Accident Prevention Based on Automatic Detection of Accident Prone Traffic Conditions: Phase I

Final Report

Prepared by:

John Hourdos
Vishnu Garg
Panos Michalopoulos

Department of Civil Engineering
University of Minnesota

CTS 08-12

Technical Report Documentation Page

1. Report No.	2.	3. Recipients Accession No.	
CTS 08-12			
4. Title and Subtitle Accident Prevention Based on Automatic Detection of Accident Prone Traffic Conditions: Phase I		5. Report Date September 2008 6.	
7. Author(s) John Hourdos, Vishnu Garg, Panos Michalopoulos		8. Performing Organization Report No.	
9. Performing Organization Name and Address Department of Civil Engineering University of Minnesota 500 Pillsbury Drive S.E. Minneapolis, MN 55455		10. Project/Task/Work Unit No. CTS Project # 2003031 11. Contract (C) or Grant (G) No.	
12. Sponsoring Organization Name and Address Intelligent Transportation Systems Institute University of Minnesota 200 Transportation and Safety Building 511 Washington Ave. SE Minneapolis, Minnesota 55455		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code	
15. Supplementary Notes http://www.cts.umn.edu/Publications/ResearchReports/			

16. Abstract (Limit: 200 words)

Growing concern over traffic safety as well as rising congestion costs have been recently redirecting research effort from the traditional crash detection and clearance reactive traffic management towards online, proactive crash prevention solutions. In this project such a solution, specifically for high crash areas, is explored by identifying the most relevant real time traffic metrics and incorporating them in a crash likelihood estimation model. Unlike earlier attempts, this one is based on a unique detection and surveillance infrastructure deployed on the freeway section experiencing the highest crash rate in the state of Minnesota. This state-of-the-art infrastructure allowed video recording of 110 live crashes, crash related traffic events, as well as contributing factors while simultaneously measuring traffic variables such as individual vehicle speeds and headways over each lane in several places inside the study area. This crash rich database was combined with visual observations and analyzed extensively to identify the most relevant real-time traffic measurements for detecting crash prone conditions and develop an online crash prone conditions model. This model successfully established a relationship between fast evolving real time traffic conditions and the likelihood of a crash. Testing was performed in real time during 10 days not previously used in the model development, under varying weather and traffic conditions.

17. Document Analysis/Descriptors	18. Availability Statement		
Traffic safety, rear-end collisions, crash prevention, freeway traffic, driver warning systems	No restrictions. Document available from: National Technical Information Services, Springfield, Virginia 22161		
19. Security Class (this report) Unclassified	20. Security Class (this page) Unclassified	21. No. of Pages 169	22. Price

Accident Prevention Based on Automatic Detection of Accident Prone Traffic Conditions: Phase I

Final Report

Prepared by:

John Hourdos
Vishnu Garg
Panos Michalopoulos

Department of Civil Engineering
University of Minnesota

September 2008

Published by:

Intelligent Transportation Systems Institute
Center for Transportation Studies
University of Minnesota
200 Transportation and Safety Building
511 Washington Ave SE
Minneapolis, MN 55455

Table of Contents

List of Tables

List of Figures

Executive Summary

Driving may be the most dangerous activity with which we are involved. According to the most recent USDOT records, at the end of 2003, vehicle-related crashes in the U.S. exceeded 6 million with 3 million injuries and 42,643 deaths [NHTSA, 2004]. Based on a 2000 study of crash costs performed by the National Highway Traffic Safety Administration (NHTSA) [Blincoe et al., 2002] and extrapolating for 2003, the total cost from the aforementioned crashes was in excess of $240 billion. While the majority of crashes, especially fatal ones, concentrate in rural areas, for example, in Minnesota in 2002 only 23,676 crashes occurred in urban freeways and trunk highways of which 83 fatal, urban freeway crashes affect a much larger percentage of the population.

According to the Minnesota Motor Vehicle Crash Facts report of 2002 [Mn/DOT, 2003], on urban freeways and trunk highways "property damage only" crashes lead arithmetically with a total of 16,992 as compared to 83 "fatal" and 6,601 "personal injury" crashes. Although no specific data are available for freeways only, assuming state wide economic figures, the average cost of a fatal crash is $1,040,000 (the FHWA value for life is greater than $3M but Mn/DOT selected a lower value in order to avoid fatal crashes overshadowing non-fatal one), the average cost of an injury is $16,500 and the average property only damage cost is $6,500. Based on these numbers, the property damage only crashes on Minnesota freeways in the year 2002 reached a total of $110,448,000 while the fatal ones account for $86,320,000 and injuries for $108,916,500. Interestingly, the cost of highway crashes is three times the annual Mn/DOT budget for road maintenance and operation.

Traditional measures to reduce crashes include improved geometric design, congestion management strategies, as well as better driver education and enforcement. While such measures are generally effective, they are often not feasible or prohibitively expensive to implement. Prior to the widespread use of cell phones, considerable research effort was spent in developing automated incident detection systems. To be sure, fast detection and clearance is currently the most important function in any traffic management center because it reduces congestion and helps prevent secondary crashes. In spite of this, incident detection is still a reactive measure.

A 1997 USDOT/NHTSA campaign [Martinez, 1997] encouraged the removal of the word "accident" from the traffic management vocabulary and suggested replacing it with the word "crash." Some of the reasons stated were:

- "Motor vehicle crashes and injuries are predictable, preventable events. Continued use of the word "accident" promotes the concept that these events are outside of human influence or control. In fact, they are predictable results of specific actions";
- "We can identify their causes and take action to avoid them. These are not 'acts of God,' but predictable results of the laws of physics".

This realization along with the increasing need to reduce crashes and their side effects has recently prompted research for proactive approaches to avoid crash occurrence. One of the most promising options gaining wide acceptance in recent years is the concept of detecting crash-

prone traffic conditions in real-time and warning drivers when the likelihood of a crash is high to increase their attentiveness, thereby reducing the number of crashes.

The majority of the most recent studies explore the estimation of crash probability in a large area i.e., an entire freeway or freeway network. Although this does not reduce the benefit from the associations presented between traffic conditions and crash likelihood, it does little toward accounting for the existence of high-crash areas. It is a known fact that, in the majority of freeways, certain locations exhibit considerably higher crash frequencies than the overall average. For example, in the freeway network of the Twin Cities metro region in Minnesota, approximately 35% of the crashes concentrate in 10 areas, each no longer than one mile (approximately 7% of freeway network with 10.93% of total vehicle miles traveled). To be sure, the highest crash area is a 0.46 mile (.74 km) long section of I-94 westbound in the south of downtown Minneapolis. This area in 2002, exhibited 4.81 crashes per million vehicle miles (MVM) while the network average is 0.96 crashes/MVM. During the PM peak period the aforementioned site exhibits an average of 15.43 crashes/MVM while in comparison the entire I-94 freeway experiences only 3.29 crashes/MVM [Mn/DOT, 2002]. From these statistics it is clear that this and similar high-crash freeway sections require special treatment not only for estimating crash likelihood but also for eventually identifying what are the specific crash causal factors. The former serves as the basis for developing effective real-time driver warning and/or traffic calming systems while, the latter, not only assists in their targeted deployment but can also guide more traditional traffic management interventions.

The research presented in this document attempts to establish reliable associations between real-time traffic measurements and crashes. These associations should help demonstrate that crash-prone traffic conditions exist, are detectable, and possibly manageable for avoiding crashes.

RESEARCH OBJECTIVES AND GENERAL APPROACH

Crashes cause million dollar losses, take lives, and drastically reduce the productivity of the transportation infrastructure. To find means to prevent them, the objectives of this research are:

1. Determine if crash-prone conditions (CPCs) are detectable by real-time traffic measurements which can be used reliably to warn drivers in advance and prevent crashes.
2. Develop effective and efficient crash-prone condition detection algorithm(s).
3. Propose Possible Implementation Schemes.

To accomplish these objectives, the effort was divided into the following tasks:

1. Selection of a crash-prone location.
2. Instrumentation of the site to capture live crashes on video and simultaneously collect detailed traffic measurements as well as weather and other environmental information.
3. Identification of potential measurements and metrics able to manifest distinct patterns during pre-crash periods.
4. Employment of measurements and metrics in a crash likelihood model.
5. Based on the resulting model(s), development of a CPC detection algorithm.
6. Testing the above algorithm under varying traffic and weather conditions.

Test Site Description

According to Mn/DOT records [Mn/DOT, 2002] the site with the highest crash rate in the state is a 1.7-mile-long (2.73km) section of I-94 westbound in Downtown Minneapolis between 11th Ave and the Lowry Hill tunnel. This site has a rate of 3.81 Crashes/MVM (million vehicle miles). This rate roughly translates to an average of, one crash every two days. I-94 is a connector type freeway joining the cities of St. Paul and Minneapolis. This freeway carries an average daily traffic in excess of 80,000 vehicles per direction and it is congested for at least five hours daily, especially during the afternoon peak period. The crash prone section runs parallel to I-35W (another major freeway) and a number of short ramps allow transfer from one freeway to the other. The site includes two entrance and three exit ramps; the average number of lanes is three with two 3000-foot (914 meter) auxiliary lanes in two weaving areas. Excessive weaving takes place in areas of the site due to the high volumes entering from the rightmost ramp which in fact combines traffic from 35W, HW55, and the downtown business area.

For the purposes of this study, the aforementioned site presented the highest potential for capturing crashes on video as well as analyzing complex traffic flow dynamics. From an economic standpoint, a crash prevention solution here will have a high potential return since any incident on this section results in hundreds of vehicle-hours of delay due to the high volumes it carries. Regardless, for this site to be selected, it had to have additional features allowing easy and affordable deployment of instrumentation.

Data Collected

The four detection and surveillance stations overseeing the deployment site provide detailed data and large amounts of wide-area measurements. In conjunction with the video recordings, this unique database can facilitate advanced studies of traffic characteristics, flow dynamics, and safety and traffic management concepts. Specifically, the sensors collect and transmit in real time individual vehicle *speed, length*, and *time headway* as well as every 10 seconds aggregated measurements of *flow rate*, *volume count, time mean speed, space mean speed, space occupancy, density, LOS* and *vehicle class count*. The aggregated measurements are saved in the sensor for later download. The aforementioned data are collected on a 24/7 basis in six areas of the freeway site. Each of these areas is 300 to 500 feet (91 to 152 meter) long and measurements are extracted approximately every 100 feet (30.5 meter) for a total of 51 detection points (18 series times 3 lanes, plus exit and entrance ramps).

To date, the laboratory has been operational during three Minnesota winters and two stormy summer seasons, providing detailed measurements during extreme weather conditions. For purposes of data management, video recording takes place between 7:00 A.M. and 8:00 P.M. during weekdays and 12:00 P.M. until 9:00 P.M. on Saturdays and Sundays. Not all video is stored long term but selectively as needed. Apart from this, a collection of video from all 11 cameras at important times, days, and weather conditions is stored for future research needs. During the periods of the initial and final deployments a verification of the systems data-collection performance was undertaken by comparing the collected information with the video records. Calibration was carried out by comparing speed measurements taken manually with a laser gun with those automatically collected.

Qualitative Analysis

Analysis of the large amount of traffic information collected during this study was addressed in two ways: qualitatively and quantitatively. The qualitative analysis discussed in the report deals primarily with the information extracted from the visual observation of the video records of crashes and the general periods of crash and non-crash traffic conditions.

Collisions

Regardless of severity, the event in which two or more vehicles collide with each other is considered a crash. Collisions, in this site, can be further divided into rear-ends or sideswipes, based on the type of the collision. All video recordings were scanned to identify collisions; therefore, a number of collisions that were not reported by the police were also included.

Near Misses

In an effort to avoid a collision, a driver may intentionally or unintentionally steer the vehicle off the road. In such situations drivers perceive that their vehicle is likely to hit the leading vehicle and just slowing down will not be sufficient to avoid collision. In, these cases, if not for the driver's evasive maneuvers these events would have resulted in rear-end collisions. Near misses were defined as sudden braking and rapid steering of a vehicle resulting in **complete departure** from the original lane of travel and entry onto the shoulder.

Traditionally, safety studies are based only on actual collisions and very little or no attention is given to near-misses. The main goal of this study is to identify crash-prone conditions i.e., traffic conditions that favor the occurrence of crashes (increase crash probability). Taking this further, since near-misses are the result of individual driver's last minute actions, it is further hypothesized they also occur disproportionally during, or are the result of, crash-prone conditions. To test this assumption, near-misses were treated as crashes.

CRASH SCENARIO

Based on the analysis of the video records as well as several test runs (An instrumented vehicle was driven several times collecting video records from the point of view of the driver, the side, and rear view mirrors) in the study section, a hypothetical scenario of the crash sequence can be formulated. Specifically:

1. Due to weaving between the 3rd Ave/I-94 flyover entrance and the Hennepin exit a compression wave or queue propagates backwards.
2. Upstream of Portland, at the joint entrance ramp from I-35W, TH-55, and downtown, heavy weaving produces friction, distraction, and the desire of the drivers to move away from the right lane to the considerably faster moving middle lane.
3. Around the Portland overpass the roadway declines and turns, limiting visibility.

RESULT: Distracted and distressed drivers passing the 11th Street weave fail to notice the sudden stopped traffic caused by the upcoming wave or queue and collide.

Detection of Crash-Prone Traffic Conditions

We capitalize on the crash likelihood model methodology to use these additional traffic metrics in the development of an effective and efficient crash prone condition detection methodology. The process is outlined in the following steps:

1. Identify a set of crash events.
2. Identify a set of normal traffic conditions (controls).
3. Select traffic metrics and parameters describing environmental conditions.
4. Define time and space variations of the aforementioned metrics.
5. Define modeling procedure and prepare utilities for streamlining the process.
6. Identify a number of digital filters to preprocess the raw data.
7. Identify alternative combinations of filters and metric variants.
8. Generate the necessary cases and controls required for the statistical modeling.
9. Develop one logit based model of crash likelihood for each alternative.
10. Incorporate the models in alarm producing algorithms.
11. Identify a set of days, not previously used in the analysis, to be used for testing and evaluating each algorithm. Produce a performance chart for each algorithm.
12. Select the best one(s).

TRAFFIC METRICS AND METRIC VARIANTS

The majority of the metrics utilized in this study have already been presented in chapter three. To summarize, the following traffic metrics are considered:

1. Temporal metrics
   - Average Speed ($\overline{S}$)
   - Coefficient of Variation of Speed (S_{CV})
   - Traffic Pressure (P_T)
   - Kinetic Energy (E_k)
   - Coefficient of Variation of Time Headway (H_{CV})
2. Spatial Metrics
   - Acceleration Noise (A_N)
   - Mean Velocity Gradient (MVG)
   - Quality of Flow Index (Q_F)
3. Empirical Metrics
   - Maximum/Minimum Speed difference on a location
   - Upstream/Downstream speed difference.
   - Right lane/Middle lane speed difference.

RESULTS OF MODEL DEVELOPMENT

There are some interesting observations stemming from the results.

1. On all models the position of the sun is considered significant.
2. In regard to speed, one can observe that all the alternatives, except those preprocessed with a low-pass filter, mainly favor speed metric variants with very short time shifts. In contrast, the two low-pass filter alternatives, in general, favor considerably larger time shifts as well as window sizes. One can expect such a behavior since the low pass filters have removed all high frequency patterns and noise that otherwise reduce the resolution as the window gets bigger. The removal of the unwanted elements also accentuates the remaining patterns, allowing for stronger correlation farther from the actual crash time.
3. From the coefficient sign the following can be derived:

a. The higher the speed in the upstream station, the greater the crash likelihood. Likelihood increases as the speed in the downstream station decreases. This is an indication that the model is influenced by the presence of compression waves.
b. Crash likelihood increases as the speed in the middle lane increases. This is observed on both stations. Additionally, the likelihood increases as the speed difference between the right and middle lanes increases.
c. As time headways increase, the crash likelihood decreases.
d. Wet pavement, as well as reduced visibility due to rain or snow, marginally increases the crash likelihood. This contradicts some of the earlier findings by other researchers (see chapter 2 of report) stating that bad weather reduces crash probability. Further analysis to discover why this is the case in this location is warranted.

The model behavior as described by the speed level variables is indicative of the correlation between compression waves and crash occurrence. Additionally, prior to a breakdown, the stream operates under "synchronized flow" [Jiang et al., 2003]. During synchronized flow vehicles move very fast, too close to each other, which can explain the low variability in speed right before the crash. However, during the period before the initiation of synchronized flow the speed variability is higher. Synchronized flow does not last long since very small disturbances in flow due to vehicles merging in or sudden braking by a single vehicle result in the formation of compression waves that do not dissipate but travel quickly upstream.

As shown in earlier human factor studies [Marshal et al., 1998], during low light conditions drivers perceive the stop lights of leading vehicles faster even when distracted. However, when the sun shines on the windshield, it adds glare and may temporary blind drivers, making it more difficult to keep the proper distance.

Preliminary Implementation Framework

Although the objectives of this study were to present evidence that crash prone conditions exist and if so, develop a detection methodology based on real-time data, the work would not be complete if at least a preliminary implementation framework were not proposed. In the literature review chapter, a few simple, straightforward approaches were presented. Those were the only ones that passed the point of research and actually implemented something in the field. During the course of the study, visions and plans of a crash prevention system were often considered, reaching always the same realization; it is not a simple thing to design an effective and efficient crash prevention solution, especially one that involves interaction with the drivers.
Based on the analysis the specific factors that need to be influenced, changed, and/or controlled to succeed in preventing crashes are:

1. Flow breakdown
2. Driver inattention
3. Dangerous driving

Although these factors can be influenced in many ways, in this work, only the ones related to real-time actions driven by a crash prone condition detection system are discussed.

FLOW BREAKDOWN FACTOR

Based on the available evidence presented, the cause of the flow breakdown and the initiation of compression wave activity is the merging traffic from the I-94 flyover part of the 3rd Ave combination entrance ramp. There are three interventions one can explore in order to change the course of the breakdown.

1. Regulate traffic entering from the ramp in order to prevent the breakdown.
2. Increase vehicle headways in the right lane prior and at the merge point to allow for easier lane changes.
3. Regulate the traffic at the right lane after the occurrence of a breakdown to halt the backward propagation of the compression wave.

DRIVER INATTENTION FACTOR

The second crash causal factor deals with driver inattention and more specifically, with the circumstances in the road that attract the driver's focus away from the observation of the leading vehicle in its lane. The data suggest that, the causal traffic pattern is the speed differential between the right and middle lanes. The hypothesis explaining this correlation states that, as the speed difference increases, lane changing becomes harder due to the decrease in the number of acceptable gaps. The driver's attention/time devoted in the search for an appropriate gap is taken from the attention/time he/she must spend to maintain a safe following distance with the leading vehicle in its lane. There are four possible interventions that can affect the aforementioned pattern.

1. Regulate/reduce the flow of the vehicles crossing the right lane towards the middle and left lanes.
2. Reduce the speed difference between the right and middle lanes to assist lane changing.
3. Prohibit lane changes between the right and middle lanes while there is a compression wave traveling upstream or even when a flow breakdown is imminent.
4. Implement a driver warning system aimed at increasing driver attention to conditions in their own lane.

MINIMUM SYSTEM REQUIREMENTS

Two separate systems were utilized in this research. The first system allowed the collection of detailed information at several locations along the roadway as well as a large number of crash video records. This information was instrumental in understanding the dynamics of the problem. The second system is the one collecting the necessary information for the detection algorithm. From the original four detection and surveillance stations, only one is finally required for real-time operation.

During the crash causal factor analysis the majority of measurements were extracted from ordinary loop sensors. It was though the addition of individually collected speeds and headways that unraveled the mystery at the end. Information is vital to the investigation of the casual factors in a high-crash area, whether they on a freeway or at an intersection. The detection and surveillance systems developed in this research are able to operate remotely and with minimum power requirements. It is therefore possible to deploy several portable, temporary ones on a high-crash area, collect detailed information for a limited period of time and then, following an analysis similar to the one presented in this work, decide what the specific requirements for permanent instrumentation are. One fact that became clear during this research is that high-crash

areas are complicated, with individual characteristics and problems requiring custom design of both data collection and analysis methodologies. For sure the statistical nature of the crash likelihood model utilized in the detection algorithm is fitted for the specific combination of geometry, traffic conditions, and data collection characteristics. For the same model to operate in any other case a model calibration is required.

It is too early to speculate on detailed system requirements for a crash prevention system. The few alternatives mentioned in this work are examples of how current systems can be utilized in the case of this high-crash area. Industrious engineers, once enough detailed information is collected/derived about the particular problem, can devise new, custom solutions.

1 Introduction

Driving may be the most dangerous activity with which we are involved. According to the most recent USDOT records, at the end of 2003, vehicle-related crashes in the U.S. exceeded 6 million with 3 million injuries and 42,643 deaths [NHTSA, 2004]. Based on a 2000 study of Crash Costs performed by the National Highway Traffic Safety Administration (NHTSA) [Blincoe et al., 2002] and extrapolating for 2003, the total cost from the aforementioned crashes was in excess of $240 billion. In the state of Minnesota alone, during 2002 (latest available data), 94,969 crashes occurred, resulting in 40,677 injuries and 657 deaths [Mn/DOT, 2003]. Minnesota ranks low in the national statistics for road crashes, while the U.S. is fourth in the world in number of crashes per 100,000 people [BASt, 2003].

While the majority of crashes, especially fatal ones, concentrate in rural areas, for example, in Minnesota in 2002 only 23,676 crashes occurred in urban freeways and trunk highways of which 83 fatal, urban freeway crashes affect a much larger percent of the population. According to the latest TTI mobility report, non-recurrent congestion in urban freeways (mostly caused by crashes) has increased and is in excess of 2.9 million hours annually [Schrank et al., 2004]. Specifically, it has been estimated that, on average, every minute during which a freeway crash is not cleared causes five minutes of delay for motorists [Helman et al., 2004]. It is clear that a lot can be gained by focusing our attention only on freeways and trunk highways.

According to the *Minnesota Motor Vehicle Crash Facts* report of 2002 [Mn/DOT, 2003], on urban freeways and trunk highways "*property damage only"* crashes lead arithmetically with a total of 16,992 as compared to 83 "*fatal"* and 6,601 "*personal injury"* crashes. Although no specific data are available for freeways only, assuming state wide economic figures, the average cost of a fatal crash is $1,040,000 (the FHWA value for life is greater than $3M but Mn/DOT selected a lower value in order to avoid fatal crashes overshadowing non-fatal one), the average cost of an injury is $16,500 and the average property only damage cost is $6,500. Based on these numbers, the property damage only crashes on Minnesota freeways in the year 2002 reached a total of $110,448,000 while the fatal ones account for $86,320,000 and injuries for $108,916,500. Interestingly, the cost of highway crashes is three times the annual Mn/DOT budget for road maintenance and operation.

"*Property damage only"* crashes are the only ones that have increased in the last three years whereas all other types have generally declined. In single vehicle crashes the primary factors are "Speeding" and "Driver Inattention." In multi-vehicle crashes "Driver Inattention" is the primary factor followed by "Following too closely (tailgating)" and "Speeding." Specifically, when single and multi-vehicle crashes are combined "driver inattention" is the primary factor for less severe crashes (the majority). Finally, by analyzing these data it can be confirmed that, in less severe crashes, the leading type is rear-end collision and this same type is concentrated mainly in the peak hour where congestion is observed [Mn/DOT, 2003].

Traditional measures to reduce crashes include improved geometric design, congestion management strategies, as well as better driver education and enforcement. While such measures are generally effective, they are often not feasible or prohibitively expensive to implement. Prior to the widespread use of cell phones, considerable research effort was spent in developing automated incident detection systems. To be sure, fast detection and clearance is currently the most important function in any traffic management center because it reduces congestion and helps prevent secondary crashes. In spite of this, incident detection is still a reactive measure.

A 1997 USDOT/NHTSA campaign [Martinez, 1997] encouraged the removal of the word "accident" from the traffic management vocabulary and suggested replacing it with the word "crash." Some of the reasons stated were:

- "Motor vehicle crashes and injuries are predictable, preventable events. Continued use of the word "accident" promotes the concept that these events are outside of human influence or control. In fact, they are predictable results of specific actions";
- "We can identify their causes and take action to avoid them. These are not 'acts of God,' but predictable results of the laws of physics".

This realization along with the increasing need to reduce crashes and their side effects has recently prompted research for proactive approaches in order to avoid crash occurrence in the first place. One of the most promising options gaining wide acceptance in recent years is the concept of detecting crash-prone traffic conditions in real-time and warning drivers when the likelihood of a crash is high in order to increase their attentiveness, thereby reducing the number of crashes. Evidence suggests that when driver attentiveness increases, crashes decline in spite of poor driving and environmental conditions [COMPANION, 2001]. Towards that goal, recent research findings have presented strong evidence of an association between traffic conditions and crash probability. For example, vehicle speed has been associated with crashes in numerous studies; Lee, et al. (2002), found that as variability in speed, exposure (demand levels) and density increases, the likelihood of a crash increases. Further, higher crash frequency was observed during the peak period and, counter intuitively, with normal weather conditions. Additionally, attempts have been made in the creation of real-time models for the estimation of crash probability, ultimately aiming in the development of automated driver warning systems for crash prevention/reduction [Madanat et al., 1995, Oh et al., 2001].

A common factor on the aforementioned studies is the fact that they explore crash events in a large area i.e., an entire freeway or freeway network. Although this does not reduce the benefit from the associations presented between traffic conditions and crash likelihood, it does little towards accounting for the existence of high crash areas. It is a known fact that, in the majority of freeways, certain locations exhibit considerably higher crash frequencies than the overall average. For example, in the freeway network of the Twin Cities metro region in Minnesota,

approximately 35% of the crashes concentrate in 10 areas, each no longer than one mile (approximately 7% of freeway network with 10.93% of total Vehicle Miles Traveled). To be sure, the highest crash area is a 0.46 mile (.74 km) long section of I-94 westbound in the south of downtown Minneapolis. This area in 2002, exhibited 4.81 crashes per million vehicle miles (MVM) while the network average is 0.96 crashes/MVM. During the PM peak period the aforementioned site exhibits an average of 15.43 crashes/MVM while in comparison the entire I-94 freeway experiences only 3.29 crashes/MVM [Mn/DOT, 2002]. From these statistics it is clear that this and similar high crash freeway sections require special treatment not only for estimating crash likelihood but also for eventually identifying what are the specific crash causal factors. The former serves as the basis for developing effective real-time driver warning and/or traffic calming systems while, the latter, not only assists in their targeted deployment but can also guide more traditional traffic management interventions.

The research presented in this document attempts to establish reliable associations between real-time traffic measurements and crashes. These associations should help demonstrate that crash-prone traffic conditions exist, are detectable, and possibly manageable for avoiding crashes.

1.1 RESEARCH OBJECTIVES AND GENERAL APPROACH

Crashes cause million dollar losses, take lives, and drastically reduce the productivity of the transportation infrastructure. To find means to prevent them, the objectives of this research are:

1. Determine if crash-prone conditions (CPCs) are detectable by real-time traffic measurements which can be used reliably to warn drivers in advance and prevent crashes.
2. Develop effective and efficient crash-prone condition detection algorithms.
3. Propose Possible Implementation Schemes.

To accomplish these objectives, the effort was divided into the following tasks.:

1. Selection of a crash-prone location.
2. Instrumentation of the site to capture live crashes on video and simultaneously collect detailed traffic measurements as well as weather and other environmental information.
3. Identification of potential measurements and metrics able to manifest distinct patterns during pre-crash periods.
4. Employment of measurements and metrics in a crash likelihood model.
5. Based on the resulting model(s), development of a CPC detection algorithm.
6. Testing the above algorithm under varying traffic and weather conditions.

As already mentioned and explained in detail in the following chapter, this is not the first attempt to establish a relationship between crashes and traffic conditions. The unique aspect of this research in comparison to similar attempts is the collection of extensive crash data, the collection of detailed microscopic traffic measurements, and the focus on high crash locations. Specifically, this research did not rely solely on police records for collecting crash cases but was based on the collection of detailed wide-area video recordings of complete live crash events as well as normal conditions; this was combined with simultaneous measurements of individual vehicle speeds, headways and other derived traffic measurements in time and space. The selected site has the highest crash rate in the state of Minnesota; this allowed the collection of a fairly large number of events under different traffic and environmental conditions.

In this document the results of extensive observations and analysis of crash video recordings guided the formulation, estimation, and selection of the most significant parameters explaining the relationship between traffic conditions and crash likelihood. Based on these observations, early on, defining what constitutes a crash-prone condition was a big challenge. There can be many definitions depending on the conditions one considers sufficient in producing a crash. For the purposes of this study, we define as a "crash-prone condition" a set or range of traffic variables values such that, given them, the probability of a crash exceeds some threshold. The challenge now is to find this set or range of measurable traffic variables that increase the probability of a crash. Following the example of earlier research efforts, a case-control study methodology is selected. Cases are sets of traffic parameters collected prior to observed crashes while controls are sets of parameters from periods where no crash was observed. It is important to note here that in this work crashes are not only collisions but include cases of near-misses. This is an enhancement on earlier research where only time periods prior to collisions where considered crash prone. Since crash-free does not necessary means non-crash-prone the resulting models will be calibrated by examining the fit of each individual case and control. Following this analysis, a number of models for real-time estimation of crash likelihood are presented. These models are incorporated into automated, crash-prone conditions detection algorithms. Finally, the performance of the proposed algorithms is evaluated over several days in real-time.

2 Background

Preventing crashes is a real and urgent problem. Currently transportation organizations depend on traditional measures to reduce crashes. Such measures include improved geometric design, congestion management strategies, as well as better driver education and enforcement. While such measures are effective they often are not feasible or are prohibitively expensive to implement. This realization along with the increasing need to reduce crashes and their side effects has recently led to proactive approaches in order to avoid their occurrence in the first place.

One of the most promising options gaining wide acceptance in recent years is the concept of detecting crash-prone flow conditions in real-time and warning drivers when the probability of a crash is high in order to increase their attentiveness, thereby reducing the number of crashes. Evidence suggests that when driver attentiveness increases, crashes decline in spite of poor driving and environmental conditions [Dinges et al., 1998; Holohan et al., 1978; Kiefer et al., 1999 and 2003; Lee et al., 2002]. However, research in identification and detection of crash prone traffic conditions is embryonic. Sufficient proof that such conditions actually exist is still lacking, let alone a methodology for effective detection and system deployment. Based on the available literature, research on the subject can be divided into two major areas. The first covers studies aimed at developing models for the estimation of safety metrics, such as crash rate and severity. This is accomplished by associating such metrics with traffic, roadway, and driver/vehicle characteristics. The second area of research covers studies aimed at developing models for real-time estimation of crash likelihood. Potentially the latter can be used in real-time driver warning systems.

2.1 CRASH RELATED FACTORS

Williams et al. (1982) support the theory that road crashes result from combinations of a variety of interacting factors rather than from single identifiable causes. Chira-Chavala et al. (1986) state that crashes are complex phenomena and the problems at each location are likely to be different or site specific. However, Baldwin (1966) supports the theory that all crash causes are equal if, by the elimination of any one of the crash causal factors, the crash might have been prevented. According to Baker (1990) a "causal factor" is any circumstance "contributing to a result without which the result could not have occurred." This definition, in essence, is used in a variety of fields, i.e. the Air Force has used "A cause is an act, omission, condition, or circumstance which if corrected, eliminated, or avoided would have prevented the mishap." This definition does not exclude the possibility of multiple, independent, causal factors and therefore supports Baldwins' approach, which differs from Williams, who supports the notion that elimination of any single interacting factor might not be sufficient to prevent the crash. Regardless, researchers agree that there are factors that can be associated with crash events. Such factors can be broadly categorized as traffic characteristics, road geometry, vehicle characteristics, driver behavior, and

environmental conditions. Several studies have been devoted to the identification of such factors and the understanding of their relationship with crash rates.

For example, vehicle speed has been associated with crashes in numerous studies. Specifically, Liu and Popoff (1997) point out that approximately 60% of the errors resulting in crashes are speed related. Their study finds that average travel speed, 85th-percentile speed and speed differential are correlated to traffic crashes. A Davis et. al., (2004) case control study supports the theory that higher speeds are associated with an increase in crash risk, while in a study from Bahrain, Aljanahi et al. (1999) found that average speed has a significant influence on crash rate. In contrast, an earlier study by Garber et al., (1989) models the association of crash rate (crashes per vehicle distance traveled) with average speed and variance. The data were collected from 31 different sites including interstates, arterials, and rural major connectors. According to the authors, crash rate increases as the variability in speed increases for all classes of roads and it does not necessarily depend on the average speed. Additionally, they demonstrate that the difference between the design and posted speed limits significantly affects speed variance. Recent work by Lee, et al. (2002), investigated the potential for crashes along a 10-km (6.21 miles) stretch in Canada using loop detector measurements. A first-order log-linear model was developed relating variation of speed, traffic density, environmental characteristics, and exposure to frequency of crashes. To test the overall goodness-of-fit, a Pearson chi-square test was performed and it was found that the observed and predicted crash frequencies were not significantly different at a 95% significance level. Also they found that as variability in speed, exposure and density increases the likelihood of a crash increases. Further, higher crash frequency was observed during the peak period and with normal weather conditions. They conclude by suggesting the model might be used to predict crash potential in real-time.

Several studies have been conducted to observe the effect of environmental factors on crashes, such as weather, time, day of week, visibility, pavement and light conditions. Edwards (1998) found that the crash severity decreases in rainy conditions, perhaps because drivers tend to adjust (lower) their speed. Additionally, Lee et al. (2002) found that crashes are more likely to occur in clear and dry weather. Similarly, Duncan et al. (1998) concluded that icy/snowy conditions do not have a significant effect on crash occurrence. The “day of the week” association with crash rate and severity has been explored in a number of studies. Specifically, Cerrelli (1996) determined that there is considerable variation in traffic crashes during different days of the week. The average daily crashes were less on Monday, Tuesday and Wednesday as compared to other days, with Thursday having a slightly higher level. Cerrelli also pointed out that the trend on Friday and Sunday is similar; with Saturday exhibiting the highest crash rate. Yau (2003) concluded that, as compared to the rest, Friday, Saturday, and Sunday exhibit crashes of higher severity.

Since the early days of transportation engineering, roadway geometry has been known to affect traffic safety. Traditionally, geometric design has been driven by the physical characteristics of the vehicle-pavement system. Specifically, alignment, inclination, and pavement material are the

usual factors associated with safety. Only recently, other roadway features, not physically participating in the driving task, been found to have identifiable relationships with crash rates and severity. Design elements such as pavement width, shoulder width, shoulder type, median type and width, and others have been proven to have an impact on driving behavior and on extending safety [Wong et al., 1992]. Specifically, Hadi et al., (1995) developed Poisson and negative binomial regression models to estimate the association of geometric design on crash rates for rural and urban highways. The highways were classified into nine categories by location, access type and number of lanes. Separate models were developed for each category. The variables included in the study were Annual Average Daily Traffic (AADT), speed limit and geometric design factors. They found that increased lane width, median width and shoulder width are associated with reduced crashes depending on the highway type. Also higher levels of AADT as well as non-uniform cross-section designs were found to be associated with higher crash frequencies. The authors concluded the study with the potential use of the models for assessing the cost effectiveness of geometric design standards and diverting attention to adverse geometric locations. Additionally, Fink and Krammes (1995) used linear regression analysis to analyze the relationship between crash rate and degree of curvature, approach tangent length and sight distance to the curve. They concluded that mean crash rate and degree of curvature has an r^2 value of 0.94 which suggests a strong relationship. In addition, tangent length and sight distance also have high correlation with crash rate. In a more recent study, Aty and Radwan (2000) used a Negative Binomial model to estimate crash frequency over 566 highway segments. The crash data were collected on SR50 in Central Florida. The indicator variables included in the analysis were annual average daily traffic and geometric variables such as degree of horizontal curvature, lane, shoulder, and median width, as well as section length. They found that heavy traffic volume had the maximum value of parameter elasticity of 0.62 indicating its strong association with crash frequency. Other variables that were statistically significant were narrow lane width, large number of lanes, urban roadway sections, narrow shoulder width and reduced median width. To conclude, according to Sabey and Taylor (1980) the human factor, either alone or in combination with other factors, contributes to most of the crashes. In addition to the roadway environment there are several in-vehicle or driver related factors that affect crash rate and severity. Such factors are age, seating position, seat belt use, fatigue, and driving under the influence of alcohol or drugs.

2.2 REAL-TIME CRASH PROBABILITY

Estimating crash probability based on real-time information, for crash prevention or traffic management in general, only recently received the attention of researchers. The studies discussed in the following sections cover the greater part of the effort devoted to the subject. Needless to say, none of these studies have reached the point of field test or implementation. The general theme in all these studies is the use of statistical methodologies to associate real-time traffic measurements with past crash events recorded on several locations of an urban freeway or highway.

For example, in an earlier study, Madanat and Liu (1995) developed an incident-likelihood prediction model based on traffic, weather and incident data. Two types of incidents—vehicle crashes and overheating vehicles—are predicted with a high goodness of fit to data and high prediction accuracy. The incident-likelihood prediction models were binary logit models, tuned with eight and a half months of incident data combined with a sample of non-incident data from the same period. The models give coefficients for the contribution of different explanatory variables like rain, speed variability etc., to the probability of a crash or an overheated vehicle.

Oh et al. (2001) estimated the freeway crash likelihood from inductive loop detector measurements and police records. The data used for the analysis were collected from I-880 freeway in Hayward, California. Two traffic conditions were defined, crash prone and normal traffic conditions. The normal traffic condition was assumed to be a 5-minute period, 30 minutes prior to the crash and the crash prone condition as the 5 minutes prior to actual crash time. Six variables were initially considered to identify difference between disruptive and normal traffic conditions, i.e. mean and standard deviation of occupancy, flow, and speed. Based on a t-test the researchers decided to use only standard deviation of speed as an indicator for distinguishing between crash prone and no-crash prone conditions. The methodology followed in estimating the likelihood of crash occurrence is based in the ability to classify any given 5-minute interval, presented as a set of random variables of traffic conditions $\vec{X}=[x_1,x_2,\ldots x_p]$, into two patterns, crash prone and no-crash prone based on the aforementioned indicator. Assuming that the probability density functions (PDFs) associated with the input vector $\vec{X}$ for the two populations are known, let normal traffic and disruptive traffic can be denoted by Π_1 and Π_2 respectively. A reasonable classification rule is provided as:

$$\pi_1 : \frac{f_1(\vec{X})}{f_2(\vec{X})} \geq \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)$$

$$\pi_2 : \frac{f_1(\vec{X})}{f_2(\vec{X})} < \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)$$

Where c(i|j) is the cost of misclassifying a given traffic condition, that is classifying $\vec{X}$ as belonging to population Π_i while it belongs to population Π_j. $p_i=p(\Pi_i)$, the prior probability of occurrence of population Π_i. the estimated PDFs, $f_i(\vec{X})$ are used to estimate the posterior probability that x belongs to Π_i.

The authors show that, based on Bayes decision theory the best classifier minimizes the probability of classification error. The a priori estimate of the probability of a certain class can be converted to the a posteriori probability by Bayesian classification. The posterior probabilities are:

$$P(\pi_1 \mid \vec{X}_0) = \frac{P(\pi_1 \text{ occurs and observe } \vec{X}_0)}{P(\text{observe } \vec{X}_0)}$$

$$= \frac{P(\text{observe } \vec{X}_0 \mid \pi_1)P(\pi_1)}{P(\text{observe } \vec{X}_0 \mid \pi_1)P(\pi_1) + P(\text{observe } \vec{X}_0 \mid \pi_2)P(\pi_2)}$$

$$= \frac{\mathrm{p}_1 f_1(\vec{X}_0)}{\mathrm{p}_1 f_1(\vec{X}_0) + \mathrm{p}_2 f_2(\vec{X}_0)}$$

$$P(\pi_2 \mid \vec{X}_0) = 1 - P(\pi_1 \mid \vec{X}_0)$$

The problem at this point is to estimate the PDFs. The authors used a density estimation technique called Kernel smoothing, in order to obtain the probability density functions, which were used to classify standard deviation of speed as either normal or disruptive. The Kernel density estimator is

$$\hat{f}_{h_n}(x) = n^{-1} \sum_{i=1}^{n} K_{h_n}(x - X_i)$$

where $K_{h_n}(u) = h_n^{-1} K(u / h_n)$ is the kernel, with scale factor h_n.

The above PDFs estimated were used in the Bayesian model to classify that traffic conditions are normal or disruptive. The model is given by the equation:

$$P(A \mid X) = \frac{P(A) * f_{disruptive}(X)}{P(A) * f_{disruptive}(X) + P(N) * f_{normal}(X)}$$

where,

$P(A|X)$ = Posterior probability of crash occurrence

$P(A)$ = Prior probability that the traffic conditions are disruptive

$P(N)$ = Prior probability that the conditions are normal

$f_{disruptive}$ = probability density function of crash occurrence estimated using kernel smoothing

f_{normal} = probability density function of normal conditions estimated using kernel smoothing

Following a real-time application the authors concluded that standard deviation of speed is a good indicator of the presence of crash prone conditions in real-time applications.

In Canada, Lee et al. (2003) studied the effect of traffic conditions on crash occurrence using real-time traffic flow characteristics defined as crash precursors. Coefficient of variation of speed,

traffic density and difference of speed between the upstream and downstream stations (U) were the traffic flow variables considered in this study. The variable U was defined as

$$\overline{U} = |\,\overline{u}_1 - \overline{u}_2\,| = \left|\frac{t_p}{\Delta t}\sum_{t=t^*-\Delta t}^{t^*}\left(\frac{1}{n_1}\sum_{i=1}^{n_1}\overline{u}_{1i}(t)\right) - \frac{t_p}{\Delta t}\sum_{t=t^*-\Delta t}^{t^*}\left(\frac{1}{n_2}\sum_{i=1}^{n_2}\overline{u}_{2i}(t)\right)\right|$$

Where,

$\overline{U}$ =Average speed difference between upstream and downstream station;

$\overline{u}_1$ and $\overline{u}_2$ = Average speed at upstream and downstream station;

t_p= observation time interval of speed (sec);

t^*= Time of crash;

n_1, n_2= number of lanes at upstream and downstream end.

An aggregate log-linear model was developed to determine the effect of crash precursors on crash frequency. The actual time of crash was not known, therefore it was estimated from the change in speed profile between the upstream and downstream station at the crash site. They found that the speed difference at the upstream and downstream station was relatively higher during crashes. The authors proposed using the model for predicting real time crash potential.

In a recent study, Aty et al., (2004) developed a crash prediction model using case control logit methodology with loop detector measurements. A total of 670 crashes were collected on I-4 for the period of April 1999 to November 1999 from Orlando Police department records. Since the actual time of crash was not known, it was determined from the police records and traffic flow diagrams. Data for normal traffic conditions were extracted during all days corresponding to the day of each crash. The variables used for analysis include average and standard deviation of volume and occupancy, and coefficient of variation of speed. Matched crash–non crash analysis was performed by varying the ratio of crash and non crash from 1 to 5. They reported that the five-minute average occupancy at the upstream station during the 5-10 minutes prior to the crash and the five-minute coefficient of variation in speed at the downstream station during the same time were the most significant variables. The log odds ratio threshold selected resulted in identifying 69% of the crashes.

Additionally, Aty et al., (2004) developed another model to predict crash likelihood using Generalized Estimating Equations from the aforementioned data set. The model output was adjusted to reflect the actual proportion of crash and non-crash events in the study over the 8-month period. They found that bad pavement conditions, the existence of horizontal curvature, and high variability in speed 15 minutes prior to crashes increase crash likelihood. Further, it was demonstrated that the existence of an on-ramp within a 0.5 mile (0.8 km) downstream of the crash location increases the likelihood of crashes.

