Design and Evaluation of Freeway Incident Detection Algorithms

by

Sreeram Thirukkonda

B. Tech. (Civil Engineering)

Indian Institute of Technology, Madras, India (1997)

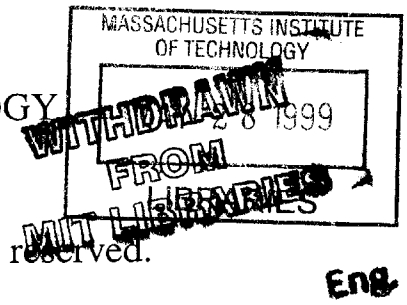
Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1999



Eng.

Author .....
Department of Civil and Environmental Engineering
May 21, 1999

Certified by .....
Moshe E. Ben-Akiva
Edmund K.Turner Professor of Civil and Environmental Engg
Thesis Supervisor

Certified by .....
Mithilesh Jha
Research Associate, Center for Transportation Studies
Thesis Supervisor

Accepted by .....
Andrew J.Whittle
Chairman, Department Committee on Graduate Students

Design and Evaluation of Freeway Incident Detection Algorithms

by

Sreeram Thirukkonda

B. Tech. (Civil Engineering)

Indian Institute of Technology, Madras, India (1997)

Submitted to the Department of Civil and Environmental Engineering
on May 21, 1999, in partial fulfillment of the
requirements for the degree of
Master of Science in Transportation

Abstract

A reliable, automatic freeway incident detection system is an important component of an effective incident management system. A traffic surveillance system equipped with loop detectors and speed traps can detect an incident by analyzing traffic measurements. Incident detection algorithms(IDAs) are the tools used to analyze these measurements. This work studies the limitations of current IDA evaluation methods and introduces some new approaches to incident detection. Case studies on I-93 North and I-90 West highlight the limitations of existing evaluation approaches. The algorithms perform optimally in a small subset of scenarios. The results of the case study indicate that it is imperative to evaluate IDAs over a broad range of scenarios to fully understand their applicability and transferability. The results also show that it may be advantageous to combine algorithms which have complementary characteristics under different scenarios to optimize overall performance. The feasibility of this approach has been demonstrated by combining a simple algorithm based on observing speed drops at a sensor with the McMaster algorithm.

A new approach to incident detection based on discrete choice models has been introduced and evaluated. This approach uses logit models based on traffic flow variables to identify congestion caused by incident. The performance of these algorithms has been studied over different scenarios. Using already existing methods, the logit models can be estimated with both simulation and real data to make the parameters both robust (simulation data from different scenarios) and easily transferable to the field (real data). A case study based on real incident data from I-880 near Hayward, California illustrates the advantage of this approach. A choice model based framework to combine different algorithms has also been presented. This framework uses the decisions of different algorithms to identify incidents. A model based on this framework has been estimated and evaluated over a set of scenarios.

Thesis Supervisor: Moshe E. Ben-Akiva
Title: Edmund K.Turner Professor of Civil and Environmental Engineering

Thesis Supervisor: Mithilesh Jha
Title: Research Associate, Center for Transportation Studies

Acknowledgments

I would like to thank my supervisors Professor Moshe Ben-Akiva and Mithilesh Jha for their guidance, support and encouragement at all stages of research. I would like to think Qi, Kazi and other ITS staff for their guidance and support in implementation. I would like to thank Alan for all the technical discussions and suggestions.

I would like to thank Bechtel,PB for funding this research. I would like to thank Cynthia Stewart and her staff for their support and cooperation.

I would like to thank my lab mates and ex-lab mates: Masroor, Dave, Sridevi, Mark for his interesting opinions of life, Mathew, Atul for all the coursework discussions, Bruno, Tomer, Jillian and Joan & Scott for the choice model discussions. I would also like to thank my friends at MIT for those lighter moments of life - Ninja, Harish, Megs, GV, and OP.

Last but not the least, I would like to thank my family.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

An incident is any unexpected event, like an accident or a breakdown that causes capacity reduction that affects traffic flow. Depending on their severity, incidents can disrupt traffic flow causing congestion and delays. As a freeway system approaches capacity, it becomes more sensitive to delays caused by non-recurring incidents. Cost estimates of such delays have been predicted to be 35 billion/year by the year 2005[28]. The FHWA estimates 61% of the total delays on freeways to be caused by incidents[25].

