred times a second. These inductive measurements can be captured and integrated to form an "inductive signature" for each passing vehicle. Inductive signature based classification seeks to identify characteristic features of these signatures to classify vehicles. Inductive signature based classification has been a subject of research for 30 years, but like automated highways, while this area has shown promise it has not entered

mainstream practice. It requires new detector hardware in the controller cabinets and is only partially compatible with the existing infrastructure. Several papers present effectively a proof of concept employing a very small validation data set of fewer than 100 vehicles, e.g., Reijmers (1979), Gajda et al (2001), Cheung et al (2005). Only slightly more ambitious, Sun and Ritchie (2000) compared performance against a manually validated data set of 300 vehicles from two detector stations. Oh et al (2002) continued the work at a new detector station and used a validation set of 340 vehicles. Separately, Ki and Baik (2005, 2006) developed a similar classification tool and validated it against a set of 622 vehicles, which were apparently manually validated. For these three latter studies, it appears that all of the data come from uncongested conditions, with long vehicles making up less than 10% of the flow.

A.2 Data Cleaning at Dual-Loop Detectors

Conventional loop detector stations measure individual vehicle actuations and then aggregate these data to flow, occupancy and average speed over fixed time periods, typically ranging from 20 sec to 5 min. The individual actuations are then typically discarded. Several researchers have developed statistical tests to evaluate whether the time series aggregate data are within statistical tolerance (e.g., Jacobson et al, 1990, Cleghorn et al, 1991, Nihan, 1997). Because these automated systems only use aggregated data, they must accept a large sample variance and potentially miss problems altogether. For example, the systems have to tolerate a variable percentage of long vehicles in the sample population. As the percentage of long vehicles increases, the occupancy/flow ratio should increase simply because a long vehicle occupies the detector for more time compared to a shorter vehicle traveling at the same velocity, see Coifman (2001) for examples.

Chen and May (1987) developed a new approach for verifying detector data using event data, i.e., individual vehicle actuations. Their methodology examines the distribution of vehicles' on-time, i.e., the time the detector is occupied by a vehicle. Unlike conventional aggregate measures, their approach is sensitive to errors such as

"pulse breakups", where a single vehicle registers multiple actuations because the sensor output flickers off and back on, i.e., dropping out. Coifman (1999) went a step further and compared the measured on-times from each loop in a dual loop detector on a vehicle-by-vehicle basis. At free flow velocities the on-times from the two loops should be virtually identical regardless of vehicle length, even allowing for hard decelerations. Many hardware and software errors will cause the two on-times to differ. At lower velocities, vehicle acceleration can cause the two on-times to differ even though both loops are functioning properly and thus, congested periods were excluded from the earlier analysis. Coifman and Dhoorjaty (2004) developed a suite of event data based tests to catch several detector errors based on physical constraints (feasible vehicle length, feasible headways, etc.). Zhang et al (2005) and Cheevarunothai et al (2007) continued the research, setting out the specific objective of, "identifying possible causes of dual-loop errors and developing a new dual-loop algorithm that could tolerate erroneous loop actuation signals."

A.3 Estimates from Conventional Single-Loop Detectors

While dual-loop detector length based vehicle classification is well within conventional practice, the fact remains that most loop detector stations are only equipped with single-loop detectors and cannot measure vehicle lengths. There are only three research groups seeking to improve speed and length estimates from conventional single-loop detectors to the point where they can be used to classify vehicles.

The first group, Kwon et al (2003), developed a method employing aggregate flow and occupancy from single-loop detectors to estimate the percentage of long vehicles that passed. The work depends on two fundamental assumptions, the presence of a truck-free lane and that the detector station exhibits high lane-to-lane speed correlation. They employed conventional detectors, used many days, using several stations from three facilities. The work only validated the results against aggregate dual loop measurements and weigh-in-motion (WIM) data. The former yielded good results, while the latter had 20% overestimation, highlighting the importance of employing a truly independent

measure of ground truth. The research studied facilities with low to moderate truck volumes (under 10% of the fleet) and did not explicitly single out performance in congested conditions. In fact they note that, "[i]t was observed that the estimate of truck volume is biased and unstable at the start of the congestion period."