The weakness of the aforementioned studies is that they are based on the same basic information, i.e., police crash records and traditional macroscopic traffic measurements such as average speeds and flow rates, rather than detailed crash observations and individual (microscopic) traffic measurements. Specifically, all real-time systems based on Generalized Likelihood Estimation (GLE) i.e., logit or GEE methodologies are, as discussed above, in need of adjustment in order to provide future probabilities of crash risk. These adjustments require that the modeler knows the difference between the ratios of crashes to normal conditions in the sample and the whole population. Experience suggests that the police records do not contain all crashes; it has been estimated that almost 30% of crashes remain unreported [Hu et al., 1994]. Therefore, studies that are based on police records for their ground truth have the inherent risk of mislabeling crash conditions as normal, ending up with improper fit and reduced model reliability. In addition, since the sample employed in the model development does not include all crashes in a certain period, crash frequencies between road sections can be miscalculated.

A second problem of using only police records stems from the fact that they rarely reflect the actual time of collisions. Usually the reported time when the dispatch was notified or the Traffic Management Center noticed the crash on the surveillance cameras. This introduces a problem in studies that depend on having all crash cases synchronized in order to determine significance of time dependent variables. The latter might be avoided through very close scrutiny of the measured data to detect the exact moment of the disturbance, making the procedure very difficult, especially with 30 sec measurements taken every 0.5 mile (0.8 km).

In this study both problems have been avoided by using continuous uninterrupted video recordings to identify, locate, and classify the crashes as well as collect individual vehicle measurements. In addition, the aforementioned studies considered the aggregation time interval as 5 minutes for estimation of metrics. It is important to consider various aggregation periods in the analysis.

Finally, all the aforementioned studies discussed the use of the model in real time prediction. However, none of the studies performed a real-time assessment of the developed model over a continuous period of time. Such a test is important for validating and demonstrating the model performance in real time.

2.3 CRASH PRONE CONDITIONS DETECTION ALGORITHMS

The models described in the previous section have not passed the research stage. Although all were developed to be part of a real-time system, no implementation of a working algorithm was presented. The complexity of the step between a model and the algorithm that employs it depends on the type of the model. In a majority of the aforementioned models the algorithm was simply the systematic calculation of crash probability every time new measurements were

available and the comparison of this probability to a predetermined threshold. Other types of implementation, as explained later in this section, bypass the use of a model and create an algorithm based on heuristic, fuzzy logic, or other equivalent methods to produce alarms when crash prone conditions are detected. This section presents studies reaching the algorithm development and evaluation step. It should be noted that none of the algorithms described have ever been implemented in the field.

IN–RESPONSE project

The majority of the algorithms were developed under European Union collaborative projects. One of these projects is called IN-RESPONSE and it is a combined approach for several incident management approaches, one of them being Incident Prediction. The incident prediction system proposed in IN-RESPONSE was designed to estimate incident frequency on a given stretch of road during a given period. The number of likely incidents was related to the road geometry, prevailing traffic and weather conditions. This prediction model was to be used in several ways:

- To provide long-term "average incident frequencies" to be used in safety analysis.
- To give short-term real-time "expected incident frequencies" for use in traffic management systems. (What is the current probability of an incident occurring, should measures be taken now to reduce this likelihood, and what should these corrective measures be?) Prevailing conditions would be constantly monitored and potentially dangerous situations counteracted through ramp metering, information provision and speed control.
- To provide incident probabilities for models to determine the benefits of different incident management strategies. Simulation models can compare several possible measures and give indications as to which measures would have most impact.

The IN-RESPONSE incident prediction algorithm uses a multi-proportional Poisson model adapted to cope with variable conditions. It compares the numbers of incidents occurring under certain circumstances with the amount of time these circumstances prevail on a given road section. The prediction model considered information like roadway characteristics (number of lanes, weaving sections, gradients, presence of lighting, presence of hard shoulder and emergency phones), traffic characteristics (traffic volume, road capacity, truck percentage, speed distribution, headway distribution, congestion levels), and weather conditions (good/poor visibility, dry/wet road surface, wind/sleet/freezing conditions) for current and past events.

The IN-RESPONSE incident probability model is a logit model, quantifying the influence of several variables (for instance: rain, fog, traffic volume, speeds) on the probability of an incident. The model parameters are estimated using a data set with records of incidents and the state of the variables at the instant of the incident. The outcome of the model estimation is a set of parameters indicating the statistical influence of each variable on incident probability. This model was supposed to be the core element of an incident probability based Decision Support

System (DSS) which was designed for use in traffic control centers as one of the modules of the integral control system.

PRIME project

Another European project with similar scope is called PRIME. In PRIME three separate algorithms were developed, two based on statistical modeling and the third based on fuzzy logic theory.

HLOGIT ALGORITHIM

The first of the PRIME algorithms establishes an association between the probability of occurrence of an incident and a number of explanatory variables by means of a hierarchical logit statistical model (HLOGIT). The model's parameters were estimated by generalized linear regression models, based on traffic and weather conditions, road geometry as independent variables and the presence of incidents as a response variable, given as a result of the probability of incident occurrence for each section on the road network for which the data are available. The model was developed based on data collected in Barcelona, Spain during 2000 and 2001. According to the final report of PRIME this algorithm did not satisfy the project requirement of 10% False Alarm Rate (FAR) (it exhibited a rate of 20 to 40%).

LOGIT ALGORITHIM

The second algorithm developed in the PRIME project was based on a logit model associating crash probability with the following variables:

SPEEDVOL: average speed over chosen period over all lanes, weighted by volume

SCHOM_VC: variation in chosen period of average volume/capacity ratio over all lanes

D1_D2: average difference in density between two leftmost lanes (of three lanes) over chosen period

V1_V2: average difference in volume between two leftmost lanes (of three lanes) over chosen period.

Apart from SPEEDVOL, all variables included describe dynamics in the traffic flow: either in time (SCHOM_VC) or between lanes (D1_D2 and V1_V2). Only the coefficient for V1_V2 had a positive sign, indicating that the larger the difference in traffic volumes between lanes 1 and 2 (leftmost lanes), the higher the incident probability. The other coefficients had negative signs, indicating that the larger the value, the lower the incident probability. For SPEEDVOL this is logical: higher speeds are associated with higher levels of service and lower incident risk.

For SCHOM_VC and D1_D2 it is more complicated. The negative sign for SCHOM_VC implies that the higher the variation in chosen period of average volume/capacity ratio over all lanes, the lower the incident probability. A high variation in volume/capacity ratio would probably be found in less dense traffic, because there is less need to use the left lane. And when

traffic gets more dense, the variation in the average volume/capacity ratio over all lanes would decrease, as traffic is then likely to be more evenly distributed. So again, higher levels of service would lead to lower incident risk.

The same reasoning can be applied to D1_D2: the difference in density between the two left lanes would decrease as traffic gets denser (leading to a higher incident probability), and would increase as the level of service increases (leading to a lower incident probability). Alternatively, the results might be explained by the level of awareness of drivers. Perhaps the variation itself is not the problem, but the perception of the drivers thereof: variation is only a problem when they are unprepared for it. This algorithm also did not satisfy the PRIME target for false alarm rates, it exhibited FAR between 17% and 51%.

According to the report both of the statistical algorithms encountered problems with the collected measurements and the amount of the available events. Specifically, inaccuracies in the reported time of the crash raised the question whether the crash was responsible for the variability in traffic conditions or whether these conditions caused the crash.

Fuzzy logic algorithm

The third and final algorithm or more appropriately "expert system" developed under the PRIME project followed a fuzzy logic approach. The fuzzy logic approach was based less on data from the field, and more on expert judgment with regard to which (traffic) conditions are more likely to produce incidents. It used fuzzy logic to capture the experience of traffic engineers and traffic police and produce estimates of the incident probability. This approach was initiated because of the aforementioned problems with the data required for the statistical models.

The algorithm was based on expert opinion from sessions with police and traffic engineers. The interview responses and knowledge were then transcribed into fuzzy rules. The steps taken in this process were:

- First, it was established which variables were considered to be the most important for the incident probability.
- Then, it was decided which variables were to be taken into account as inputs to the model.
- Next, the most important part of the knowledge required from the experts was the fuzzy rules for the rule base.
- Finally, each of the variables was divided into a number of fuzzy sets, which was done in accordance with the suggestions of the experts.

In the de-fuzzyfication process, the output of all the fuzzy rules in the rule base was combined into a single output. That output was a representation of the risk incurred by a single driver when traveling the roadway link during the time interval under consideration.

The fuzzy rule base contained 119 rules, of which 14 related to time variables, 27 to traffic conditions, 27 to weather conditions and 15 to road conditions. The combination of variables between classes resulted in another 36 rules. There was no clear indication as to the performance of the fuzzy algorithm, or even if it has been evaluated.

2.4 CRASH PREVENTION SYSTEMS

Development and implementation of crash prevention systems can be divided into two categories. The first category deals with In-vehicle crash avoidance systems like adaptive cruise control, forward and backward collision avoidance systems, run off-road driver assist systems, as well as the larger Intelligent Vehicle Highway System program aimed at removing the driver from the equation. The majority of these systems depend on instruments in the vehicle which measure speed, headway, acceleration/deceleration, and distances from the surrounding vehicles and use these to influence the vehicles trajectory and avoid a collision or running off the road. In general, none of these systems depend on information concerning the general traffic flow conditions or even what happens 2-3 vehicles ahead. Although a number of these systems have reached the market, very few vehicles in the road are currently equipped with such systems. In a study by Rajamani and Levinson (2005) it was illustrated through microscopic simulation that unless the market penetration of, in their case, adaptive cruise control systems exceeds 90% there will be no visible improvement in the traffic stream. Considering the scope of this research no further discussion will be devoted to in-vehicle systems.

The second category, the one where this works belongs, covers infrastructure crash prevention systems. Unlike the first category, very few advances have taken place in the development and implementation of infrastructure-based crash prevention systems. The few that have been developed and deployed are either manual in operation (require human intervention) or very simplistic in the underlying automation. For completeness, three examples of infrastructure based systems are presented in this section.

2.4.1 Manual Driver Alert Systems

The most common example of manually operated driver alert system is the decades old Highway Advisory Radio (HAR) system. In the United States, this system is comprised of short range FM transmitters manually activated by the traffic operators. Electronic signs or Variable Message Signs (VMS) on the side of the road alert the drivers about a problem downstream and prompt them to tune-in to the prescribed frequency for more details. In general HAR is used for incident management alerting vehicles upstream of the location of a crash, the approaching congestion, and of possible alternate routes. In Europe HAR-like systems, called RDS-TMC, are more sophisticated in the sense that each vehicle's radio is capable of automatically tuning to the emergency frequency, therefore not requiring the use of special roadside signs. As an incident management tool HAR has proven very successful but is still a reactive measure in need of human control (I-95 Corridor Coalition Report, 2001).

Another manually operated system commonly encountered in tunnels involves Lane Signs. As can be seen in Figure 2.1 lane signs are dynamic displays capable of changing from a green arrow indicating no problem ahead, to a yellow arrow signaling congestion in the lane downstream of the present location, or a red X signaling that the lane is blocked either in the tunnel or immediately after it. Lane signs are usually activated manually by the traffic operators,

although a few automated systems exist in European tunnels. Such implementation depends on sensors measuring the speed of the stream in each lane and/or the existence of stopped vehicles.

Figure 2. 1 Lowry Hill tunnel Lane Signs. Minneapolis, MN.

In Europe, as part of the TABASCO and INFOTEN projects, a freeway driver alert system has been deployed. This system, called COMPANION, depends on Light Sticks on the side of the road. These light sticks resemble the traditional reflective sticks usually found on the roadside in locations where heavy snowfall is frequent, but are equipped with dynamic LEDs capable of changing color and flashing pattern. In their current state the light sticks are manually activated and change from yellow to red signals varying congestion and lane blocks downstream.

2.4.2 Automated Driver Alert Systems

Automated driver warning and crash prevention systems based on dynamic weather and pavement information have been deployed in the field during the last ten years. These systems depend on sensor information on rain, fog, and pavement ice formation to trigger warning mechanisms like VMS, variable speed limit signs, entrance ramp gates, and de-icing systems. None of the deployed systems take into account the prevailing traffic conditions.

The most notable effort in the development and deployment of crash prone traffic condition detection and alert systems is currently active in Japan. Under the Advanced cruise-assist Highway System (AHS) program, two systems relevant to this research are currently in the field deployment and evaluation stage. The first is an automated roll-over alert system. This system is comprised of several sensors providing information on individual vehicle speed, height, and weight and estimates the probability this vehicle might roll-over in the approaching sharp curve. If this probability is greater than a defined threshold variable, message signs and HAR-like radio alert the specific driver of the danger and require him to slow down. These systems are based on simple physical models for their determination of the roll-over probability and are mainly aimed at preventing large truck crashes.

The second system developed under the AHS program is more relevant to the scope of this research. It is an automated rear-end crash prevention driver alert system aimed at roadway sections with geometric designs that limit visibility. Two such systems have been deployed in high crash areas in Japan. The first system is on the freeway connecting Tokyo and Osaka (Figure 2.2) while the second system is located in the Sangubashi section of the Shinjuku Line on Tokyo Metropolitan Expressway No. 4 (Figure 2.3). Both deployments are very similar. The problem is that the sharp curve of the roadway limits visibility. Additionally, in both locations frequent congestion, either recurrent or non-recurrent, causes vehicles to queue-up immediately after the curve. The forming queue is not visible to the approaching vehicles, resulting in frequent and some times fatal rear-end crashes. The solution in both cases involves traffic sensors detecting the formation of a queue, or in the case of a previous crash, lane blockages and the automatic activation of several driver warning devices. The method of the warning depends on the distance to the curve and can be seen in Figures 2.2 and 2.3.

The logic in these systems is simple. If there is a queue in the predefined position or any other lane obstruction, the warning mechanism is activated. These systems are not sensitive to the level of congestion or the existence of compression waves in the traffic stream.

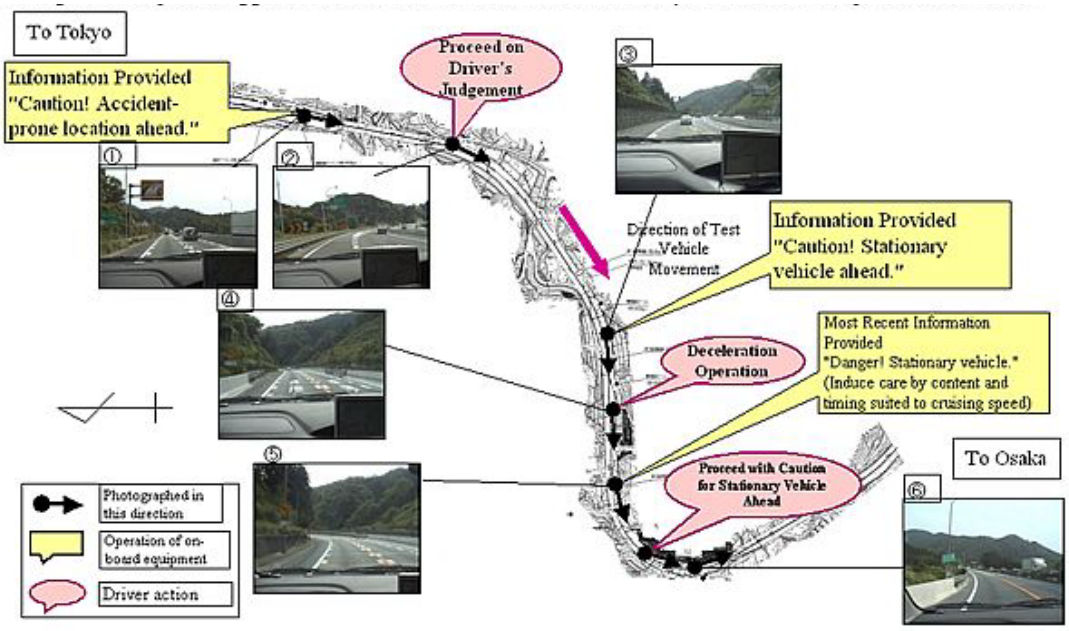

Figure 2. 2 Tokyo - Osaka Freeway Rear-end crash prevention system.

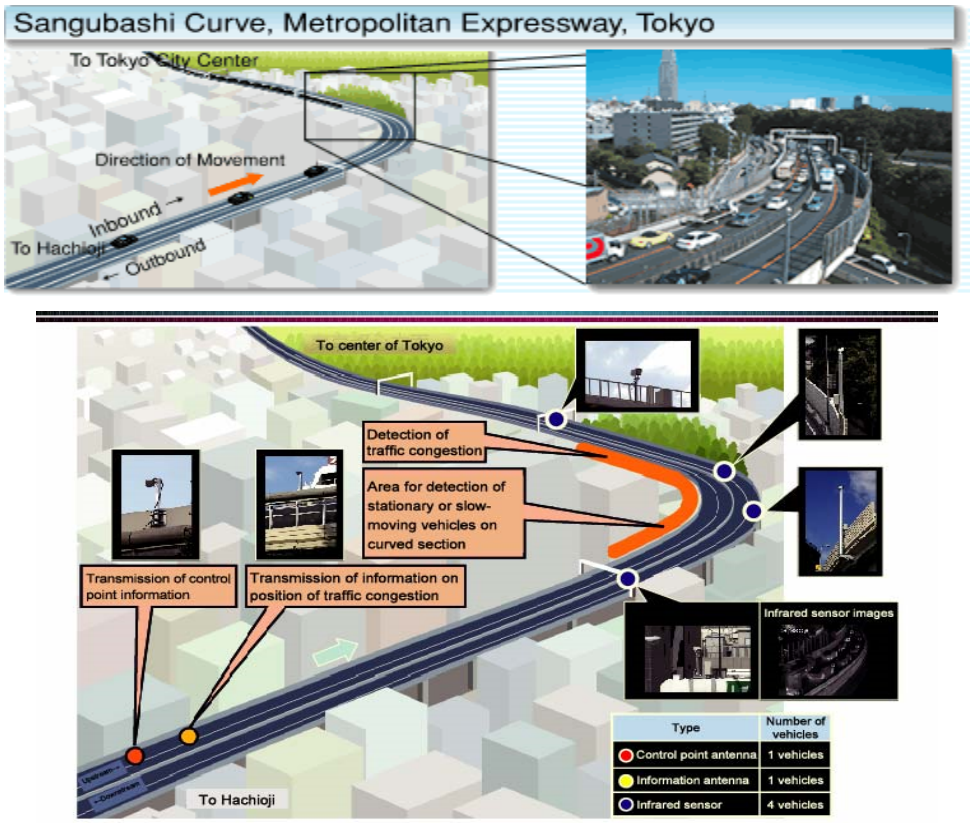

Figure 2. 3 Sangubashi Curve Rear-end Crash prevention system.

2.5 SUMMARY OF LITERATURE REVIEW

Table 2. 1 Literature.

Crash related factors				
Liu and Popoff (1997)	Average speed 85th percentile of speed Speed difference	Based on loop detector data and police records	Statistical correlation analysis	Increase in any of the three factors causes increase in crash frequency.
Davis et al. (2004) and Aljanahi et al. (1999)	Average speed	Reconstructed speeds		Higher speeds increase crash risk.
Garber et al. (1989)	Speed variance	Loop detector data	Statistical correlation analysis	Speed variance positively related to crash risk. Average speed not related to crash risk.
Lee et al. (2002)	Speed variance Exposure (volume) Density, clear/dry weather	Loop detector data	Log-linear model	Increase in any of the three factors causes increase in crash frequency.
Edwards et al. (1998)	Rain	Loop detector data	Statistical correlation analysis	Rain decreases crash severety.
Duncan et al. (1998)	Ice/snow	Loop detector data	Statistical correlation analysis	No effect on crash risk.
Cerrelli (1996) Yau (2003)	Day of the week	Loop detector data	Statistical correlation analysis	Significant association with crash frequency.
Hadi et al. (1995)	# of lanes, access type, location, AADT, speed limit, lane, shoulder, and median width	Loop detector data	Poison and negative binomial regression models.	All factors associated with crash frequency.
Fink and Krammes (1995)	Degree of curvature, approach tangent length, sight distance	Loop detector data	Linear Regression	Noted associations.
Aty and Radwan (2000)	AADT, degree of curvature, shoulder and median width.	Loop detector data	Negative binomial model	Heavy volume, narrow shoulder and reduced median increase crash frequency.

Real-time crash probability				
Madanat and Liu (1995)	Volume, occupancy, speed, weather	Loop detector data and police records	Binary Logit	Two models: Crash probability, Overheating Prob.
Oh et al. (2001)	Mean and standard deviation of: Speed, flow, occupancy	Loop detector data and police records	Bayes classification method	Crash conditions defined as the 5 minute period prior to crash. Normal conditions were defined as the 5 minute period, 30 minutes prior.
Lee et al. (2003)	Coefficient of variation of speed, Density, speed differential (up/downstream)	Loop detector data and police records. Crash time adjusted based on speed profile.	Log-linear model	
Aty et al. (2004)	Average and standard deviation of volume, occupancy, Coef. of var. of speed.	Loop detector data and police records.	Matched case-control logit analysis	Selected as normal 5 same days of the week for each crash (assuming no other crash was reported).
Aty et al. (2004)	Pavement condition, horizontal curvature, speed variance, presence of on-ramp.	Loop detector data and police records.	Generalized Estimation Equations	Same as above.

Crash prone conditions detection algorithms			
In-Response (2003) Europe	Incident prediction through a Multinomial Poisson model	Incident probability through a logit model	Not clear how the two models where combined in a real-time system. Tested only through simulation.
PRIME (2004) Europe	Three crash probability Algorithms: 1. Hierarchical logit 2. Logit 3. Fuzzy logic	Two Datasets: 1. Barcelona (loops, police records) 2. Athens (loops, police records)	During off-line evaluation no algorithms reached acceptable performance levels. Never implemented in the field.

Crash Prevention systems			
AHS (Japan)	No crash probability algorithm. Queue identification based on video sensors triggers driver warning systems.	Two deployment sites in Japan.	Consortium devoted in Advanced highway systems including crash prevention.

3 Data Collection

In the majority of studies related to the estimation of crash probability, the information comes from historical records. As described in chapter 2, when the information about the crash event is based on the available police records, it is not always accurate or complete and this introduces uncertainty and errors into the analysis. Additionally, the detail contained in the records is limited and does not always allow the researcher to understand the sequence of events leading to the actual collision. In this study, although the original plan was to collect and analyze historical crash records, preliminary analysis confirmed the aforementioned problems. It was therefore concluded that for the goals of this particular research, mainly because it is aimed on high crash locations, historical crash reports and loop detector measurements are insufficient. Regardless, 450 historical crash reports have been retrieved, categorized, and classified based on the reported crash time, location, and type (rear-end, sideswipe, road departure, etc.).

In order to alleviate the aforementioned problems the entire sequence of events leading to real collisions had to be visually observed while simultaneously collecting a variety of traffic measurements on and around the crash location. This approach posed additional requirements in the design of the data collection methodology and system. In this chapter these requirements will be discussed along with a description of the selected test site and the deployed instrumentation. Finally, the additional, derived metrics used in this study will be introduced along with their definitions and references in literature.

3.1 DATA AND TEST SITE REQUIREMENTS

The goal of this study is to identify crash prone traffic conditions and design an automated crash prevention/reduction system. From earlier studies in the subject it was clear that a corridor-wide system would not be feasible. Additionally, following an initial examination of the crash statistics for the Twin Cities, it was observed that the majority of freeway crashes concentrate in a relatively small number of roadway sections. These locations will be referred to as *high crash areas* throughout this document. Although a corridor-wide system is at the moment not feasible, a system targeting a specific high crash area could be successfully designed and deployed.

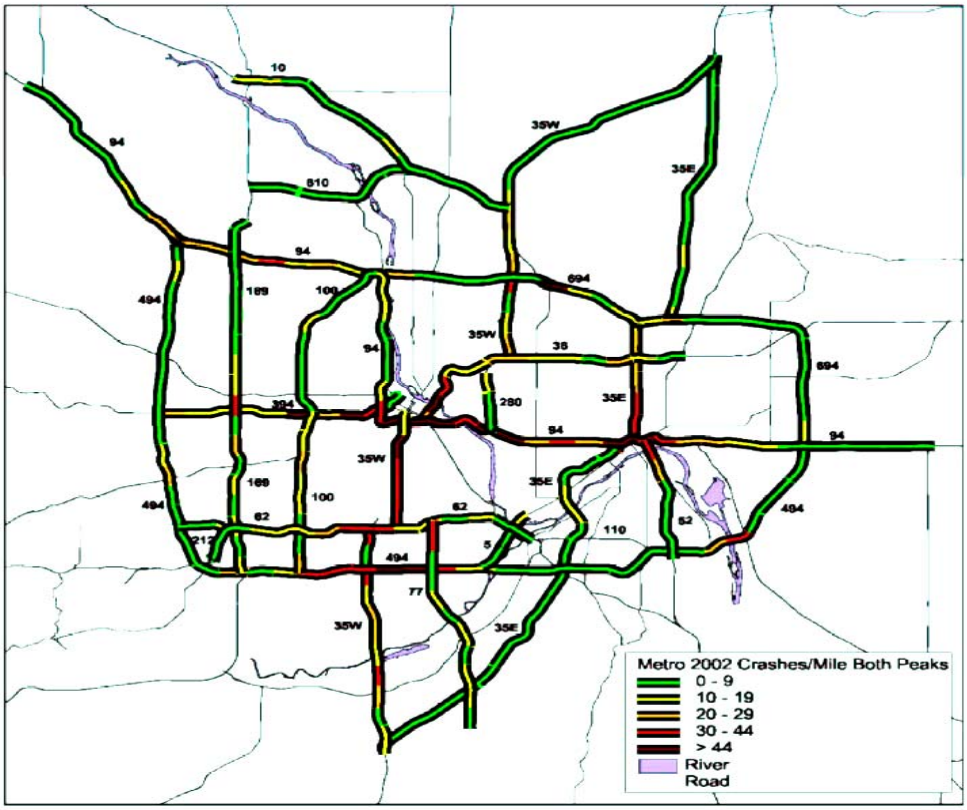

Figure 3. 1 2002 Twin Cities Freeway Crashes per mile.

(Source: Mn/DOT Crash Facts 2002)

The candidate sites for this research were the high crash areas in the Twin Cities freeway network. Mn/DOT has already studied these crash concentrations and has created maps like the one seen in Figure 3.1 ranking each freeway section by the number of reported crashes in a year; the figure is for the year 2002 (latest available).

In order to ensure that all crashes, regardless of severity, are identified and the video records saved, the surveillance coverage had to be continuous throughout the high crash section selected. Crashes are rare events and, at least in the beginning of this research, unpredictable both temporally and spatially. In order to capture a sufficiently large number of crashes on tape the surveillance area had to be relatively large.

Simultaneously with the video recording, detailed traffic measurements had to be collected in several locations throughout the area. This guarantees that at least one station will always be downstream and upstream of all crashes, regardless of their position in the high crash section. The collected traffic measurements should not restrict the resolution of the analysis since, at the start of this research, the speed which crash prone conditions progress is not known. Individual vehicle speeds, headways, and classification are the most qualified since, as it will be seen later; most flow variables and metrics can be derived from these basic ones.

In summary, this study required a high crash freeway section where it would be possible to deploy a large amount of surveillance and traffic detection equipment, allowing continuous video coverage and simultaneous collection of individual vehicle speed, headway, and length in several locations. The selection of a site was based also on the opportunity to maximize data collection with easy access for the research team and affordable communication requirements. Figure 3.2 shows the ten highest crash sections.

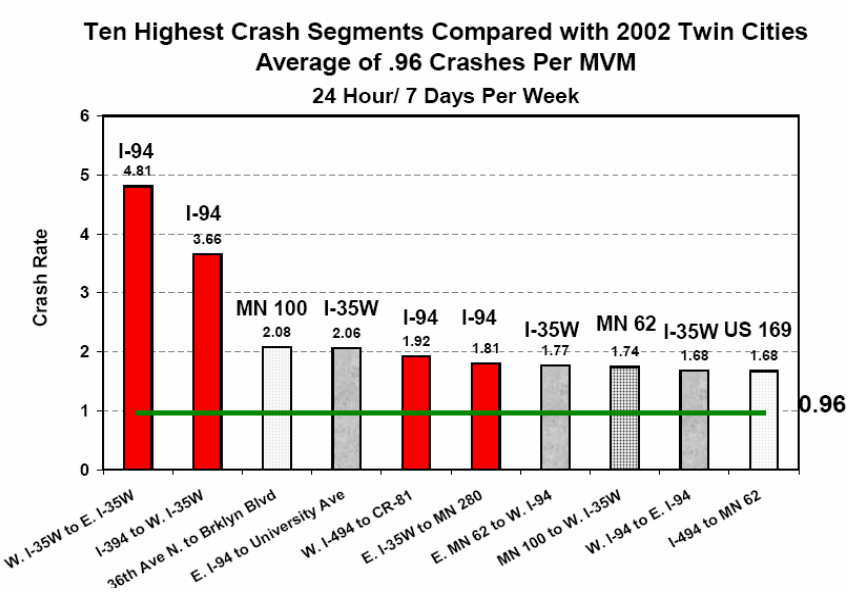

Figure 3. 2 Ten Highest Crash Sections in 2002.

(Source: Mn/DOT Crash Facts 2002.)

Additionally, the study required environmental information, synchronized to the surveillance and traffic data records. As will be discussed in greater detail later in this chapter, initially the police records were to be used to guide the search in the video records for crashes. Although such records were acquired from the State Patrol they were not particularly useful.

3.2 TEST SITE DESCRIPTION

According to Mn/DOT records [Mn/DOT, 2002] the site with the highest crash rate in the state is a 1.7-mile-long (2.73km) section of I-94 westbound in Downtown Minneapolis between 11th Ave and the Lowry Hill tunnel. This site has a rate of 3.81 Crashes/MVM (million vehicle miles). This rate roughly translates to an average of, one crash every two days. I-94 is a connector type freeway joining the cities of St. Paul and Minneapolis. This freeway carries an average daily traffic in excess of 80,000 vehicles per direction and it is congested for at least five hours daily, especially during the afternoon peak period. The crash prone section runs parallel to I-35W

(another major freeway) and a number of short ramps allow transfer from one freeway to the other. An aerial photo of the site is presented in Figure 3.3. The site includes two entrance and three exit ramps; the average number of lanes is three with two 3000-foot (914 meter) auxiliary lanes in two weaving areas. In section B, shown in Figure 3.3, excessive weaving takes place due to the high volumes entering from the rightmost ramp which in fact combines traffic from 35W, HW55, and the downtown business area.

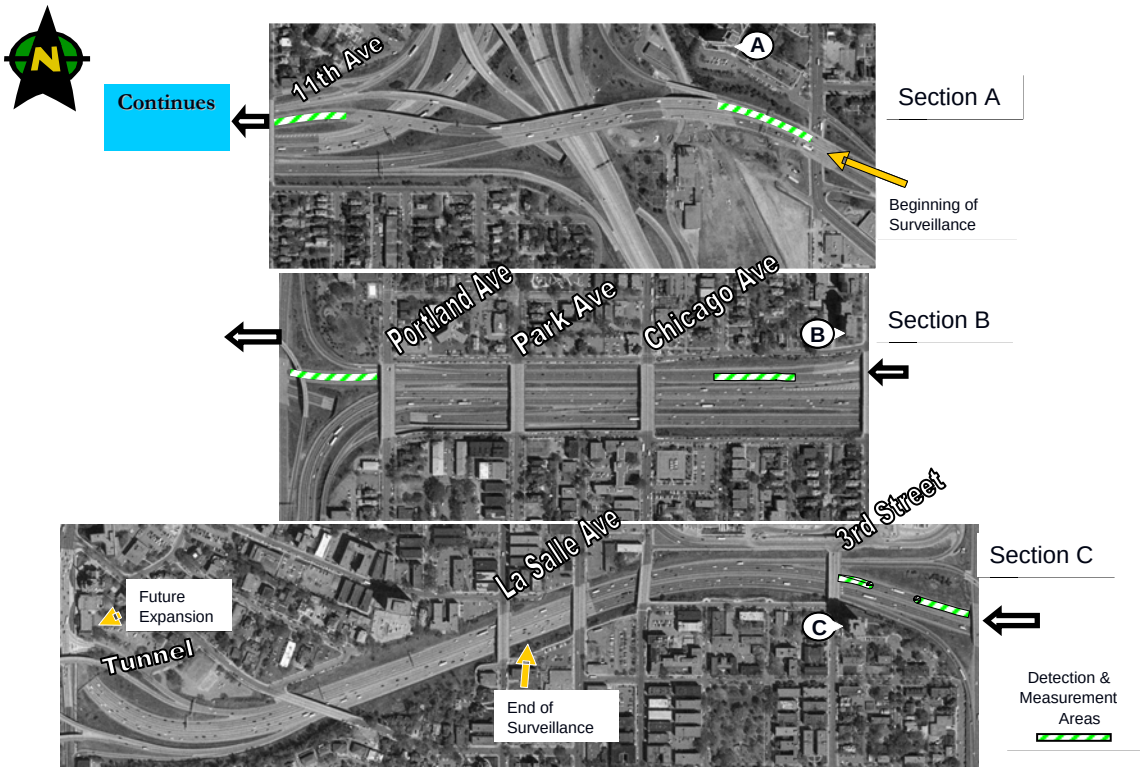

Figure 3. 3 Selected High Crash Section: I-94 Westbound.

For the purposes of this study, the aforementioned site presented the highest potential for capturing crashes on video as well as analyzing complex traffic flow dynamics. From an economic standpoint, a crash prevention solution here will have a high potential return since any incident on this section results in hundreds of vehicle-hours of delay due to the high volumes it carries. Regardless, for this site to be selected, it had to have additional features allowing easy and affordable deployment of instrumentation.

3.3 INSTRUMENTATION REQUIREMENTS

The Twin Cities freeway network is one of the most heavily instrumented in the country. Mn/DOT has deployed close to 4,000 inductive loop detectors. Loop detectors are located on the mainline of freeways approximately every ½ mile and on entrance and exit ramps. These detectors collect information on volume and time occupancy, reporting the data back to the

Regional Traffic Management Center (RTMC) every 30 seconds via a copper SONET network. Additionally, there are 285 cameras located along the Twin Cities freeway system. The standard design includes color pan/tilt zoom cameras mounted on 50 foot poles, one mile apart, with fiber optic communications. Video from all cameras is shared via a distribution network with stakeholders including Minnesota State Patrol, Mn/DOT Metro District Maintenance, Metro Transit, cities, counties, local television stations, and the Center for Transportation Studies at the University of Minnesota.

Unfortunately the available instrumentation was not enough for the purposes of this study. The traffic information collected by the loop detectors is aggregated every 30 seconds and includes only volume and lane occupancy. Additionally, the detectors are located roughly every ½ mile. Therefore, in the site under consideration there were at most four stations available. As already mentioned, this project required traffic information of higher spatial and temporal resolution. It was clear that additional traffic sensors had to be deployed and preferably allow relocation if such action was necessary during the course of the research.

In the site under consideration there were in total three Mn/DOT cameras available. The design of these cameras allows the operators to observe ½ mile of roadway upstream or downstream of the camera location. For the other ½ mile to be observed the camera has to be rotated by the operator. This functionality serves traffic operators since their primary function is to look for abnormalities in the flow, detect incidents, and verify reports from the field. For these reasons the cameras cannot be locked in one place for continuous observation and even if they did, only half of the site would be covered. Again, it was clear that the already available instrumentation was not suitable for the purposes of this project and additional surveillance cameras had to be deployed.

Keeping in mind the needs of this research study and the inadequacy of the available instrumentation, a separate project was initiated to design, assemble, and deploy the necessary equipment. The following sections present a brief description of the process and the lessons learned during that project.

3.3.1 Functional Requirements and Technologies Selected

Functional and cost requirements as well as design of the detection and surveillance system drove component selection. Many functional requirements were considered and some were ultimately dropped to minimize complexity and cost or to ensure reliability. In the end it was decided that the integrated system should have the following characteristics:

- Detect traffic and collect real-time measurements.
- Capture and transmit surveillance video.

- Disassemble and reassemble easily at new deployment sites. Minimum cabling or site wiring is required.
- Install with minimal effort and time, causing no traffic disruptions.
- Consume little power to allow operation through a solar panel when no regular power source is available.
- Have unified communication system for data and video. This minimizes the number of components and simplifies communication to achieve a real-time link between the deployment site and the supervising station.
- Be accessible remotely for component configuration, debugging and monitoring.

Based on these characteristics, specifications were developed for the technologies used in the system's three components: traffic detection, surveillance, and communication.

3.3.1.1 Detection Technology

Although widely used, loop or other in-pavement detectors are not a feasible choice for this detection system because installation of sensors requires disruption of traffic operations, is time consuming and costly. Therefore, the conclusion was that non-intrusive technologies such as radar or machine vision would be a better choice. This type of detection technology should:

- Be non-intrusive.
- Configure easily to minimize deployment time and effort.
- Be capable of robust placement to allow freedom in locating the sensors. Ideally, the sensors will function some distance from the roadway.
- Provide detailed, individual vehicle detection and measurements like speed, headway, and classification.
- Provide information with a single sensor in order to minimize installation/relocation time.
- Support real-time extraction of comprehensive microscopic and macroscopic measurements.
- Consume little power.
- Demonstrate reliability, accuracy and robustness in real-life installations.

These requirements narrowed the options for detector selection to two major technologies. The first is based on the Doppler principle and includes radar, microwave or laser and the second is based on machine vision.

The first group satisfied the cost, power consumption, and ease-of-use requirements but it was restrictive on sensor placement. Doppler-type devices need to be close to the roadway but best operation is achieved if the sensors aim perpendicular or parallel to the roadway; this implies

detection in only one direction (one dimensional) rather than simultaneously in two directions (two dimensional). In addition, some users of these technologies have expressed concerns about their reliability during adverse weather conditions. Another inherent problem of this technology apparently is poor stopped-vehicle detection [NIT Phase II report, 2002, and FHWA, 2001].

Machine vision sensors on the other hand do not have most of the aforementioned restrictions. Specifically, assuming proper camera selection (telephoto versus wide angle lens), machine vision detection equipment can be very robust in terms of placement since it can be installed hundreds of feet away from the roadway. For example, they may be placed on top of very tall buildings as long as they have a clear line of sight. The greatest advantage of machine vision is that it provides the means for checking accuracy and operation since along with the measurements it also supplies video from the scene. This allows both surveillance and recording of the traffic conditions for off-line verification or analysis. Thus, machine vision sensors can, with proper lens selection and placement, double as a surveillance camera. Finally, these sensors allow two-dimensional detection, i.e. a single sensor can cover multiple lanes over a wide area or multiple approaches and lanes at intersections. Their costs are similar to comparable Doppler-based sensors. Based on all the aforementioned requirements and characteristics, the machine vision technology was selected for deployment in the detection stations.

3.3.1.2 Surveillance Technology

The majority of permanent surveillance systems currently deployed use analog video capture and transmission. However, analog video, albeit of very good quality, is expensive to transmit over anything other than land-based copper or fiber optic lines. Specialized wireless devices can transmit analog video but they can only transmit images from a single camera and they work in dedicated pairs (transmitter/receiver). For each additional camera another radio pair is necessary. Such constraints render analog video unsuitable for a temporary, portable design. Contrary to analog, digital video can be of comparable quality but in a more robust, easier to transmit form. Based on such considerations, the major functional specifications for the surveillance technology are the following:

- Video should be in digital form when it is transmitted to the supervising station, regardless of the video-capture format.
- The refresh rate of the video images should be at least 15 frames per second (fps) (analog video is 30 fps). This requirement ensures the observer a clear view of the traffic stream, the traffic-stream speed and other flow conditions.
- The technology should allow the recording of multiple video feeds at the remote site to assist in system communications. Not all video feeds need to be watched simultaneously, so it would be a waste of bandwidth (and increased cost) to transmit video from all cameras at all times.
- The image resolution must be of sufficient quality to allow the viewer to easily distinguish vehicles for manual identification and classification as well as to confirm

detection/measurement accuracy. A minimum size field of view ensures a minimum vehicle size in the observation monitor.

- The system needs to cover as wide an area of the roadway as possible. This requirement as well as the previous one can be both satisfied if the selected technology allows the combination of video feeds from more than one camera.
- The system should be compact, requiring minimum cabling in order to be easily moved from place to place.
- The system should have low power requirements.

Digital video is superior to analog because it can be compressed at the source to a small fraction of its original size. This reduces transmission cost while preserving speed and quality. Video digitization and compression are relatively new technologies that are evolving rapidly. Originating from the security industry, a limited number of products have recently been adapted to transportation needs. Real-time video digitization typically uses proprietary hardware devices and cannot be easily customized by the end user. In contrast, because of the lack of recognized standards, video compression still uses mainly software/firmware. As a result, a plethora of proprietary, non-proprietary, and experimental CODECs (Compression-Decompression algorithms) applications are available. All of these factors make it difficult to grade and select a particular compression technology, but also allow for customization to fit specific requirements.

The most widely used video compression effort is being undertaken by the Moving Picture Experts Group (MPEG), which is a working group of the International Standards Organization (ISO/IEC) in charge of the development of standards for coded representation of digital audio and video. Established in 1988, the group has produced MPEG-1, the standard used by Video CD and MP3 formats, MPEG-2, the standard used within Digital Television set top boxes and DVD technologies, and MPEG-4, a standard for multimedia for the fixed and mobile web. All MPEG compression algorithms take advantage of the fact that video scenes change little between frames. Therefore, instead of the entire video frame, only the temporal differences are stored or transmitted. The differences between MPEG-1, MPEG-2 and MPEG-4 are on the working resolution (MPEG-2 has twice the resolution of MPEG-1) and the way they treat differences between frames (MPEG-4 compresses more than the other two, resulting in lower picture quality but with considerable reduction in size). MPEG-1 and MPEG-2 were designed for applications with a lot more bandwidth than a portable surveillance system can afford. In contrast, MPEG-4, being the newest of the three, was designed for this purpose, i.e., multimedia support for mobile equipment. Unfortunately, for the same reason, the MPEG-4 standard is still open to interpretation. Because of this, there is no interoperability or image-video format exchange between products that claim to be MPEG-4 compliant.

These issues required the exploration and testing of a number of compression solutions during the design stage. The objective was to capture and transmit video and store it in a format that either can be played back or processed through imaging algorithms for vehicle tracking, classification, or extraction of other information. The proprietary implementations of MPEG-4

that were tested, although very successful in compressing the signal and maintaining a good visual quality, lacked the translation tools and robustness that allow use of the stored information by other applications. The final implementation, as described later in this chapter, depends on in-house integration of off-the-shelf hardware components with software applications employing MPEG-4 compression that can be obtained at no cost. This solution afforded considerable latitude to experiment and fine tune the CODEC to achieve the optimal balance between picture quality and bandwidth requirements**.**

3.3.1.3 Communication Technology

Having established the requirements for traffic detection and surveillance, the key component that pulls everything together is the technology that transmits the information from the field to the supervising station. For permanent infrastructure this is accomplished through a web of cables buried in the ground. Such methods are incompatible with the temporary, portable nature of the proposed system. With the recent progress in wireless data communications, there are better solutions. The basic functional requirements of the communication technology are as follows:

- Transmits both data and video.
- Supports two-way communications, allowing remote monitoring and control of the system.
- Can be deployed in any location on or near the roadway because it does not depend solely on ground-based communications.
- Relocates easily because of compact, portable design and operates license free.
- Has low power requirements.
- Allows interaction with the system from any location with internet access (optional).

As mentioned in the previous section, a few wireless analog video-transmission devices are available in the market but their cost and the fact that they only support one way communication reduces their usefulness in this system. Therefore, alternative options were considered based on the observation that when all information is in digital form (data and digital video), wireless Ethernet radios can be used to establish a two-way link between the field and the supervising station. The basic characteristic of these radios is that they work in pairs. When communication with a single site is desired, a point-to-point wireless link is the best choice. In the case of several deployment sites a point-to-multipoint configuration, although slower, is required. In both of these configurations an Ethernet radio is located on the base side, acting as the Access Point and is connected to the Local Area Network (LAN) of the supervising station. Another Ethernet radio is located in the field, acting as the client or subscriber unit and is connected with the station LAN. In the case of a point-to-multipoint configuration one access point can be connected simultaneously with several subscriber units. In metropolitan areas where an infrastructure for wireless computing exists, there might be several access points available.

The major factor that drives equipment selection is bandwidth. While fiber and copper lines provide nearly unlimited bandwidth, wireless solutions are greatly constrained. One of the aforementioned requirements is that the selected solution does not require an FCC license, ruling out several high-gain radios (ample bandwidth). With the recent increase in demand for portable computing, the technologies supporting low-power, license-free, wireless Ethernet radios have increased dramatically (e.g., 802.11a/b/g protocols) [FCC, 2001]. These radios have the same bandwidth constraints of distance, line-of-sight, and interference. The effect of distance is straightforward; longer distances result in lower bandwidths. Radios that require line-of-sight usually achieve the highest bandwidths. The constraints due to interference are the most difficult to overcome. The location of the radio antennae plays a major role in performance. For example, large concrete structures can block the signal; large surfaces like the sides of tall buildings near the path of the signal can deflect it, reducing signal power, increasing noise, and even confusing the receiver with echoes that are out of phase. Even the ground features between receiver and transmitter can play a major role in performance.

The site under consideration fortunately lies within five miles from the University campus where the supervising station was located. The application required more bandwidth and deterministic response to large streams of data. After several trials, 802.16 radios operating within the UNI-II band (5.4 GHz) were deployed. They use QUAM (Quadrature phase Amplitude Modulation) to encode bits within the RF signal over Time Division Multiplexing (TDM). Under this scheme, each base station is given an equal short time slice (a few milliseconds) over the frequency spectrum range. Data transfer efficiency is at its highest only when there is enough to fill the entire time slice—which is most likely the case during streaming and high transfer rate applications (e.g., streaming video and data transfers from multiple stations).

3.3.2 System Architecture

A combination of three separate technologies, machine vision (for vehicle detection and measurement), digital video recording, and wireless data transmission comprise the final system. A brief description of each module along with the final integrated design is given in this section.

3.3.2.1 Detection Module

For implementation of the detection component of the system, the Autoscope Solo Pro machine vision (MV) sensors were selected [Michalopoulos, 1991]. The selection of this sensor was based on the results of the second phase of the Minnesota DOT and FHWA's Evaluation of Non-Intrusive Technologies (NIT II) study [FHWA, 2001] in which it had the highest ratings on its reputation, familiarity and support. Except for power and mounting restrictions, there is no limit on the number of MV sensors that can be placed in a single station. Depending on the location of the station (roadside versus tall building or other structure) a single sensor can cover a maximum of 600 feet (182 meters) of roadway with at most eight lanes (regardless of direction) and still maintain high accuracy on individual vehicle-speed measurements. If the latter is not required, this area can increase to a maximum of 1000 feet (304 meter) of roadway. These particular

sensors work by establishing regions (virtual detectors) on the video image in which the sensor needs to detect and measure traffic. There is practically no restriction on the number of virtual detectors that can be placed in a single camera view, but for practical purposes one every 100 feet/lane (30.5 meter/lane) for speed detection is the working limit (closer than that, the speed detectors overlap). The integration of three MV sensors in a single station is shown schematically in Figure 3.4. As the figure suggests not all sensors double as surveillance cameras. In addition, since the sensors communicate through a RS-485 port, a converter to Ethernet is required for integration with the LAN.

Two types of data are collected by the MV sensors: real-time measurements of presence, speed, headway, and vehicle length as well as aggregated measurements of flow, time mean speed, space mean speed, density and Level Of service (LOS). These are provided in aggregated total or average form, over a user-specified time period. The sensors can transmit all information in real time back to the supervising station over the wireless link and/or store the aggregated data locally for later download.

3.3.2.2 Surveillance Module

The objective of the surveillance module is twofold. First, it has to provide a clear real-time picture of the site that can be transmitted (streamed) back to the supervising station; second, it must save the video in digital form for later viewing or download. In the current application of the system long-term storage is not required for all video as only the sequences containing crashes need to be stored; after quick inspection, only the desired one-hour chunks are downloaded and burned in DVDs.

In the current design a single system can accommodate one or four CCTV cameras. When one camera is used the maximum resolution of the video is 720x480 pixels or D1 as it is known in the industry. With four cameras, each view occupies one quarter of the screen, having a resolution of 352x240 pixels or SIF. In the case of two or three cameras the configuration for four is used. Depending on the user requirements for video streaming, storing, and later manipulation, the four-camera version can either transmit four individual streams each having SIF resolution or combine all four through a quad processor and digitize the combined image on a single D1 resolution. Both versions were tested during the first phase of the implementation. The second version was finally selected based on the specific deployment requirements. It was achieved through a combination of common off-the-shelf equipment like a small form personal computer with digitizing board and free software applications for MPEG-4 compression and recording. The schematic of the integrated surveillance module can be seen in Figure 3.4.

The final configuration is the result of considerable customization, designed to meet additional needs such as seamless vehicle trajectory extraction and testing of future experimental vehicle tracking algorithms.

3.3.2.3 Communication Module

The detection and surveillance modules are integrated into a unified system through the communication module. As described in the functional specification section, the most efficient method for this integration is to establish a “mini” LAN at the deployment site. For this purpose the I/O interfaces of the detection and surveillance modules are first translated into Ethernet compatible ones. A common off-the-shelf network switch connects all devices together and forms the LAN.

During the course of the systems’ design a number of different communication solutions had to be explored for functionality and integration testing. These ranged widely in respect to bandwidth which varied from 19Kbps CDPD modems to 30Mbps Ethernet radios. However, because bandwidth is critical in the case of a fully configured system (multiple sensors and cameras), the 30Mbps Ethernet radio was chosen. The key advantage of the proposed architecture is it can be customized to meet specific deployment requirements and connectivity choices and therefore save on equipment cost and installation time. For example, a clear line of sight between the deployment location

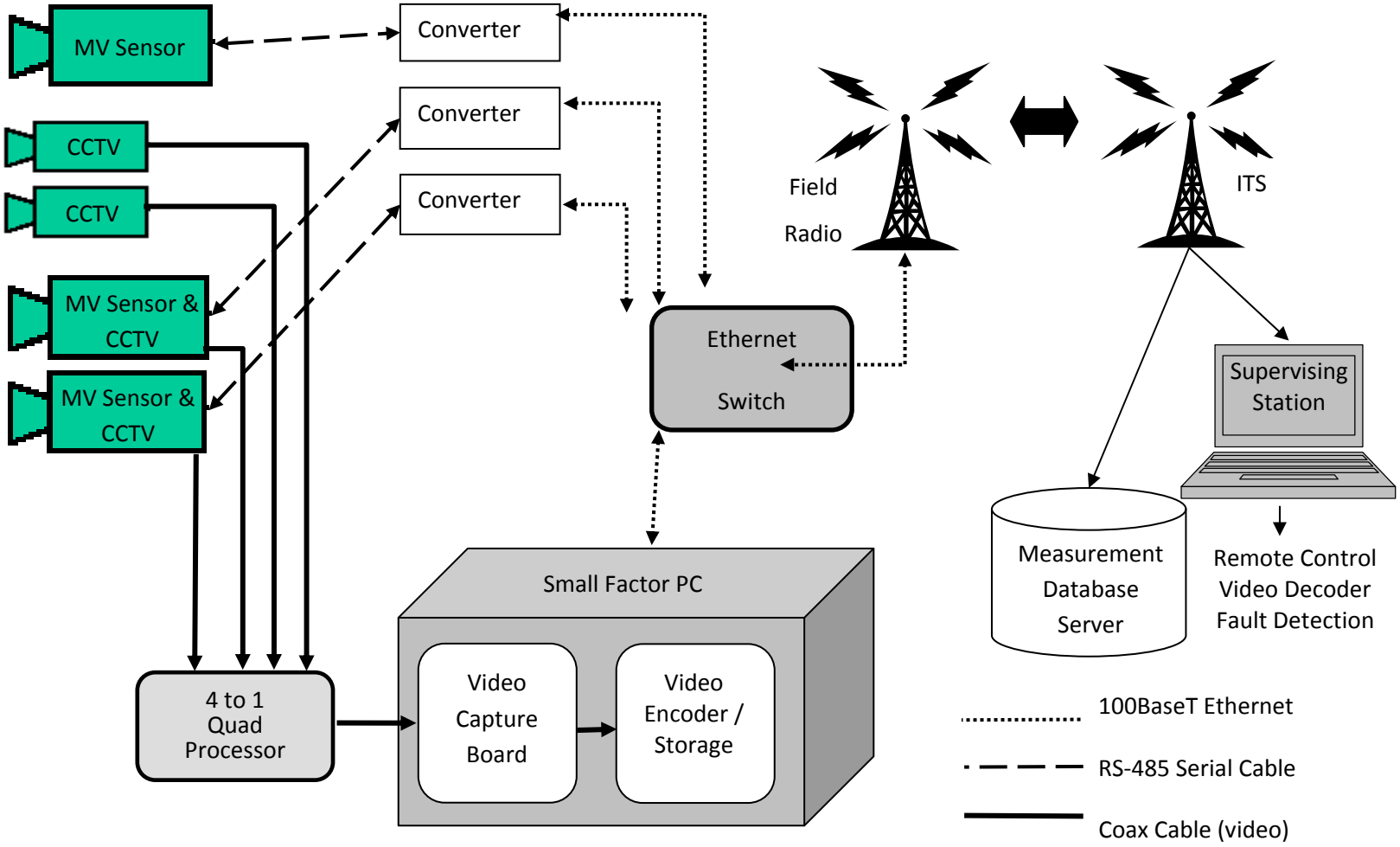

Figure 3. 4 Integrated System Architecture.

and the supervising station (or an adjacent tall building) are likely to be found in metropolitan areas. In this case a single point-to-point (or multipoint) wireless link will perform the best. On the other hand, in cases where there is no clear line of sight, a shorter range Spread Spectrum wireless link can connect the system with the nearest controller or communication cabinet. From that cabinet the information can join the permanent detection and surveillance infrastructure (usually a fiber or Copper SONET network), thereby closing the gap with the supervising station. In extremely rural installations a long distance away from the nearest telecommunication hub, a simple residential satellite Internet link can be used. Regardless of the variety and combinations of communication mediums, the basic system architecture remains the same, ensuring portability and quick deployment. This architecture was designed to leverage the fact that Internet connectivity has become ubiquitous.

As we will describe later, a single point-to-multipoint 30Mbps wireless link was selected for the current system deployment. The wireless link acts both as a router, facilitating connectivity between the field and a supervising station LANs, and as a security firewall prohibiting unauthorized access to the field system. The communication module, the integrating agent of the overall system architecture, is shown in Figure 3.4.

3.3.2.4 Data Management at the Supervising Station

Although not a part of the deployed system, the software running in the supervising station plays an essential role in the operation. When the link between the field system and the supervising station is established, traffic measurements and video begin to arrive. As explained in the surveillance module section, video is preferably stored in the field; therefore FTP connectivity is used to download only the selected files. For real-time surveillance/inspection, an off-the-shelf video streaming application is used to transmit video to the supervising station.

The most challenging problem is the traffic measurement retrieval. Specifically, in the case of the systems deployed for the crash study, individual vehicle speeds, headways, and classification information need to be collected every 500-1000 feet (152-304 meter) for a section totaling 6300 feet (1920.2 meters) and transmitted in real-time. This process is continuous, 24 hours a day, seven days a week; considering that the particular freeway site carries in excess of 80,000 vehicles daily, the amount of data is enormous. To cope with this, a relational database server has been established at the supervising station along with the necessary software that receives and stores the measurements. To develop such software it is important to select a traffic detection system that is accompanied by a well documented Applications Programmers Interface (API) and a robust Software Development Kit (SDK). The greatest risk in this process is the creation of a bottleneck in the information flow, due to inefficient software. Such an occurrence could result in the interruption of the transmission and loss of information. The architecture of the supervising station software modules can be seen in Figure 3.5. The data collection operation can be seen in Figure 3.6.

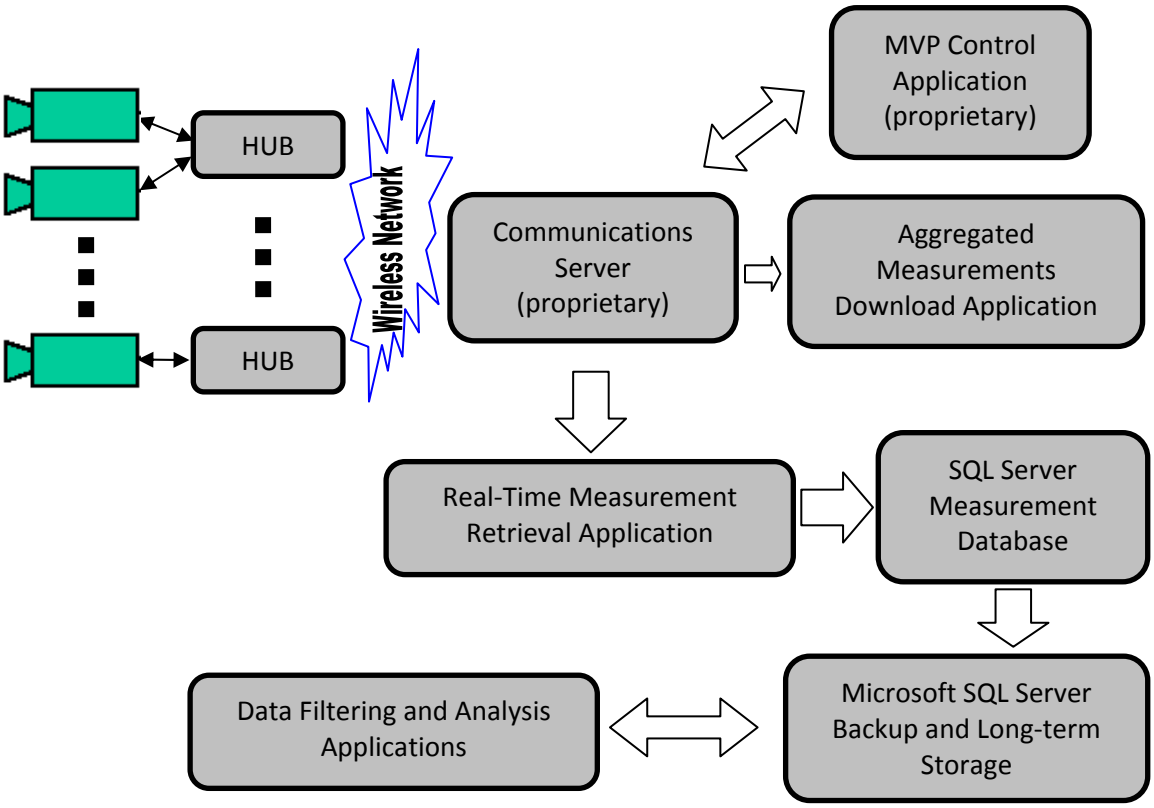

Figure 3. 5 Supervisor Station Architecture.

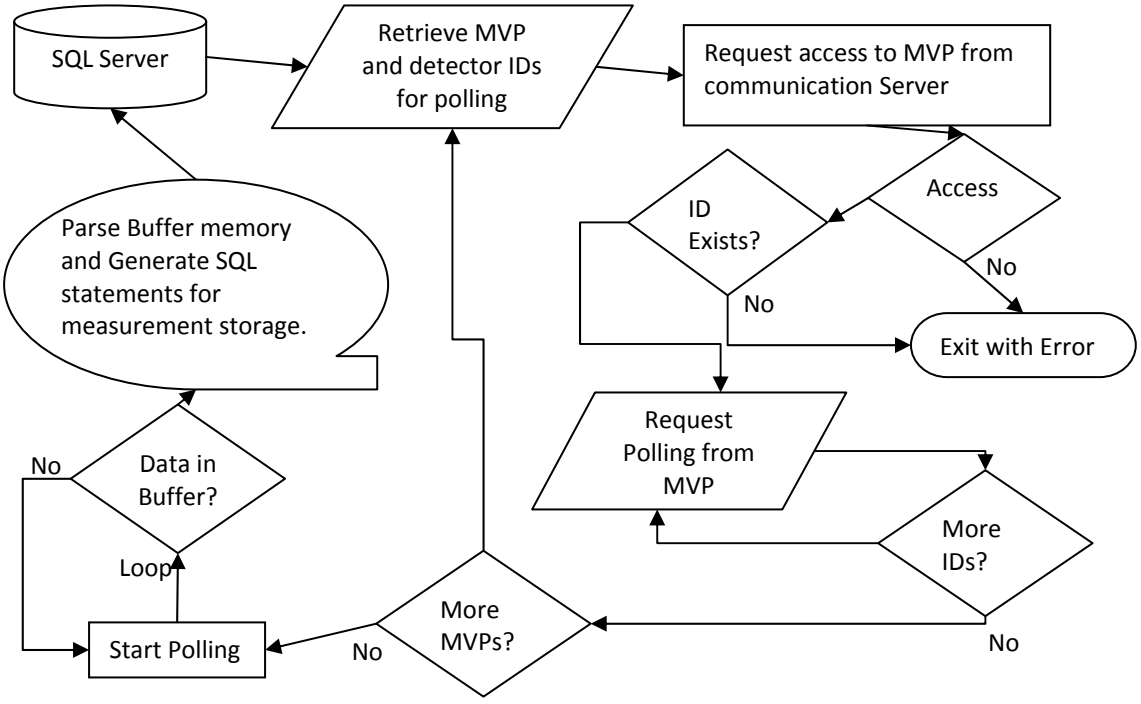

Figure 3. 6 Real-time Measurement Application Flowchart.

3.4 DEPLOYMENT OF INSTRUMENTATION

Following the initial design, the system was deployed at the selected high crash area. In total, four detection and surveillance systems were deployed in two phases for capturing live crashes on video at the selected site and to simultaneously extract vehicle traffic measurements.

3.4.1 Initial Deployment

Preliminary investigation identified sections A and B in Figure 3.3 as the ones generating the majority of the accidents. From this information two stations were deployed on the rooftops of high-rise buildings A and B, respectively. Details about the buildings and their locations are shown in Figure 3.7 while a picture of the station at building B can be seen in Figure 3.6

Figure 3. 7 Augustana Detection and Surveillance Station.

Station A is equipped with one MV sensor, doubling as a camera looking east and one regular camera looking west with a field of view that covers halfway to building B. Station B is equipped with two MV sensors and two surveillance cameras. The sensors are pointed down on the roadway east and west of the building (as indicated by the shaded areas in Figure 3.3) while the surveillance cameras are observing the same areas but with a lens configuration that provides a much longer field of view.

During this initial deployment each station was equipped with one commercially available video-streaming and digital video storage device, based on a proprietary implementation of the MPEG-4 compression algorithm by Stradient. In addition, each station was equipped with a 3Mbps BreezeAccess Ethernet radio subscriber unit by Alvarion for communication with the supervising station.

During this period of operation three problems were discovered. The first and most alarming was that the section under observation was not the point of impact for the majority of the collisions. As a result only a few crashes were recorded. Second, the realized bandwidth of the selected wireless link was in the range of 1Mbps. This was due to the nature of the particular design which assigns each direction (from and to the base station radio located in the University of Minnesota Center for Transportation Studies lab) half the available bandwidth, even if the actual data transfer is consistently higher in one direction. The result was very slow communications, prohibiting the streaming of video from all cameras simultaneously. Finally, the third problem arose from the video-streaming device. Although the algorithms employed produce a high quality picture requiring relatively small bandwidth, the video storage format was not compatible with any other image analysis software. This prohibits replaying the video through anything other than the user interface provided by the manufacturers. In addition, the manufacturers did not provide a working SDK which would have allowed in-house authoring of translation utilities to more widely accepted formats.

These particular problems could not have been anticipated at the time of equipment selection. The information gained during the initial deployment proved valuable in selecting surveillance and communications equipment for the second deployment which performed similarly but without the problematic limitations. In addition, valuable information was collected about the traffic conditions (since the detection module operated reliably). These data, in conjunction with the visual observations, assisted in selecting the next set of sensors and camera placements, substantially improving the crash recording capability. The previously installed stations were also upgraded but not moved as they continued to play an important role—complementing the continuous wide-area coverage of the entire site.

3.4.2 Second Deployment Phase

After the initial phase, a third rooftop was selected as the next deployment site to complement the area coverage. The rooftop of building C, shown in figure 3.3, is located west of the initial buildings. Because of the ideal location of the building, a large section of roadway is visible from its rooftop; therefore, two additional stations were deployed. New wireless links were selected and an in-house video streaming and storage solution was developed without affecting the system design. Specifically, all wireless links at all three sites were replaced with 30Mbps Tsunami Ethernet radios (in the final deployment one link services two stations rather than just one).

The new surveillance module, although conceptually similar to the previous one, was composed entirely of common, off-the-shelf components. Specifically for portability, this surveillance module (Figure 3.4) consists of a regular half-height PC compatible with a 1.8GHz Intel Pentium4 processor, 512MB of RAM and a 60GB HD. A Haupaugge WinTV video capture board, located in one of the two expansion slots of the motherboard, digitizes the signal from the cameras. A license-free video-recording application named VirtualVCR, written by Shaun Faulds and available from the Free Software Foundation, runs under a scheduling application. The video recording employs the DIVx MPEG-4 compression algorithm which is available in the market at minimal cost (less than $50). Since the video capture card has only one input, four video feeds are first combined into one with a video 4-1 quad processor. This video solution was subsequently implemented in all stations. In summary, at the third location, three MV sensors, four surveillance cameras, two quad processors, two video streaming PCs and one wireless link were deployed. All of the MV sensors are doubling as surveillance cameras allowing for continuous coverage of approximately 6,300 feet (1920 meter) of roadway.

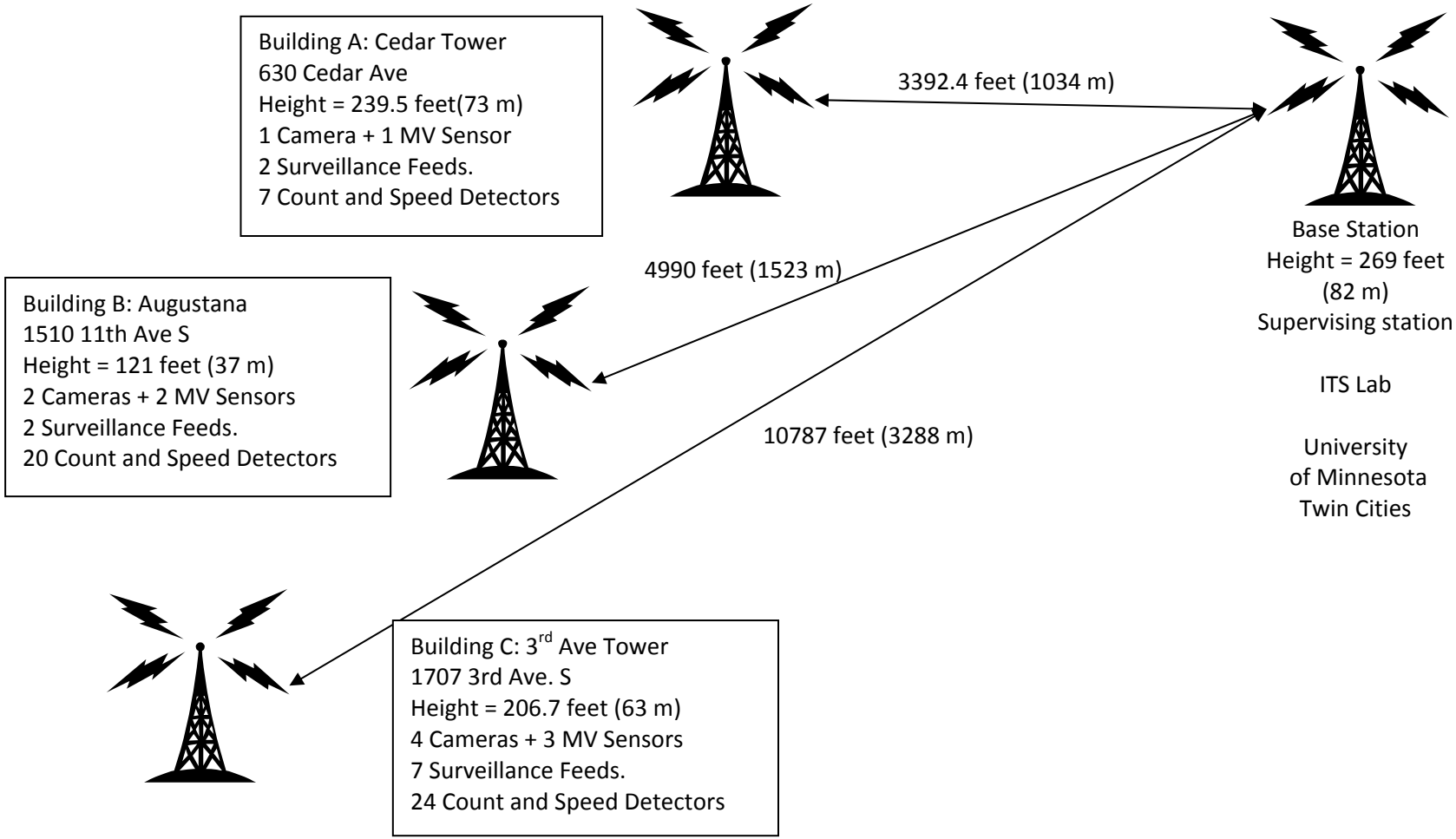

Figure 3. 8 Deployment Site Topology.

3.5 DATA COLLECTED

Several sources of information about the traffic, weather, and crash events were utilized in this study. The detection and surveillance stations were the most instrumental of all the available sources. In addition to station measurements from the Mn/DOT, we retrieved inductive loop detectors in the site as well as acquired environmental measurements from the local RWIS station, and crash records from the State police. The following sections presents a summary of the information collected from each of the aforementioned sources.

3.5.1 Detection and Surveillance Stations

The four detection and surveillance stations overseeing the deployment site provide detailed data and large amounts of wide-area measurements. In conjunction with the video recordings, this unique database can facilitate advanced studies of traffic characteristics, flow dynamics, and safety and traffic management concepts. Specifically, the sensors collect and transmit in real time individual vehicle *speed, length*, and *time headway* as well as every 10 seconds aggregated measurements of *flow rate*, *volume count, time mean speed, space mean speed, space occupancy, density, LOS* and *vehicle class count*. The aggregated measurements are saved in the sensor for later download. The aforementioned data are collected on a 24/7 basis in six areas of the freeway site as seen in Figure 3.3. Each of these areas is 300 to 500 feet (91 to 152 meter) long and measurements are extracted approximately every 100 feet (30.5 meter) for a total of 51 detection points (18 series times 3 lanes, plus exit and entrance ramps). An example of the detector setup as well as the video collected can be seen in Figure 3.9; the detectors are represented by the green bars overlaid on the video image while the number displayed next to them indicates the speed of the last vehicle passed.

To date the laboratory has been operational during three Minnesota winters and two stormy summer seasons, providing detailed measurements during extreme weather conditions. For purposes of data management, video recording takes place between 7:00 A.M. and 8:00 P.M. during weekdays and 12:00 P.M. until 9:00 P.M. on Saturdays and Sundays. Not all video is stored long term but selectively as needed. Apart from this, a collection of video from all 11 cameras at important times, days, and weather conditions is stored for future research needs. During the periods of the initial and final deployments a verification of the systems data-collection performance was undertaken by comparing the collected information with the video records. Calibration was carried out by comparing speed measurements taken manually with a laser gun with those automatically collected.

For the purposes of this study, between January 2003 and January 2004, 110 crashes have been recorded along with 350 previously unreported near misses. In May 2003 a bug was discovered in the supervisory station application software. This unfortunate event resulted in the loss of all data from three sensors for the period of January 2003 to July 2003. For this reason, although all video records were used in the qualitative analysis, only 51 crash and 122 near-miss cases were used for the subsequent time series and statistical analysis. The error resulted in the storage of

vehicle records with the wrong sensor ID. It might be possible to recover some of these records although no such action is planned for this phase of the research.

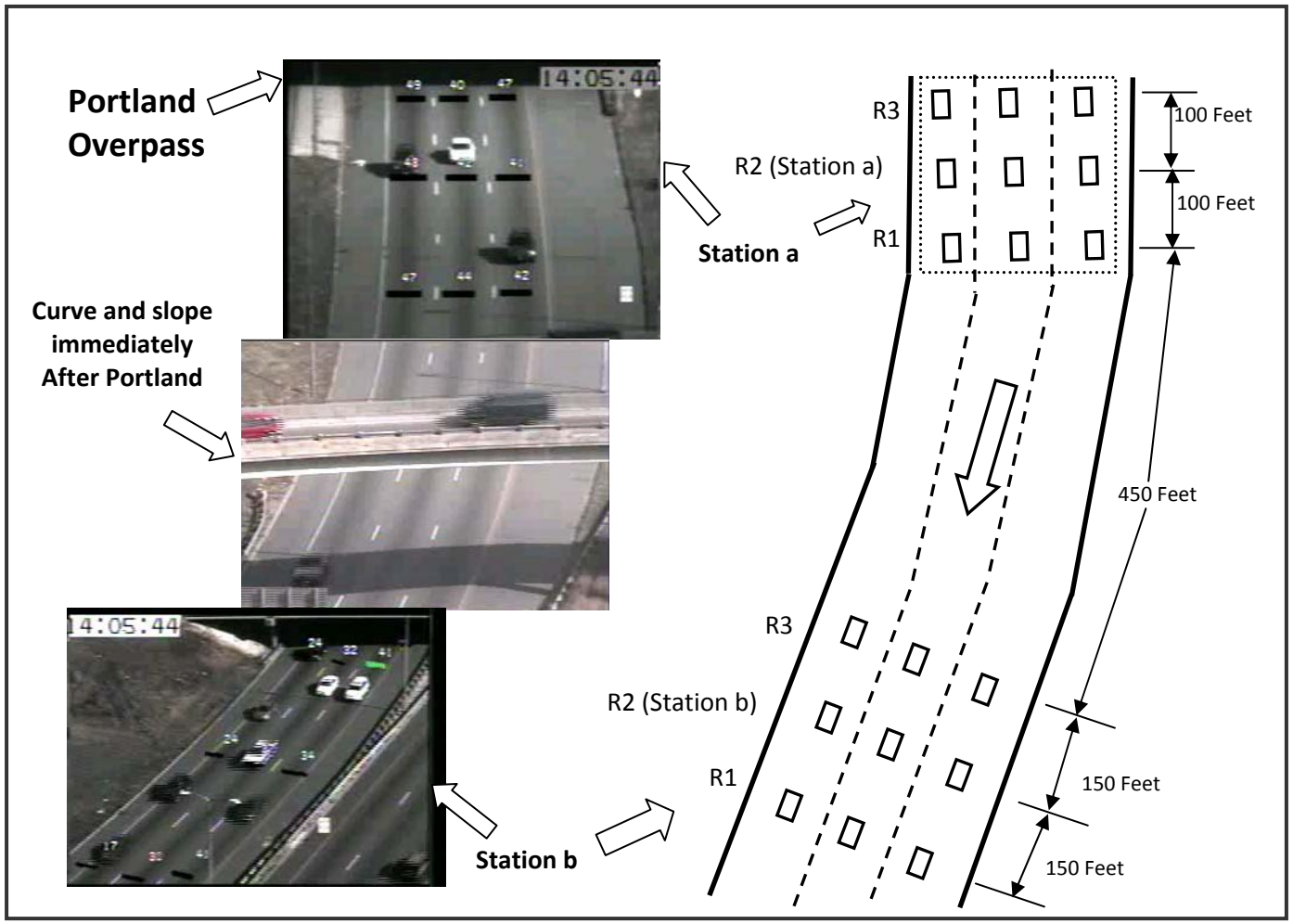

Figure 3. 9 Detector and Surveillance Setup Example.

3.5.2 Mn/DOT Loop Detectors

As mentioned earlier the Mn/DOT has over 4,000 loop detectors installed in the pavement of the Twin Cities freeway network. Loops are installed on each lane of a particular location and are referred to as loop detectors throughout this document while the set comprised of all lane loop detectors in a particular location is termed a Station. Stations are placed approximately every half a mile of roadway. In the case of the selected high crash location there were 5 mainline, 2 exit, and 2 entrance stations measuring volume and occupancy every 30 seconds. Specifically, on the mainline of I-94 westbound the following station data were acquired between January 2003 and January 2005 (refer to Figure 3.3 for street names):

- One 3 loop station on the east side of 11th Avenue, 500 feet before the entrance ramp merge.
- One 4 loop station on the east side of Park Avenue, approximately 300 feet before the 11th Street exit.
- One 3 loop station on the west side of the TH156 (downtown exit/entrance ramps of I-35W).
- One 4 loop station on the east side of La Salle Ave.
- One 3 loop station on the mouth of the Lowry Hill Tunnel entrance.

Unfortunately, as will be seen in subsequent chapters, the loop data were not suitable for the real-time detection of crash prone.

3.5.3 Weather

Information on the weather conditions observed in the crash site was acquired from Mn/DOTs Regional Weather Information Service (RWIS) records. Although detailed precipitation, temperature, and visibility records were available the final analysis reduced this information into categorical variables for reasons that will be explained later in this document.

3.5.4 Crash Information

The original plans for this research required the acquisition of all crash records in the area dating a year prior to the deployment of the surveillance stations and continuing until the end of the analysis. These records were to be used to determine the major types of crashes observed in the area as well as to reduce the work load of examining the video records. Specifically, it was the intention of the researchers to approximate the time of the crash based on the state patrol records, and to locate the appropriate video records and save them for later analysis. This process was deemed necessary since for a year worth of recordings there will be approximate 4300 hours of video with only a few hundred hours of interesting information. Unfortunately, as will be elaborated later, due to the incompleteness of the state patrol records and the discovery of near-misses, the records were not utilized this way and all video recordings were manually observed.

Regardless, the state patrol crash records offered additional information about the drivers' testimony about the causes of the crash.

3.6 TRAFFIC EVENT CLASSIFICATION

Traffic flow presents a number of inhomogeneities that have been measured and identified in the literature. It is beneficial to define them early in this document. The events discussed in this section are observed under certain prevailing flow conditions. Flow conditions can be classified either as congested or uncongested. As stated in the beginning of this chapter, during uncongested or normal flow, upstream conditions control the flow, i.e., disturbances propagate downstream. While when congested, flow is affected by disturbances in downstream conditions and therefore controlled by them. Traffic events producing traffic disturbances include bottlenecks, recurrent congestion, traffic pulses in uncongested flows, compression waves in congested flows, random fluctuations, and incidents [Chassiakos, 1992]. Sensor failure is also an event but is only related to the measurement component of traffic detection systems.

Bottlenecks are formed where the freeway cross-section changes, e.g., in a lane drop (including shoulder elimination), on entrance ramps that have a substantial on-ramp traffic volume, on freeway interchanges, also uphill sections and curves with reduced capacity. Unlike other events, bottlenecks result in longer lasting spatial density or occupancy discrepancies. Bottlenecks are evident during uncongested conditions as areas of lower average speed and higher density but, they become more pronounced during recurrent congestion. The latter is the result of demand increase above the capacity level, resulting in a traffic flow breakdown at the location of the bottleneck.

Traffic pulses are observed in uncongested flows and are created by platoons of cars moving downstream. Such disturbances may be caused by a large entrance ramp volume lasting for a specific duration.

Compression waves occur in congested traffic, usually following a small disturbance and are associated with severe slow-down, speed-up vehicle speed cycles. Waves are typically manifested by a sudden, large increase in occupancy (in the case of loop detectors) that propagates through the traffic stream in a direction counter to the traffic flow as time progresses. Compression waves can be the result of independent disturbances like merging vehicles from a ramp or lane changes due to route-choice decisions. Depending on the flow conditions as well as the initiating disturbance, compression waves might become a periodic phenomenon.

3.7 ADDITIONAL TRAFFIC METRICS

After considering observations during the first year, preliminary time series analysis of single station data, engineering judgment and other similar cases in the literature, it was concluded that the simple speed and headway measurements (from now on referred to as RAW measurements)

were too noisy for pattern recognition. In traffic-flow literature there is a variety of proposed metrics for quantifying the stability and quality of the traffic stream, although they have never been used in a real-time application or generally for identifying crash prone conditions. The detailed nature of the measurements directly obtained from the MV sensors allows derivation of such complex metrics. Such metrics take into account traffic flow fluctuations in space and time, and in some cases describe phenomena like shockwaves with greater detail. These metrics include Space Mean Speed based on measurements collected over a wide area, Standard Deviation of Speed in space and time, Coefficient of Variation of Speed [Snedecor et al., 1967], Standard Deviation of Time and Space Headways, Quality of Flow Index [Greenshields et al., 1961], Acceleration Noise [Herman et al., 1959], Mean Velocity Gradient [Helly et al., 1967], and Traffic Pressure [Phillips, 1979]. These metrics are used in the development of real-time crash-prone condition detection algorithms, presented in chapter seven.

These metrics can be divided into two groups, Temporal and Spatial. This distinction is partly based on the flow characteristics they describe as well as the amount of RAW data required for their estimation; temporal metrics depend on single station (location) raw measurements, while spatial metrics require data from two or more subsequent stations (locations).

3.7.1 Temporal Metrics

The following metrics are temporal in nature because they quantify changes in traffic characteristics over a time interval. Additionally, these metrics are directly defined and derived from point measurements. A moving window technique is employed in the calculation of these metrics from the speed and headway time series.

3.7.1.1 Average Speed ($\overline{S}$)

Average speed is the most commonly used traffic parameter. It is a natural measure of the flow quality and indicates level of service. Mathematically it is defined as the arithmetic mean of the spot speeds and is given by equation 3.1.

$$\overline{S} = \frac{1}{N}\sum_{i=1}^{N} u_i \qquad (3.1)$$

Where,

u_i = Speed of vehicle i

N = Total Number of speed observations

Based on single MVP detector measurements, the calculation of this metric is accomplished through the employment of a Moving Average filter over the individual vehicle speed time series. The size of the moving average window is defined in number of vehicles and for the purposes of this research several window sizes were tested varying from 15 to 120 vehicles. This Moving Window approach is employed in all the metrics found in this work.

3.7.1.2 Coefficient of Variation of Speed (S_{CV})

The coefficient of variation of individually measured speeds is defined as the standard deviation of the speed sample divided by the samples mean value. It is a measure of dispersion and is considered as a means of "normalizing" the standard deviation. A high value of S_{CV} indicates higher variability and vice versa. This metric was included to assist in capturing the effect of rapid speed changes (due to waves, braking maneuvers) associated with pre-crash conditions. Mathematically, S_{CV} is given by

$$S_{CV} = \frac{\sigma}{\overline{S}} \tag{3.2}$$

Where, S_{CV}= Coefficient of Variation of Speed

$\overline{S}$ = Average Speed

σ = Speed standard deviation

S_{CV} is calculated using a similar moving window approach as previously described. Standard deviation of speed (σ) refers to the distribution of individually measured speeds on a certain location, and is used to measure the variability of speed. The standard deviation of speed at any location on a roadway is calculated through the following equation:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(u_i - \mu)^2}{N}} \tag{3.3}$$

Where,

u_i = speed of vehicle i

N = Total Number of speed observations

Coefficient of variation is more commonly used than the variance because the mean and standard deviation tend to change together. For example, a standard deviation of 3 miles per hour (4.82 km/h) is considered small if the average speed was 60 miles per hour (96.5 km/h) (coefficient of variation = 0.05) but larger if the average speed was 15 miles per hour (24.14 km/h) (coefficient of variation = 0.20). Empirical studies have shown that coefficient of variation is normally in the order of 0.08 to 0.17 [May, 1990].

3.7.1.3 Traffic Pressure

Traffic Pressure is defined as the product of density and speed variance [Phillips, 1979]. It is a measure of smoothness of traffic flow and combines the properties of both density and speed variance in one metric. Simultaneous high values of both speed variance and density are generally undesirable. Mathematically traffic pressure is given by:

$$P_T = S_V * K$$

$$..... = \left(\frac{\sum_{i=1}^{N} (u_i - \mu)^2}{N-1} \right) \qquad (3.4)$$

Where,

P_T=Traffic Pressure

S_V= Speed Variance

K=Density

Density (K) is defined as number of vehicles occupying a mile of roadway at a particular instant. Since density is by nature a spatial variable it is commonly estimated from single point measurements. For an assumed small section of roadway with length dx the instantaneous density is the number of vehicles over that length N/dx. For a given time interval the average density will be the average number of vehicles over the length. The average number of vehicles inside dx in a given time period t can be estimated as follows. The probability that any given vehicle is inside dx during the interval t is $P(i|dx)=t_i/t$, where t_i is the travel time of the vehicle on dx. The mean number of vehicles inside dx equals the sum of probabilities that N vehicles will be inside dx during time t. Therefore, the density can be written as:

$$K = \frac{\sum t_i}{t * dx} = \frac{TTT}{t * \Delta x} \qquad (3.5)$$

Where,

K = density

t_i = travel time of vehicle i.

t = time

The **Total Travel Time** is the summation of travel time for all vehicles in time t. It is calculated using the product of average volume and average travel time and can be written in the form of equation as:

$$TTT = \mathbf{V} * \overline{T_t} = \frac{V * \Delta x}{\overline{u_t}} \qquad (3.6)$$

Where,

TTT= Total travel time in time t

V=Average volume in time t

$\overline{T_t}$ = Average travel time = $\Delta x / u_t$

$\overline{u_t}$ = Time Mean Speed

Substituting the value of *TTT* from Equation 3.6 into 3.5, the final form of density is given by equation:

$$K = \frac{V}{t * \bar{u}_t} \tag{3.7}$$

3.7.1.4 Kinetic Energy

Kinetic energy of the traffic stream is analogous to the classic kinetic energy as defined in physics. It is defined as the energy of motion of the traffic stream. The energy conservation law states that within the confines of a given system or control section, energy is neither created nor destroyed, although it may appear in different forms. Not extending the analysis to chemical, electrical, or atomic forms of energy, two forms can be identified in the traffic stream, kinetic energy and internal energy. The internal energy or lost energy [Drew, 1968] is either lost or erratic motion because of adverse geometrics and traffic interactions. Later in this section a measure of jerkiness of the driving in the stream is given by acceleration noise. Drew supported that acceleration noise represents the internal (undesirable) energy.

As the kinetic energy decreases, the force of friction increases due to erratic motion of vehicles. The lost energy from the traffic motion is converted into a less valuable form of internal energy, due to the effects of friction (adverse geometrics and traffic interactions).

Kinetic energy of a traffic stream equals density times the square of average speed.

$$E_K = a K \bar{u}^2 \tag{3.8}$$

Where,

a = kinetic energy correction factor, a dimensionless constant, here equals 1.

K = density of the traffic stream.

$\bar{u}$ = average speed of the stream.

Density is calculated using the same method as in calculation for traffic pressure and average speed is calculated using the speeds of vehicles passing a particular point using a moving average.

3.7.1.5 Coefficient of Variation of Time Headway

Time Headway is defined as the elapsed time between the front of the leading and following vehicle passing a point on the roadway. Time headway between vehicles is an important metric that affects safety, level of service and capacity of a roadway.

Mathematically it is given by:

$$h_t = t_{i+1} - t_i \tag{3.9}$$

Where,

h_i = Time Headway that exists between leading vehicles V_i and following vehicle V_{i+1}

t_i= Time at which the sensor detects the leading vehicle

t_{i+1}= Time at which the sensor detects the following vehicle

It is sometimes recommended to leave at least 2 seconds time headway when following a lead vehicle. If the road surface is wet or slippery, it is advised to double the "two second rule." Different studies have investigated the relationship between time headway and crashes. For example, Michael, Leeming, and Dwyer [May, 1990] found that nearly 50% of drivers follow with headway of less than 2 seconds. Initial investigation during this study confirmed similar tendencies. Such a large percentage of short headways reduce the utility of the measurement as a crash precursor. Therefore, for the purposes of this research the actual time headways, although considered, are not as important as the magnitude and rate of their change may be. The additional metric in this case is the coefficient of variation of time headways in a group of n vehicles.