The second group, Wang and Nihan (2003, 2004) also developed a method employing aggregate flow and occupancy from single-loop detectors to estimate the percentage of long vehicles that passed. Like Kwon et al, their work also depends on two fundamental assumptions, though slightly different, "constant average speed for each [three minute long] time period and at least two intervals containing only [short vehicles] in each period." They employed conventional detectors, used many days, using four detector stations. Also like Kwon et al, the work only validated the results against aggregate dual loop measurements. The research studied facilities with low to moderate truck volumes (under 10% of the fleet) and did not explicitly single out performance in congested conditions. These authors note, "[h]ence, the algorithm should work better under less congested conditions." The authors also explicitly note the limitation of the small number of test sites, "[a]lthough the algorithm performed reasonably well at the selected sites and days, future research is needed to handle the conditions when one or both of the assumptions are violated in order to reduce estimation errors.... The proposed algorithm will be more robust and accurate when the violation circumstances are properly addressed." They also state that "[f]uture research will specifically address these problems and widely check the transferability of the algorithm to other sites." More recently, this group has revised their methodology (Zhang et al, 2006). This recent study is subject to many of the same limitations as their earlier work, it employs aggregate flow and occupancy, was tested at only two detector stations (with approximately 10% truck flows), and only compared the results against aggregate dual-loop detector measurements. The final conclusion of this paper states that, "[a]lthough the proposed ANN method produced favorable bin volumes, further improvements to its performance are possible through optimizing its network design and training, especially under heavily congested conditions. Additionally, more accuracy tests with data from different types of

roads and different areas also help in understanding the spatial transferability of the proposed method." Clearly indicating both the fact that they did not explicitly examine performance under congested conditions and that the results are limited to the small number of test sites.

The third group is our group. Our research was funded by the predecessor to NEXTRANS (Coifman, 2007, Coifman and Kim, 2009). Unlike the other two groups working in this area, we did not employ aggregate data, instead, we used the individual vehicle actuations and explicitly classified each and every vehicle. This point is important, because the earlier efforts that relied on aggregate measurements from dual-loop detectors allow over-counting errors to cancel undercounting errors, so the reported results in the earlier studies may be overly optimistic. While the earlier studies compiled results against dual-loop detectors for several days, our study compiled results for 13 dual-loop detector stations over an entire month (again, we used individual vehicle comparisons while the earlier studies used aggregate data). Unlike the earlier studies, we considered truck volumes over 10% of the fleet. In fact we explicitly generated synthetic detector data to simulate truck volumes up to 90%. Furthermore, we did not rely strictly on dual-loop detectors for validation, we manually generated ground truth vehicle length data from concurrent video for approximately 25,000 vehicles While this number of vehicles is large, it still only represents a total of 6 hours of traffic, sampled at two locations, and we felt a much broader ground truth is needed to ensure the classification process is robust, especially since we discovered that detectors were "dropping out" in the middle of semi-trailer trucks, a problem that impacted both dual and single-loop detector classifications alike (see Coifman, 2007 for details). Even with a much larger data set than the second group, we came to a similar conclusion, namely that additional locations need to be tested. Hence the present research.

A.4 Classification from Non-Invasive Detectors

As noted earlier, most of the non-invasive vehicle detectors that have entered conventional practice mimic the operation of single loop detectors. The two most

prevalent examples of these detectors being the SmartSensor by Wavetronix and RTMS by EIS. Both sensors can provide length based classification data, though the specific algorithms are proprietary. While the sensors often provide reasonable counts and speed estimates in aggregate data, per-vehicle analysis has shown that the aggregate data allow over-counting errors to cancel under-counting errors and that individual vehicle on-times can be subject to large errors (see, e.g., Zwahlen et al, 2005; Coifman, 2006b). The literature is surprisingly lacking in terms of evaluating the classification performance from these sensors.

Zwahlen et al, (2005) evaluated the Wavetronix sensor in uncongested, low volume traffic, with low truck flows. While these conditions should lead to favorable performance by the sensor, after comparing the classification results against manually generated ground truth data the authors concluded that, "vehicle classification is unreliable; the fraction of trucks in a lane can be severely overestimated or underestimated." Trucks were undercounted by as much as 80% in the worst case and "at this time, the system does not reliably estimate the number of trucks in the traffic stream."

Zhang et al (2007) compared the Wavetronix sensor against the Autoscope (an image processing based system) with no independent ground truth. As a result, they conclude that "[a]n extensive calibration effort with the support of ground truth data would be required before any definitive statements can be made related to the accuracy of speed and vehicle classification capabilities. More investigation on this topic is suggested for future studies."