Mathematically it can be written as, $$Hcv = \frac{\sqrt{\sum_{i=1}^{N}(h_i - \bar{h})^2}}{\frac{N-1}{\bar{h}}} \qquad (3.10)$$

Where,

H_{CV} =coefficient of Variation of time headway

$\bar{h}$ = mean time headway

3.7.2 Spatial Metrics

For the purposes of this research, spatial metrics refer to measurements derived from the trajectory information of a single vehicle over a specified section of roadway. Such measurements are hard to collect. However, as indicated in the literature, metrics belonging in this category can be very descriptive of the state of the flow. Three such metrics were identified, Acceleration Noise, Mean Velocity Gradient and Quality of Flow Index. These can be measured by following a vehicle using an aerial speed-monitoring device or by using an instrumented vehicle that reports speed at periodic intervals of time. In this study, such measurements were not possible with the available sensors because even the machine vision sensors have a limited range of operation and do not track vehicles for distances long enough to be meaningful.

Two different methodologies are proposed in estimating the spatial metrics. The first was based on approximating speeds at multiple stations. It is assumed that the vehicle at the current station will emulate the behavior (speeds) of the vehicles at the downstream stations i.e. to estimate the speed profile of a vehicle currently passing the upstream station; the speeds of the most recent vehicles that passed the downstream stations are used. For example, when a vehicle passes

(current) over station 1, shown in Figure 3.10, the speeds of the most recent vehicles that arrived at station 1 and the downstream stations 2, 3, 4, 5 and 6, are used to estimate the speed profile of the current vehicle at different intervals as it travel across the roadway. After the speed profile of the current vehicle has been acquired, it is used to estimate the multiple station metrics.

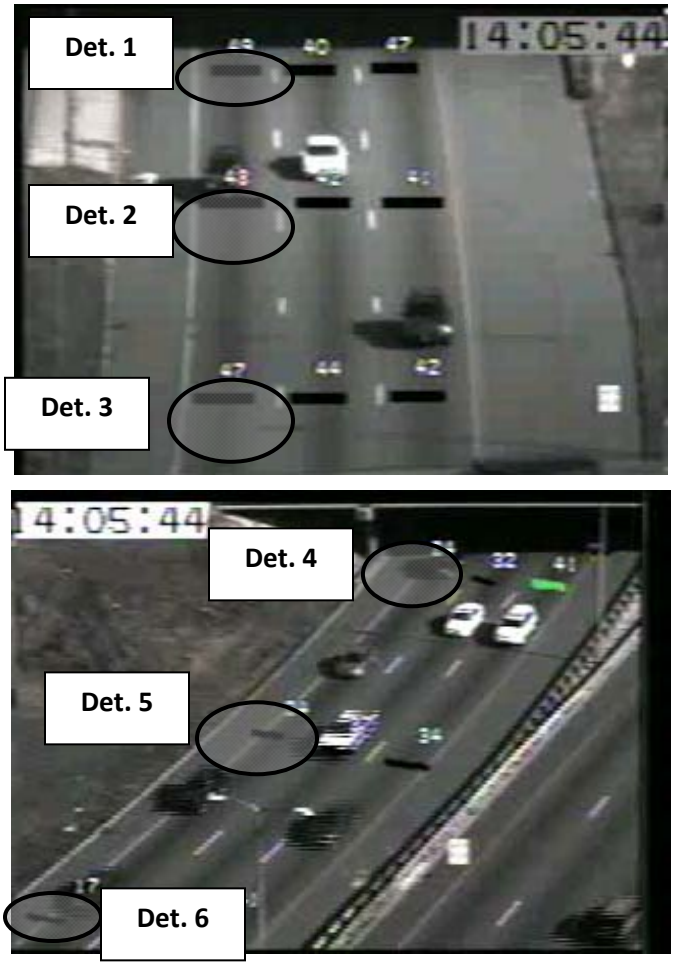

Figure 3. 10 Successive detector Locations used at two zones to estimate speed profile

However, to implement this methodology successive speed measurements must be consistent. During a short test period, individual vehicle were tracked on a roadway under free flow conditions. It was found that the speeds were inconsistent between successive stations even within a distance of 150 feet 945.72 meter). Due to this nonlinear bias between successive detectors, it was not possible to estimate the spatial metrics from multiple stations.

The second proposed methodology is using single station measurements (speed and headway) and is based on the following assumptions [Subramanian, 2001]:

- The length of the vehicles is measured as they pass through a point and assuming that the space headway between successive vehicles does not change, it is possible to identify the vehicles that can fit in a one mile section of roadway.
- The speeds of these vehicles are measured when they enter this hypothetical mile and it is assumed that they will remain constant thereafter.
- For the purpose of estimating the changes in speed of a single vehicle in the space of a mile, it is assumed that it will match the speeds of its predecessors inside that hypothetical mile.

Thus whenever a vehicle passes over the sensor, the speeds of all the vehicles prior to this vehicle that lie within an estimated distance of approximately a mile are counted to estimate the spatial metrics.

3.7.2.1 Acceleration Noise (A_N)

Acceleration Noise is defined as the standard deviation (root-mean-square deviation) of accelerations of a vehicle. It is a measurement that quantifies the smoothness of flow in a traffic stream by estimating the amount of individual acceleration dispersion from the mean acceleration. It depends upon three basic elements of the traffic stream, namely, (1) the driver, (2) the road, and (3) the traffic condition [Subramanian, 2001]. A_N varies with the amount and frequency of acceleration and deceleration. The more violent and frequent these maneuvers are the higher is the noise. Previous studies have suggested that there is a correlation between acceleration noise and crash potential. High values of A_N, indicates a potentially dangerous situation and lower values represent smoother flow. Different experiments suggest that 0.7 ft/sec^2 (0.2 m/sec^2) is a low value and 1.5 ft/sec^2 (0.45 m/sec^2) is a high value [Drew, 1968]. A_N is estimated using the accelerations or decelerations of a vehicle at different intervals as it travels across a roadway.

Mathematically, A_N is given by

$$A_N = \{\frac{1}{T}\int_0^T [a(t_i) - a_{ave}]^2 dt\}^{\frac{1}{2}} \qquad (3.12)$$

Following is the equation developed by Jones and Potts (1975) for approximating acceleration noise:

$$A_N^2 = \frac{1}{T}\int_0^T [a(t_i)]^2 dt - \frac{1}{T}\int_0^T 2a(t_i)a_{ave}dt + \frac{1}{T}\int_0^T (a_{ave})^2 dt$$

$$\Rightarrow A_N^2 = \frac{1}{T}\int_0^T [a(t_i)]^2 dt - 2a_{ave}\frac{1}{T}\int_0^T a(t_i)dt + (a_{ave})^2$$

$$\Rightarrow A_N^2 = \frac{1}{T}\int_0^T [a(t_i)]^2 dt - 2(a_{ave})^2 + (a_{ave})^2$$

$$\Rightarrow A_N^2 = \frac{1}{T}\int_0^T [a(t_i)]^2\,dt - (a_{ave})^2$$

The value of AN can be estimated by approximating [41],

$$A_N \approx \left[\frac{1}{T}\sum_{i=0}^{T}\left(\frac{\Delta u}{\Delta t}\right)^2 \Delta t - \left(\frac{u_T - u_o}{T}\right)^2\right]^{1/2} \qquad (3.13)$$

Where,

u_i= Speed of vehicle i

u_o , u_T = Speed of first and last vehicle in the hypothetical mile

$T = t_T - t_o$, where to, t_T arrival times of first and last vehicles at the measuring point

3.7.2.2 Mean Velocity Gradient (MVG)

Helly and Baker (1965) modified the acceleration noise into a new parameter, Mean Velocity Gradient, a measure of the mean change in velocity per unit distance of the trip. Baker observed that the acceleration noise alone does not describe the quality of the trip and is not always very sensitive to changing traffic conditions in a very congested environment, because it fails to distinguish adequately the differences between trips at very low speeds in dense traffic and faster trips with the same acceleration noise. Mean velocity gradient attempts to correct this deficiency by also considering the mean speed. Mathematically, MVG is the ratio of acceleration noise to mean velocity i.e.

$$MVG_i = \frac{A_{N_i}}{\bar{u}_i} \qquad (3.14)$$

Where A_{N_i} = acceleration noise and

$\bar{u}_i$ = mean velocity of vehicle i

The aforementioned methodology for estimating spatial metrics from point measurements was used to estimate MVG for an individual vehicle. The individual vehicle MVGs are then used to estimate the average MVG. A moving window of different sizes is used to calculate the average MVG as discussed for the temporal metrics. The average MVG is estimated as:

$$\text{MVG} = \frac{\sum_{i=1}^{N} MVG_i}{N} \qquad (3.15)$$

Where, MVG_i = average mean velocity gradient of vehicle i,

N = number of vehicles in the hypothetical mile.

3.7.2.3 Quality of Flow Index (Q_F)

Formulated by Greenshields (1961), QF attempts to quantify how good the traffic conditions are at a particular roadway. It is calculated from a vehicle's average speed, absolute sum of changes in speed and number of speed changes in a mile. The amount of speed changes and the frequency of speed changes are undesirable factors, which irritate drivers and increase the cost of operation and are inversely proportional to the quality of flow. Overall speed determines travel time and thus it is proportional to the quality of flow index. According to Greenshields, estimation of QF requires the use of approximated speeds of each vehicle at different intervals as it travels across a roadway.

Mathematically, quality of flow index is given by:

$$Q_F = \frac{k\bar{u}}{\Delta u}\sqrt{f} \tag{3.16}$$

Where,

$\bar{u}$ =average speed in mph.

Δu = Absolute sums of speed changes in a mile

f=number of speed changes in a mile

k= constant of 1000 when speed in mph and length of section one mile.

In this research Q_F is estimated for individual vehicles and then the average of individual vehicle Q_Fs is used as a metric. The average Q_F is estimated as:

$$Q_{OF} = \frac{\sum_{i=1}^{N} Q_{F_i}}{N} \tag{3.17}$$

Where,

Q_{OF} = Average Quality of Flow Index

Q_{Fi} = Quality of flow index of vehicle i,

N=number of vehicles in the hypothetical mile.

3.8 CONCLUSIONS

In this chapter the preliminary investigation, functional specifications, design, and final deployment of a new advanced detection and surveillance station were presented. These stations, located in the highest crash freeway location in the state of Minnesota, will supply this study with a variety of real-time measurements and crash video recordings. The video records will be utilized in the qualitative analysis described in the following chapter while the detailed traffic

measurements will be analyzed and forma the test bench for the development of a crash prone conditions detection algorithm.

4 Crash Traffic Information Analysis

4.1 QUALITATIVE ANALYSIS

Analysis of the large amount of traffic information collected during this study was addressed in two ways: qualitatively and quantitatively. The qualitative analysis discussed in this chapter deals primarily with the information extracted from the visual observation of the video records of crashes and the general periods of crash and non-crash traffic conditions.

This type of analysis depends on the observer's judgment to identify the different events and conditions leading to each crash. Additionally, it offers the opportunity to extract more detailed information on the traffic composition at the time of each crash as well as detailed environmental conditions. In previous studies the weather and pavement conditions (rain, snow, etc.) were extracted from the police reports or from the closest weather station; in this study this information was collected by directly observing the video records, reducing ambiguity. Finally, the qualitative analysis helped to weed out of the dataset crashes which, from the sequence of events, were obviously not traffic related (e.g., one case of flat tire, causing the vehicle to decelerate and a second case in which a truck stopped for no apparent reason).

The quantitative approach involved time series and statistical methodologies applied to the collected traffic measurements. During this stage specific traffic patterns were identified and their features quantified i.e., shockwave boundary speeds and frequency. Additionally, the amount of noise due to detector or other errors was quantified and the foundation set for the design of digital filters aimed at improving data quality and boosting the performance of subsequent models. Both analyses informed the researchers about the sequence of events leading to crashes and helped them identify new, previously unobserved events.

4.1.1 Video Processing

The video footage from the site was central to our qualitative analysis. It was digitally stored and downloaded at the ITS laboratory of the University of Minnesota. Later the video was played back in order to identify the various traffic events. Originally, by agreement with the Minnesota State Patrol, the crash reports for the area were acquired at regular intervals. The plan was to use the time and location reported to minimize the effort of finding each crash in the video records. From observation of the video records the original suspicion that not all crashes were reported was quickly verified. In addition to the identification of unreported crashes, it was also observed that in a large number of cases drivers were taking evasive maneuvers during which the vehicle(s) departed from the regular lane and ended up on the shoulder or even the ditch. Since no actual collision occurred, this type of event was never reported. These events were termed "near-misses" and considering the scope of the project (identify crash prone conditions) it was decided that near-misses should be treated as crashes during the analysis. This subject will be presented

in greater detail in subsequent sections. The net result of these discoveries was that the original plan to visually analyze only the relevant video records based on the state patrol reports would have excluded a great number of important cases from the analysis. Therefore, all collected videos (2200 hours) were visually inspected in order to identify as many events and traffic conditions as possible. For a year the members of the research team followed a strict schedule and method in the hope of minimizing the loss of valuable information. Specifically, a special digital video player (VirtualDub application by Avery Lee, member of the GNU free software coalition) was used that allowed easy and clear replay of the video at higher than normal speeds both forward and in reverse. This method allowed the observation of one hour of video in less than 15 minutes if no event was observed. In the case an event was observed additional time was spent observing the event sequence several times at normal or reduced speed.

4.1.2 Event Classification

4.1.2.1 Collisions

Regardless of severity, the event in which two or more vehicles collide with each other is considered a crash. Collisions, in this site, can be further divided into rear-ends or sideswipes, based on the type of the collision. All video recordings were scanned to identify collisions; therefore, a number of collisions that were not reported by the police were also included. This is an improvement over previous efforts relying only on police records for the selection of cases. Figure 4.1 shows snapshots captured from crash video recorded at the site.

Figure 4. 1 Snapshots from two Collisions observed at the test site.

Videos were found to be the most accurate way to determine location, exact time of collision and in general the sequence of events that lead to each collision. For practical reasons, only segments of the video containing collisions were finally kept in a video library. These are the full-hour

segments containing each collision, while separately shorter versions, starting one minute before the event and ending approximately 30 seconds after, were also saved for quick reference. Using the video segments, vehicle measurements were extracted during the 30 minutes preceding each collision and stored separately.

4.1.2.2 Near Misses

In an effort to avoid a collision, a driver may intentionally or unintentionally steer the vehicle off the road. In such situations drivers perceive that their vehicle is likely to hit the leading vehicle and just slowing down will not be sufficient to avoid collision. In, these cases, if not for the driver's evasive maneuvers these events would have resulted in rear-end collisions. Near misses were defined as sudden braking **and** rapid steering of the vehicle towards the shoulder in order to avoid a crash. According to Arai and Ezaka et al., (2001) sudden braking is characterized by longitudinal deceleration of 6.0 m/s^2 or more, while rapid steering indicates a situation when the vehicle has lateral acceleration of 19.68 ft/s^2 (6.0 $m/s^{2)}$ or more. Such detailed information is not available from the video records; therefore a more tangible definition had to be given to reduce the implied ambiguity. So, for the purposes of this study a near-miss is defined as sudden braking and rapid steering of a vehicle resulting in **complete departure** from the original lane of travel and entry onto the shoulder.

Figure 4.2 shows snapshots captured from two near-miss video recordings observed at the site. Traditionally, safety studies are based only on actual collisions and very little or no attention is given to near-misses. Near-misses are not reported to the police since they do not result in a collision. Therefore, normally, no action is taken to investigate their causes or to prevent recurrence. The main goal of this study is to identify crash-prone conditions i.e., traffic conditions that favor the occurrence of crashes (increase crash probability). Crash non-occurrence does not imply non-crash-prone conditions while it is hypothesized that crash occurrence accompanies a high probability of crash-prone conditions. Taking this further, since near-misses are the result of individual driver's last minute actions, it is further hypothesized they also occur disproportionally during, or are the result of, crash-prone conditions. To test this assumption, near-misses were treated as crashes. Furthermore, from this point forward, unless specified differently, mention of crash cases implies the joint group of collisions and near-miss cases. In later sections describing the quantitative analysis, attention is given to evidence corroborating or refuting the aforementioned hypothesis.

Figure 4. 2 Snapshots from two Near Misses observed at the test site.

4.1.2.3 Crash-Free Conditions

It is important to consider crash-free traffic behavior as a reference condition in order to understand the factors that are observed during crashes. In this study crash-free conditions are primarily selected from periods where neither a collision nor a near-miss was observed. This definition has its dangers. Since the ultimate goal is to detect crash-prone conditions, we want to define periods where no crash-prone conditions were identified. Considering that a-priori the actual nature of crash-prone conditions is not clear, it is impossible to decisively identify them by just looking at the video. A crash, even a near-miss, is the combined result of many causal factors including, to a large extent, human factors. Therefore, one can argue that, despite all the necessary crash-prone traffic conditions being present, drivers may still avoid collisions or even near-misses. For the purposes of this research, as mentioned in the introduction, we define as crash-prone any set or range of measurable traffic variables values sufficient in raising the probability of a crash. Since a case–control methodology is followed, observed crashes (collisions and near-misses) are selected as cases while controls initially are selected from the aforementioned group of periods where no crash was observed. The selected cases and controls belong in the four categories outlined in Table 4.1. Approximately five hundred 20-30 minute periods were randomly extracted (video and measurements). From these 500 sequences, due to the equipment malfunction described in chapter 3, 277 were useable for subsequent analysis. 10 sequences involved situations where rapid slow down and evasive steering was observed but the vehicle did not enter the shoulder. Visually these cases resembled near-misses but did not fit the strict definition therefore not included in the case set. Based on a preliminary analysis, these cases were considered outliers where not included in the control group either. As can be seen in Table 4.1, three crash cases were also considered outliers and removed from the analysis. These cases involved crashes caused by erratic movements of semi-trailers, movements that had no clear reason based on downstream conditions.

Table 4. 1 Case-control matrix.

	No crash	Crash (Collision or Near-Miss)
Crash prone condition	10	460 (173)
Non-crash prone	500 (267)	3

Videos of the selected periods were stored to enable comparison between crash-free and crash-prone conditions. The date and time of the conditions was also noted for later traffic measurement extraction. Since crashes occur during different periods of the day, the selected crash-free conditions vary with the prevailing traffic conditions. Approximately 50% of these videos were taped during uncongested conditions; while the other 50% were during periods when the freeway was congested and experiencing low average speeds and large queues. The uncongested 50% does not include any sequences where the traffic was free flowing (speeds >60mph (96.5 kph)).

4.1.3 Case Statistics

4.1.3.1 Environmental Conditions

Primarily from the video records but correlated with officially collected measurements, different environmental conditions were identified and their relation with crashes examined. All conditions were translated into categorical variables defining distinct states e.g., dry and wet pavement conditions. The variables considered are visibility, pavement conditions, and sun position relative to the driver.

Visibility Conditions (weather)

During rain or snow conditions visibility is reduced, interfering with the ability to see ahead, see the edge of the road, and observe other vehicles from a safe distance. Therefore, two different visibility conditions were considered: Clear (good) and Poor (due to bad weather such as rain, snow) as seen in Figure 4.3.

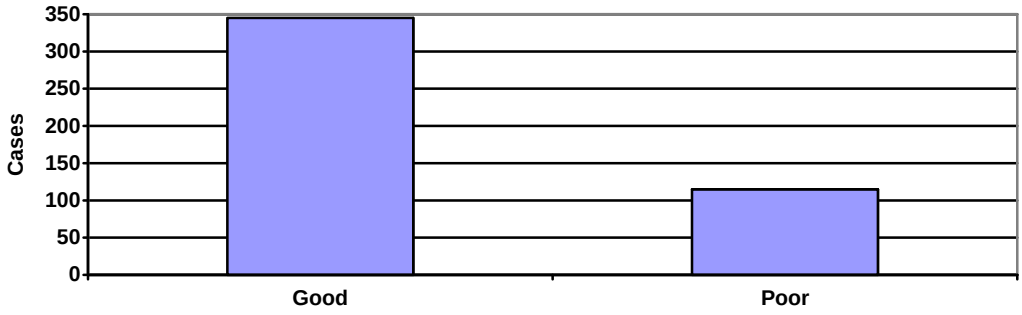

Figure 4. 3 Number of Crashes vs. Visibility.

Since the above chart does not take into consideration the percent of poor visibility days during the study year the following chart might add some clarification on the possible effect visibility conditions have on crashes.

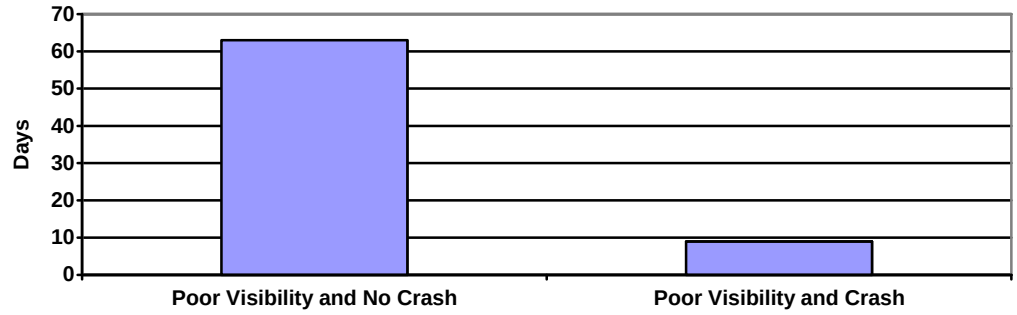

Figure 4. 4 Poor Visibility Days.

Pavement Conditions

The pavement condition affects the surface friction between the tires and road and thus affects the braking distance. In addition, driving behavior changes in locations with wet and icy pavement. Each case was categorized as wet or dry pavement as shown in Figure 4.5. It was observed that less than 10% of the crashes and near-misses occurred during rain/snow or with wet pavement.

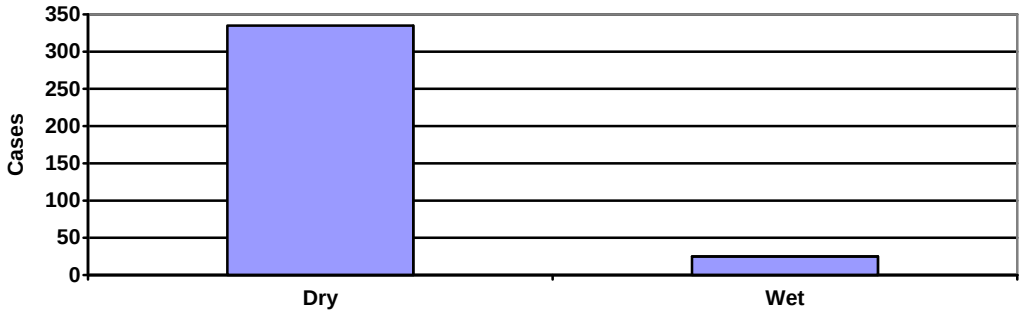

Figure 4. 5 Crashes vs. Pavement Condition.

Sun Position Relative to Driver

Early on in the study, from the video records as well as preliminary test runs in the high crash area, it was observed that in certain periods the sun is facing the driver and causes glare in the windshield. This glare significantly reduces the ability of the driver to notice the leading vehicle's brake lights. In order to investigate this phenomenon further, for each crash case the location of the sun was recorded with respect to the vehicle. This was accomplished by observing the length and direction of the shadows cast by the nearby light poles. Observation of the sun's position was recorded using the following four categories: night (dark), no sun (cloudy), sun at back (180^o) or side (90^o) of driver, and sun at 45^o or at 0^o (front) relative to driver. Crash case numbers for each category can be seen in Figure 4.6.

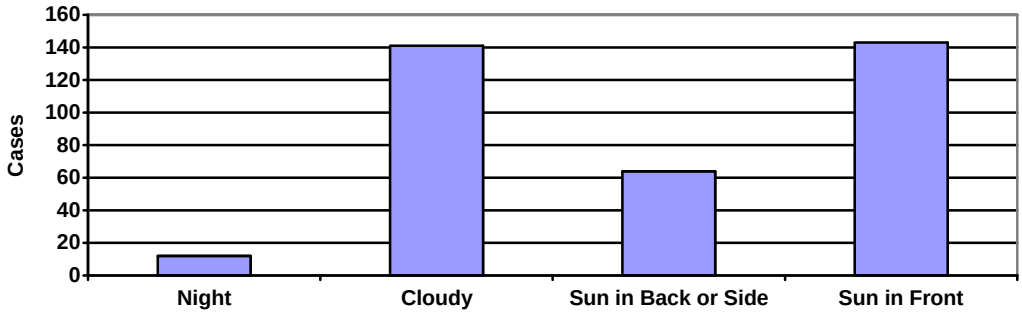

Figure 4. 6 Crash cases vs. Sun position.

From these statistics one can observe a tendency for poor visibility/wet pavement conditions to discourage crashes. This observation is in agreement with previous findings by Edwards (1998) and Lee et al. (2002). The low numbers during low light conditions can be attributed in the fact that the leading vehicles' break lights can be noticed faster even from the drivers' peripheral vision (not looking straight ahead). Similar findings can be found in several human factors studies on crashes [Marshal et al., 1998]. Of course one has to be careful in generally associating

bad weather with low crash risk since crashes are not always caused by inattention and the severity of a crash has been proven to increase during adverse conditions [Khattak et. al., 1998].

4.1.3.2 Crash Location

From the video records it was possible to identify the crash location with great accuracy. In many cases the vehicles involved in a crash continue driving until they find a safe spot on the shoulder to pull over; this fact causes frequent crash location errors in the police reports. Ninety percent of the collisions and 100 percent of the near –misses occur in the right lane. As will be discussed later in this chapter, the right lane traffic in this freeway section exhibits some unique dynamics.

From the total number of cases, 18% took place upstream of Portland Avenue (leftmost overpass on section B, Figure 3.3) while 80% took place **at** Portland and only 2% downstream. Following this observation, the study focused on the measurements collected from two areas (out of six). The first area located between Portland and Chicago and the second approximately 1000 feet (304 meters) downstream of that (Figure 3.9). The location of collisions relative to the Portland overpass can be seen in Figure 4.7.

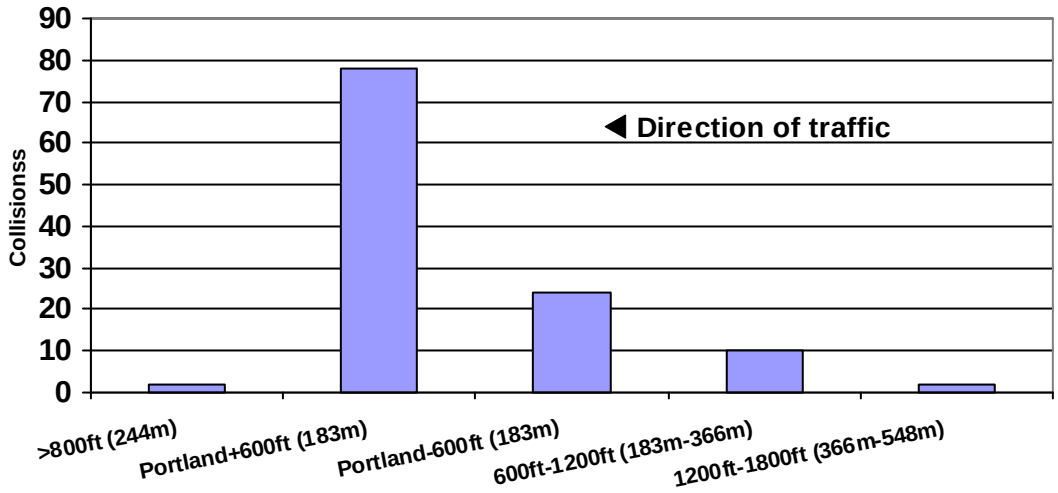

Figure 4. 7 Collision totals per location.

It is interesting to notice that the downgrade and turn of the roadway in this section occurs after the 600 feet (183 meters) where the majority of the crashes take place. At the beginning of the downgrade the driving visibility improves a little since the driver sits in a higher position relative to the vehicle in front. Additionally, for a brief instance it is possible to see traffic ahead as the driver approaches the turn.

From the video records the high crash area is observed to be more than 500 feet (152m) downstream of the lane drop. No crash was observed due to late merge from the auxiliary lane on

the right. The collisions on the middle and left lanes generally occur slightly upstream of the majority of crashes in the right lane; but no farther than 1000 feet (304m) from the Portland overpass.

4.1.3.3 Crash Time

This particular freeway section exhibits high volumes both during the morning and evening peak periods. The evening peak period is considerably longer than the morning and the congestion can extend to well after 7 p.m. Although congestion does not begin until 3 or 4 p.m., as can be seen in Figure 4.8, a large number of crashes occur early in the afternoon before the official start of the peak period (the freeway ramp control begins at 15:00 and lasts until 18:00). The number of crashes during the morning peak is considerably fewer. This can be attributed to the different traffic patterns observed during the morning. These and other general observations will be discussed in the following section.

All the equipment, used in this study, was regularly time synchronized with the official time server of the University of Minnesota. This provided assurance that all location recordings and sensors were synchronized, allowing direct correlation analysis of measurements and visual observations from different locations. Apart from hardware time mismatch there is another source of error that can affect pattern identification. Crashes (collisions and near-misses), although concentrated in specific locations, vary in their exact distance from the sensors. At the speeds where the majority of crashes are observed, two or three car lengths difference can result in 2 to 5 seconds time difference. The more upstream the crash occurs the larger the time difference. Therefore, although the measurements are all synchronized spatially, if one uses the crash time for a base between cases, patterns will be shifted in time. To minimize such an error and increase the pattern identification accuracy, it was decided not to use the exact crash time as the base. From the video records the shockwave generating each crash was identified and traced back in space up to the detector location near the Portland overpass. All cases were synchronized using as a base the time the rear boundary of the shockwave passed the detector. For simplification this corrected time is used as the crash time throughout this document.

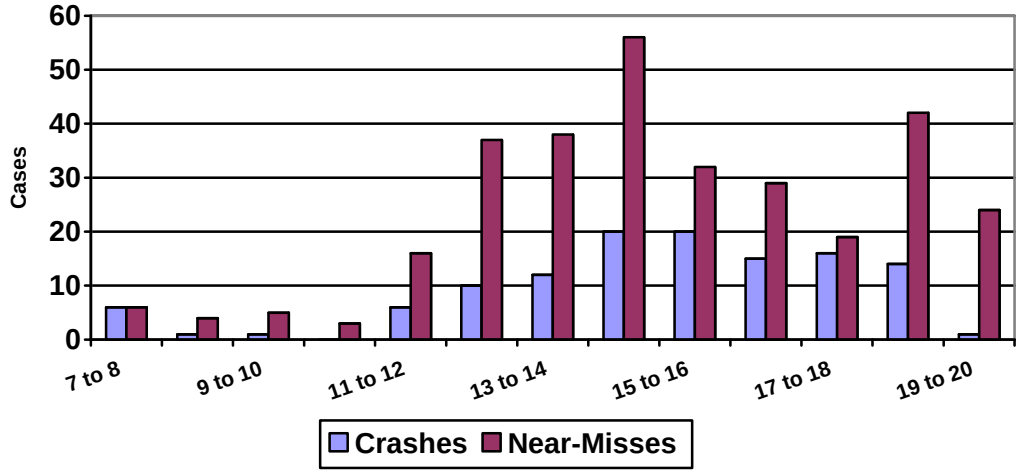

Figure 4. 8 Crashes vs Time

4.1.4 Traffic Dynamics Observed

The study section of I-94 westbound presents a number of interesting traffic dynamics due to complex geometries and complicated O/D paths. In particular, due to the different origins and destinations served through this section a considerable amount of weaving takes place in a relatively short distance. To better understand the different traffic dynamics in this section, the origins and destinations served can be broken down as follows:

Destinations in succession according to traffic direction (Westbound):

1. I-35W southbound exit. This left-hand exit terminates the short auxiliary lane in the first part of the study section. With a very short ramp it leads to the I-35W section that runs in parallel with the I-94 freeway. This is the only access toward the south of the metro and it is heavily used throughout the day.
2. 11th Street. This is a right-hand exit to the local streets of the south downtown area. This exit is usually congested during the morning peak period with a queue frequently extending into the auxiliary lane of the freeway mainline. During the evening peak this exit exhibits very low volumes. The auxiliary, fourth lane on the second part of the study section ends at this exit.
3. Hennepin/Lyndale Avenues exit. This is a multiple destination two lane right-hand exit leading both south, towards the uptown area, as well as towards the southwest area of downtown. This exit exhibits high volumes during both morning and evening peak periods. This exit can also be used as an alternate route towards I-394 or even I-94 west for vehicles prohibited from using the Lowry Hill Tunnel. The auxiliary, fourth lane on the third part of the study section ends at this exit.
4. I-394 exit. Immediately after the Lowry Hill Tunnel there is a right hand, two lane exit towards the I-394 freeway. Although generally a high volume destination, this exit exhibits even higher volumes during the evening peak period.

Origins in succession according to traffic direction (Westbound):

1. I-94 mainline from St. Paul. I-94 is the only freeway directly connecting downtown St Paul with downtown Minneapolis. The only other choice would involve a considerable trip north to I-35W and then south towards Minneapolis. Additionally, the three preceding entrances to I-94 serve the University of Minnesota campus, a considerable producer of work, school, and leisure related trips. During the morning peak period, traffic from this part of the freeway is directed towards the downtown area through the 11th Street and Hennepin exits. During the evening peak a large percentage of the traffic of I-94 is directed towards the I-394 freeway and the west side of the metro region. Especially in the case of the I-394 destination there are no other alternative routes that involve a freeway. The next choice would involve the 494/694 ring road. This puts a lot of pressure on the right lane of the entire section since this is the only one leading towards I-394. Signs to that exit start half way before the study section.
2. Southbound I-35W, TH-55, and local streets combined entrance ramp. This entrance ramp joins the mainline through the fourth, auxiliary lane on the middle part of the study section. This entrance combines traffic from three major origins. First origin is the I-35W southbound freeway. Traffic from this origin is mainly directed towards the I-394 freeway destination and the northwest part of the metro region. Additionally, traffic from North I-35W can also be directed towards the Hennepin/Lyndale exit or the 11th Street exit, but for both of these destinations there are other, alternate, routes that are usually more attractive. In fact no other freeway routes connect I-35W with either I-394 or the north part of I-94. The second origin is TH-55 or Hiawatha Avenue. Considering the other available routes from TH-55 the entrance ramp to the study section is primarily used for traffic that is either directed towards I-394 or to north I-94. It leaves no doubt that these are also the primary destinations of the local roads using this entrance ramp. The result is that, the vast majority of the traffic using this ramp must perform one lane change before the drop of the 1,200 feet (366m) long auxiliary lane and one more if they are not headed towards I-394.
3. Northbound I-35W, 3rd Avenue combined entrance ramp. Approximately 1,500 feet (457m) after the drop of the aforementioned auxiliary lane, another one begins at the combined entrances of I-35W northbound and 3rd Avenue downtown, ending approximately 2,000 feet (609m) later at the Hennepin/Lyndale exit. The traffic using the I-94 flyover, the name given to the ramp coming from I-35W because it passes above I-94, will most likely not exit on Hennepin since an alternate route towards the downtown area is available and is considerably shorter. In contrast, the only alternate route to I-394 or north I-94 will require traversing the entire downtown area. Traffic from 3rd Avenue, being already in the Downtown area will probably not be directed towards the I-394 or the Hennepin/Lyndale exit. Similar to the previous auxiliary lane, the traffic using this one primarily will have to make one or even two lane changes before the tunnel entrance.

From reviewing thousands of hours of video records of the study section, a number of interesting traffic patterns have been observed. For most of these observations corroborating evidence based

on measurements will be presented in the next section. Specifically, the following observations added to the general knowledge of the study section:

- In the majority of crashes observed, 98% to be exact, a compression wave was observed propagating backwards. Originally it was assumed that these waves are caused by the I-394 exit, since it is known that it generates congestion in the tunnel. When the video coverage was extended beyond 3rd Avenue, along with data from the detector measurements, it was confirmed that the section between that entrance and the tunnel exhibits lower densities than the traffic in the tunnel or the merge area of 3rd Avenue. Additionally, no significant number of compression waves were observed traversing this section. After further observation it was evident that the bottleneck causing the compression waves is the merge immediately after the 3rd Avenue entrance ramp. Evidence supporting this observation can be found further on in this chapter.
- Furthermore, it was also observed that the auxiliary lane between 3rd Avenue and the Hennepin/Lyndale exit is underutilized by the merging traffic. There is an obvious tendency by the drivers using the I-94 flyover to merge too quickly to the right regardless of whether there are more favorable merging conditions downstream. In contrast, drivers originating from the 3rd Avenue part of the ramp (right side) are merging much later and with fewer consequences to the mainline traffic.
- Prior to 67% of the crashes and near misses, the middle lane in the Portland Ave area had twice or more the speed of the right lane (left lane was even faster). In all cases where the crash was in the right lane the middle and left had higher speeds. In general, it was observed that the right lane becomes congested earlier that the other two lanes in the three lane section downstream of Portland. This can be attributed to the fact that all the traffic from the first combined origin must first merge in the right lane. In many situations risky merges were observed between the right and middle lanes. In spite of this, very few sideswipe collisions have been recorded. Ninety-eight percent of the crashes are rear end collisions (or begin as such).
- In the majority of the crash cases, the traffic stream speed in the right lane was below 45 mph (72.4kph). This is one reason why no fatal or very serious crashes have been observed.
- Available video records and measurements show the right lane downstream of Portland having very short time headways. Regardless, in a lot of crashes it appeared that enough space was available between the two colliding vehicles for the follower to make a safe emergency stop. One possible explanation is that the driver of the second vehicle was not attentive to the driving task or had his focus elsewhere. Considering the speed difference between the lanes, the lane changing task is considerably harder than normal. Drivers must spend more time looking to the left and behind for an appropriate gap, and less time looking forward.

4.1.5 Crash Scenario

Based on the analysis of the video records as well as several test runs (An instrumented vehicle was driven several times collecting video records from the point of view of the driver, the side, and rear view mirrors) in the study section, a hypothetical scenario of the crash sequence can be formulated. Specifically:

1. Due to weaving between the 3rd Ave/I-94 flyover entrance and the Hennepin exit (Section C, Figure 4.3) a compression wave or queue propagates backwards.
2. Upstream of Portland, at the joint entrance ramp from I-35W, TH-55, and downtown (rightmost area of Section B), heavy weaving produces friction, distraction, and the desire of the drivers to move away from the right lane to the considerably faster moving middle lane.
3. Around the Portland overpass the roadway declines and turns, limiting visibility.

RESULT: Distracted and distressed drivers passing the 11th Street weave fail to notice the sudden stopped traffic caused by the upcoming wave or queue and collide.

4.2 QUANTITATIVE ANALYSIS

In science, measurements are the cornerstone of knowledge acquisition. In earlier sections the qualitative analysis described a real, visual link between the researcher and conditions in the field. Valuable observations were made that shaped the understanding of why, how, and when crashes occurred at this location. However, these are based on the eye of the observer so an attempt was made to explain and rationalize the observed sequence of events. In the remaining of this chapter the goal is to support these observations using measurements. A priori it is neither clear which measurements best describe the sequence of events leading to a crash, nor the way these measurements are to be analyzed. To cope with these problems, we collected measurements from a diverse array of instruments covering the whole study section at various levels of resolution. Several methodologies were employed for the analysis and visualization of these measurements. In the interest of space economy, this chapter contains the cases where the analysis did produce relevant and useful information.

The measurements analyzed originate from two main sources, inductive loop detectors and machine vision sensors. As mentioned in the data collection chapter the loops were located approximately every half mile on the mainline, on every entrance and every exit from the study section. Approximate locations of the loop detectors can be found in Figure 4.9. The loops provide volume counts and lane occupancy values per lane, every 30 seconds.

The machine vision sensors were located on roof tops and collected measurements from the nearby freeway sections (coverage locations can be found on Figures 3.3 and 3.9). As can be

seen in Figure 3.3 several areas of the study section are covered by the Machine Vision (MV) sensors. Based on the findings of the qualitative analysis concerning the crash locations and prevailing traffic dynamics, measurements from two MV locations (Figure 4.9) were selected for further analysis in this chapter.

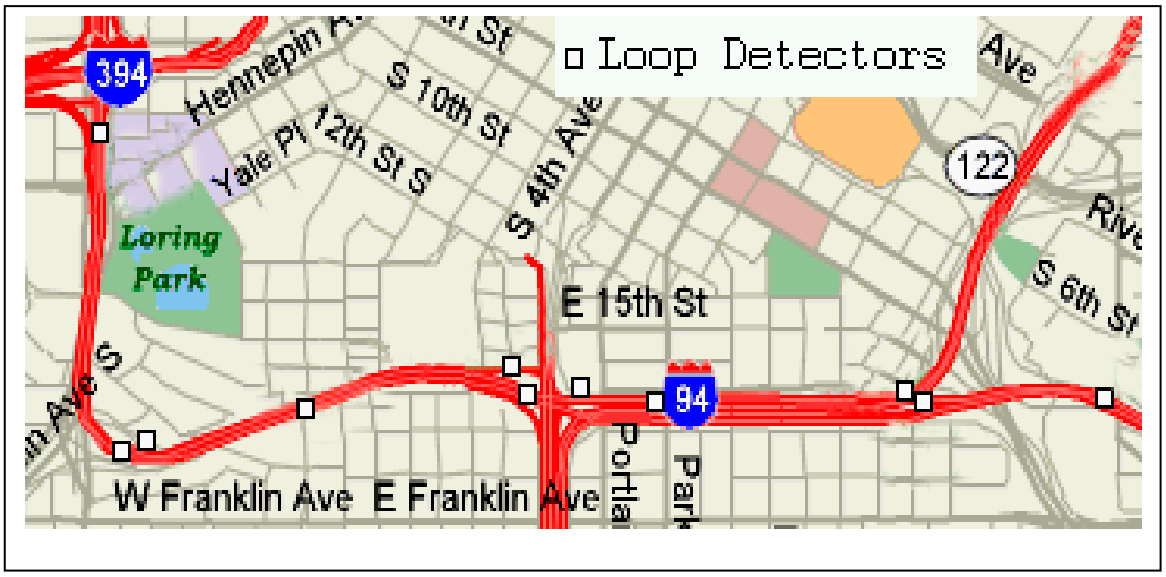

Figure 4. 9 Inductive loop detector locations.

(Source: Mn/DOT live traffic report website.)

4.2.1 Volume/Occupancy Over Space

As noted in previous sections, the breakdown of flow leading to the creation of compression waves was one of the most common patterns observed during pre-crash periods. Unfortunately there was not sufficient video coverage beyond the 3rd Avenue/I-94 flyover combination entrance ramp to securely conclude which location was the bottleneck responsible for the flow breakdown. Furthermore, it was not guaranteed that a bottleneck is responsible for the flow breakdown and is not a result of Phantom Congestion (flow breakdown in locations where there is no road geometry change or other external disturbances. Caused by steady increase in flow density (Kerner, "Physics of Traffic" (2004)). Finding the answer to these questions will be a process of elimination of possible scenarios based on the available measurements. The possibilities for the cause of a flow breakdown are the following:

1. The bottleneck or some other reason for a breakdown is situated after the Lowry Hill Tunnel.
2. The Entrance to the Lowry Hill Tunnel is the cause of the flow breakdown.

3. The area in between the I-94 flyover ramp and the entrance to the tunnel trigger the breakdown possibly due to the lane changes between the right and the auxiliary lanes.
4. The merge area of the I-94 flyover ramp is the bottleneck responsible for the flow breakdown.
5. The breakdown occurs in the space between the entrance ramp and Portland Avenue.

In the proceeding analysis three basic assumptions were made. These assumptions are fundamental in traffic flow theory [Lighthill, et al., 1995].

1. During uncongested periods traffic conditions change in response to events happening upstream of a given position i.e. state changes propagate in the same direction as the traffic stream. The shockwaves that accompany these state changes are called Traffic Pulses.
2. During congested periods traffic conditions change in response to events happening downstream of a given position i.e. state changes propagate in a direction opposite to the traffic flow. The shockwaves that accompany these state changes are called compression waves.
3. Following the breakdown of flow at a bottleneck location, downstream speeds are higher that the ones encountered upstream of it.

Corollary:

- The result of a flow breakdown causing the transition of the state of the flow from uncongested to congested will propagate in a direction opposite to the flow. Chronologically, if there are a number of measuring stations along a roadway, state changes will be recorded in sequence from downstream to upstream with a time lag determined by the speed of the compression waves' upstream boundary.

In the case concerning the section between the I-394 exit and Portland Avenue there are three loop detector stations that can solve the mystery of the location of the flow breakdown.

In the case concerning the section between the I-394 exit and Portland Avenue there are three loop detector stations that can help solve the mystery of the location of the flow breakdown.

Figure 4.10 presents 30 second average speeds per lane (estimated from volume/occupancy measurements) on three loop detector stations. Station 76 is located 550 feet (167 m) downstream of the Portland avenue overpass and marks the boundary beyond which very few crashes have been observed. Station 86 is located midway between the I-94 flyover entrance ramp and the Tunnel entrance and covers four lanes, three mainline ones and the auxiliary that

ends at the Hennepin exit. Finally, station 89 is located in the tunnel entrance. This figure shows the process of locating the flow breakdown responsible for the compression waves associated with crashes. These findings are based on the analysis of more than a year's worth of measurements from the aforementioned stations. The condition encountered during this particular day presented in Figure 4.10 can be considered typical for this study section.

In the first graph of Figure 4.10 a rapid speed drop is observed on the right lane, approximately at 13:45. If this speed drop was the result of congestion propagating backwards from downstream of Station 86 then the speed drop would have appeared at station 86 earlier in time. What is observed from the middle graph is that, for the right lane, the speed drop appears at least 15-20 minutes later than the drop at station 76. Furthermore a speed drop in the tunnel entrance takes place an hour later than the

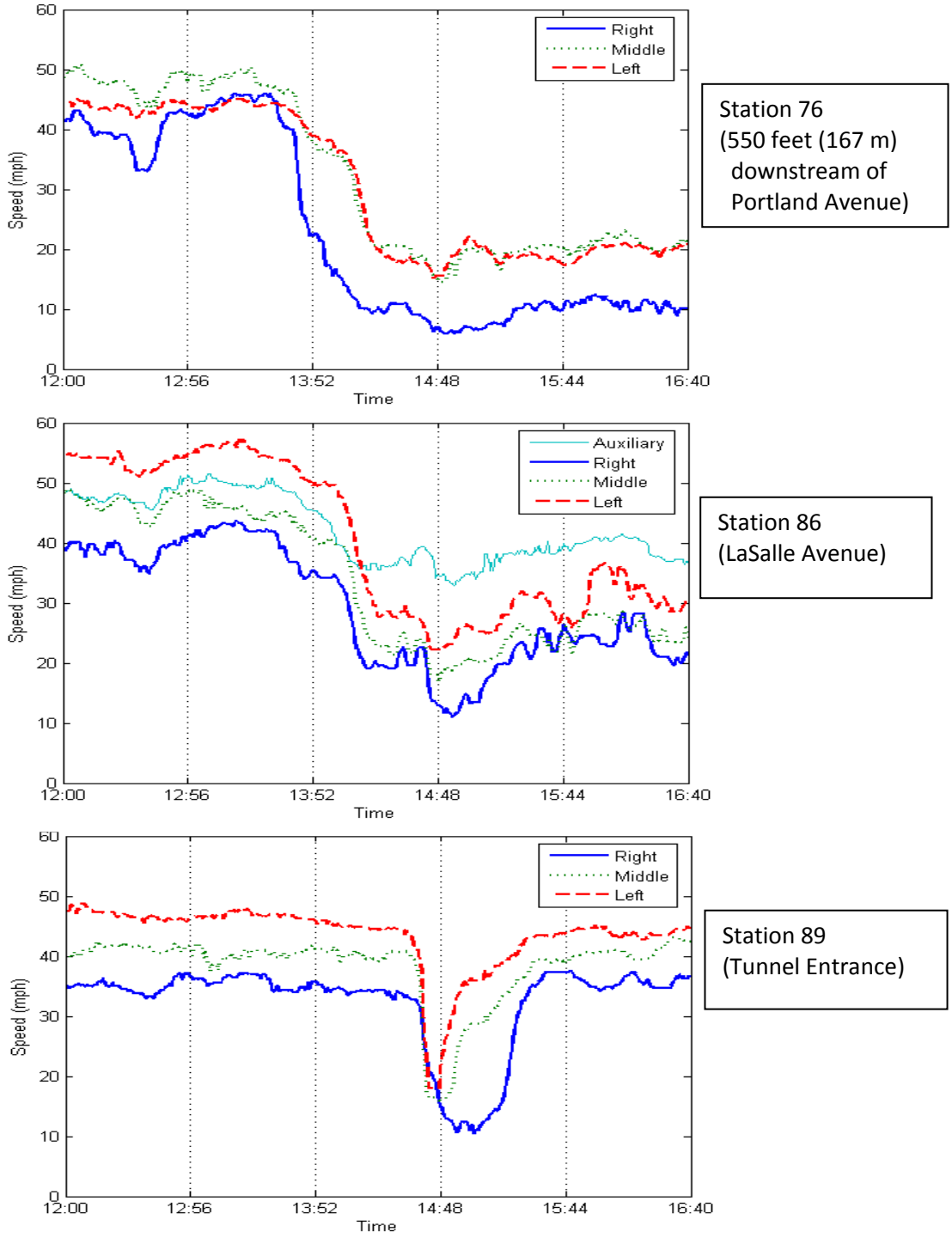

Figure 4. 10 Vehicle speed per lane at different stations. August 26th, 2003.

upstream station. Strictly in reference to conditions in the right lane, the measurements suggest that the flow breakdown causing the congestion at Portland is located somewhere between stations 76 and 86. The other two lanes, middle and left, exhibit a slightly different behavior. In those lanes the speed drop is encountered first on station 86 and approximately 5-10 minutes later on station 76. The speed drop in the tunnel entrance is encountered even later, approximately 45 minutes after station 86. The measurements on the middle and left lanes suggest that the flow breakdown is located somewhere in between station 86 and the tunnel entrance.

It is important to stress the difference in traffic conditions between the right and the rest of the mainline lanes. As suggested by Kerner (2004) in his Three-Phase traffic flow theory, the first state transition is from uncongested to synchronized flow (Kerner in his book *Physics of Traffic* refers to the uncongested state as free flow state. In this work the term free-flow is reserved for the conditions where traffic is so light that it can be assumed that vehicles are driving totally independent of each other. From Kerner's description of free-flow it is believed that the term un-congested better describes the traffic state where the majority of vehicles are able to drive at high speed but their behavior is influenced by interactions). Following this transition, all lanes in the roadway exhibit approximately the same speed changes. The logic behind the above theory is that due to increased volume in the rightmost lane speeds are decreasing. Subsequently, drivers will change to the next lane on the left therefore increasing the volume there and lowering speeds. Shortly after, through a chain reaction, all lanes will be synchronized in regard to speed and density. In the section described in Figure 4.10 it is observed that the right lane acts independently from the other two at least in the area of the Portland overpass. An explanation for this behavior stems from the previous description of origins and destinations in the area. The synchronized flow theory assumes that all lanes lead to the same point. In the case of the study section, the right lane not only leads to a major destination approximately a mile down the road but due to the earlier weaving area it has a lot of pass-through traffic heading for the other two lanes. It's a combination of vehicles that need to be in the right lane and vehicles that are trapped in the right lane because they cannot merge on to the middle. The latter case will be discussed in more detail on subsequent sections. Interestingly, from Figure 4.10, station 86 is a good example of synchronized flow. As illustrated, the speed drop first appears in the right lane, then shortly after in the middle and left lanes. The slope (degree of reduction) of the speed drop is approximately the same in all lanes. This observation suggests that the majority of the lane changing maneuvers due to origin destination relationships have been accomplished upstream of this point.

Another way of verifying the location of the flow breakdown or at least narrow it down in sections between loop detector stations can be accomplished by looking at the volume/occupancy relationship in the three stations. Figure 4.11 presents an example of the Volume-occupancy graph from the right lane of station 76 (Portland). Each point corresponds on a 30sec period and the colors indicate speed regions from >40mph to 30-40mph, 25-30mph, 20-25mph, 15-20mph, 10-15mph, and <10mph (>64km/h to 48-64km/h, 40-48km/h, 32-40km/h, 24-32km/h, 16-

24km/h, and <16km/h). The two leftmost regions belong to uncongested traffic while the rest signal congested conditions. Similar graphs have been examined for the other stations/lanes.

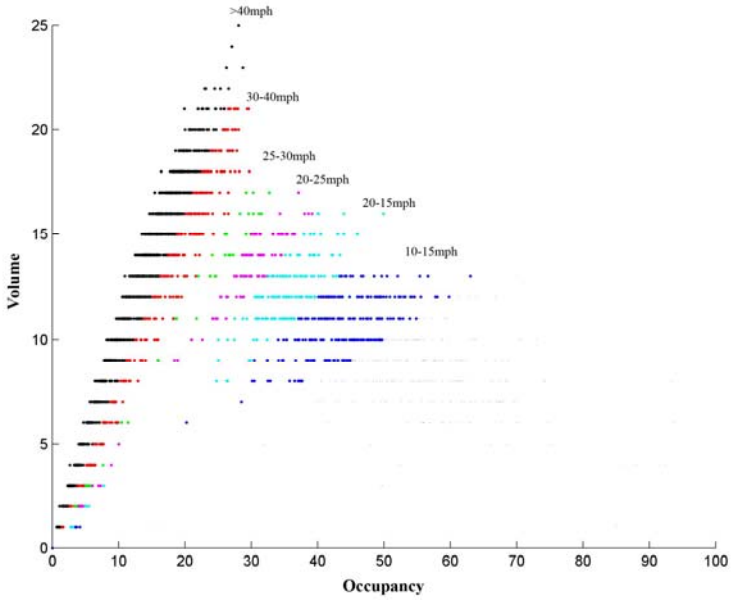

Figure 4. 11 Volume/occupancy graph with speed regions. August 26th, 2003 @ det 76 right lane.

Based on such graphs, traffic was classified as uncongested, mildly congested, and three groups of very congested traffic depending on the duration of the congestion. Table 4.2 presents the results of this classification for the periods 5-10 minutes prior to each crash during the one year of the experiment.

Table 4. 2 Traffic State on the right lane 5-10 minutes prior to crashes (total crashes: 460).

Location / Traffic state	Uncongested	Mild congestion	Congestion	Very congested	Jammed
Detector 356 Portland Ave	121	206	78	54	1
Detector 366 Midway	365	65	17	13	0
Detector 378 Tunnel entrance	91	321	30	11	7

From Table 4.2 it is noted that the conditions in the three locations are different during the periods preceding the majority of the crashes. Specifically, while the Portland Avenue area experiences congested conditions varying from mild to severe, downstream midpoint to the Tunnel is generally uncongested, and the entrance to the tunnel usually exhibits mild congestion. Another interesting observation is that prior to 121 crashes, the Portland Avenue exhibited uncongested conditions. This leads to the belief that a crash can occur immediately after (temporally) the flow breakdown which affects the selection of a crash prone condition detection method.

4.2.2 Time Series Analysis

From visual analysis of the surveillance video a number of possible crash causes were identified. In addition specific traffic patterns appeared suspiciously frequently during pre-crash periods. Compression waves are a prime example of one of these traffic patterns. Regardless of other causes, 98% of the crashes and near-misses involved a compression wave. Qualitatively this information is useful. Are all compression waves indicative of crash prone conditions? Considering the frequent nature of such traffic patterns, if the answer were yes, crashes would not be a rare event and hundreds would happen in a single location everyday. It would be more accurate to say that compression waves of certain magnitude are dangerous and therefore indicators of crash prone conditions. With the help of time series analysis techniques, the wave characteristics are quantified to identify their relationship with crashes. Additionally, time series analysis can decompose the traffic measurements to identify additional patterns correlated to pre-crash conditions. Finally, having at least identified which parts of the measurements contain useful information, the rest (noise) can be removed or reduced with the use of digital filters. The design of these filters is in effect based on the results of the time series analysis.

4.2.2.1 The Signals

Time series are defined as a series of measurements taken at consecutive points in time. These series are traditionally referred to as signals. Signals usually describe or measure a specific physical quantity e.g., speed at a point. In this study the sensors provide several signals, measuring different quantities. The loop detector stations provide the volume and time occupancy per lane every 30 seconds. The MVPs provide the speed of every individual vehicle, time stamped with the arrival time at the detector, the vehicle's length, as well as measurements of volume, speed, and time occupancy aggregated over variable time periods. Individual vehicle time headways can be extracted from the arrival time stamps.

Based on preliminary analysis as well as prior traffic measurement analysis studies found in the literature, volume either from loops or from the MVPs is not a particularly informative measurement [Chen et al., 2001]. On the other hand, speed is an accurate and fast indicator of the state of traffic. Two signals were selected for analysis, individual vehicle speeds from the MVPs and 30-second time occupancy from the loop detectors. Aggregated measurements from the MVPs were not analyzed since any characteristics in them will be also included in the root signal (individual speeds).

4.2.2.2 Analysis Methods

There are two main goals in time series analysis: (a) identifying the nature of the phenomenon represented by the sequence of measurements, and (b) forecasting (predicting future values of the time series variable). The latter is irrelevant to this work. Accomplishing the former requires that the pattern or patterns in the signal (time series) are identified and more or less formally described. Two major methods are available for decomposing a signal and quantifying patterns contained within: *Cross-Correlation* and *Spectral* analysis. In the case of a signal originating from one source the first method is usually referred to as auto-correlation, while in the case of two signals being simultaneously analyzed the two methods are refer to as cross-correlation and cross-spectral analysis. Actually auto-correlation is a special case of cross-correlation.

4.2.2.3 Traffic Event Characteristics in Time and Frequency

As described in this chapter, traffic flow presents a number of inhomogeneities that have been measured and identified in the literature. Traffic events like bottlenecks, recurrent congestion, traffic pulses in uncongested flows, compression waves in congested flows, random fluctuations, and incidents have specific identifiable characteristics in their time and frequency domain representations [Chassiakos, 1992].

Bottlenecks are formed where the freeway cross-section changes, Unlike other events, bottlenecks result in longer lasting spatial density or occupancy discrepancies. Bottlenecks are evident during uncongested conditions as areas of lower average speed and higher density but, they become more pronounced during recurrent congestion. The latter is the result of demand increase above the capacity level, resulting in a traffic flow breakdown at the location of the bottleneck.

Traffic pulses occur in uncongested traffic. In the case of traffic pulses the cross-correlogram will show a high negative lag relative to the speed of the traffic stream. The lag is the travel time of the pulse from the upstream sensor to the downstream one (upstream leads). Unless the traffic pulse is periodic, like the one created due to a signalized intersection upstream of an un-metered entrance ramp, there will be no clear indication in the spectrogram. The same applies for the auto-correlogram.

Compression waves occur in congested traffic. The pattern observed in the cross-correlogram is similar to the one from the traffic pulse with the difference that the lag now is positive since the event is first observed in the downstream sensor. The lag represents the travel time of the wave between sensors and if the distance is known the speed of the wave can be measured. Depending on flow conditions, as well as the initiating disturbance, compression waves might become a periodic phenomenon and therefore appear in the spectrogram as low frequency spikes. Considering the randomness of the compression wave causes, this is a rare phenomenon. More

often compression waves are identified in the spectrogram as high magnitude regions between zero and 5mHz.

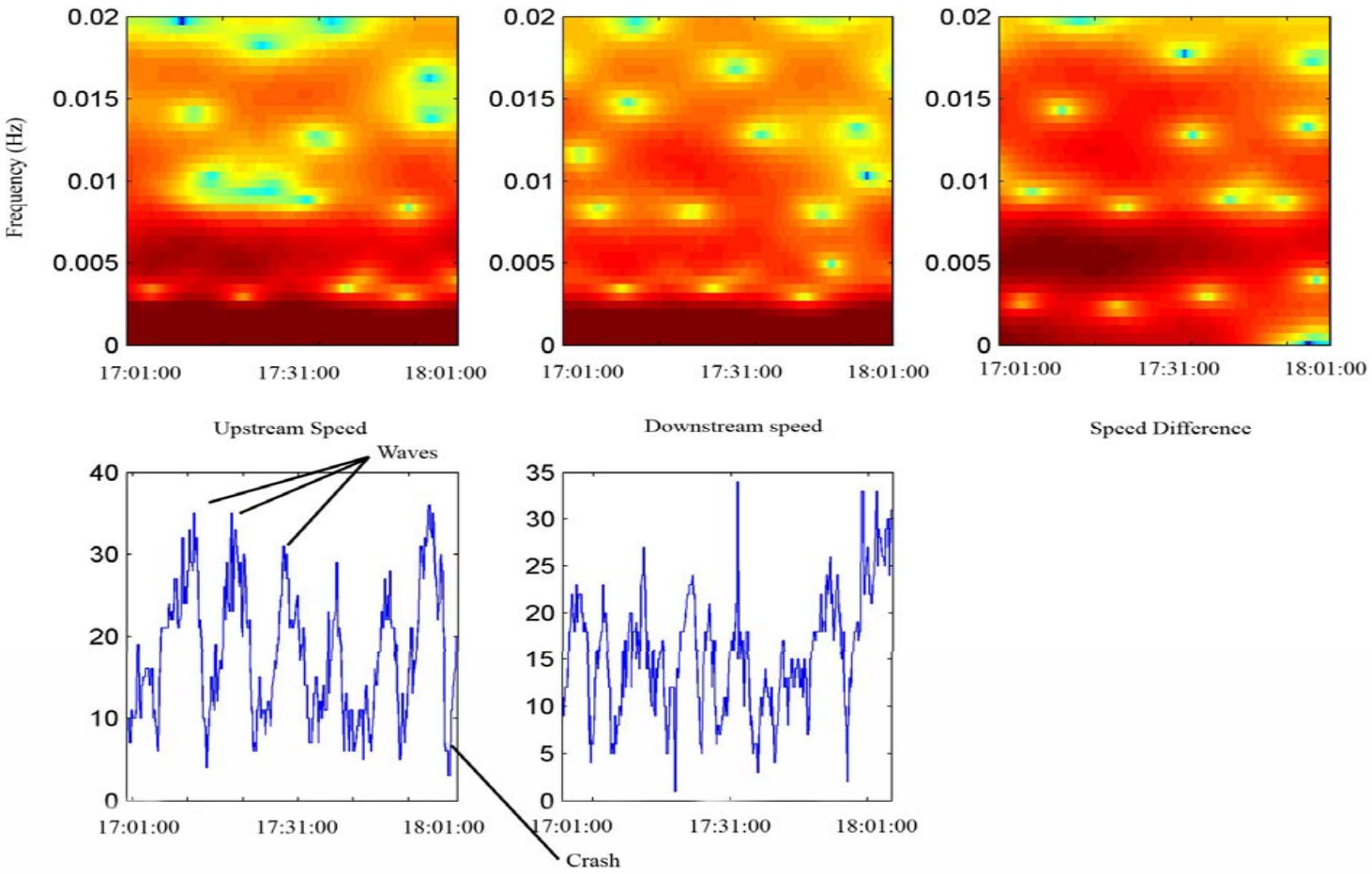

Figure 4. 12 Spectrograms of 3600 seconds of Individual Vehicle Speed Measurements.

Recurrent congestion produces a similar picture but with considerably lower magnitudes. Both compression waves as well as recurrent congestion present high frequency content in the whole range as compared to uncongested conditions. These differences in the frequency domain are summarized in Figure 4.12.

4.2.2.4 Cross-Correlation Analysis of Detector Pair

When multiple stations are available it is possible to explore the correlation of measurements on sensor pairs. As mentioned earlier such analysis of two signals is called cross-correlation. The day used in the previous sections as an example will also be used here to present the results of the analysis. Figure 4.14 and 4.15 present the cross-correlogram of individual vehicle speeds collected in two subsequent detectors located on the right lane immediately downstream from Portland Ave (circled red in Figure 4.13). The two graphs present the entire day (7 a.m. to 8 p.m.) and the start of the evening peak period respectively. Similar to the auto-correlogram presented in the previous section the horizontal axis represents the time of day, while the vertical axis represents the time lag of patterns as they appear in the two detectors. Positive time lags indicate that the traffic patterns move upstream, contrary to the flow (appear first on the downstream detector then on the upstream) while, negative lags represent the opposite. Compression waves are expected to produce patterns with positive lag while traffic pulses for example, have negative ones.

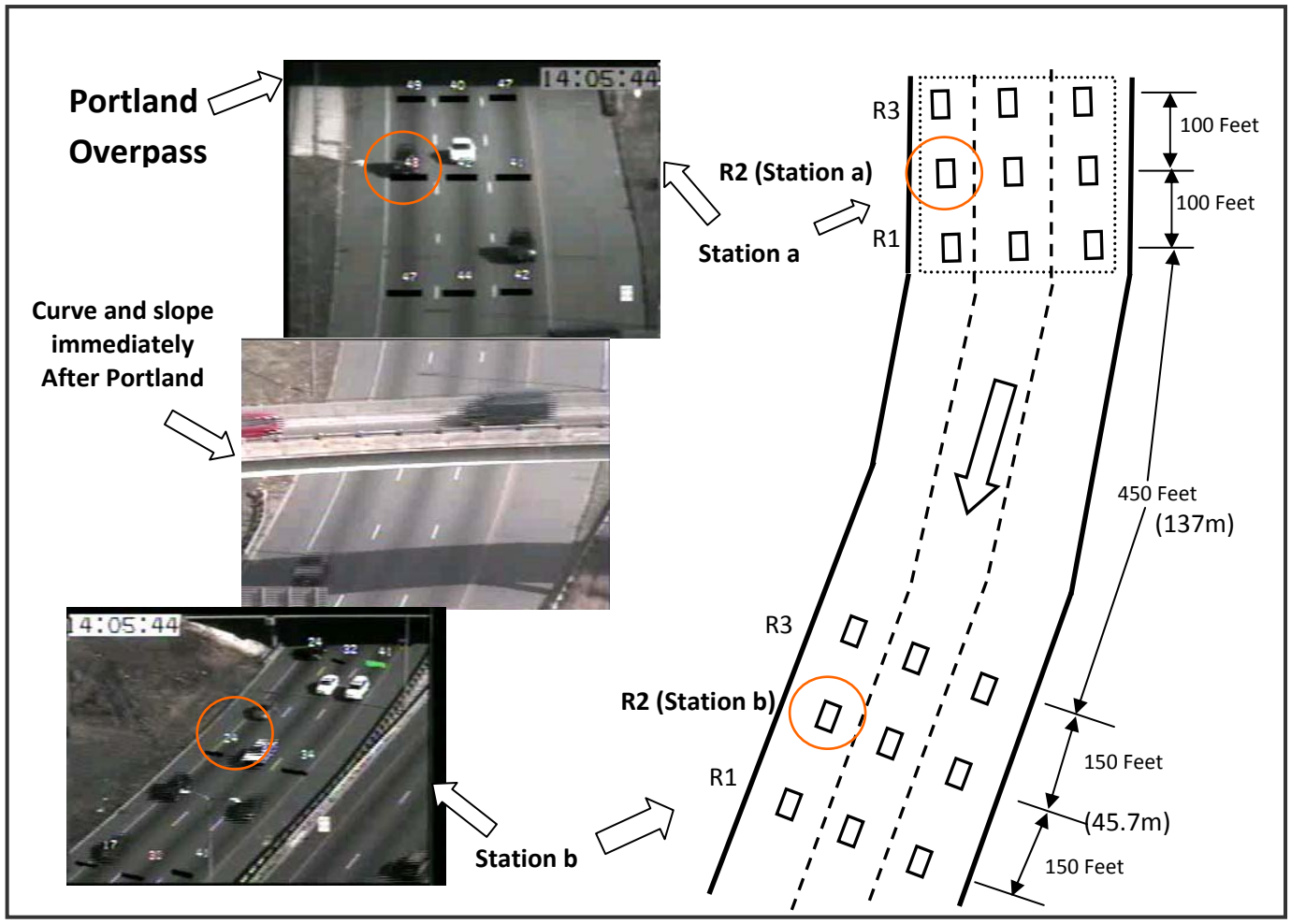

Figure 4. 13 Machine vision detector setup for Multi-station analysis.

As noted in Figure 4.14 there are several clearly described patterns in the correlogram. During uncongested periods the platoons of vehicles create a constant stream of traffic pulses. The speed which these pulses propagate is equal to the average speed of the stream at each time period. Since such patterns are not of particular interest in this work the resolution of the graph does not allow detail inspection. In contrast, on the positive lag side during congested conditions there are clear signs of compression wave activity. During the transition stage between uncongested and congested states the patterns generated vary greatly in their characteristics therefore the picture produced in the graph is that of simultaneous negative and positive lags (pockets of uncongested traffic intermingled with high density, slower moving platoons). As traffic conditions deteriorate, density increases while speed remains relatively high. Traffic at this point is in a **metastable** state. The term metastable is used mainly in physics and recently has been introduced in traffic flow theory. A metastable state is any dynamic system condition that is stable only because a disturbance of high enough amplitude has not yet occurred [Helbing et al., 2002]. In traffic such disturbances can be merging flow from a ramp, lane changing maneuvers, or any other event that will cause large enough drops in speed or increases in density. The deeper the traffic stream goes into the metastable state the smaller is the amplitude of the required disturbance. Once such a disturbance occurs, the traffic stream becomes unstable and congested. Following this transformation compression waves start to propagate upstream creating the clear high correlation areas marked on Figure 4.14.

In Figure 4.15 the aforementioned transition can be seen in greater detail. Early on in the congested period the compression waves are traveling backwards with relatively high speeds. As traffic worsens and the overall stream speed drops the wave speeds also decrease. It is possible to measure the wave speeds considering that the distance between the two detectors is known (700 feet (213m)) and the time period it takes for each wave to travel from one detector to the other is taken from the correlogram.

At this point it is prudent to explain some of the features found in the correlogram. Any distinct traffic pattern traveling between the two detectors will be represented by a dark region in the graph. When the graph is drawn in 3D it shows as a peak. The height of the peak (darker red or darker grey) signifies the similarity between the patterns measured between the two detectors (it is possible for these patterns to mutate as they progress in space; this will decrease the correlation). The width of the peak (horizontal dimension in the graph) represents correlation in time and the existence of several waves in regular intervals.

As seen in Figure 4.15, immediately after the breakdown (transition) the compression waves are propagating backwards with a speed of 25mph (40km/h). Initially there are few waves and in random intervals. Later when density increases the speed of the waves drop to 15mph (24km/h) and repetitive appearances take place in intervals ranging from 110 to 300 seconds. Even later in the day the speed of the stream drops to an average of 15mph (24km/h) and the speed of the waves drop to approximately 12mph (19.3km/h). Very few

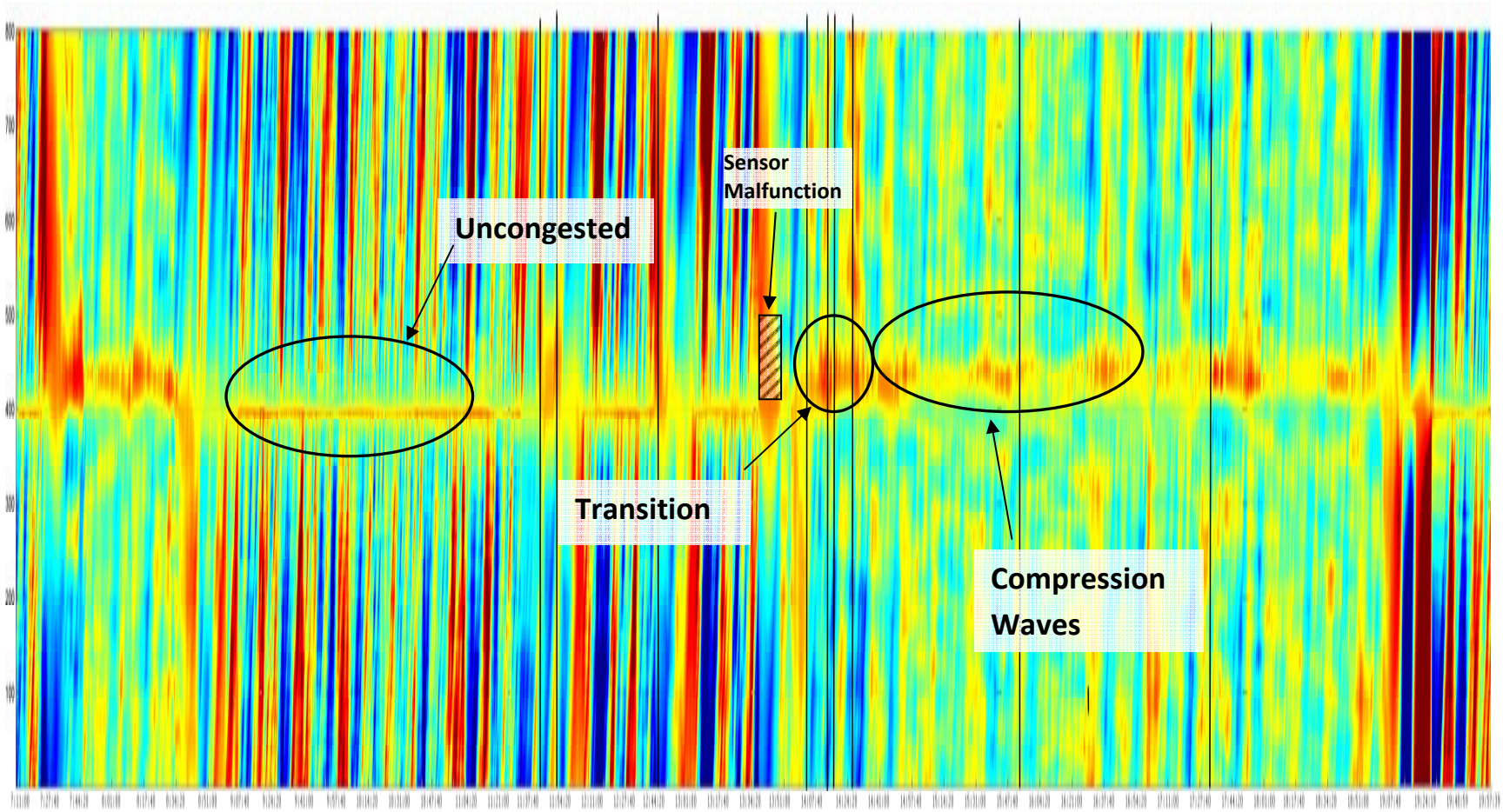

Figure 4. 14 Cross Correlation of individual vehicle speed measurements on August 26th, 2003.

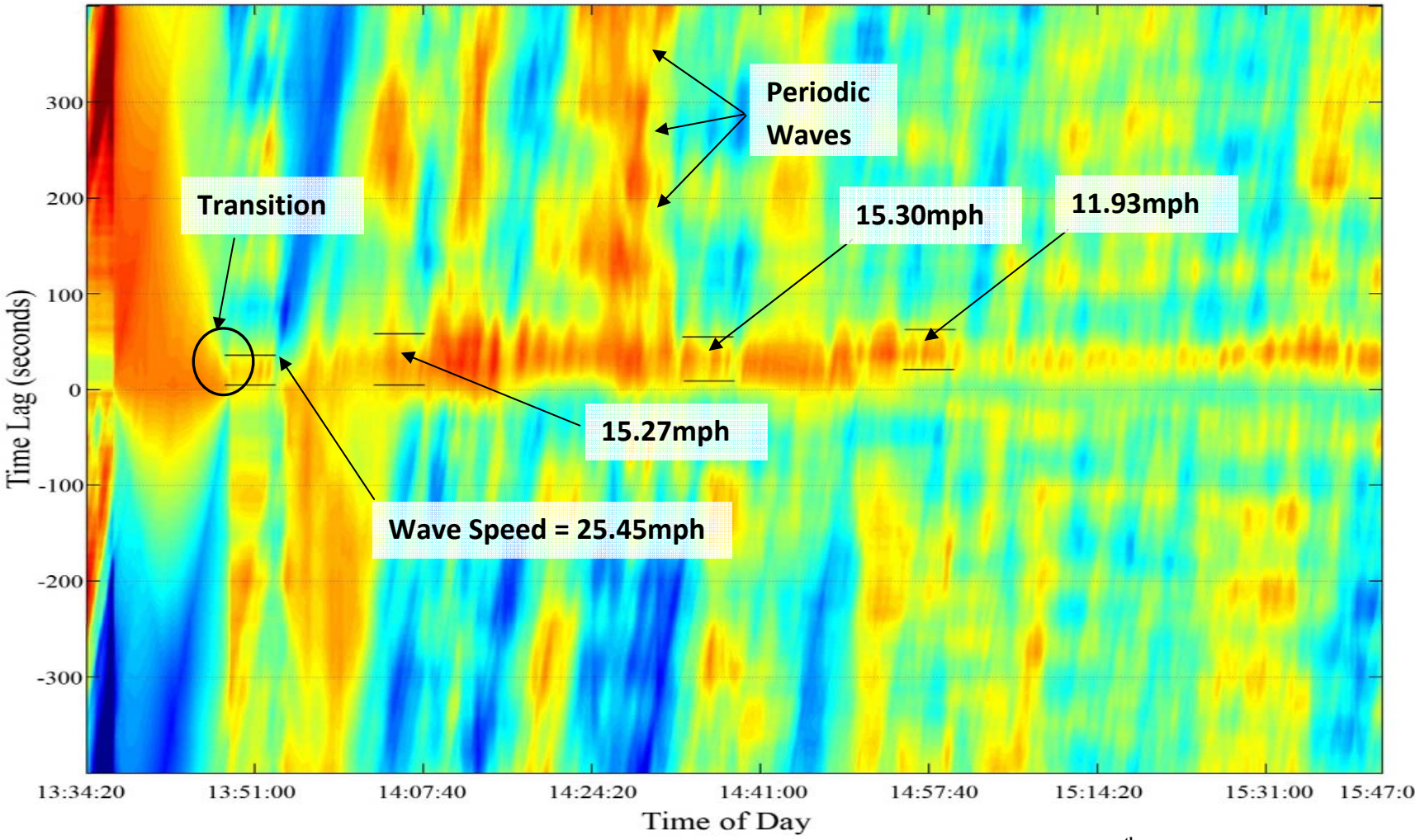

Figure 4. 15 Detail of Cross Correlation of individual vehicle speed measurements on August 26th, 20.

crashes have been observed when the wave speed drops below 15mph (24km/h). The information about the compression wave characteristics is useful in order to understand the conditions under which the majority of the crashes take place. The existence of compression waves does not guarantee that a crash is imminent. In all the days analyzed, compression waves are frequent and with characteristics that don't change a lot throughout the year. Crashes are not nearly as common as compression waves. This leads to the conclusion that a crash prone conditions detection system cannot be based solely on the detection and measurement of compression waves. Compression waves seem to be a necessary but not sufficient cause for crashes.

Analysis similar to the one described in this section has been performed with other detector pairs in the rest of the lanes in the same location as well as on several points upstream of Portland Ave. Although the majority of the conclusions reached here were verified there also, no additional useful information was acquired.

4.2.2.5 Spectral Analysis of Detector Pairs

When two adjacent detectors are available it is possible to generate a speed difference signal. For the purposes of spectral analysis there is no issue in how to calculate the difference since the frequency domain representation of the signal will be the same. Figure 4.16 presents the spectrogram of the difference of speeds collected on the right lane in the two machine vision sensors used during the cross-correlation analysis in the previous section. In this case, similar to the single detector measurements, the two general traffic states, congested and uncongested, can be easily distinguished. Specifically, the uncongested state exhibits high power frequency content between 0 and 3 mHz with negligible power on higher frequencies. The congested state exhibits similar power levels between 0 and 30 mHz. Additionally, mainly on the boundaries between the two states (transition periods), higher power peaks are observed mainly between 0 and 3mHz but in certain cases extending up to 10mHz. Eight out of the nine near-misses observed during that day follow in time such extended peaks. This is also a general observation for all days analyzed. Although the majority of the peaks that extend up to the 10 mHz level are followed by near-miss sightings, not all do. Again, it is observed that during the morning peak such peaks on the boundaries of state changes are not followed by any crash events. In previous work dealing with the frequency domain representation of traffic states, Chassiakos (1992) analyzed occupancy differences between two subsequent loop detectors. In that work it was found that a distinguishable difference between recurrent congestion and compression waves is presented as a secondary peak centered around 2.47 mHz. In the case of the speed difference the aforementioned extended peaks could be explained by the existence of a secondary peak around 5 mHz. The difference in frequencies is normal since the signals have different origins. This observation enforces the association between compression waves and crash events. Considering that the goal of this research is to find a way to detect crash prone conditions, a rough calculation, based for example on the Figure 4.16, shows a possible detection rate of 88% with a false alarm rate of more than 50%. Basically, a detection scheme based on the spectrum of speed difference

will raise the alarm in all instances a fast compression wave is propagating upstream through the detector pair. This is not good enough for a robust and efficient system.

Considering the machine vision collected measurements, all the important features in the spectrogram can be found below 25-30 mHz. This information will be utilized later in this section when noise filters will be developed and implemented. During the spectral analysis step, measurements from two loop detectors on the right lane where used to generate the time occupancy difference. The spectrogram for this signal can be seen in Figure 4.17. Unfortunately no new information is revealed from this step; therefore no further discussion follows.

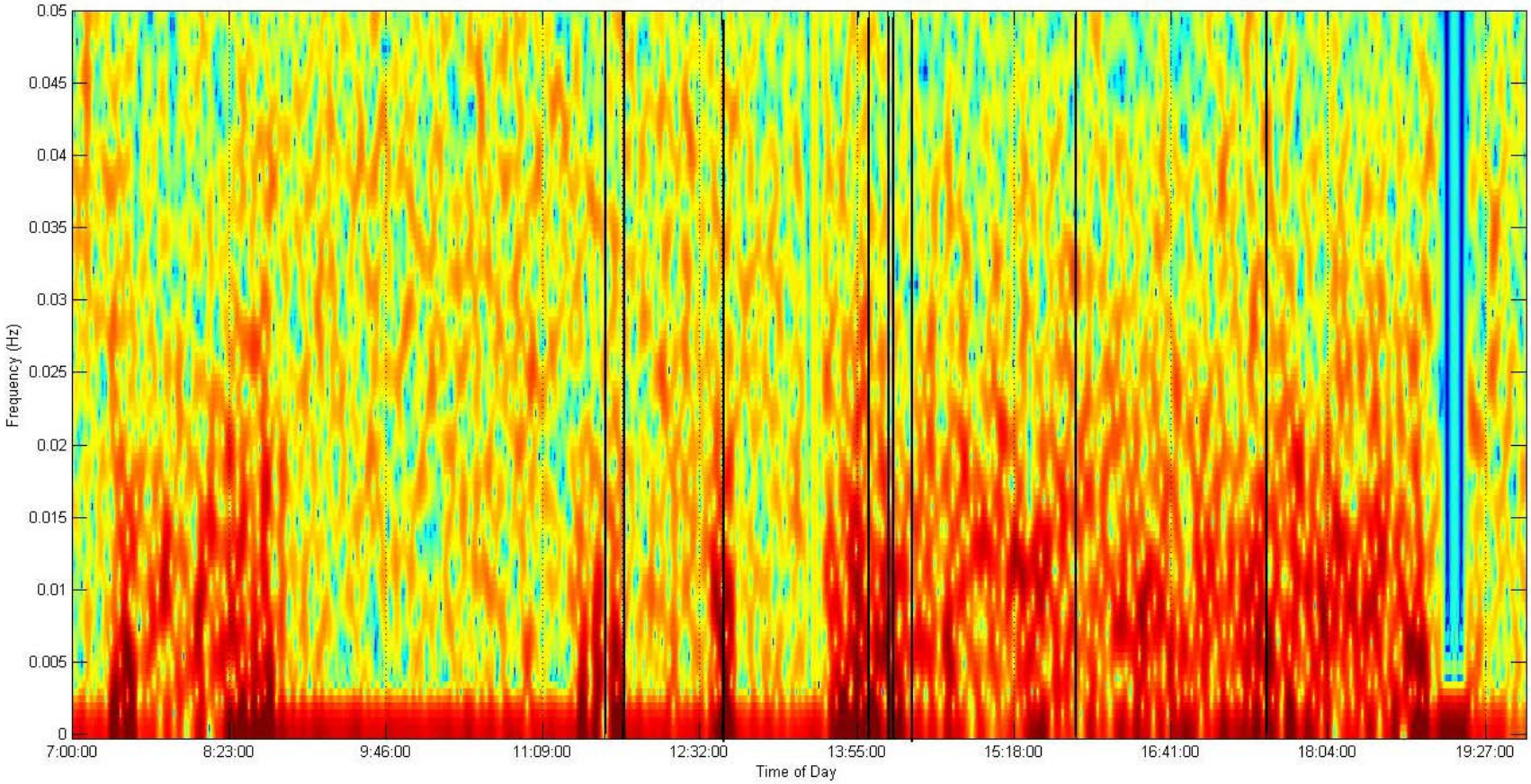

Figure 4. 16 Spectral Analysis of speed difference measurements on August 26th, 2003.

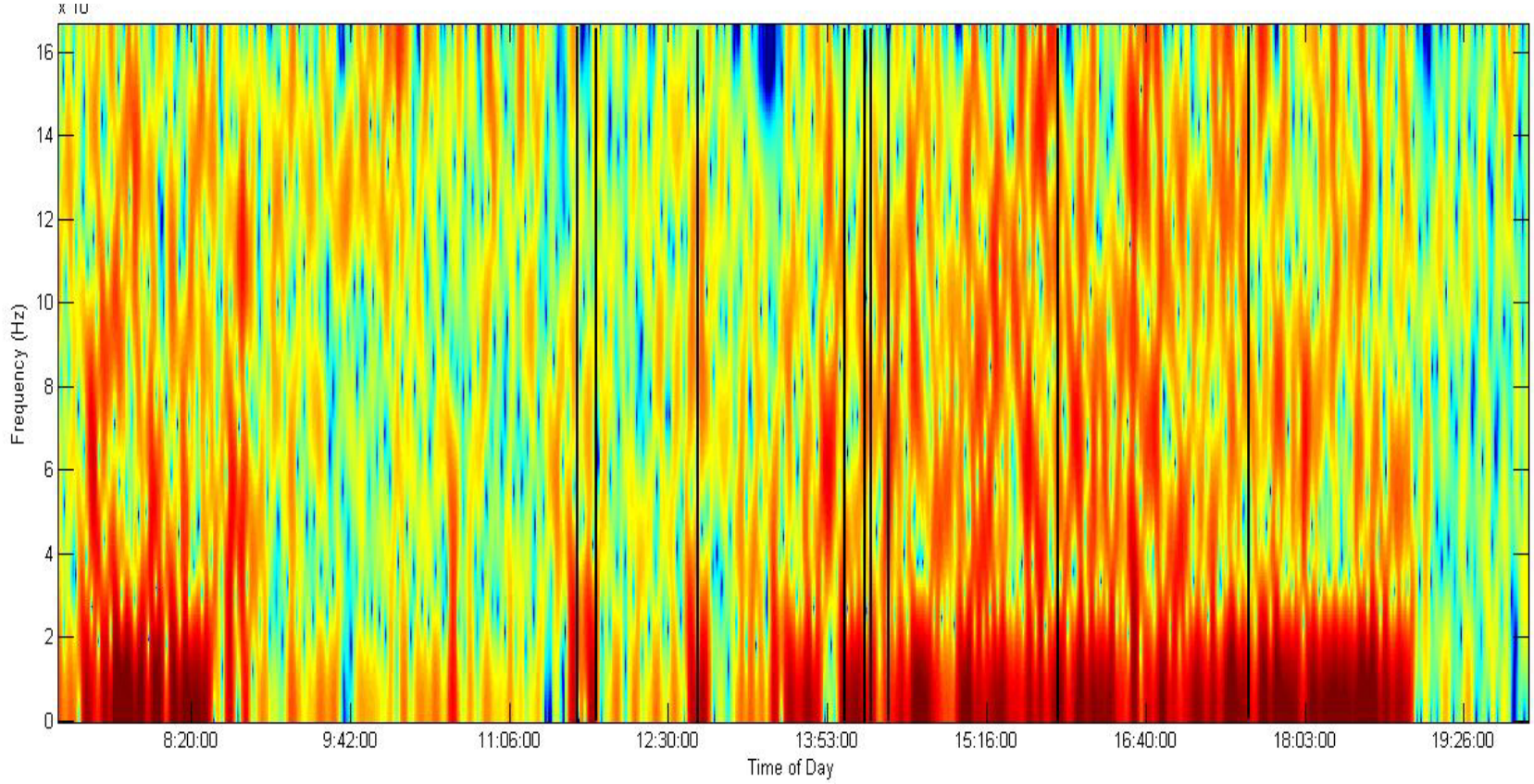

Figure 4. 17 Spectral Analysis of loop occupancy difference measurements on August 26th, 2003.