Finally, the EIS web site provides excerpts of a research study conducted for the Pennsylvania Department of Transportation (EIS, 2007). This two-page summary includes a table reporting classification performance from the RTMS and several other non-invasive sensors tested at three sites, all with truck volumes of 10% or less. It provided no indication of the traffic conditions. While the apparent objective of the document is to show that the RTMS by EIS performs better than a "competitive radar", the statistics included in this promotional flyer clearly show that at all three sites RTMS

counts for single unit trucks and tractor-trailer trucks were consistently 50% below manual counts (the "competitive radar" had counts that were frequently 2-5 times larger than the manual counts).

So while the manufacturers offer vehicle classification from these non-invasive sensors, the specific algorithms are undocumented and to the extent that they have been evaluated in the literature, the performance is poor.

APPENDIX B

Speed trends underlying the ground truth data in congestion

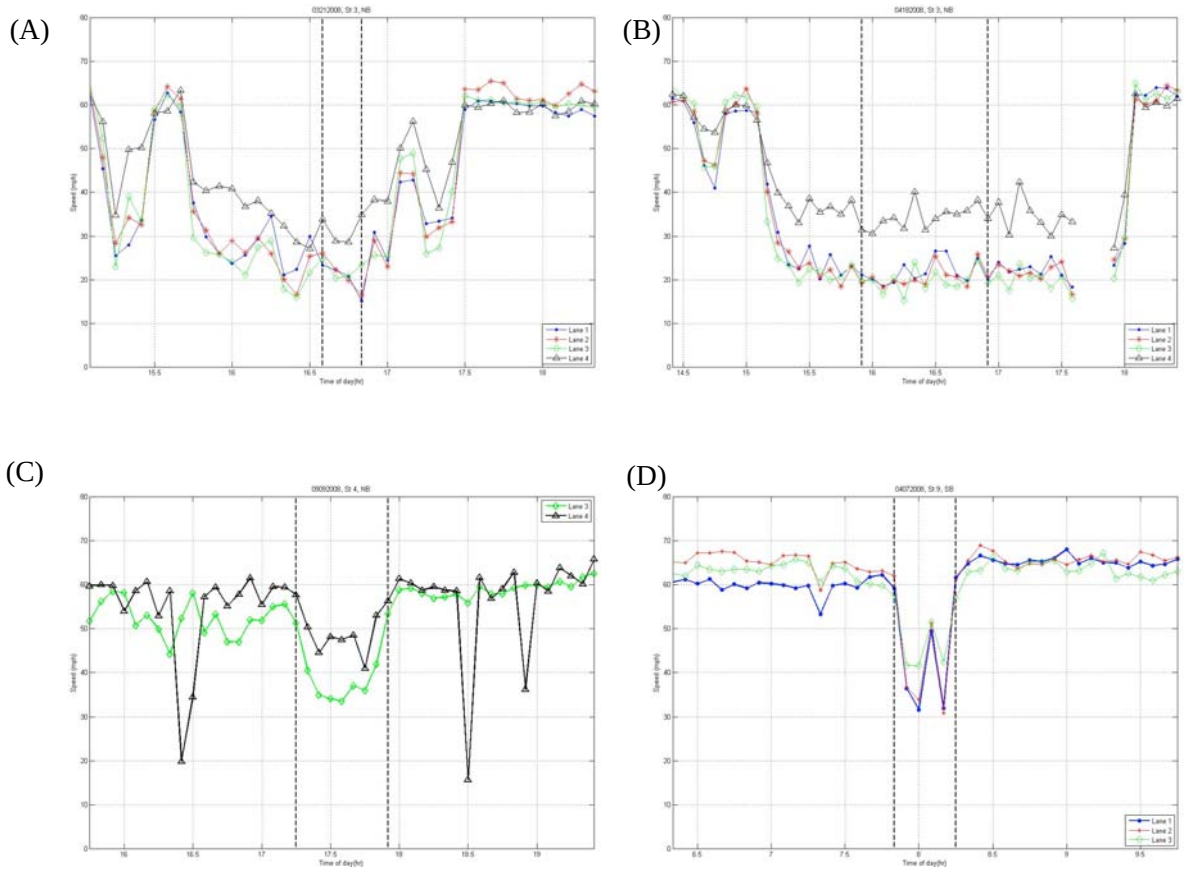
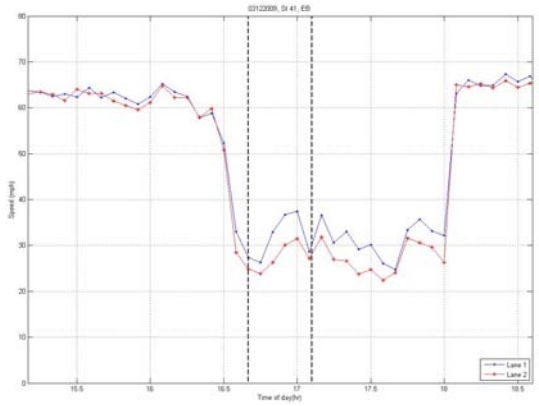
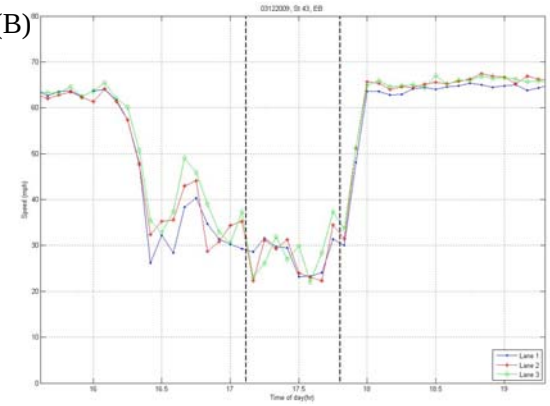