4.2.3 Analysis Conclusions

This chapter so far has explored in various ways the information collected at the study site. When this analysis stage was planned, the goal was not only to increase our understanding of the traffic dynamics in the high crash area but also to uncover a way to automatically detect them in real-time. Although a fair number of patterns were uncovered in the process (summarized in the rest of this section), rough examination of their frequencies in comparison with crash events revealed that the facts do not favor an efficient detection system. The main realization stemming from this analysis is that purely traffic flow related events like compression waves, flow breakdowns, etc. are not sufficient to cause a crash although they're clearly common precursors. There must be additional factors, primarily dealing with driving behavior, which in conjunction with the aforementioned traffic patterns generate all the necessary conditions for crashes. From the literature search, it is clear that most of the current research, aimed in real-time detection of crash prone conditions, in its majority, only deals with the aforementioned traffic flow dynamics.

Nevertheless, the qualitative analysis succeeded in uncovering certain detectable patterns that can better describe the state of the traffic stream in this high crash area. For example, from the investigation of the speed (estimated from volume and occupancy) as measured by the loop stations in the section between Portland and the Lowry Hill Tunnel, the most probable area for the traffic breakdown was identified. This not only helps in disproving some of the original and popular views about the breakdown but also allowed the research team to reconfigure the surveillance equipment and investigate the breakdown area in greater detail. In summary, the popular view that the cause of the crash related congestion is the Lowry Hill Tunnel and the I-394 exit is unsupported by the available measurements. Instead, the measurements point to the merge area of the I-94 combination ramp as the friction point and one of the causes of flow breakdown. Further investigation presented in later chapter revealed additional evidence corroborating this result.

The time series analysis performed on the available measurements succeeded in visualizing the various traffic states encountered in the study area. In addition, traffic patterns like compression waves, traffic pulses, and others had their magnitude and frequency measured. The fact that the compression waves closely associated with crash events propagate backwards with speeds up to 25 mph (40 km/h) constitutes a vital piece of information and sets the standards for the design of any crash prone condition detection scheme. Specifically, one of the measures of effectiveness for detection systems is Time-to-Detection. Based on the geometry of the site and the respective positions of crashes and flow breakdown, the detection system must have a time-to-detect smaller than 2 minutes to be able to provide sufficient warning time. Additionally, this knowledge also affects the design of any automated crash prevention or driver warning systems since they need to react this fast to prevent a possible crash.

Finally, the results of the time series analysis and in particular that of the spectral analysis assist in the design of linear and non-linear digital filters that can potentially improve the quality of the raw measurements and improve the performance of the final detection scheme. The following

section describes the fundamentals of designing such filters. The final selected filters can be found in chapter five.

4.3 DETECTION AND REMOVAL OF IMPULSIVE NOISE

There are several reasons for the introduction of noise in traffic measurements and each sensor technology has its own noise characteristics. In general though the most common effect of impulse noise is that it does not corrupt each data point, but at the points where it corrupts the signal, it suppresses the signal value by a high (or low) value. Loop detector output, averaged every 30 seconds, includes isolated high magnitude impulses in volume and occupancy measurements. Similar noise is encountered in machine vision measurements.

Removal of impulsive noise has been extensively used in speech processing applications. Linear filters have received considerable attention in digital signal processing theory and application because of their inherent smoothing properties. A simple linear filter for such application is a running average smoother of the form

$$y[m] = \frac{1}{2N+1}\{x[m-N] + \cdots + x[m] + \cdots + x[m+N]\}$$

where {x[.]} and {y[.]} are the input and output respectively of the linear filter with size 2N+1. During this research, in one of the earlier stages in the statistical analysis, the raw data were filtered with the help of a linear smoother in an attempt to improve the quality of the result. Specifically, a weighted Moving Average Filter was utilized. The weighted moving average filter tracks the weighted average of the N sample observations. Figure 4.18 shows the estimation process of the weighted average filter. This filter implements an 8-point window.

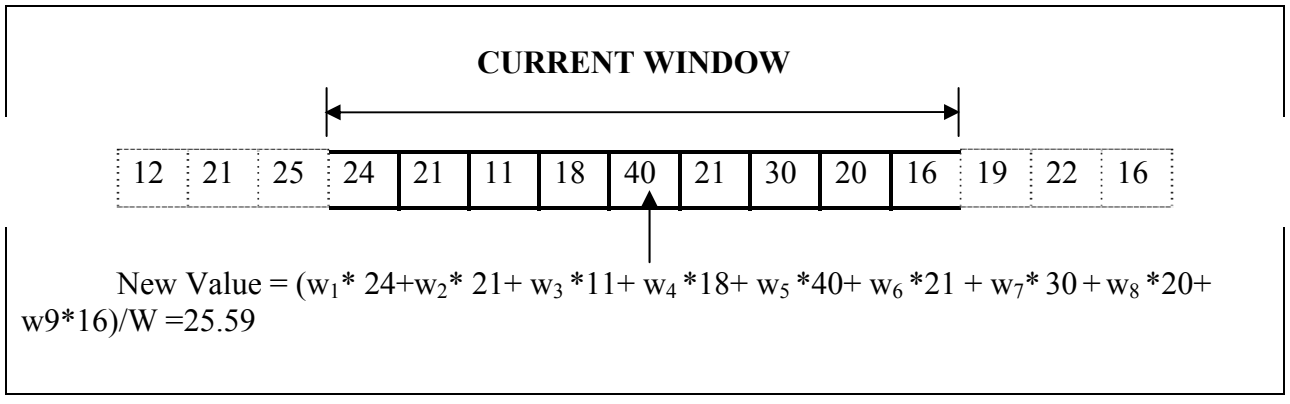

Figure 4. 18 Weighted average linear filter.

Linear smoothing performs well when the desired signal spectrum is significantly different from that of the interference. In many situations, however, the signal displays noticeable sharp

discontinuities. Such discontinuities contain much high-frequency energy, and their spectral content is essentially indistinguishable from that of the noise content. A linear smoother would, therefore, smear out the sharp edges in the data while filtering the noise. Considering that one of the goals is to retain any information concerned with compression wave activity additional filters were researched.

For example, nonlinear smoothing could be used to preserve sharp discontinuities in the data and still filter out noise superimposed on the data. Running medians of the form

$$y[m] = median\{x[m-N], \cdots, x[m], \cdots, x[m+N]\}$$

are a good and simple design. An important property of running medians is that they do not smear out sharp discontinuities in the data, as long as the duration of the discontinuity exceeds a critical value. So, the size of the running medians is designed based on the minimum duration of the discontinuity that the user wishes to preserve. All the raw measurements both from machine vision sensors as well as loops used earlier in this chapter for the time series analysis were first filtered with appropriate running median smoothers. In the case of individual vehicle speeds running medians with sizes between 10 and 15 were used, while for loop detector measurements sizes between 3 and 5 were more appropriate.

4.3.1 Filter Design for Pattern Enhancement

From the spectral analysis results presented in earlier sections it became clear that the traffic patterns associated with crash events have specific spectral characteristics. These patterns in the signal co-exist with impulsive noise as well as other types of noise. It is possible to design specialized filters to remove or suppress higher frequency content in the hope of improving the clarity of the patterns we are interested in detecting. Regardless of the final method utilized to detect crash prone conditions such filters will improve performance. Several filters were designed and tested. The final filter selection depends on the detection methodology finally utilized. The rest of this chapter presents the process of the filter design; while the actual testing will be presented in the next chapter along with the detection methodology.

4.3.1.1 Filter Design Methods

Most of the interesting patterns are encountered in relatively low frequency ranges. To suppress the influence of higher frequency content a lowpass filter is required. Regardless of the type, there are several different methods for designing digital filters. Each method and its derivatives produce filters with the desired general characteristics but with small variations in the actual structure. These variations can be critical to the end result. There are two filter families, IIR (Infinite Impulse Response) and FIR (Finite Impulse Response). In each family several alternative design methods were explored. For brevity, the final filters selected for implementation and evaluation are presented.

4.3.1.1.1 Infinite Impulse Response (IIR) Filters

For the purposes of this work the sampling rate of the interpolated machine vision individual speed measurements is fs=1Hz. Therefore a filter with following characteristics is desired:

Table 4. 3 Filter characteristics.

Type	Low-Pass
Pass band	0.05Hz
Stop Band	0.1Hz
Minimum order	Calculated at 9
Cut off frequency	Calculated at 0.108Hz

The resulting Numerator and Denominator of the transfer function are presented in table 4.4.

Table 4. 4 Digital IIR Butterworth filter coefficients.

Numerator	Denominator
$0.00477274430788*10^{-5}$	1.00000000000000
$0.04295469877090*10^{-5}$	-7.04571211346482
$0.17181879508361*10^{-5}$	22.23652897152238
$0.40091052186176*10^{-5}$	-41.23152442859955
$0.60136578279264*10^{-5}$	49.47233373966907
$0.60136578279264*10^{-5}$	-39.81527312986237
$0.40091052186176*10^{-5}$	21.48371343617754
$0.17181879508361*10^{-5}$	-7.49183134998919
$0.04295469877090*10^{-5}$	1.53160738004123
$0.00477274430788*10^{-5}$	-0.13981806904343

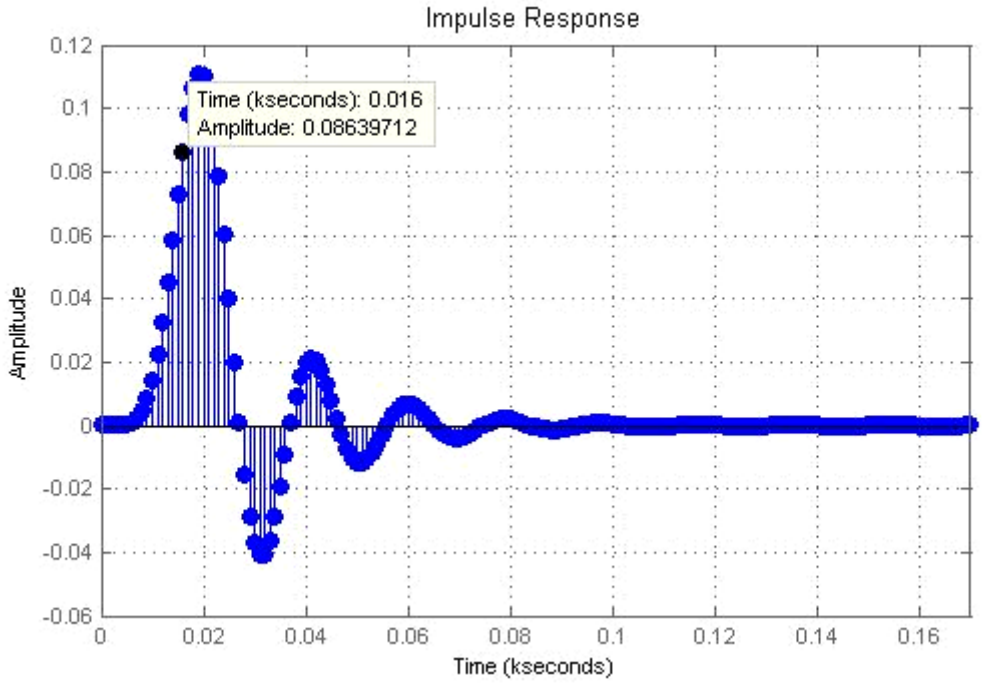

Figure 4. 19 Impulse response of designed Digital IIR Butterworth filter.

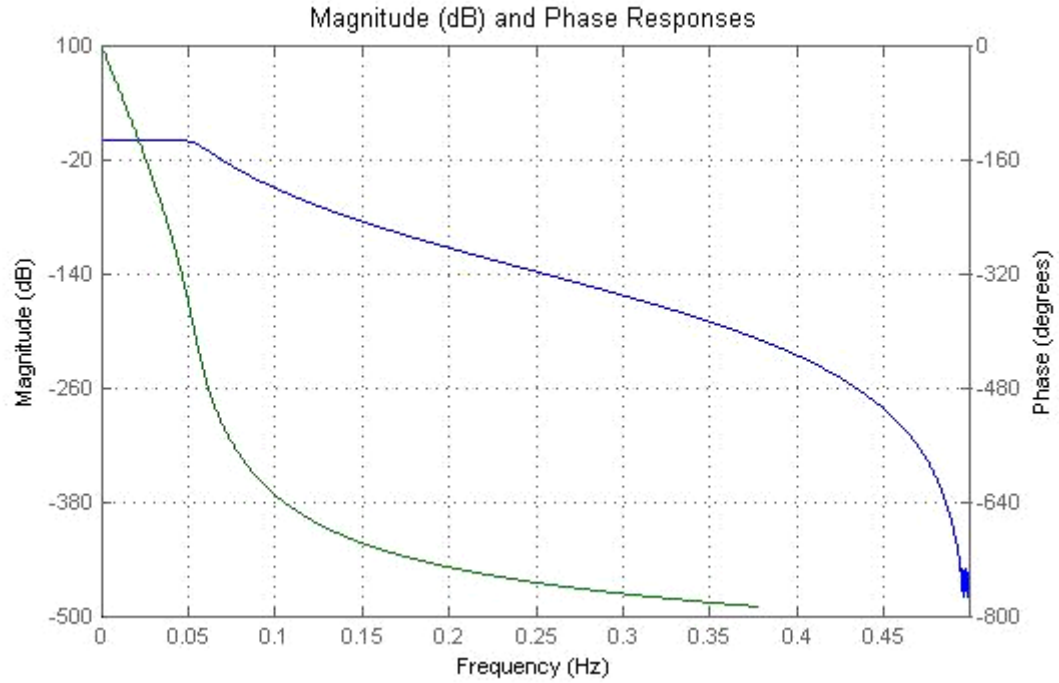

Figure 4. 20 Magnitude and Phase responses of designed Digital IIR Butterworth filter.

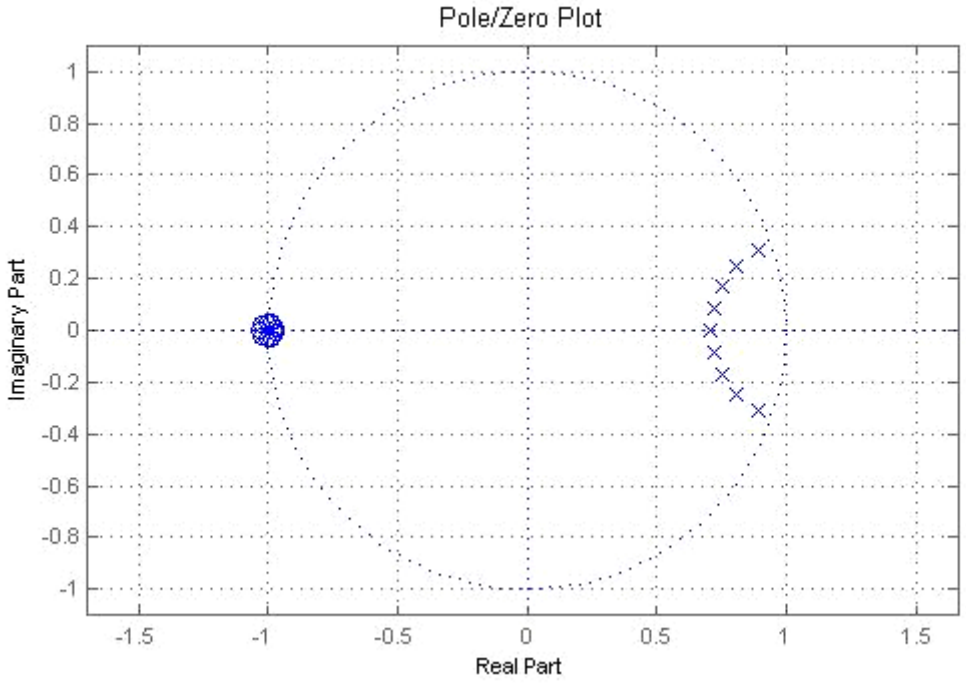

Figure 4. 21 Pole/Zero configuration of designed Digital IIR Butterworth filter.

4.3.1.1.2 Finite Impulse Response (FIR) Filters

FIR filters need to gave considerably larger lengths to match the performance of the equivalent IIR filter. In the case of the filter designed for the preprocessing of the machine vision raw measurements (sampling frequency 1Hz) a filter with the following characteristics is selected:

Type	FIR Low-Pass
Filter order	50
Cut off frequency	0.03Hz

The resulting Numerator (Denominator=1) of the transfer function are presented in table 4.5.

Table 4. 5 Digital FIR Hamming filter coefficients.

Numerator
-0.00101859163579
-0.00108949475321
-0.00121537651446
-0.00137192145588
-0.00151265440124
-0.00157023918071
-0.00145997242744
-0.00108531159893
-0.00034509009909
0.00085809124142
0.00260893330709
0.00497068241703
0.00797728232361
0.01162712743101
0.01587901067640
0.02065070804407
0.02582044950378
0.03123130816805
0.03669831296072
0.04201787339149
0.04697891592186
0.05137498565096
0.05501647704677
0.05774213117385
0.05942897716575
0.06000000000000
0.05942897716575
0.05774213117385
0.05501647704677
0.05137498565096
0.04697891592186
0.04201787339149
0.03669831296072
0.03123130816805

0.02582044950378
0.02065070804407
0.01587901067640
0.01162712743101
0.00797728232361
0.00497068241703
0.00260893330709
0.00085809124142
-0.00034509009909
-0.00108531159893
-0.00145997242744
-0.00157023918071
-0.00151265440124
-0.00137192145588
-0.00121537651446
-0.00108949475321
-0.00101859163579

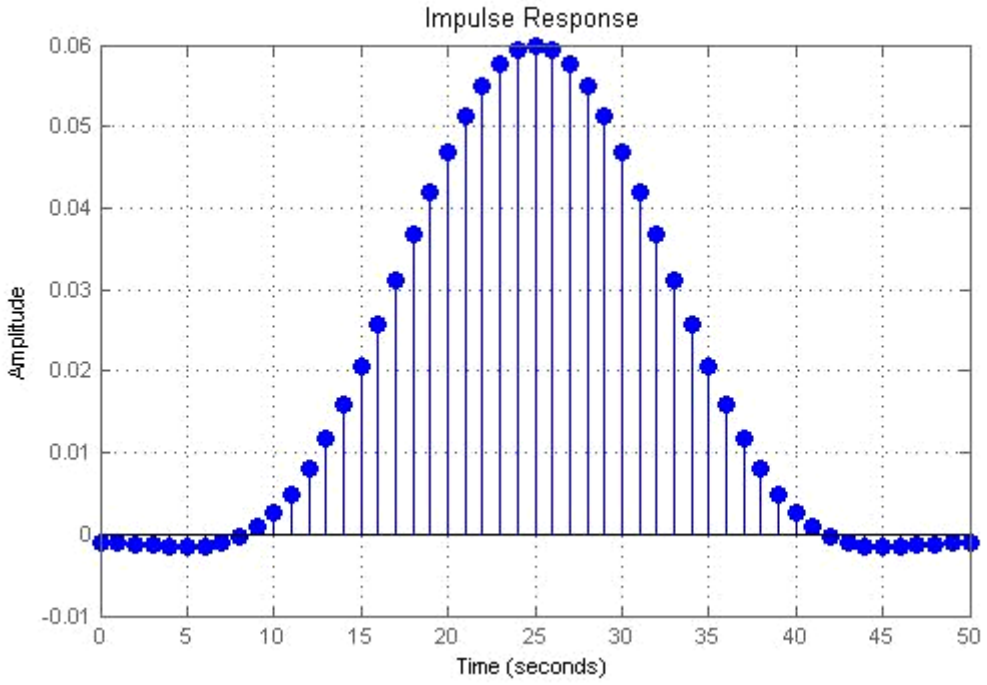

Figure 4. 22 Impulse response of designed Digital FIR Hamming filter.

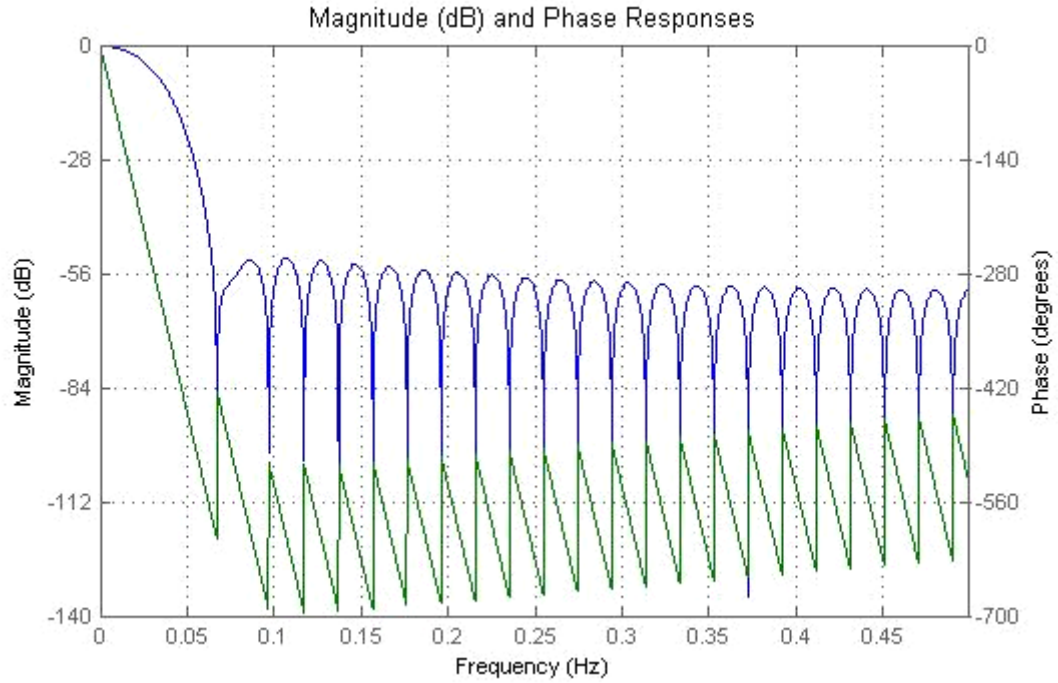

Figure 4. 23 Magnitude and phase responses of designed Digital FIR Hamming filter.

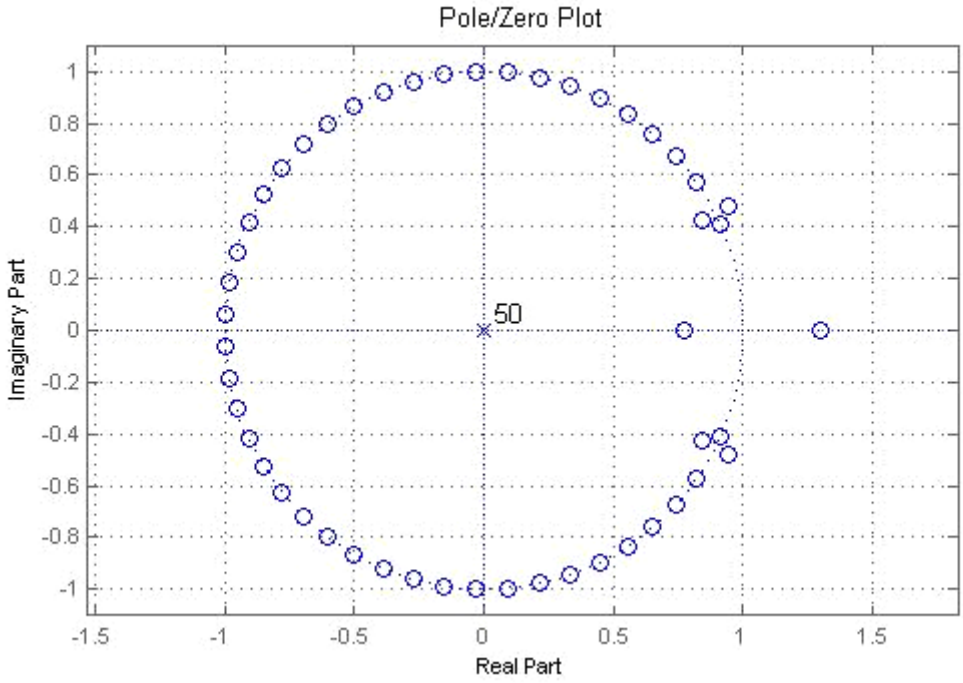

Figure 4. 24 Pole/Zero configuration of designed Digital FIR Hamming filter.

5 Detection of Crash-Prone Traffic Conditions

This research has two main objectives. The first is to verify the existence of crash prone traffic conditions and to establish an automated detection methodology based on real-time information. The second, building on these two, is to, propose a crash prevention/reduction system and/or solutions. Designing a reliable detection system does not necessarily require understanding the nature of crash prone conditions. In contrast, in order to design an effective and efficient crash prevention/reduction solution one needs to attack the actual crash causes. However, the statistical modeling employed in the search for reliable associations and empirical causal connections showed promise in detecting crash prone conditions in real time. This chapter answers the question of whether there is a way to improve its performance.

In the previous chapter, the goal was to associate crash events with real-time measurements that have physical meaning and can help us understand the underlying traffic flow dynamics and driving behaviors involved. For a detection algorithm such knowledge, albeit helpful, is not required; instead, measurements and metrics that maximize performance are needed. The metrics described in chapter three are more suitable for evaluating the state and progression of traffic in time and space. In this chapter, we will capitalize on the crash likelihood model methodology to use these additional traffic metrics in the development of an effective and efficient crash prone condition detection methodology. The process is outlined in the following steps:

1. Identify a set of crash events.
2. Identify a set of normal traffic conditions (controls).
3. Select traffic metrics and parameters describing environmental conditions.
4. Define time and space variations of the aforementioned metrics.
5. Define modeling procedure and prepare utilities for streamlining the process.
6. Identify a number of digital filters to preprocess the raw data.
7. Identify alternative combinations of filters and metric variants.
8. Generate the necessary cases and controls required for the statistical modeling.
9. Develop one logit based model of crash likelihood for each alternative.
10. Incorporate the models in alarm producing algorithms.
11. Identify a set of days, not previously used in the analysis, to be used for testing and evaluating each algorithm. Produce a performance chart for each algorithm.
12. Select the best one(s).

Step one does not need any introduction. The crash cases used in this process are the same used in the chapter 4. A number of cases are set aside for step eleven. In the previous chapter the

majority of the measurements originated from loop detectors. Because of this, the selection of cases and controls simply depended in the labeling of each 30 second measurement group as a case or a control. Although this generated a very large number of samples, it was workable because the number of the parameters considered was relatively small. In the current process the number of potential metrics and their variants was very large. Therefore, to make the process computationally feasible a case to control ratio of 1:2 was selected. This translates to the selection of approximately 240 crash-free traffic periods.

All the metrics described in chapter 3 were considered. In addition a number of environmental conditions were taken into account through the use of a number of categorical parameters. For reasons explained in detail later, a large number of variants of the aforementioned metrics are identified. These variants aim to maximize the models performance. Since all these calculations are to be repeated many times with different combinations of metric variants, environmental parameters, as well as raw data preprocessing, it is prudent to streamline the process with the help of computer utilities. Some of the utilities are created from scratch in C++ while others use MATLAB. All data are stored in relational databases.

In chapter 4, following the time series analysis of the raw measurements, we proposed the possibility of enhancing the data quality by processing them through a number of digital filters. Although specific time and frequency domain characteristics were identified in the raw measurements, still the selection of the most appropriate filter must be an iterative process involving the development of the model and testing it in pseudo real-time conditions. In chapter four the underlying filter design methodology was described mostly for the reader to understand the filter's nature and characteristics. This methodology has been successfully incorporated in a number of computer applications like MATLAB, providing the user with graphical tools that accelerate the filter design and evaluation process. Most of the filters used in the development of the detection algorithm were produced this way.

Raw measurements were separately preprocessed with each filter, which generates all the metric variants used in the development of the model. Since there are a number of filters as well as metric variants, a large number of alternative combinations are identified. Each combination constitutes a set of proposed variables to be fitted into the logistic regression model. Specific cases and controls are identified and prepared for the fitting process. Several computer applications are available for the development of the logic model of crash likelihood. In the case of this study ARC [Cook et al., 1999] was selected because it had the best combination of features and streamlining required. Following development, each model is evaluated in regard to its statistical validity. Unacceptable models are discarded.

Each acceptable model is incorporated into an algorithmic framework that streamlines the testing process. This process includes the estimation of the necessary metric variants from real-time continuous raw measurements for the days set aside in step one and the generation of the alarms

when crash prone conditions are detected. The performance of each model-algorithm is evaluated based on a pre-selected set of thresholds. A performance curve displaying detection rates over false decision rates for each threshold is created. This process was iteratively executed until a set of acceptable filter-model-thresholds combinations was defined. Later in the document it will become clear why a single answer is not optimal in this case.

5.1 IDENTIFICATION OF CASES AND CONTROLS

Due to the data collection mishap explained in chapter 3, 51 collisions and 122 near misses were used to develop the logistic regression models. As described in chapter 4, the road segment immediately downstream of Portland Avenue is the focus point of all of the crashes. Therefore, measurements from the two machine vision stations used in the quantitative analysis are utilized in the calculation of the traffic metrics necessary for this stage. For each crash a single case is generated. As it will explained later in this chapter, the metrics at this stage can describe conditions up to 5 minutes before the crash time. Additionally, it is left for the model to decide which metrics are significant in separating crash prone from non-crash prone conditions.

During the qualitative analysis stage a large number of no-crash time periods were randomly selected. These periods were scrutinized to ensure that no collision or near miss occurred during or shortly after. From the set of crash-free conditions, 267 periods were selected. Each of the selected periods represented 20 minutes worth of MV individual speeds, headways, and classification. Approximately half of these controls were during uncongested conditions but no cases with speeds above 55 mph (88 km/h) (free flow) while the other half were during congested conditions (average speed below 20 mph (32 km/h)).

5.2 TRAFFIC METRICS AND METRIC VARIANTS

The majority of the metrics utilized in this study have already been presented in chapter three. To summarize, the following traffic metrics are considered:

1. Temporal metrics
    - Average Speed ($\underline{S}$)
    - Coefficient of Variation of Speed (S_{CV})
    - Traffic Pressure (P_T)
    - Kinetic Energy (E_k)
    - Coefficient of Variation of Time Headway (H_{CV})
2. Spatial Metrics
    - Acceleration Noise (A_N)

- Mean Velocity Gradient (MVG)
- Quality of Flow Index (Q_F)

In addition to this list of metrics, the data analysis was used to explore a number of heuristic variables. Finally, two heuristic variables were found to be descriptive of the patterns observed before each crash; one is temporal and the other is spatial. The following sections describe these two heuristic metrics.

5.2.1 Max-Min-Diff

Based on the association between crash events and compression waves, a variable was conceived quantifying the speed drop caused by the wave. Towards that effect, Max-Min-Diff is defined as the speed percent difference between the maximum and minimum speeds within a group of vehicles passing over a single detector. Figure 5.1 shows how the two speeds are identified from a group of vehicles. Using these speeds, the relative difference is calculated and is multiplied by 100 to convert it to a percent. The greater the percentage is, the larger the change in speeds. Different sizes of vehicle groups (15, 20, and 30) were considered for estimating this metric.

Mathematically it is given by,

$$\text{Max-Min-Diff} = (\text{maxU} - \text{minU})*100/\text{maxU} \quad (5.1)$$

Where,

maxU= max speed in n vehicles

minU = min speed in n vehicle

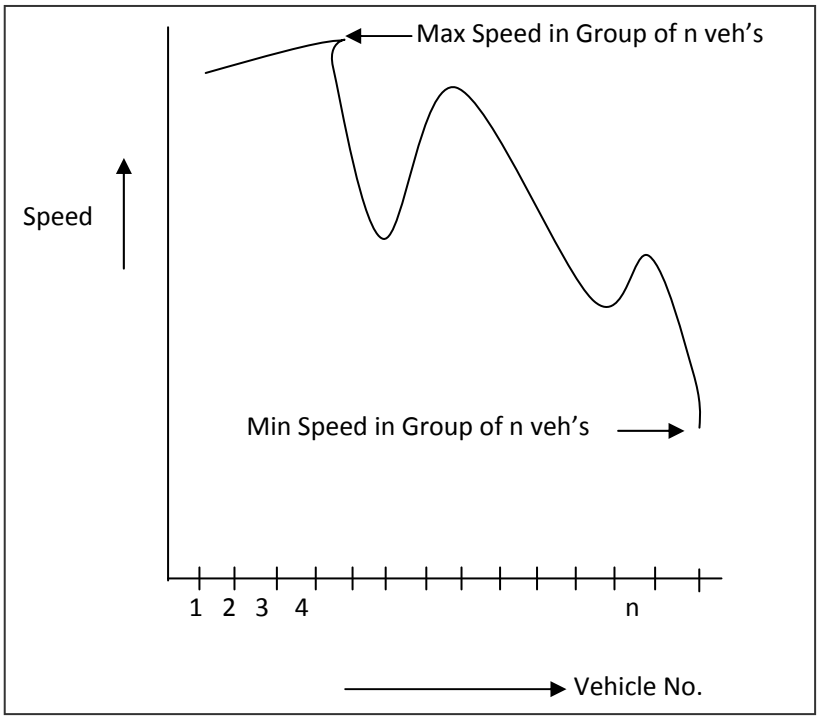

Figure 5. 1 Max-Min-Diff estimation from a group of n vehicles.

5.2.2 Up-Down-Diff

This metric attempts to capture the effect associated with state transitions. Considering two separate measuring locations, we collect at each of them speeds from a group of n vehicles. Up-Down-Diff is defined as, the speed percent difference between the maximum speed at the upstream station and the minimum speed at the downstream station. Different sizes of vehicle groups (15, 20, and 30) were considered for estimating the metric.

Mathematically it can be written as:

$$\text{Up-Down-Diff} = (\max_U(a) - \min_U(b))*100/\max_U(a) \quad (5.2)$$

Where,

$\max_U(a)$ = max speed at station a in group of n vehicles

$\min_U(b)$ = min speed at station b in group of n vehicles

5.2.3 Right-Middle-Percent Difference (RM)

These metric attempts to capture the effect associated with the speed differential between lanes. The metric is simply the difference of the corresponding average speed metrics in the right and middle lane. It is calculated for all speed metric variants.

5.2.4 Crash Time Adjustment

As described in detail in chapter three, individual vehicle speeds were collected in 6 stations along the study section. Since the majority (98%) of the crashes and near misses occurred at or upstream of Portland Ave, only two stations were used in the calculation of the aforementioned metrics. From the video record analysis, the exact time and location of each crash is noted. Of course, the events occurred in variable distances away from the selected stations, which lead to the problem of having the measurements, and by extension the metrics, not synchronized among all cases. The following example attempts to illustrate why this is a problem.

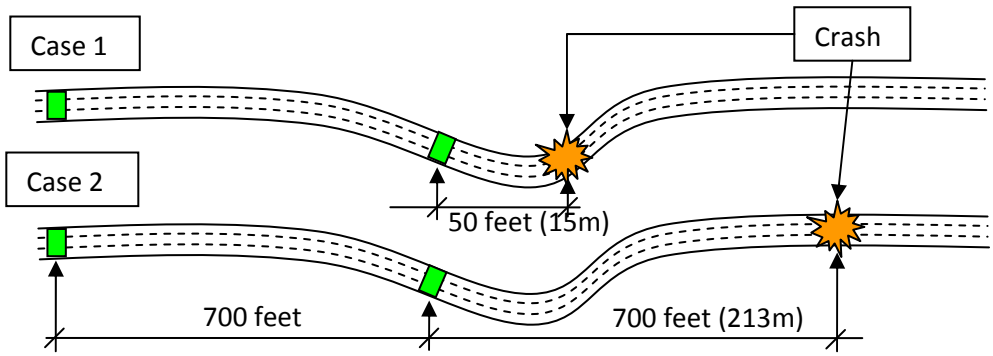

Figure 5. 2 Compression wave pattern mismatch example.

First, assume that in two separate cases (Figure 5.2) a crash directly followed a compression wave. In the first case the crash occurred 50 feet before the upstream station and in the second case the crash occurred 700 feet upstream, both times on the right lane. Assuming the tail of the wave traveled backwards with a speed of 23feet/sec (7m/s) (average observed speed), in the first case the tail passed the sensor 2 seconds before the crash while in the second case the tail passed the sensor 30 seconds before. If in these two cases the metrics are calculated based on the exact time of the crash, assuming there is a unique pattern associated with pre-crash conditions, the pattern evidence will not match. In the case where these data will be used to build a statistical model this mismatch will reduce its performance. To minimize this problem, after closer inspection of the collected measurements, the crash (or near miss) time was adjusted to the approximate time the tail of the wave responsible for the crash (or near miss) passed the upstream station.

5.2.5 Traffic Metric Variants

As described in chapter three, the majority of the metrics use a moving window method in their calculation. Varying the size of the moving window produces a number of different alternatives or variants. In particular, the temporal and spatial metrics used the following window sizes in vehicles:

15, 30, 40, 50, 60, 70, 80, 100, 110, 120

Min-Max-Diff and Up-Down-Diff, described earlier in the chapter, are computed over 15, 20, and 30 vehicles. This variable aggregation size is important because, a priori, it is not clear how much information is required in each metric calculation to reveal a unique, strong pattern during the pre-crash period.

Additionally, in respect to the time of the crash, the time interval the pre-crash patterns are stronger in the metric time series is unknown. To cope with this, a variable time offset is introduced in the calculation of each metric. Specifically, all of the temporal and spatial metric variants (due to window size) are computed 10, 30, 60, 120, 180, 240, and 300 seconds prior to the adjusted crash time. No time offset was used on Min-Max-Diff and Up-Down-Diff since they were direct product of pre-crash visual observations. Table 5.2 summarizes the window sizes and time shifts used in the generation of the metric variants. With this final addition in the metric calculation methodology, for each of the temporal and spatial metrics, 70 variants are created per station. Including the empirical and counting for both stations, 1132 metric variants in total are candidates for describing the relationship between traffic conditions and crash likelihood.

A procedure is followed to represent traffic metrics. All the traffic metrics except the derived ones were denoted by:

$$X - n - \Delta t - i$$

Where,

X = traffic metric (Table 5.1)

n = moving window (cars) (Table 5.2)

Δt = Prior Time Offset (seconds) (Table 5.2)

i = station name, a (upstream) and b (downstream)

For example, avgsp15-120-a indicates average speed estimated from a group of 15 vehicles with the last one passing over station 120 seconds prior to occurrence of the event. Table 5.1 presents all the traffic metrics and their corresponding abbreviations.

Table 5. 1 Traffic Metrics and corresponding symbols.

No.	Traffic Metrics (X)	Abbreviation
1.	Right Lane Average Speed	AvgSp
2.	Middle Lane Average Speed	AvgSpM*
3.	Coefficient of Variation of Speed	CV
4.	Traffic Pressure	TP
5.	Kinetic Energy	KE
6.	Coefficient of Variation of Time Headway	TH
7.	Average Quality of Flow Index	QOF
8.	Average Mean Velocity Gradient	MVG
9.	Derived Metric: Max and Min Speed Difference using one Station	MaxMinDiff
10.	Derived Metric: Max and Min Speed Difference using two Stations	UpDownDiff

- Only metric estimated for Middle Lane

Table 5. 2 Vehicles Window size and prior time shift.

Vehicle Window size (n) (Vehicles)	15,30,40,50,60,70,80,100,110,120
Prior time shift (Δt) (seconds)	10,30,60,120,180,240,300

Table 5. 3 Environmental Variables and their description.

Variables	Description
Visibility Condition	(0) Clear; (1) Poor
Pavement Condition	(0) Dry; (1) Wet
Sun Facing Driver	(0) Night; (1) No sun; (2) Sun at back(180^o) or at side(90^o); (3) Sun at 45^o or at 0^o

5.3 MODEL DEVELOPMENT PROCEDURE

As an introduction the next section presents a brief theoretical background on the modeling methodologies. Logistic regression is an established statistical modeling methodology, documented in relevant textbooks [Hosmer et al., 1989, Cook et al., 1999, Demaris, 1992, Eliason, 1992, Pampel, 2000]. Here I offer enough information for the reader to understand which areas of the methodology are used in this study and how to interpret the results, given later in this chapter. Following the theoretical background, I show the process of modeling crash likelihood and its integration in a crash prone conditions detection algorithm. This process as already mentioned earlier in the chapter is an iterative one and therefore, a flow chart is provided for assistance. Finally, this chapter will conclude with the presentation of the results from the various stages of the process as well as the final performance chart of the most promising algorithm alternatives.

5.3.1 Logistic Regression: A Summary

Regression methods have become an integral component of any data analysis concerned with describing the relationship between a response variable and one or more explanatory variables. It is often the case that the outcome variable is categorical, taking on two or more possible values. Over the last decade the logistic regression model has become, in many fields, the standard method of analysis in such a situation. In this study the outcome variable is binary or dichotomous, crash=1 vs. no-crash=0 and therefore, logistic regression fits nicely in this problem. In addition, logistic regression also originated from the desire to estimate the posterior probabilities of events like traffic crashes or disease outcomes.

Logit models are commonly used in epidemiology for estimating an individual's risk or probability of a disease as a function of disease risk factors. Logit regression is based on a logistic curve, a symmetric function whose values range between 0 and 1. The logistic transformation or logit of a success probability p is given by $\log\{p/(1-p)\}$, which is defined as logit (p). For n explanatory variables the logit model is given as:

$$p = \frac{\exp(\beta_0+\beta_1 x_1+\beta_2 x_2+....+\beta_n x_n)}{1+\exp(\beta_0+\beta_1 x_1+\beta_2 x_2+....+\beta_n x_n)} \quad (5.3)$$

$$\text{or, } \log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n \quad (5.4)$$

Where,

β_0 =Constant of the logit regression equation

β_i = Coefficient estimate of variable x_i

x_i = Independent variables or predictors.

exp(x) = exponential function, equivalent to e^x

Fitting a model to a set of data first entails estimating the unknown parameters $\beta_0, \beta_1, \beta_2, ..., \beta_n$ in the model. The most widely used method of estimation is the method of maximum likelihood. In a very general sense the method of maximum likelihood yields values for the unknown parameters, which maximize the probability of obtaining the observed set of data. Maximum likelihood estimators are consistent, asymptotically efficient, and asymptotically normal.

In order to apply this method we must first construct a function, called the likelihood function. This function expresses the probability of the observed data as a function of the unknown parameters. The maximum likelihood estimators of these parameters are chosen to be those values that maximize this function. Thus, the resulting estimators are those which agree most closely with the observed data. If y (observation) is coded as 0 or 1 then the expression of p given in equation 5.3 provides the conditional probability that y is equal to 1 given x (explanatory variables), This is denoted as P(y=1|x). it follows that the quantity 1-p gives the conditional probability that y is equal to zero given x, P(y=0|x). Thus, for those pairs (x,y) where y=1. the contribution to the likelihood function is p and for those pairs where y=0, the contribution to the likelihood function is 1-p. A convenient way to express the contribution to the likelihood function for a given pair (x,y) is

$$p^y(1-p)^{1-y}$$

Since the observations are assumed to be independent, the likelihood function is obtained as the product of the above terms:

$$l(\beta) = \prod p^y(1-p)^{1-y}$$

The principle of maximum likelihood states that we use as our estimate of β the value which maximizes the aforementioned expression. However, it is easier mathematically to work with the log of that equation. The expression of log likelihood ($L(\beta)$), is given as

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^{N} \{y_i \ln[p_i] + (1 - y_i) \ln[1 - p_i]\} \qquad (5.5)$$

Where,

y_i = Observed value and

p_i = Predicted value for the i^{th} event

N= number of observations

Estimation of parameters using the aforementioned method is a well known process and is already implemented in several statistical packages like SAS, STATA, SPSS and ARC. ARC was used throughout this study.

Goodness-of-Fit Measures: After fitting data to a model, the goodness of their fit is evaluated. There are several summary statistics able to measure the discrepancy between observed and fitted values. These measures aim to quantify how well the developed model fits the data. In this study we utilized deviance as explained in the next section.

Deviance: One of the most important and commonly used goodness-of-fit statistics for binary data is deviance (D). The deviance in logit regression modeling plays the same role as the residual sum of squares does in linear regression. Deviance is the likelihood-ratio statistic for comparing a current model with a saturated model. This later model is taken to be the model for which the fitted values coincide with the actual observations, that is, a model that fits the data perfectly. Such a model will have the same number of unknown parameters as there are observations.

Mathematically deviance is given as:

$$D = -2 \ln\left[\frac{\text{likelihood of the current model}}{\text{likelihood of the saturated model}}\right] \qquad (5.6)$$

The log likelihood of the current model and saturated model is estimated using equation 5.5. For the saturated model the predicted value is equal to the observed value i.e. $\hat{y}_i$ = y_i. Hence the likelihood for saturated model ($l(\beta)_{saturated}$) can be written as:

$$l(\beta) = \prod_{i=1}^{N} y_i^{y_i} (1 - y_i)^{(1-y_i)} = 1$$

Using the likelihood of both current and saturated model in equation 5.6, i.e.

$$D = -2\ln[\prod_{i=1}^{N} p_i^{y_i}(1-p_i)^{1-y_i}]$$

(5.7)

5.3.2 Selection of Explanatory Variables

Model building requires excluding extraneous variables that increase the standard error of coefficients and reduce prediction accuracy. The main aim of the modeling process is to identify variables that need to be included in the model. When the number of potential explanatory variables is large, the number of possible models is correspondingly large. The decision on whether or not to include a particular variable is based on the difference in deviance that results from inclusion or exclusion of a variable from the model. The different methods used to determine a sub model with only significant variables are:

1. Backward Elimination: The model begins with a full model (all explanatory variables). In a step-by-step process, an independent variable is dropped if its inclusion in the model gives the smallest increase in the deviance. At every step after removing the variable from the model, the change in deviance is compared with the appropriate value of the χ^2 distribution. The process is repeated until only significant variables remain in the model.

2. Forward Selection: In this method variables are added sequentially to a base model. The base model can have any number of variables and a new model is obtained by adding one of the remaining ones. The process of fitting each model with the remaining variables is repeated until including an additional variable does not improve the performance of the model.

In both methods difference in deviance is used to decide whether a variable is deleted or added to the model. The deviance difference of two nested models measures the extent to which the additional terms improve the fit of the model to the observed response variable. This procedure is used to select a sub model with significant variables from a full model. As variables are added to the model, deviance difference is compared to χ^2, which tests for the significance of that particular combination of predictor variables.

For example: If there are two models, of which one is the sub model of other, the deviance difference from the two models can be compared to the χ^2 distribution.

$D(X_1) = 922.451$

$D(X_1, X_2) = 918.180$

Where, $D(X_1, X_2, X_3 \ldots X_n)$ indicates deviance of model with n terms. To test whether to include X_2, the difference in deviance of two models with and without X_2 is compared to x^2 distribution.

Hypothesis Test:

H_o: $\beta_2=0$ (*Coefficient of X_2 is zero)*

H_a: $\beta_2\neq0$ (*Coefficient of X_2 is not zero)*

The difference of deviance from the two models can be estimated as,

$$G = D(\text{Model without } \beta_2) - D(\text{Model with } \beta_2)$$

$$= 922.451-918.180$$

$$= 4.271 < \chi^2_{(0.975,1)} = 5.00$$

Thus, using the deviance difference it is concluded that null hypothesis (H_o) is true and X_2 is not included in the model. This procedure was adopted in the study in all variable selection steps.

5.3.3 Model/Algorithm Development Process

The large number of metric variants, coupled with different types of filtering of the raw data, generates a very large number of possible models. The process followed in this study is just one of the many alternatives. The aim was to devise a regimented process that can be streamlined to allow exploration of as many different combinations as possible. The steps are described below.

Step 1: Preprocessing of raw data. Based on the time series analysis presented in chapter four, a number of linear and non-linear filters can be designed to help reduce noise in the measurement time series. Naturally, for comparison the case of unfiltered data is also included. In this first step the raw data stored in four relational databases are filtered, resulting in four new databases. The four databases contain measurements taken during collisions, near-misses, uncongested normal conditions, and congested normal conditions. Programs written for MATLAB streamlined the process. Due to the size of the databases the filtering process can take several hours.

Step 2: Metric variant calculation. As presented in tables 5.1 and 5.3 there are 8 metrics per station, 2 heuristic metrics, and 3 environmental factors. The 8 metrics combined with the 10 moving average window sizes as well as the 7 possible time offsets (Table 5.2) generate a total of 560 traffic metric variants per station. The two heuristic ones coupled with 3 window sizes each generate 6 heuristic metric variants. Finally, the environmental factors are introduced as categorical variables, 5 in total. The total number of possible explanatory variables comes to 560*2+6+5=2811. All these variables are calculated automatically and stored in text files. All the variants of a single metric for all four groups are each stored in a single file. The heuristic metrics as well as the environmental variables are stored in separate files.

Step 3: Identification of significant traffic metric variants. Even if it were computationally feasible to fit a logit model with 2811 variables, it is not advisable.

Therefore the procedure must be divided into smaller steps. The eight groups, of 70 metric variants per station, corresponding to the eight temporal and spatial metrics are first fitted separately. Each group was taken through the model fitting process in order to identify the most significant variants that are able to distinguish between cases and controls. This was accomplished with the help of the forward selection method. Metric variants are considered significant if the p-value is less than 0.01. This step results in a considerable reduction of the variables from 1120 to something between 20 and 40 variables. The exact number depends on the particular preprocessing of the raw data. The process takes place in ARC.

Step 4: Explanatory variable transformations. For all the selected metrics from the two stations, transformations are applied to the metrics based on the density distribution. There are two primary reasons for performing transformations: To make the fit of the model linear i.e. normalize the variables and also to achieve homogeneity or constant variance. A transformation of a variable v is simply v^{λ}, where λ is the transformation parameter. The transformation parameter can take any value, but the most useful values are in the interval $-1 \leq \lambda \leq 2$.[Hosmer et al., 1989]. To find the appropriate terms, the conditional density function $f(x|y=j)$, $j=0,1$ of a variable x is used by constructing a histogram. Table 5.4 summarizes appropriate terms for various densities. In the case of normal densities with the same variance then there is no need for transformations whereas in case of normal but with different variance, it is required to include both x and x^2. If the densities $f(x|y=j)$, $j=0,1$ are skewed, both x and log(x) are included in the model so that their relative contributions can be accessed directly.[Hosmer et al., 1989]

Table 5. 4 Appropriate Terms for various Densities function.

Density Function	Terms
Normal, common variance	x
Normal, different variance	x , x^2
Skew	x , log(x)

Step 5: Logit model fitting. At this step, three alternative models are considered. The first two models correspond to the selected from steps 3 and 4 single station metric variants, one set per station including the environmental variables. The third alternative model is the combination of metric variants from both stations, the environmental, and six heuristic ones. The single station models are developed as a final system design alternative (a single station detection system has approximately half the cost of a dual station one). The selection of the most significant variables here is accomplished through the backward elimination method removing all variables with p-values greater than 0.05. Based on the deviance of each model the best fitting the data is selected. As it happened

for all filters as well as for the unfiltered case the model which combines metrics from both stations exhibits the best fit. Due to time limitations and based on the aforementioned fact only two experiments were conducted for all three models i.e., the unfiltered case and the linear smoothing window filter case. For all other experiments only the full, two station, model was developed.

Step 6: Integration with detection algorithm. The model produced in the previous step is integrated with an algorithm capable of operating in real time. The integration is simple. Each time a new vehicle passes over the upstream detector; all relevant variables are calculated and fed to the model. Since it is not efficient to produce an algorithmic decision for every car, several alternative designs that reduce the decision frequency were explored. These designs will be described in detail in a later section.

Step 7: Algorithm testing. The algorithm of step six is evaluated over 10 selected days in pseudo real-time. A performance curve is plotted with one point per threshold considered.

Step 8: Repeat. Select another filter and repeat the process.

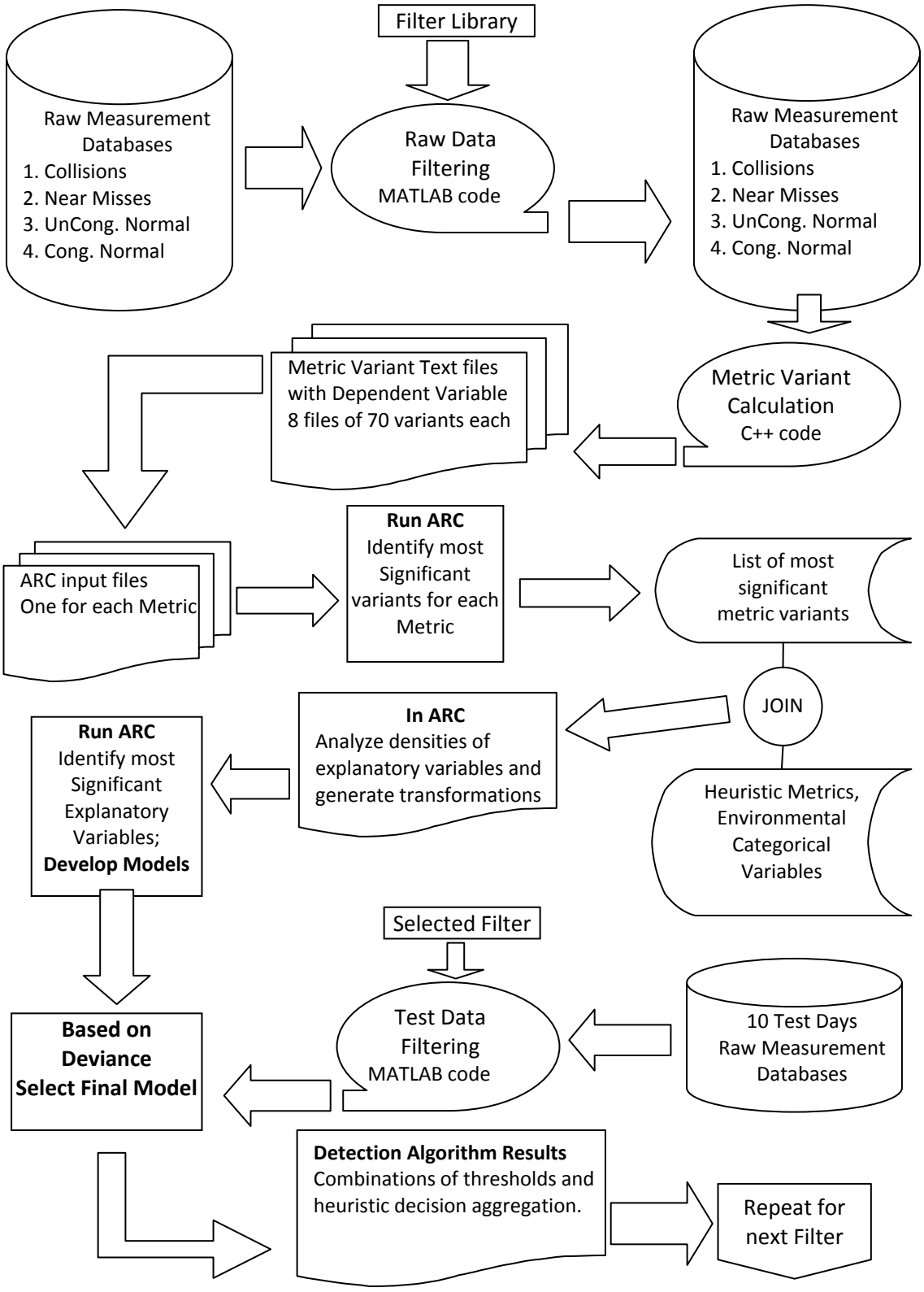

Figure 5. 3 Model development and algorithm testing procedure.

5.4 RESULTS OF MODEL DEVELOPMENT

In total there are five filter alternatives, including the unfiltered case. For each of the alternatives, three models can be developed i.e., two single station models, one using the upstream station and the other the downstream, and the combined dual station model. After some experimentation it was clear that the combined dual station model is superior on all filter alternatives. For brevity, results for all three combinations are presented only for the unfiltered case while for the rest only the dual version is presented.

The results of the model development are summarized in three table groups. The first group has one table presenting the goodness-of-fit measures for each of the models (Table 5.4). The second group of tables covers Tables 5.5 to 5.11. These tables present the model details; estimate, standard error, and p-value for the variable coefficients. Finally, Table 5.12 presents an overall view of the different metric variants and the models in which each one is used. It's worth noting that from the original 2811 possible variables only about 62 are finally considered significant in distinguishing the difference between crash and no-crash prone conditions.

Due to high correlation between variables, it is not possible to assess each variable's contribution to the crash likelihood; however, there are some interesting observations stemming from the results.

1. On all models the position of the sun is considered significant.
2. In regard to speed, one can observe that all the alternatives, except those preprocessed with a low-pass filter, mainly favor speed metric variants with very short time shifts. In contrast, the two low-pass filter alternatives, in general, favor considerably larger time shifts as well as window sizes. One can expect such a behavior since the low pass filters have removed all high frequency patterns and noise that otherwise reduce the resolution as the window gets bigger. The removal of the unwanted elements also accentuates the remaining patterns, allowing for stronger correlation farther from the actual crash time.
3. From the coefficient sign the following can be derived:
    a. The higher the speed in the upstream station, the greater the crash likelihood. Likelihood increases as the speed in the downstream station decreases. This is an indication that the model is influenced by the presence of compression waves.
    b. Crash likelihood increases as the speed in the middle lane increases. This is observed on both stations. Additionally, the likelihood increases as the speed difference between the right and middle lanes increases.
    c. As time headways increase, the crash likelihood decreases.
    d. Wet pavement, as well as reduced visibility due to rain or snow, marginally increases the crash likelihood. This contradicts some of the earlier findings by other researchers (see chapter 2) stating that bad weather reduces crash

probability. Further analysis to discover why this is the case in this location is warranted.

The model behavior as described by the speed level variables is indicative of the correlation between compression waves and crash occurrence. Additionally, prior to a breakdown, the stream operates under "synchronized flow" [Jiang et al., 2003]. During synchronized flow vehicles move very fast, too close to each other, which can explain the low variability in speed right before the crash. However, during the period before the initiation of synchronized flow the speed variability is higher. Synchronized flow does not last long since very small disturbances in flow due to vehicles merging in or sudden braking by a single vehicle result in the formation of compression waves that do not dissipate but travel quickly upstream.

As shown in earlier human factor studies [Marshal et al., 1998], during low light conditions drivers perceive the stop lights of leading vehicles faster even when distracted. However, when the sun shines on the windshield, it adds glare and may temporary blind drivers, making it more difficult to keep the proper distance.

Table 5. 5 Deviance and Max-rescaled R-square for developed models.

No.	Model	Deviance
1.	Single Station Model - Upstream (Un-Filtered Measurements)	253.553
2.	Single Station Model – Downstream (Un-Filtered Measurements)	178.848
3.	Dual Station Model (Un-Filtered Measurements)	158.246
4.	Dual Station Model (Linear smoothing Filter)	134.300
5.	Dual Station Model (Median (7) Filter)	106.068
6.	Dual Station Model (Low-pass IIR Butterworth)	123.702
7.	Dual Station Model (Low-pass FIR Hamming)	97.126

Table 5. 6 Un-filtered single station (upstr.) model.

Variable name	Estimate	Standard error	p-value
Constant	-14.4825	1.7016	0.0000
AS-15-10-a	0.6389	0.1003	0.0000
AS-15-10-a^2	-0.0124	0.0018	0.0000
AS-M-30-30-a	0.1724	0.0335	0.0000
TH-120-300-a	-1.5073	0.5210	0.0038
TP-100-30-a	0.0003	0.0001	0.0134
MaxMinDiff-30-a	0.0340	0.0101	0.0007
Sun_{night}	-0.9511	0.7561	0.2085
Sun_{side}	0.1457	0.4497	0.7460
Sun_{front}	0.7862	0.4319	0.0687
Visibility	0.2665	0.7387	0.7183
Pavement	0.7475	0.8039	0.3524

Table 5. 7 Un-filt single station (downst.) model.

Variable name	Estimate	Standard error	p-value
Constant	-79.3105	16.0849	0.0000
AS-M-120-300-b^2	0.0000001	0.0000	0.7645
log[CV-15-10-b]	2.3343	0.5081	0.0000
log[KE-15-10-b]	7.7151	1.7477	0.0000
TH-100-300-b	11.7311	3.2597	0.0003
TH-120-300-b	-11.9298	3.2510	0.0002
AS-120-300-b^2	-0.0177	0.0033	0.0000
KE-15-10-b	-0.0003	0.0001	0.0000
AS-120-300-b	1.0947	0.1792	0.0000
Sun_{night}	-1.2205	0.9084	0.1791
Sun_{side}	1.0039	0.6545	0.1250
Sun_{front}	1.2917	0.5080	0.0110
Visibility	-0.4606	0.9814	0.6388
Pavement	2.0039	1.0629	0.0594

Table 5. 8 Un-filtered dual station model.

Variable name	Estimate	Standard error	p-value
Constant	-16.0627	9.5712	0.0933
AS-15-10-b^2	-0.0071	0.0015	0.0000
AS-M-30-30-a^2	0.0040	0.0006	0.0000
CV-100-10-a	16.0520	6.3615	0.0116
CV-100-30-b	-21.4343	9.6833	0.0269
CV-70-60-a	-13.6168	5.8046	0.0190
log[CV-100-30-b]	6.4464	3.0972	0.0374
log[CV-15-10-b]	1.4162	0.5193	0.0064
log[KE-15-30-b]	2.9222	0.7561	0.0001
TH-120-300-a	-1.5655	0.5302	0.0032
TP-120-300-b	0.0004	0.0001	0.0087
Sun_{night}	-0.8702	0.8958	0.3313
Sun_{side}	0.3078	0.6583	0.6401
Sun_{front}	1.1494	0.5582	0.0395
Visibility	0.4520	0.9618	0.6384
Pavement	0.8466	0.8425	0.3149

Table 5. 9 Linear smoothing filter model.

Variable name	**Estimate**	**Standard error**	**p-value**
Constant	-32.7814	17.116	0.0962
AS-120-300-b	0.22867	0.06191	0.00020
AS-15-120-a	0.32461	0.12515	0.00950
AS-15-120-a^2	-0.00704	0.00262	0.00710
AS-15-30-b	-0.27744	0.10227	0.00670
AS-M-30-10-a^2	0.00421	0.00091	0.00000
CV-110-10-b	-49.1112	21.9112	0.02500
KE-15-10-b	-0.00008	0.00004	0.02790
log[CV-110-10-b]	12.5625	5.58787	0.02460
log[KE-15-30-b]	4.62125	1.31443	0.00040
MaxMinDiff-30-b	0.05188	0.01637	0.00150
MVG-120-300-b	134.746	40.6807	0.00090
TP-100-30-a	-0.00195	0.00098	0.04740
TP-110-10-a	0.00260	0.00102	0.01060
Sun_{night}	-1.89274	1.15225	0.1005
Sun_{side}	0.658823	0.7118	0.3547
Sun_{front}	1.19134	0.5953	0.0454
Visibility	0.39735	1.18592	0.73760
Pavement	0.80218	1.08341	0.45900

Table 5. 10 Median(7) filter model.

Variable name	**Estimate**	**Standard error**	**p-value**
Constant	-35.5278	9.49426	0.0002
AS-15-10-b	-0.227267	0.0901220	0.0117
AS-M-30-30-a	0.211445	0.0577941	0.0003
log[AN-120-300-a]	5.38940	1.24463	0.0000
log[QF-15-10-b]	1.87225	0.883870	0.0342
log[TH-120-300-a]	-9.68875	3.75785	0.0099
log[TP-110-10-a]	1.56533	0.494205	0.0015
MinMax-15-b	-0.0237586	0.0351859	0.4995
MinMax-15-b^2	0.000572170	0.000423355	0.1765
MVG-60-300-a^2	-730.982	258.704	0.0047
QF-120-300-b^2	-7.078303E-8	4.379100E-8	0.1060
RM-15-10-a	1.15962	1.93215	0.5484
TH-120-300-a	5.94426	3.51171	0.0905
UpDown-15-R	2.29297	2.93872	0.4352
Sun_{night}	0.257255	1.37905	0.8520
Sun_{side}	-0.837523	0.745696	0.2614
Sun_{front}	2.59843	0.845205	0.0021
Visibility	0.0324266	1.37168	0.9811
Pavement	3.71372	1.48834	0.0126

Table 5. 11 Low Pass "Butterworth" filter model.

Variable name	**Estimate**	**Standard error**	**p-value**
Constant	-85.8426	17.7586	0.0000
AN-120-300-a	6.25212	1.9197	0.0011
AS-15-120-a	0.179198	0.117597	0.1276
AS-15-120-a^2	-0.00353	0.002408	0.1422
AS-M-15-180-a^2	0.002851	0.000892	0.0014
KE-15-10-b^2	-1.35E-09	4.34E-10	0.0019
log[KE-15-10-b]	5.49572	1.36315	0.0001
log[MVG-80-240-b]	-2.42846	1.03661	0.0191
log[TH-120-300-b]	-3.98227	1.89546	0.0356
MinMax-30-a	0.094965	0.042089	0.0241
MinMax-30-a^2	-0.0009	0.00053	0.0877
MVG-50-300-a	539.497	155.555	0.0005
MVG-50-300-a^2	-8642.48	2173.52	0.0001
MVG-120-300-a^2	977.149	506.621	0.0538
SV-60-10-b^2	11.6378	3.90795	0.0029
TH-120-300-a	-6.7466	2.59566	0.0093
TH-120-300-b^2	2.60853	0.787413	0.0009
TP-110-10-a	0.000467	0.0002	0.0196
TP-110-10-b	-0.00017	4.88E-05	0.0005
UpDown-15-R	8.09169	1.76835	0.0000
Sun_{night}	0.380087	1.35297	0.7788
Sun_{side}	-0.79372	0.643246	0.2172
Sun_{front}	1.32544	0.708112	0.0612
Visibility	-0.07961	1.39059	0.9543
Pavement	2.85318	1.47566	0.0532

Table 5. 12 Low Pass "Hamming" filter model.

Variable name	**Estimate**	**Standard error**	**p-value**
Constant	-90.0408	25.2009	0.0004
AN-120-300-b	137.664	58.5105	0.0186
AN-120-300- b^2	-120.600	55.3432	0.0293
AS-110-30- a^2	-0.00639662	0.00242728	0.0084
AS-M-120-60-a^2	0.00833956	0.00213203	0.0001
AS-50-10-a	-0.239594	0.128247	0.0617
KE-70-10-a	0.0000984530	0.0000592931	0.0968
KE-70-10-b	0.000488608	0.000139616	0.0005
KE-70-10- b^2	-5.271554E-9	1.381270E-9	0.0001
log[AN-120-300-b]	-12.9692	5.95459	0.0294
log[QF-120-300-b]	2.15917	1.40134	0.1234
log[TH-120-300-a]	-11.4075	3.72697	0.0022
MinMax-15-b	0.0434565	0.0156707	0.0056
MVG-120-300- a^2	4761.88	1424.88	0.0008
QF-100-300-a	0.00164756	0.000770096	0.0324
QF-100-300- a^2	-1.294356E-7	4.493798E-8	0.0040
RM-40-10-b	3.64025	2.62734	0.1659
TH-120-300-a	6.25472	2.24819	0.0054
TP-120-300-a	-0.000318061	0.000222849	0.1535
TP-15-10-b	-0.00100693	0.000346782	0.0037
UpDown-30-R	5.63527	2.13610	0.0083
Sun_{night}	1.07476	1.49339	0.4717
Sun_{side}	-1.35631	0.870132	0.1191
Sun_{front}	1.71887	0.897545	0.0555
Visibility	-0.939466	1.66126	0.5717
Pavement	2.61431	1.71002	0.1263

Table 5. 13 Summary table of all predictor variables per model (spans three pages, X indicates variable use in model).

Explanatory Variables	Un-filtered Downstream Model	Un-filtered Upstream Model	Un-filtered Two Stations Model	Linear Smoothing Filter Model	Median Filter Model	Low Pass 1 Model	Low Pass 2 Model
Constant	X	X	X	X	X	X	X
AvgSp-15-10-a		X					
AvgSp-15-120-a				X		X	
AvgSp-15-240-a							
AvgSp-50-10-a							X
AvgSp-110-30-a							X
AvgSp-120-300-a						X	
AvgSp-15-10-b			X		X		
AvgSp-15-30-b				X			
AvgSp-120-300-b	X			X	X		
AvgSp-M-15-180-a						X	
AvgSp-M-30-10-a				X			
AvgSp-M-30-30-a		X	X		X		
AvgSp-M-120-60-a							X
AvgSp-M-120-300-b	X						

Explanatory Variables	Un-filtered Downstream Model	Un-filtered Upstream Model	Un-filtered Two Stations Model	Linear Smoothing Filter Model	Median Filter Model	Low Pass 1 Model	Low Pass 2 Model
AN-120-300-a					X		
AN-120-300-b							X
CV-70-60-a			X				
CV-100-10-a			X				
CV-15-10-b	X		X				
CV-100-30-b			X				
CV-110-10-b				X			
KE-15-120-a					X		
KE-70-10-a							X
KE-15-10-b	X			X		X	
KE-15-30-b			X	X			
KE-70-10-b							X
QOF-100-300-a							X
QOF-120-300-a						X	
QOF-15-10-b					X		
QOF-120-300-b					X		X

Explanatory Variables	Un-filtered Downstream Model	Un-filtered Upstream Model	Un-filtered Two Stations Model	Linear Smoothing Filter Model	Median Filter Model	Low Pass 1 Model	Low Pass 2 Model
TP-100-30-a		X		X			
TP-110-10-a				X	X	X	
TP-120-300-a							X
TP-15-10-b							X
TP-30-120-b							
TP-120-300-b			X				
TH-120-300-a		X	X		X	X	X
TH-100-300-b	X						
TH-120-300-b	X					X	
MVG-50-300-a						X	
MVG-60-300-a					X		
MVG-120-300-a					X	X	X
MVG-15-10-b							
MVG-80-240-b						X	
MVG-120-300-b				X			

Explanatory Variables	Un-filtered Downstream Model	Un-filtered Upstream Model	Un-filtered Two Stations Model	Linear Smoothing Filter Model	Median Filter Model	Low Pass 1 Model	Low Pass 2 Model
UpDownDiff-15-R					X	X	
UpDownDiff-30-R							X
MaxMinDiff-15-a						X	
MaxMinDiff-15-b					X		X
MaxMinDiff-30-a		X		X		X	
RM-15-10-a					X		
RM-15-10-b					X		
RM-15-60-b					X		
RM-40-10-b							X
RM-40-60-a						X	
SV-60-10-b						X	
Sun_{night}							
Sun_{side}					X		
Sun_{front}	X	X	X	X	X	X	X

Explanatory Variables	Un-filtered Downstream Model	Un-filtered Upstream Model	Un-filtered Two Stations Model	Linear Smoothing Filter Model	Median Filter Model	Low Pass 1 Model	Low Pass 2 Model
Visibility	X						
Pavement					X	X	

5.5 MODEL ADJUSTMENT FACTOR

In the previous section a number of logistic regression models were developed to estimate crash likelihood based on real-time information. As is the case with any statistical model the process is based on a sample containing as many crashes as was possible to collect in the allotted time and an even bigger number of crash-free conditions. Computational and other restrictions do not allow the use of the entire population in the development of the model, so case-control sampling was used. Therefore, it is not guaranteed that the ratio between crashes and non-crash cases between the sample and the entire population is the same. Actually, in the dataset used in this work the ratio between crashes and normal conditions is approximately 1:2, while in reality crashes are relatively rare events even in high crash locations like the study area. This discrepancy results in the overestimation of the probability of a crash, prohibiting the use of the model to predict crashes. For the model to be used prospectively in real-time estimation of crash likelihood it needs to be adjusted [Collett, 1991, Hosmer et al., 1989]. During the course of this study two adjustment methods were proposed. Due to the very different nature of both methods we will provide a short explanation of each although the second one was the one finally utilized.

Method 1: Model Invariant Process [Garg, 2004]

The conditional probability of a crash sampled in the study is given by,

$$p_o(x) = \frac{\pi_1 p(x)}{\pi_1 p(x) + \pi_2 \{1 - p(x)\}} \qquad (5.\ 8)$$

Where,

π_1 = Sampling fraction for the crashes i.e., proportion of crashes sampled from the crash population

π_2 = Sampling fraction for the controls (normal conditions) i.e., proportion of crash free sampled from the crash free population.

$p_o(x)$ = Probability of a crash in the sample included in the study (estimated).

p(x)=Probability an crash in the population (unknown).

If the linear logistic model is adopted for the probability in the population, then

$$\text{Logit } \{p(\mathbf{x})\} = \beta_o + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k \qquad (5.\ 9)$$

Equation 5.8 can be written as

$$p_o(x) = \frac{1}{1 + \frac{\pi_2}{\pi_1}\left[\frac{1 - p(x)}{p(x)}\right]} \qquad (5.\ 10)$$

similarly,

$$\frac{p_o(x)}{1-p_o(x)} = \frac{\dfrac{1}{1+\dfrac{\pi_2}{\pi_1}\left[\dfrac{1-p(x)}{p(x)}\right]}}{1-\dfrac{1}{1+\dfrac{\pi_2}{\pi_1}\left[\dfrac{1-p(x)}{p(x)}\right]}}$$

or,

$$\left[\frac{p_o(x)}{1-p_o(x)}\right] = \left[\frac{\pi_1}{\pi_2}\right]\left[\frac{p(x)}{\{1-p(x)\}}\right] \qquad (5.\ 11)$$

Take log on both sides in Equation 5.11,

$$\log\frac{p_o(x)}{1-p_o(x)} = \log\frac{\pi_1}{\pi_2} + \log\frac{p\ (x)}{1-p\ (x)}$$

$$\text{Logit}\{p_o(\mathbf{x})) = \log(\frac{\pi_1}{\pi_2}) + \text{logit}\{p(x)\} \qquad (5.\ 12)$$

Therefore, the probability that the crash is an event in the study is given by

$$\text{Logit}\{p_o(\mathbf{x})\} = \log(\frac{\pi_1}{\pi_2}) + \beta_o + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k \qquad (5.\ 13)$$

Or, $\text{Logit}\{p_o(\mathbf{x})\} = \beta_o' + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$

Where, $\beta_o^{'} = \beta_o + \log(\frac{\pi_1}{\pi_2})$ (5.14)

From Equation 5.9 and 5.14, it is implied

$$\text{Logit}\{p(\mathbf{x})) = \beta_o^{'} - \log(\frac{\pi_1}{\pi_2}) + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$$

Hence the corrected estimate is given by: $\beta_o = \beta_o^{'} - \log(\frac{\pi_1}{\pi_2})$ (5.15)

Where,

β_o = Adjusted coefficient estimate

$\beta_o^{'}$ = Original coefficient estimate

β_0 can be estimated using the two sampling fractions π_1 and π_2, assuming the information of how many crashes and crash free conditions in the population are been sampled. Therefore the next step is to calculate the sampling fraction of crashes and crash free conditions.

Method to Estimate π_1 and π_2 :

$$\pi_1 = \frac{\text{Total crashes Sampled}}{\text{Total crashes in population}} = \frac{n_1 * N}{n * N}$$

$$\pi_2 = \frac{\text{Total Non crashes Sampled}}{\text{Total Non crashes in population}} = \frac{n_2 * N}{T - (n * N)}$$

Where,

n_1=Number of crashes sampled from the population

n=Total number of crashes in the population

n_2=Number of non-crash events sampled

T=Total number of vehicles in the population

N=Number of vehicles involved in each Event

Using data from this chapter

n_1= 148

n=224 (crashes that occurred during the analysis period)

n_2= 267

T=5*30*25,000= 3,750,000 (Assuming 25,000 vehicles per day in one lane)

N=120 (Since the maximum number of vehicles used to estimate the longest traffic metric in the model is 120, it is assumed that approximately 120 vehicles are likely to effect the likelihood of crash occurrence. It implies that each crash case involves 120 vehicles prior to the crash)

$$\pi_1 = \frac{148 * 120}{224 * 120} = 0.6601714$$

$$\pi_2 = \frac{267 * 120}{3750000 - (224 * 120)} = \frac{32040}{3750000 - 26880} = 0.008606$$

Therefore by using above values of π_1 and π_2 , the adjustment factor is calculated as:

$$\ln(\frac{0.6601714}{0.008606}) = 4.340898$$

Initially different values of N are used to estimate the sampling fraction ratio. Table 5.14 shows the sampling factors for different values of N. The value of π_1 remains constant for all values of N but π_2 changes from 0.001069 to 0.008606 as N varies from 5 to 120 vehicles. The adjustment factor range from 6.427 to 4.340 estimated by varying π_1 and π_2 . For the aforementioned assumption the sampling fraction corresponding to the value of N as 120 is used to adjust the constant estimate.

Table 5. 14 Adjustment Factor for Different values of N.

N	15	30	60	90	120
π1	0.660714	0.660714	0.660714	0.660714	0.660714
π2	0.001069	0.00214	0.004287	0.006443	0.008606
ln(π1/π2)	6.426637	5.732593	5.037649	4.630384	4.340898

For comparison, the output produced by the model in Table 5.8 (linear filter) after applying the adjustment factor is presented in table 5.15 for various N along with the original constant estimate model results. In the case of events when the chances of crash prone conditions are certain, both prior and posteriori likelihood are close to 1. Table 5.15 shows that the adjusted likelihood increases as N increases from 15 to 120. Thus the selection of N affects the adjusted likelihood. However, the relative trend is the same for all values of N i.e., the adjusted likelihood increasing trend is the same with the unadjusted likelihood one.

Table 5. 15 Comparison of Likelihood after applying adjustment factor for different values of N.

Without Adj. Factor	With adjustment factor				
	N=15	N=30	N=60	N=90	N=120
0.00010	0.00000	0.00000	0.00000	0.00000	0.00000
0.00256	0.00000	0.00001	0.00002	0.00003	0.00003
0.02515	0.00004	0.00008	0.00017	0.00025	0.00034
0.03579	0.00006	0.00012	0.00024	0.00036	0.00048
0.12806	0.00024	0.00047	0.00095	0.00143	0.00191
0.61396	0.00257	0.00511	0.01021	0.01527	0.02029
0.63402	0.00279	0.00556	0.01112	0.01661	0.02207
0.78663	0.00593	0.01176	0.02336	0.03470	0.04582
0.85072	0.00914	0.01806	0.03566	0.05264	0.06910
0.90145	0.01458	0.02868	0.05603	0.08189	0.10645
0.94224	0.02572	0.05002	0.09573	0.13724	0.17525
0.98101	0.07713	0.14292	0.25106	0.33499	0.40222
0.99122	0.15441	0.26703	0.42276	0.52394	0.59515
0.99495	0.24174	0.38878	0.56115	0.65771	0.71962
0.99671	0.32880	0.49428	0.66271	0.74700	0.79773
0.99956	0.78482	0.87918	0.93601	0.95649	0.96706
0.99980	0.88946	0.94136	0.96995	0.97980	0.98480

This method makes a number of assumptions some readers might find difficult to accept or even have the information available. Regardless, one characteristic of the method is that the estimate adjustment is irrelevant to the actual model it will be applied to. Basically, the underlying assumption is that there is only one true model. As presented in the previous section, different input information, in our case filtered measurements, will produce models with different coefficient estimates. Which is correct? They all produce estimates of the likelihood, each from slightly different perspective. This method might be mathematically valid but needs to be used with extreme caution.

Method 2: Poisson Process

Assuming that, during a time interval (0,t) there are N(t) crashes occurring in the study area and that the crash occurrence follows a point process with no history (future crash occurrence does not depend on past ones) then, we can treat N(t) as if is generated by a non-stationary Poisson

process. Further, we can define the Poisson process intensity function as the result of a generalized linear model,

$$\lambda(t) = e^{(\beta_o + \beta' x(t))}$$

If the duration of the interval is short enough to guarantee that x(t) can be considered constant, meaning x(t)=x, the number of crashes occurring during the interval is a Poisson random variable with expected value

$$E[N(t)] = e^{\beta_o + \beta' x(t)} t$$

If we condition N(t) to be either 0 or 1, then

$P[N(t)=1] = \dfrac{e^{\beta_o + \ln(t)\beta' x(t)}}{1 + e^{\beta_o + \ln(t)\beta' x(t)}}$ which is the logit model. As we did in the previous section, based on a sample collection of cases and controls we can compute consistent estimates of the parameters vector **b**, but not of the constant term b_0.

Assume data are collected each day during an interval (0,T] for a total of p days indexed with k=1,…,p. If T is long enough so x(t) is not a constant, then the expected value would be

$$m_k = E[N_k] = \int_0^T e^{\beta_o + \beta' x_k(s)} ds$$

Where N_k is the total crash count for day k. if $N = \sum N_k$, then N is also Poisson with expected value $\sum m_k$. If n is the actual crash count, and if the vector **b** is known, then

$$n = \sum_k \int_0^T e^{\beta_o + \beta' x_k(s)} ds = e^{\beta_o} \sum_k \int_o^T e^{\beta' x_k(s)} ds$$

Therefore the method of moments estimate will be

$$\beta_o = \ln\left[\frac{n}{\sum_k \int_0^T e^{\beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_k x_k} ds}\right]$$

Considering that measurements are collected for every vehicle, we can assume that between measurements the metric remains static, therefore, the integral can be approximated with a sum. So,

$$\beta_o = \ln\left[\frac{n}{\sum_k \sum_0^T e^{\beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_k x_k}}\right].$$

In difference to the previous method the adjusted coefficient depends not only on the number of crashes in the population but also on the crash likelihood estimation model. The only negative

part of this method is that all the calculation must take place over the entire population or at least a large enough chunk so the ratio is correct.

The models developed in this study are adjusted through this methodology for the entire 5 months of the study with data for each day between 7am and 8pm. In the following table the adjusted coefficient estimates are presented.

Table 5. 16 Adjusted coefficient estimates for Crash likelihood models.

Model	Linear Filter	Median Filter	IIR Butterworth	FIR Hamming
Unadjusted Coefficient	-32.7814	-35.5278	-85.8426	-90.0408
Adjusted Coefficient	-49.6541	-82.0407	-261.1701	-113.1529

5.6 ALGORITHM DEVELOPMENT AND TESTING

There are countless ways to integrate a prediction model in a detection algorithm. One can simply streamline the model into a program producing a crash likelihood estimate every time a new vehicle is measured, compare this estimate with a threshold and decide to raise or not to raise the alarm. Considering that the crash likelihood will not be used for actual prediction of risk level, alternatively, one can program the algorithm to produce an estimate every second or minute, even take the mean of 10 or 30 consecutive estimates and compare that to a threshold. The goal is to have an efficient and effective detection algorithm. In the course of this study several alternatives were tested but after a while it became clear that this subject is closely linked with the intended use of the detection algorithm. Although in the next chapter a preliminary assessment of requirements and designs for a crash prevention system are discussed, this study was not intended to investigate this subject in detail. For the purpose of testing the algorithms/models in this work we choose to implement two alternatives that showed promise during the preliminary investigation. The crash likelihood is estimated every time a new vehicle crosses the upstream detector. Based on this the two integration methods are:

1. Calculate the median of the crash likelihood of N consecutive vehicles and compare it to a threshold.
2. In addition to the above, compare the speeds in the middle and right lane and raise the alarm when both comparisons pass their respective thresholds.

The selection of the number N for the median window has a very large impact on false alarm rate and a smaller effect on the detection rate scores. Specifically, following observations of the operation of the algorithm it became clear that, at least for the versions integrating the FIR and IIR filtered models, the crash likelihood is raised and stays high for a period of time. This differs

from the response observed from the linear and specifically the median filter versions where sharp peaks of short duration were more common. The smaller the N, the more alarms are raised in the first two cases while for the same N there might be no alarms at all for the latter since the estimate does not stay high long enough to affect the median. Therefore the selection of N was decided on a trial basis separately for each algorithm version.

As noted in the previous paragraph, the two digital filters prompt the model to produce large increases of the crash likelihood for longer periods of time. This is actually a good sign since we would not expect the safety indicator of the traffic stream to fluctuate greatly between successive vehicles. This demonstrates that the digital filters succeeded in removing a much larger amount of higher frequency patterns that can cause a more erratic behavior of the model (as can be seen in the median filter). This stable behavior on behalf of the digitally filtered models poses some serious problems with integration in the algorithm. The crash likelihood reaches such high levels that it is very difficult to find appropriate thresholds to calibrate the response. Here is where the likelihood adjustment factor was instrumental. As explained earlier, the adjustment factor reduces the crash likelihood to its normal levels expressing the fact that crashes are rare events even in high crash areas. As was illustrated, the response of the model to different levels of the adjustment factor is not linear therefore it helps to narrow the peaks produced by the model as well as keep the smooth response. Again it should be stressed in this point that these model/algorithm behavior adjustments are beneficial to the selected methodology for raising the alarm. Other implementations might not benefit therefore the reader should consider them as suggestions instead of the best proposed method.

The algorithms/models were tested in real time with measurements collected over a number of days in order to assess its performance. These days were not used in the development of the models. Observations and measurements from 10 selected days were collected. Although, any number of days would be acceptable, the days summarized in Table 5.17 were selected based on the prevailing weather conditions as well as the number of observed crashes. Specifically, days with rain, snow, and bright sunshine were selected during summer, fall, and winter 2003. Additionally, in this group there are days that experienced a lot of crashes and days with none. The corresponding video records were closely scrutinized prior to the tests to extract the complete and accurate ground truth for the evaluation. The number of crashes that occurred on each day is shown in Table 5.17.

Table 5. 17 Days selected for evaluation and general conditions.

Day	**Date selected**	**General weather conditions**	Number of Collisions and Near Misses
1.	Aug 15, 2003	Clear and Sunny	8
2.	Aug 26, 2003	Clear and Sunny	11
3.	Sep 05, 2003	Clear and Sunny	6
4.	Sep 18, 2003	Rainy	4
5.	Oct 11, 2003	Rainy	10
6.	Nov 11, 2003	Clear and Sunny	4
7.	Nov 15, 2003	Cloudy and Rainy	7
8.	Nov 19, 2003	Bright and Clear	3
9.	Nov 26, 2003	Cloudy and Sunny	4
10.	Jan 22, 2004	Cloudy and Sunny	3

For the 10 days, between 7 a.m. and 8 p.m., the model estimates the likelihood of a crash every time a new vehicle passes over the right lane detector on the upstream station. A number of thresholds were selected in order to evaluate the performance of the algorithms. Each implementation requires its own set of thresholds.

In order to evaluate the model, detection rate, false decision rate and false alarm rate were selected as the measures of the effectiveness of the model. Following are the measures selected:

5.6.1 Detection Rate

The detection rate is the ratio of the number of crashes detected by the model to the actual number of crashes in the data set. First, each of the selected day's total crashes detected by model is compared with actual number of crashes. The detection rate is calculated by combining the performance from all testing days. An event is classified as valid detection if the crash occurs after the probability rises above a given threshold.