Figure B.1. Speed trend at detector stations that are selected from the ground truth data with pulse breakup, (a) St 3 NB 3/21/08, (b) St 3 NB 4/18/08, (c) St 4 NB 9/09/08, and (d) St 9 SB 4/07/08

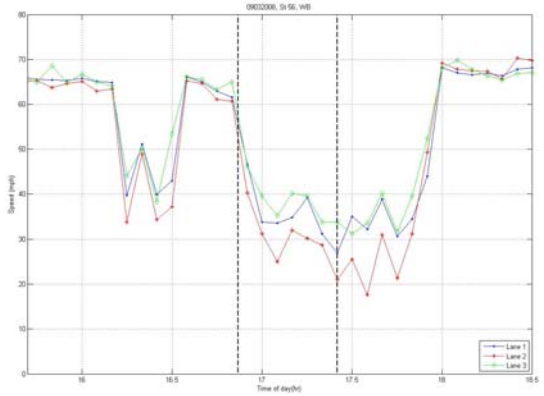
(A)



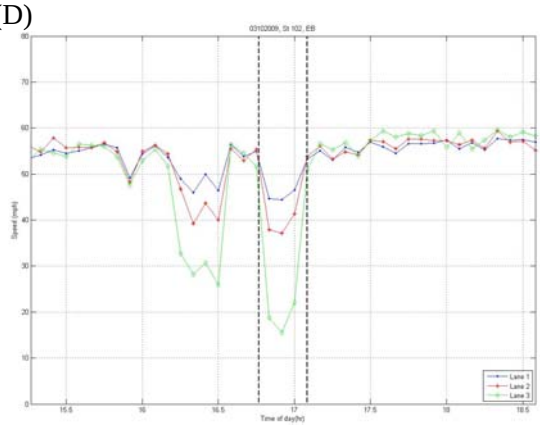
(B)



(C)



(D)



(E)

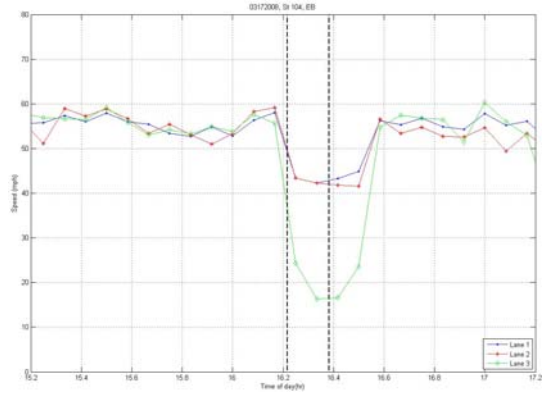


Figure B.2, Speed trend at detector stations that are selected from the ground truth data without pulse breakup during congested condition: (a) St 41 EB 3/12/09, (b) St 43 EB 3/12/09, (c) St 56 WB 9/03/08, (d) St 102 EB 3/10/09, and (e) St 104 EB 3/17/08