Mathematically,

$$DR = \frac{C_D}{C_T} \times 100 \quad (5.16)$$

Where,

DR = Detection rate (%).

C_D = Crashes detected by the model

C_T = Total number of crashes

5.6.2 False Decision Rate

False decision rate can be defined in various ways. In this study, false decision rate is the fraction of false alarms to the total number of model decisions. A false alarm occurs when the model signals a crash and there is none. The total number of false alarms is calculated by subtracting the number of crashes detected by the model from the number of alarms produced. The total number of decisions equals the number of vehicles that passes over the right lane detector on upstream station A, from 7 a.m. to 8 p.m. for all testing days.

Mathematically,

$$FDR = \frac{A_F}{D_T} \times 100 \quad (5.17)$$

Where,

FDR = False decision rate (%).

A_F =Number of false alarms.

D_T = Total number of decisions by the model.

5.6.3 False Alarm Rate

False Alarm Rate has been defined differently, following a different perspective of what a false alarm is. It is the fraction of false alarms to the total number of alarms produced by the model.

Mathematically,

$$FAR = \frac{A_F}{A_T} \times 100 \quad (5.18)$$

Where,

FAR = False alarm rate (%).

A_F = Number of false alarms.

A_T = Total number of alarms given by the model.

The evaluation results are presented with the help of a performance chart graph of Detection Rate vs. False Decision Rate. Since different sets of thresholds are selected for each alternative the different performance curves are lined up according to False Decision Rate. The performance graph for the version using the median of N vehicles can be found on figure 5.4 while the performance graph of the double test of Median and speed difference between the right and middle lanes is presented on figure 5.5. Both graphs display the expected trade-of between detection and false alarm rates. The algorithms based only on crash likelihood achieve higher detection rates but also experience high false decision rates. On the contrary, with the introduction of the speed difference heuristic test, a cap is enforced on detection performance but the FDR is considerably lower. In both cases the FIR filter was the one with the most stable performance.

5.7 CONCLUSION

The data analysis presented in previous chapters produced strong evidence on the existence of traffic conditions that favor crashes. The next logical step is to find a way to detect these conditions and trigger actions that will reduce the likelihood of a crash. The method that showed the greatest potential was based on statistical analysis and specifically in the development of models for the estimation of crash likelihood. This chapter described the process and results of this development.

Based on the real-time measurements and metrics presented in chapter 3, a very large number of metric variants were developed. Through the process of logistic regression the variants that exhibit the largest correlation with pre-crash conditions were incorporated into crash likelihood estimation models. Based on the data filters presented in chapter 6, one such model per filter was developed. These models were then integrated into crash-prone condition detection algorithms combining them with more realistic detection aggregation periods as well as a heuristic test that enhanced performance. The crash-prone traffic condition detection algorithms were tested for additional 10 days of varying weather and traffic conditions not earlier employed in the model & algorithm development. From a number of threshold groups identified through a trial and error process, detection performance curves were developed and presented. From the performance curves it was found that the model with the most successful balance between detection rate and false decision rate was the one developed from data filtered by the FIR Hamming low-pass digital filter. Finally, it was confirmed that in general the models developed with filtered data outperform the ones developed with unfiltered measurements.

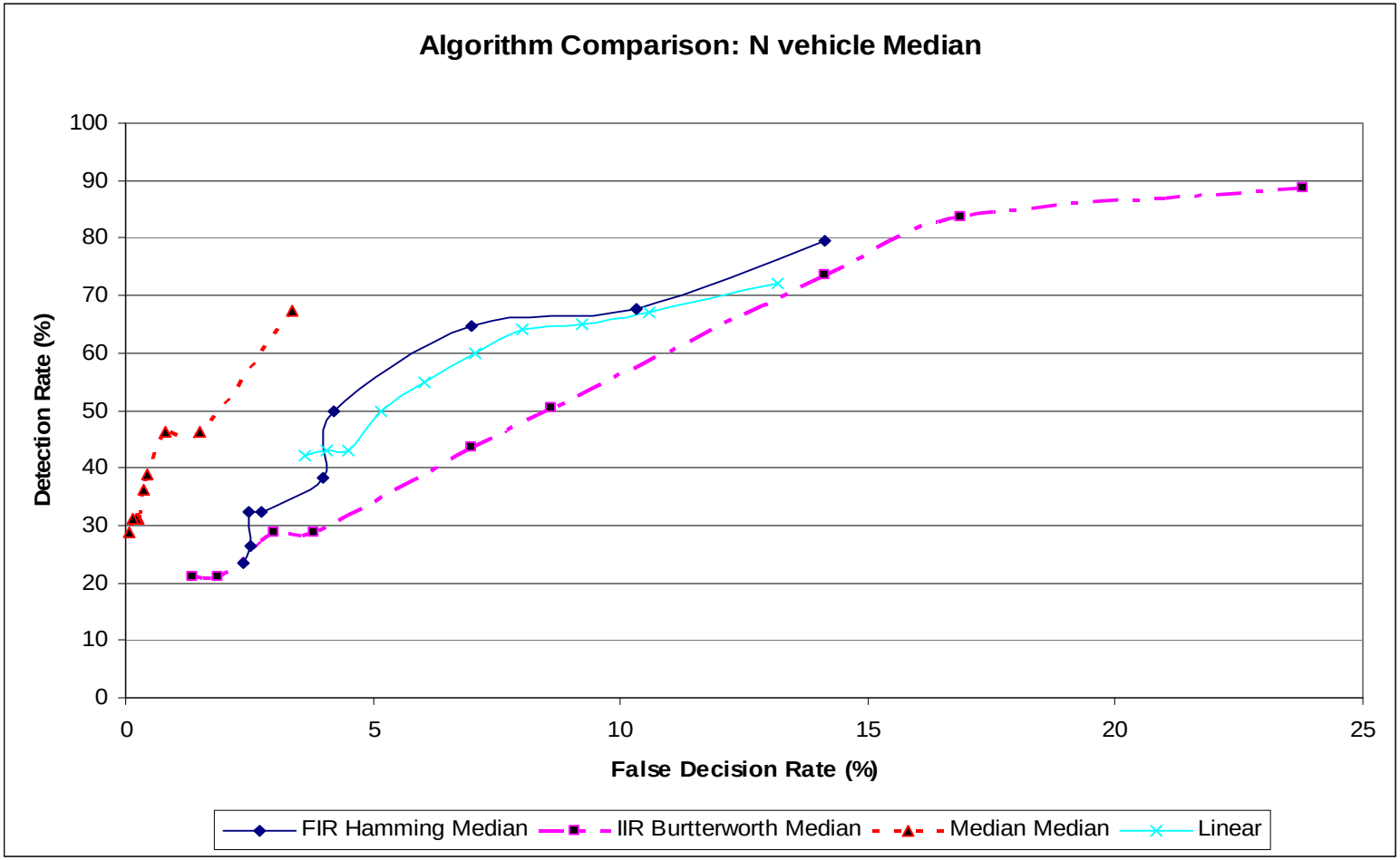

Figure 5. 4 Crash prone conditions detection algorithm performance curves: N vehicle median.

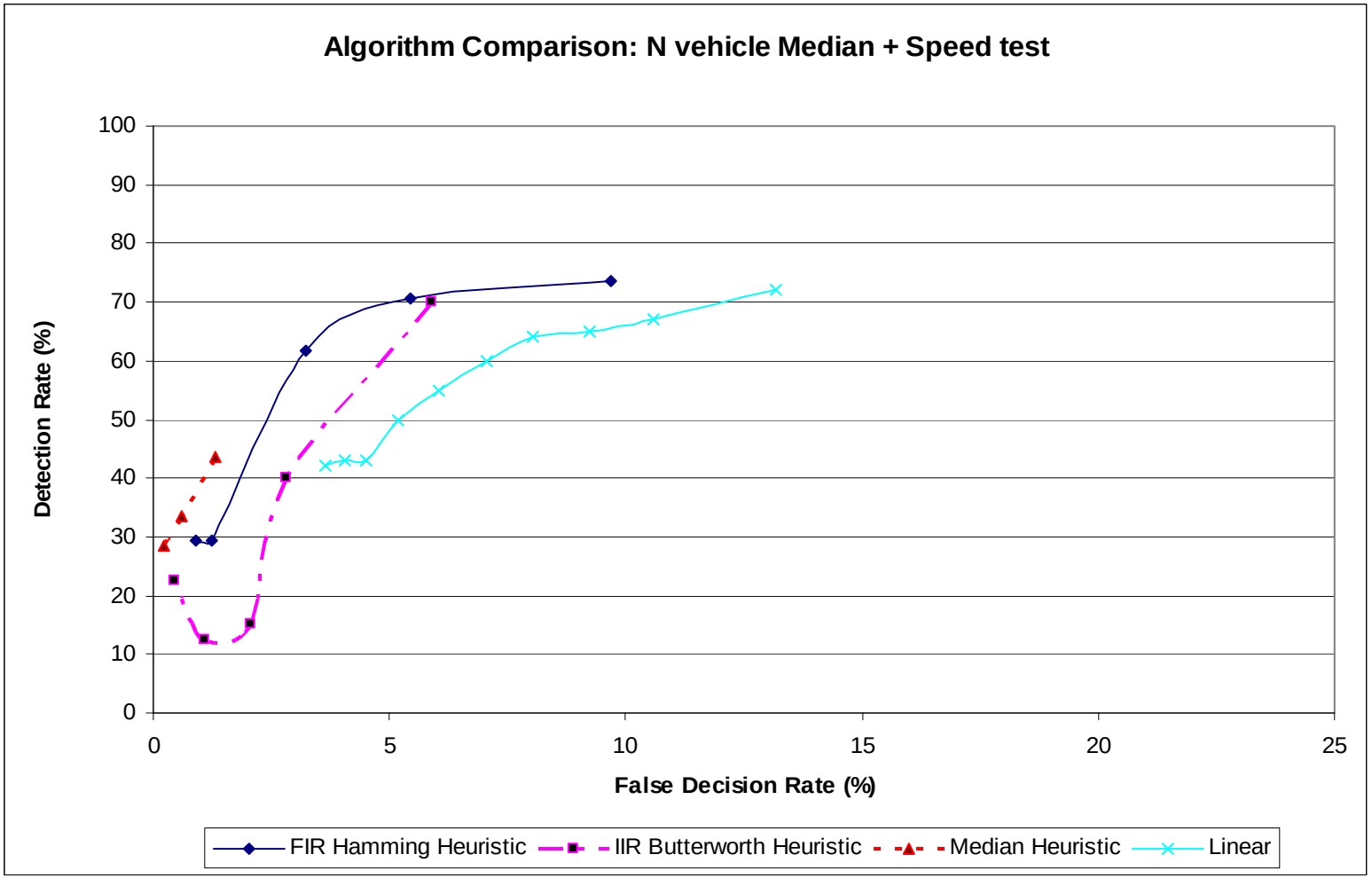

Figure 5. 5 Crash prone conditions detection algorithm performance curves: N vehicle median + Speed.

6 Preliminary Implementation Framework

Although the objectives of this study were to present evidence that crash prone conditions exist and if so, develop a detection methodology based on real-time data, the work will not be complete if at least a preliminary implementation framework is proposed. In the literature review chapter, a few simple, straightforward approaches were presented. Those were the only ones that passed the point of research and actually implemented something in the field. During the course of the study, visions and plans of a crash prevention system were often considered, reaching always the same realization; it is not a simple thing to design an effective and efficient crash prevention solution, especially one that involves interaction with the drivers.

Based on the analysis the specific factors that need to be influenced, changed, and/or controlled to succeed in preventing crashes are:

1. Flow breakdown
2. Driver inattention
3. Dangerous driving

Although these factors can be influenced in many ways, in this chapter only the ones related to real-time actions driven by a crash prone condition detection system are discussed.

6.1 FLOW BREAKDOWN FACTOR

Based on the available evidence presented in chapter seven the cause of the flow breakdown and the initiation of compression wave activity is the merging traffic from the I-94 flyover part of the 3rd Ave combination entrance ramp. There are three interventions one can explore in order to change the course of the breakdown.

1. Regulate traffic entering from the ramp in order to prevent the breakdown.
2. Increase vehicle headways in the right lane prior and at the merge point to allow for easier lane changes.
3. Regulate the traffic at the right lane after the occurrence of a breakdown to halt the backward propagation of the compression wave.

The first intervention can be accomplished by on-demand ramp metering. It has been shown in several simulation and field studies that ramp metering is successful in reducing the friction at the merge point and reducing crashes by up to 22% [Salman, 2001]. There are two issues involved with such a solution.

1. Ramp metering is primarily a peak period traffic management strategy. Even at those times drivers do not appreciate waiting in line while the freeway is un-congested.

2. Specifically in this location, a problem stems from the fact that the I-94 flyover is a freeway-to-freeway ramp carrying a lot of traffic. Regulating this flow may considerably increase the overall total travel time and delay of the system.
3. Since the flow breakdown is only one of the causal factors, preventing it will prevent the crash but preventing it while the other factors are not present is a waste of time. Unfortunately, the other two factors are usually encountered chronologically after the flow breakdown therefore by the time they are detected it might be too late for ramp metering to help prevent the crash.

Regardless, considering that more than half of the reported crashes did not occur due to the very first compression wave, if one does not care about the increase in delay, ramp metering will most likely prevent the crash here. However depending on the degree of metering, spillback may cause problems on I-35W.

The second intervention involves influencing drivers into increasing their following distances as they approach the merge area. As with all solutions involving influencing human actions, the human factors involved are important and beyond the scope of this dissertation. Regardless, from a search in the literature there are a few examples of systems aimed in influencing following distances.

1. A system, currently in operation in French interstates, involves special road markings and static signs instructing and helping the drivers to judge their following distances. In France this is a static solution implemented in areas where traffic is usually flowing with or above the speed limit. Only when the speed is known, static markings can generate the proper time headway of more than 2 seconds. Considering the traffic patterns encountered in this high crash area such a solution will not be applicable.
2. Another approach that has been proposed and currently implemented is designed to influence following distances inside tunnels. In this implementation the drivers are instructed to follow a racing spotlight shining in front of their vehicle. As long as the light spot is not passed, the time headway is equal or more than 2 seconds. This idea is superior since the distance between light spots can be dynamically adjusted to account for vehicle speed and length; assuming sufficient instrumentation is available. In the case of the study area such a solution can involve in-pavement lights generating racing stripes the drivers can follow. As will be explained later in this section this idea can also serve in altering the other two causal factors.

The third intervention involves changing the traffic pattern on the right lane after a flow breakdown has been detected, aimed in halting the backward propagation of the compression wave(s). According to car-following model theory, a compression wave (asymptotic flow instability) can be stopped if one of the vehicles in the chain has a sufficiently large following headway and/or the driver has sufficiently fast reaction time [Davis, 2003]. Both cases allow for a smoother slowing down affecting all subsequent vehicles. Other than in-vehicle adaptive cruise control systems there is no established method for accomplishing such an intervention. Even if

all the upstream entrance ramp flows are regulated, resulting in lower flow levels on the right lane, there is no guarantee that the remaining vehicles will be sufficiently spread out to halt the progression of the compression wave. Indeed, as was illustrated in chapter seven, the data suggest that it is not the flow/density of the upstream entrance ramp that is the direct causal factor; instead, these conditions increase the speed differential between the right and middle lanes and that in turn increases the crash probability. Therefore, it is more appropriate to discuss an intervention aimed at driver reaction time later when the driver inattention factor is explored.

6.2 DRIVER INATTENTION FACTOR

The second crash causal factor presented in chapter seven deals with driver inattention and more specifically, with the circumstances in the road that attract the driver's focus away from the observation of the leading vehicle in its lane. The data suggest that, the causal traffic pattern is the speed differential between the right and middle lanes. The hypothesis explaining this correlation states that, as the speed difference increases, lane changing becomes harder due to the decrease in the number of acceptable gaps. The driver's attention/time devoted in the search for an appropriate gap is taken from the attention/time he/she must spend to maintain a safe following distance with the leading vehicle in its lane. There are four possible interventions that can affect the aforementioned pattern.

1. Regulate/reduce the flow of the vehicles crossing the right lane towards the middle and left lanes.
2. Reduce the speed difference between the right and middle lanes to assist lane changing.
3. Prohibit lane changes between the right and middle lanes while there is a compression wave traveling upstream or even when a flow breakdown is imminent.
4. Implement a driver warning system aimed at increasing driver attention to conditions in their own lane.

The first intervention involves some sort of on-demand ramp metering on the upstream entrance combination ramp of I-35W/HW-55. Some of the issues previously discussed about ramp metering are also applicable in this case. In addition, the mechanism that connects the entrance flow with the speed differential between lanes and the level of difficulty in changing lanes is not yet clear. Therefore, there isn't yet available an analytical way to implement such an alternative. Further research in O/D patterns in the area as well as gap acceptance characteristics is required.

In several European cities, as well as a few in the U.S, systems allowing for individual lane variable speed limits have been implemented and field tested. Such systems are primarily aimed in regimenting traffic is such a way that separates slow from fast moving vehicles, therefore reducing acceleration noise and increasing quality of the flow. All these systems deal with speed limits on each lane as well as variable minimum allowable speeds. From the observations and analysis conducted in this study it is clear that during crash prone conditions all lanes are flowing

with average speeds that are bellow the set speed limit of 55 mph (88 km/h). Implementing variable speed limit signs instructing drivers in the middle lane to reduce speed to 45 even 40 so they can match that of the right lane, apart from the resistance from the drivers, might increase the danger. Specifically, if such low individual lane speed limits are enforced, drivers' compliance will result in increasing middle lane density (since the leaders will decrease speed before the followers) reducing even more the available gaps. Without a proper human factors study as well as use of better car-following models, the result of such an intervention is not clear.

The third intervention proposes a selective prohibition of lane changes between the right and middle lanes. Such an action is possible with the use of variable message signs instructing the drivers to stay in their lane or better with in-pavement lights creating a solid line between the two lanes. Conceptually such an intervention might be successful although a proper human factors study is required for uncovering potentially harmful side effects as well as the levels of driver compliance. Regardless, such a dynamic solution can potentially serve a dual purpose by also being a driver warning system.

The fourth intervention, already hinted many times throughout this document, is that of a driver warning system. Such a system along with other effects, aims in increasing the drivers attention levels on conditions ahead in their lane as well as possibly informing them on the nature of the danger. There are several alternatives for such a system. The closest, currently working, system is the AHS driver warning system in Japan, already discussed in the literature review. Although, the problem it is designed to solve is substantially simpler from the ones concerning the I-94 crash area, the method of passing the information to the drivers is state-of-the-art. The system depends on Variable Message Signs (VMS) upstream of the crash zone to communicate to the drivers the danger as well as the existence of the system itself (not a common traffic control measure). Closer to the crash zone it relies heavily on in-vehicle navigational systems to pass the warning directly to the drivers. This way it succeeds in passing more information while presumably minimizing the load it adds on the driving task. If a similar design is sought for the study area, an issue will arise with the timing of the warning i.e., how much time exists between the initiation of the alarm/information transfer and the time the drivers will encounter the problem. When designing automated incident detection algorithms one of the performance measures is Mean-time-to-detect. This is simply the average time between the incident and the raising of the alarm. In the case of crash-prone conditions, the ground truth is not complete enough to allow similar evaluation. Specifically, we do not have records of crash prone conditions, only of actual crashes and near-misses. The design of the driver warning system must be balanced i.e., it must succeed in warning the drivers early enough to avoid crashes but at the same time it must allow for the drivers to realize what they had just avoided. If the warning, albeit successful, is transmitted too early therefore, the warned drivers might not encounter any fluctuation in traffic; soon they will loose faith in the system thinking that it produces too many false alarms.

As mentioned in the introduction of this chapter, developing a crash prone conditions detection algorithm, once the methodology is established, is straightforward. Designing a driver warning system involves human factors engineering issues beyond the scope of this research. The next phase of this research deals with the development of a "visual workbench" aimed at assisting in the design of such systems. Additionally, in order to avoid implementing systems when we are unsure of their effectiveness, an experimental setup is required. The microscopic simulators currently available are by definition "safe". Specifically, the underlying models, all are based on the assumption that drivers will keep a safe distance and will act to maintain this safe distance. As was pointed out earlier in this document, one of the suspected causal factors is dangerous driving. This factor, which is supported by observations and measurements, contradicts the basic premise of all commercially available car-following models. Therefore, in addition to a visual environment allowing researchers to experiment with drivers reaction to the system, a new breed of simulators is needed to generate realistic traffic conditions. In other words, we need to construct the right tools before we start building a driver warning system.

To conclude the discussion of interventions targeting causal factors, dangerous driving is a matter of perspective. Very few drivers will admit dangerous driving. From their perspective they are simply driving on the limit of their skills. It is not easy to theorize about an infrastructure system preventing dangerous driving (apart from state patrol troopers). In vehicle driver assist technology when becomes mainstream theoretically will protect us from ourselves. Until then, traffic control and driver warning solutions, like the ones described briefly in the previous sections, might succeed in reducing a considerable amount of crashes.

6.3 MINIMUM SYSTEM REQUIREMENTS

Two separate systems were utilized in this research. The first system allowed the collection of detailed information at several locations along the roadway as well as a large number of crash video records. This information was instrumental in understanding the dynamics of the problem. The second system is the one collecting the necessary information for the detection algorithm. From the original four detection and surveillance stations, only one is finally required for real-time operation.

During the crash causal factor analysis the majority of measurements were extracted from ordinary loop sensors. It was though the addition of individually collected speeds and headways that unraveled the mystery at the end. Information is vital to the investigation of the casual factors in a high crash area, be they on a freeway or intersections. The detection and surveillance systems developed in this research are able to operate remotely and with minimum power requirements. It is therefore possible to deploy several portable, temporary ones on a high crash area, collect detailed information for a limited period of time and then, following an analysis similar to the one presented in this work, decide what the specific requirements for permanent instrumentation are. For the study are, based on the information the developed algorithm requires, in at least two separate locations information on individual vehicle speeds and headways is

required. Further experiments, in other high crash areas, are needed to identify the systems sensitivity on detection point spacing. In this location the detectors could not have been spaced farther away because we are bounded by the distance between the flow breakdown and crash locations. For the sensors there is no specific reason to finally use machine vision since there are other sensor technologies able to provide individual vehicle speeds and headways (loops can do it if they are connected appropriately). It is the opinion of the author that machine vision sensors are best suited for the initial investigatory period instrumentation because they combine detection with surveillance. To conclude, one fact that became clear during this research is that high crash areas are complicated, with individual characteristics and problems requiring custom design of both data collection and analysis methodologies. For sure the statistical nature of the crash likelihood model utilized in the detection algorithm is fitted for the specific combination of geometry, traffic conditions, and data collection characteristics. For the same model to operate in any other case a model calibration is required.

It is too early to speculate on detailed system requirements for a crash prevention system. The few alternatives mentioned in this chapter are examples of how current systems can be utilized in the case of this high crash area. Industrious engineers, once enough detailed information is collected/derived about the particular problem, can devise new, custom solutions.

6.4 CONCLUSION

In this chapter we attempted to offer some insight about possible solutions to this particular high crash area. The solutions target two out of the three identified crash causal factors. Each intervention needs to be researched in detail before implementation because there is always the possibility of creating more problems than it solves. We believe that currently we lack the proper tools to investigate, measure and visualize such solutions therefore creating these tools has priority over the creation of expensive crash prevention solutions. One area of intervention that was not discussed in this chapter involves permanent geometric changes in the roadway design to influence/eliminate the crash causal factors. Such solutions can include entrance redirection or closure, extra lanes, and others. Even such hard solutions cannot be guaranteed without proper investigation using a microscopic simulator capable of emulating in detail the identified crash causal factors.

7 Summary and Conclusions

Freeway crashes are the main source of non-recurring congestion. Finding ways of reducing them is critical for improving freeway operations and lowering costs to the traveling public without having to provide new capacity or major reconstruction. Because of this, recent efforts are focused towards the identification of traffic conditions leading to crashes. This combined with detection schemes based on real-time traffic measurements can be used in the development of interventions aimed at reducing crash occurrence. So far such attempts are limited to associating general traffic conditions with crash occurrence over a large area i.e., an entire freeway. However, evidence suggests that a large number of crashes accumulate in a few, short, high crash freeway sections. This implies that such locations differ considerably from the norm regarding the traffic conditions associated with crash likelihood. The work described in this document focused on detecting crash prone conditions at high crash freeway sections based on information collected by state of the art sensor technology. This was accomplished by developing models/algorithms for real time estimation of crash likelihood. Although the models/algorithms developed apply only in the high crash area considered, the methodology developed here as well as the traffic metrics identified are general and can be replicated in other freeway high crash areas.

Following observation of a large number of recorded crashes as well as other traffic events a fair understanding of traffic dynamics leading to crashes was acquired. Based on this knowledge, several general undesirable traffic flow conditions were identified i.e., large speed differences among lanes, compression waves leading to abrupt changes in the traffic flow, and others. Additionally, a number of traffic metrics related to crashes as well as their time and space variants were identified. Based on statistical modeling the most significant of these metrics were selected and their behavior explained. Specific patterns in speed variability as well as environmental factors such as lighting and sun position were confirmed to affect crash likelihood. The model performance, evaluated during an independent 10 day real-time test, was found to be promising as it combined high detection rates with relatively low false alarm rates.

Conceptually, any crash prevention system can be divided into two major components. The first is the trigger or alarm mechanism while the second is the set of actions taken to reduce the crash danger. The crash prone condition detection algorithms developed in this work can be used as the trigger. In order to design efficient and effective crash prevention actions the sequence of events and conditions producing the crashes must be understood. As a continuation to this research a process can be investigated to identify the crash causal factors as well as the crash mechanism. A preliminary qualitative investigation has shown that in this location, three crash causal factors can be identified, flow breakdown, short headways, and driver distraction. For any countermeasures to be effective at this high crash location, and similarly to others, one or more of these causal factors must be addressed. A preliminary exploration of possible crash prevention

actions is also provided, recognizing that proper action must also include detailed human factors analysis.

In conclusion, the research presented in this report succeeded in demonstrating that crash-prone traffic conditions exist and is detectable. Because any crash prevention/reduction solution needs to target the crash causes in order to be effective a methodology for identifying crash causes at high crash locations was also developed. Even though, both the detection and the causal analysis methodologies, are based on measurements and observations at a specific high crash location, they are general in nature i.e., can be applied/replicated in other high crash freeway locations.

References

Abdel-Aty M. A. and Radwan E. A. (2000) "Modeling traffic accident occurrence and involvement" Accident Analysis and Prevention Vol 32, pp. 633-642.

Abdel-Aty M., Uddin N., Abdalle M. F., Pande A. and Hisa L. (2004) "Predicting Freeway Crashes Based on Loop Detector Data using Matched Case-Control Logistic Regression." Presented at the 83rd Annual Meeting of Transportation Research Board, Washington, DC.

Abdel-Aty M., and Abdalle M. F., (2004) "Linking Roadway Geometrics and Real-Time Traffic Characteristics to model Freeway Crashes using generalized estimating equations for correlated data" Presented at the 83rd Annual Meeting of Transportation Research Board, Washington, D.C.

AHS:(Advanced Cruise-Assist Highway Systems)
http://www.its.go.jp/ITS/1998HBook/chapter3/3-3e.html

Aldrich, J. H. and Nelson F. D. (1984) "Linear probability, Logit and Probit models" Series: Quantitative application in the social sciences. SAGE.

Aljanahi A.A, Rhodes A.H, Metcalfe A.V. (1999). "Speed, speed limits, and road traffic accidents under free flow conditions". Accident Analysis and Prevention, 31(1-2): 161-168.

Arai, Y; Nishimoto, T; Ezak, Y; Yoshmoto, K., (2001). "Accidents and Near-Misses Analysis by Using Video Drive-Recorders in a Fleet Test". Proceedings of the 17th International Technical Conference on the Enhanced Safety of Vehicles (ESV) Conference, Amsterdam, The Netherlands. National Highway Traffic Safety Administration, Washington, DC. DOT HS 809 220, June 2001. Paper Number 225.

Baker, S. (1990). "Causes and contributing factors in traffic accidents". In L. Fricke (ed.) Traffic accident reconstruction, Northwestern University Traffic Institute.

Baldwin David M. (1966).Accident causes and Countermeasures, Traffic Engineering

BASt, German Federal Highway Research Institute (2003). "International Road Traffic and Accident Database, 2002". German Federal highway Research Institute.

Blincoe, LJ, Seay, AG, Zaloshnja, E, Miller, TR, Romano, EO, Luchter, S, Spicer, RS. (2002) "The Economic impact of Motor Vehicle Crashes, 2000" NHTSA Technical Report HS-809 446. Washington DC.

Cartwright, N. (1989) "Nature's capacities and their measurement". Oxford, U.K.

Cerrelli, E.C. (1996) "Research Note: Trends in Daily Traffic Fatalities 1975-1995"

Chassiakos A. (1992) "Spatial-temporal filtering and correlation techniques for detection of incidents and traffic disturbances". Ph.D. Thesis University of Minnesota,

Chen, C., K. Petty, A. Skabardonis, P. Varaiya, and Z. Jia, (2001) "Freeway Performance Measurement System: Mining Loop Detector Data," paper 01-2354, presented at the 80th TRB Annual Meeting, Washington DC.

Chira-Chavala T. and Mak, K. K. (1986). "Identification of Accident Factors on Highway Segments: A Method and Applications". Transportation Research Record 1068, TRB, National Research Council, Washington, D.C., pp 52-58.

Collett D. (1991) "Modelling Binary Data", Chapman & Hall.

Cook D.R. and Weisberg S. (1999) "Applied Regression Including Computing and Graphics", John Wiley and Sons, Inc.

Davis G.A. and Swenson T. (2003) "Causal Determination in Road Accidents: An Application of the Halpern/Pearl Notion of 'Actual Cause'.". Proceedings of the Second Workshop on the Investigation and Reporting of Incidents and Accidents (IRIA 2003), Kelly J. Hayhurst and C. Michael Holloway (compilers).

Davis G., Davuluri S. and Pei Jianping (2001), A Case Controlled Study of Driving Speed and Crash Risk. Research Project Report, ITS Institute, University of Minnesota.

Demaris, A. (1992) "Logit modeling. Practical Applications" Series: Quantitative application in the social sciences. SAGE.

Dinges, D. F. & Mallis, M. M. (1998). "Managing fatigue by drowsiness detection: Can technological promises be realized?" In HARTLEY, L. R. (Ed.) Managing Fatigue in Transportation. Proceedings of the Third International Conference on Fatigue and Transportation, Fremantle, Western Australia. Elsevier Science Ltd., Oxford UK.

Dobson, A. (1945). "An introduction to generalized linear models. Second edition" Texts in statistical science, Chapman and Hall/CRC.

Drew, Donald. R, (1968) "Traffic Flow Control and Theory", McGraw Hill.

Duncan C. S., Khattak A. J. and Council F. M. (1998) "Applying the Ordered Probit Model to Injury Severity in Truck-Passenger Car Rear-End Collisions" Transportation Research Record 1635, TRB, Washington DC. pp 63-71.

Edwards J. B. (1998). "The relationship Between Road Accident Severity and Recorded Weather". Journal of Safety Research, 29(4), 249-262.

Eliason, S. R. (1993) "Maximum likelihood estimation. Logic and practice" Series: Quantitative application in the social sciences. SAGE.

Federal Communications Commission, (2001) "Radio Frequency Devices, Sec. 15.247 Operation within the bands 902-928 MHz, 2400-2483.5 MHz, and 5725-5850 MHz." U.S. Government Printing Office.

Fink Kenneth L. and Krammes Raymond A. (1995) "Tangent Length and Sight Distance Effects on Accident Rates at Horizontal Curves on Rural Two-Lane Highways" Transportation Research Record 1500, TRB, Washington DC. pp 162-168.

Fisher, R.A., (1958), "Lung Cancer and Cigarettes?" Nature 182.

Fisher, R.A., (1958) "Cancer and Smoking" Nature 182.

Fisher, R.A. (1953) "Design of Experiments". Oliver and Boyd, London.

Garber, N. J. and Gadiraju, R. (1989) " Factors affecting speed variance and its influence on accidents" Transportation Research Record 1213, TRB, Washington DC.

Glymour, C., Scheines, R., Spirte,s P., and Kelly, K. (1987) "Discovering Causal Structure: Artificial Intelligence, Philosophy of Science, and Statistical Modeling". Academic Press, Orlando, FL.

Greenshields, B. D. (1961) "Quality of Traffic Flow, Quality and Theory of Traffic Flow." Symposium, Bureau of Highway Traffic, Yale University, New Haven, Conn. pp. 3-40.

Hadi, M. A., Aruldhas, J., Chow, L. and Wattleworth, J. A. (1995) "Estimating Safety Effects of Cross-Section Design for Various Highway Types Using Negative Binomial Regression." Transportation Research Record 1500, TRB, Washington DC. pp. 169-177

Helbing D. and Treiber M. (2002) "Critical Discussion of "Synchronized Flow"" Cooper@tive Tr@nsport@tion Dyn@mics 1, 2.1{2.24}

Helly, W., and Baker, P.G., (1965). "Acceleration noise in a congested signalized environment." Vehicular Traffic Science. Proceedings of the Third International Symposium on the Theory of Traffic Flow. American Elsevier, New York. pp. 55-61.

Helman, D.L. (2004) "Traffic Incident Management". Public Roads Vol. 68, Is. 3. Washington DC.

Herman R., Montroll E.W., Potts R.B., Rothery R.W. (1959). "Traffic Dynamics: Analysis of Stability in Car Following," Operations Research 7, pp. 86-106

Heylighen F. (2000) "Referencing pages in Principia Cybernetica Web", in: F. Heylighen, C. Joslyn and V. Turchin (ed.): *Principia Cybernetica Web* (Principia Cybernetica, Brussels), URL: http://pespmc1.vub.ac.be/REFERPCP.html.

Hijar M, Carrillo C, Flores M, Anaya R, Lopez V. (2000) "Risk factors in highway traffic accidents: a case control study". Accident Analysis and Prevention; Vol 32, pp 703-709.

Holohan, C and Culler, R, Wilcox, B (1978). "Effects of Visual Distraction on Reaction Time in a Simulated Traffic Environment," Human Factors, Vol. 20, No. 4, pp. 409-413
Hu, P., and Young, J. (1994) "1990 NPTS Databook: Volume II, Chapter 10". NPTS Databook, U.S. DOT, Office of highway Information Management.

Hosmer D.W. and Lemeshow S., (1989) "Applied Logistic Regression", John Wiley and Sons, Inc.

I-95 Corridor Coalition. (2001) Report "HAR Operational test evaluation". http://144.202.240.28/pman/projectmanagement/Upfiles/reports/full123.PDF.

INFOTEN, (1997) "Demonstrators Validation Report". EU Project TR 1032, Deliverable D9.

IN-RESPONSE, (1996) "State of the Art and State of Practice" Report. EU Project TR1030, Deliverable D3.2.

IN-RESPONSE, (1997) "Module Design" EU Project TR 1030, Deliverable D5.1.

IN-RESPONSE, (1999) "Report on Evaluation" EU Project TR 1030, Deliverable D9.2

Ioannou P. (2003) "Evaluation of the Effects of Intelligent Cruise Control Vehicles in Mixed Traffic" UCB-ITS-PRR-2003-2 California PATH Research Report University of Southern California

Jiang R., Wu Q. (2003) "First- and second-order phase transitions from free flow to synchronized flow" Physica A 322. pp.676 – 684.

Kerner B.S. (2004) "The physics of traffic". Springer-Verlag Berlin Heidelberg.

Kerner B.S. and Konhäuser P. (1993) "Cluster effect in initially homogeneous traffic flow" Phys. Rev. E 48, R2335–R2338 [Issue 4]

Kerner B. S. (2004) "Three-phase traffic theory and highway capacity". Physica A 333, 379-440

Khattak A., P. Kantor, and F. Council, (1998) "The role of adverse weather in key crash types on limited access roadways: Implications for Advanced Weather Systems." Presented at 1998 Annual Transportation Research Board Meeting, Preprint No. 981132, Washington, D.C. Transportation Research Record 1621, TRB, National Research Council, Washington, D.C., pp. 10-19.

Kiefer, R., LeBlanc, D., Palmer, M., Salinger, J., Deering, R., and M. Shulman, (1999) "Development and Validation of Functional Definitions and Evaluation Procedures for Collision Warning/Avoidance Systems". DOT HS 808 964, NHTSA, USDOT, Washington, DC.

Kiefer, R.J., Cassar, M.T., Flannagan, C.A., LeBlanc, D.J., Palmer, M.D., Deering, R.K., and M.A. Shulman, (2003). "Forward Collision Warning Requirements Project Task 1 Final Report: Refining the CAMP Crash Alert Timing Approach by Examining 'Last-Second' Braking and Lane-Change Maneuvers Under Various Kinematic Conditions." DOT HS 809 574, NHTSA, USDOT, Washington, DC.

Lagnado, D., Waldmann, M., Hagmayer, Y., and Sloman, S. (2005) "Beyond covariation: Cues to causal structure". A. Gopnik & L. Schulz, (eds). Causal Learning: Psychology, Philosophy, and Computation. Oxford: Oxford University (in-press) http://www2.warwick.ac.uk/fac/soc/philosophy/research/conandselfcon/causal/seminars/lagnado_etc_2005.pdf

Lagnado, D.A., and Sloman, S.A. (2004). "The advantage of timely intervention". Journal of Experimental Psychology: Learning, Memory, and cognition, 30, 856-876.

Lee, C., Saccomanno,F., and Hellinga, B., (2002) "Analysis of crash precursors on instrumented freeways". Presented at the 81st annual meeting of Transportation Research Board, Washington DC.

Lighthill M.J. and Whitham G.B. (1955) "On kinematic waves II. A theory of traffic flow on long crowded roads". Proc. Royal Society of London, series A 229, 317-345.

Lin, G. X. and Popoff, A. (1997). "Provincial-Wide Travel Speed and Traffic Safety Study in Saskatchewan". Transportation Research Record 1595, TRB, National Research Council, Washington, D.C., pp 8-13.

Lee, J.D., McGehee, D.V., Brown, T.L., and M.L. Reyes, (2002). "Driver Distraction, Warning Algorithm Parameters, and Driver Response to Imminent Rear-End Collisions in a High-Fidelity Driving Simulator". DOT HS 809 448, NHTSA, USDOT, Washington, DC.

Lee C., Hellinga B., and Saccomanno F., (2003) " Real Time Prediction Model for the Application to Crash Prevention in Freeway Traffic" Presented at the 82nd Annual Meeting of Transportation Research Board, Washington, DC.

Rajamani R., Levinson D. Michalopoulos P., Wang J., Santhanakrishnan K., and Zou X. (2005) Adaptive Cruise Control System Design And It's Impact on Traffic Flow" CTS Project Number: 2001040, Report No. CTS 05-01 University of Minnesota.

Madanat, S., and Liu, P. (1995) "A prototype system for real-time incident likelihood prediction" IDEA project final report (ITS-2), Transportation Research Board, National Research Council, Washington DC.

Martinez, R. (1997) "Crashes aren't Accidents". NHTSA Campaign material.

Marshal, R, E., and Mahack, K. (1998) "Determinants of Headway Selection". 77th Annual Meeting, TRB, Washington D.C.

May, D. (1990) "Traffic Flow Fundamentals", Prentice Hall.

McGehee V., Mazzae E., and Baldwin S.G. (2000). "Driver Reaction Time in Crash Avoidance Research: Validation of a Driving Simulator Study on a Test Track". Proceedings of the IEA2000/HFES 2000 Congress. *Santa Monica, CA: Human Factor*s and Ergonomics Society. www-nrd.nhtsa.dot.gov/vrtc/ca/capubs/IEA2000_ABS51.pdf

Michalopoulos, Panos G. (1991) "Vehicle detection video through image processing: The autoscope system." IEEE Transactions on Vehicular Technology. v. 40, n. 1, pt. 1, p 21-29.

Minnesota Department of Safety (2003) "2002 Minnesota Motor Vehicle Crash Facts". Minnesota Department of Safety, St. Paul, Minnesota.

Minnesota Dept. of Transportation. (2002) "Freeway Volume-Crash Summary, Twin Cities Metropolitan area". Systems and Research Section, Office of Traffic Engineering, Minnesota Dept. of Transportation. St. Paul.

NHTSA, (2004) "2003 Annual Assessment of Motor Vehicle Crashes". NHTSA Technical Report HS-809 755. Washington DC.

Oh, C., Oh J. and Ritchie S. G. (2001) "Real Time Estimation of Freeway Accident Likelihood" Presented at the 80th Annual Meeting of Transportation Research Board, Washington, DC.

Oppenheim and Schafer. (1989) "Discrete-Time Signal Processing" 0-13-216771-9; Prentice-Hall.

Pampel, F. C. (2000) "Logistic regression. A primer" Series: quantitative applications in the social sciences, SAGE.

Pearl, J. and Verma, T.S. (1991) "A Theory of Inferred Causation" UCLA Cognitive Systems Laboratory, Technical Report (R-156). In J.A Allen, R. Fikes, and E. Sandewall (Eds.), Principles of Knowledge Representation and Reasoning: Proceeding of the Second International Conference, San Mateo, CA. Morgan Kaufmann, 441-452.

Pearl J. (2000) "Causality : Models, Reasoning, and Inference" Cambridge University Press.

Phillips, W. F., (1979). "A kinetic model for traffic flow with continuum implications". Transportation Planning and Technology 5, 131-138.

PRIME, (2002) "Prediction Of Congestion And Incidents In Real Time, For Intelligent Incident Management And Emergency Traffic Management, Deliverable 6.1. Prototype Testing Part III-Estimation of Incident Probability" EU Project IST-13036 (Version 2), February

Proctor, R.N. (1995). "Cancer wars: How politics shapes what we know and don't know about cancer". New York: Basic Books

Sabey, B. and H. Taylor. (1980) "The Known Risks We Run: The Highway. Supplementary Report SR 567", Transport and Research Laboratory, Crowthrone, U.K.,.

Salman D. (2001). "Twin Cities ramp Meter Evaluation", Final Report, Minnesota Department of Transportation and CAMBRIDGE SYSTEMATICS.

Schrank, D. and Lomax T. (2004) "The 2004 urban Mobility Report". Texas Transportation Institute.

Shinar D. and Shacham M. (2002) "Evaluation of Different Methods for Keeping a Safe Headway in a Tunnel". "Safe Tunnel" Project Presentation Turin, Italy.

Snedecor and Cochran, (1967) "Statistical Methods," Sixth Edition. The Iowa State University Press, (pp. 62-64).

SRF Consulting and Mn/DOT. (2002) "NIT Phase II, Evaluation of Non-Intrusive Technologies for Traffic Detection" Final Report, Minnesota Department of Transportation.

SRF Consulting and Mn/DOT. (2001) "Field Test of Monitoring of Urban Vehicle Operations Using Non-Intrusive Technologies." Final Report, FHWA publication number FHWA-PL-97-018. St. Paul, MN.

Subramaniam V. K. (2001) "Detection of Undesirable Traffic Flow Conditions". M.S. Thesis, University of Minnesota.

TABASCO, (1998) "Final Evaluation Report and Exploitation Plan" EU Project TR 1054, Deliverable D10.3.

Thaggard, P. (1999) "How scientists explain disease". Princeton University press Princeton, New Jersey.

Thagard, P. (1998). "Explaining disease: Correlations, causes, and mechanisms". Minds and Machines, 8. pp.61-78.

Waldmann, M. R. (1996) "Knowledge-based causal induction". In D. R. Shanks, K.J. Holyoak, and D.L. Medin (ed.). The psychology of learning and motivation (Vol. 34, pp. 47-88). Academic Press. San Diego, CA.

Wong, Y. D.; Nicholson, A. (1992) "Driver behavior at horizontal curves: risk compensation and the margin of safety". Accident Analysis and Prevention. 24: 425-436.

Yau K.K.W. (2004) "Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong" Accident Analysis & Prevention, 36 pp. 333-340