Freeway Traffic Management Systems(FTMS) use incident detection and effective incident management practices to reduce delays due to incidents. Automatic incident detection is an important component of an FTMS. Incident detection algorithms(IDAs) analyze the data from the network to identify incident. The performance of Incident Detection Algorithms is typically evaluated with three main Measures of Effectiveness (MOEs), namely Detection Rate (DTR-fraction of all incidents that the algorithm detects), the False Alarms (FAS), the False Alarm Rate (FAR) and the Time taken to Detect(TTD) the incident. FAR is typically defined as the percentage of checks made by the algorithm that result in incident declarations. The IDAs use measurement of traffic variables from the system (using different types of sensors) to detect incidents.

Using the general Neyman-Person formulation for decision theory[38], a false alarm can be considered as declaration of incident when the null hypothesis (H_0) of no incident is true. Thus, a false

alarm can be considered as a *type I error*. Similarly, an undetected incident can be considered as declaring no incident when the hypothesis H_1(incident) is true. Therefore, a missed incident would be a *type II error*. The trade-off between the different MOEs is intuitive. For example, An increase in the detection rate (decrease in *type II* error) can generally be associated with increase in the FAR (increase in *type I* error) and decrease in TTD. A decrease in FAR (decrease in *type I* error) implies a decrease in DTR (increase in *type II* error) and an increase in TTD. Therefore, the problem of incident detection typically has been the determination of the different algorithmic thresholds so that the *type-I* and *type-II* errors are balanced.

1.1 Research Motivation

Different factors affect the performance of incident detection algorithms. They can be classified into two main groups - system characteristics and algorithmic parameters & thresholds. The system characteristics include network features like the flow rate(high, medium or low), road geometry(entry or exit ramps, curvature, add or drop lanes) and properties of the sensor system such as type of sensor(loop detector, speed traps, roadside to vehicle communication-RVC sensor etc.), the sensor data averaging period and the spacing of the sensors. Algorithmic parameters are thresholds used by different algorithms, persistency checks to confirm incidents and persistency checks used to clear incidents. These factors influence the trade-offs between the MOEs in different ways. A detailed analysis of the trade-off between the different MOEs will be presented in subsequent sections.

The *robustness* of an incident detection algorithm can be defined as consistent performance under different scenarios. Although there have been a number of studies applying different theories and techniques to incident detection, there have been few studies directed towards evaluating the robustness of the algorithms. Although a new algorithm may have excellent performance under the controlled conditions used in the study, the performance is much different when it is actually applied in the field. Therefore, any new algorithm must be tested over a wide range of scenarios to identify not only the ideal operating conditions but also the limitations.

1.2 Literature Review

Incident detection algorithms can be classified into two main types - Model based algorithms and Data processing algorithms.

Model based algorithms use models to represent freeway traffic states using sensor data. The most basic model based incident detection algorithms are the California algorithms([18],[17]). These algorithms compare measurements of traffic variables and spatial and temporal differences in traffic variables against calibrated thresholds. The extensions to the basic California algorithms incorporate compression wave testing and other features to identify duration and severity of incidents. The All Purpose Incident Detection(APID) algorithm[34] uses the basic California algorithm as one of its subroutines. The HIOCC algorithm proposed by Collins et al.[22] looks for persistent high occupancy measurements by loop detectors. This algorithm basically identifies congestion and is ineffective in identifying the cause of congestion. The Multiple model method proposed by Willsky et al.[3] uses explicit models to predict different traffic states - incident, compression waves, recurrent congestion etc. The traffic state at any detector station at any instant is a vector containing the speed and occupancy as its elements. The model chooses the component model that has the minimum error of prediction to produce the overall result. The McMaster algorithm(Forbes[20],Persaud et al.[7],Hall et al.[8]) based on Catastrophe theory, proceeds in two steps. In the first step, congestion is identified. In the second step incident congestion is distinguished from recurrent congestion.

Data processing algorithms use statistical techniques to identify incidents. Exponential algorithms developed by Cook and Cleveland[2] use a forecasting technique called exponential smoothing to forecast the values of traffic variables. The forecasted value is compared against the measured value to detect incidents. Lin and Daganzo[45] present a simple scheme based on comparing cumulative occupancy measurements at two adjoining stations to detect incidents. The Standard Normal Deviate(SND) algorithm proposed by Dudek et al.[11] uses the normalized deviation of a traffic variable from its mean over the last n time steps to detect incidents. Ahmed and Cook[40] use a Box-Jenkins time series forecasting technique to model the occupancy at a single loop detec-

tor. They use an autoregressive integrated moving average model (ARIMA) to provide short-term forecasts of occupancy and associated confidence levels. If the occupancy is out of the confidence bound, incident is detected. Neural Network based incident detection algorithms[16] recognize traffic patterns under incident conditions. These algorithms require a large amount of training data before they can be used. Ritchie and Cheu[41] use simulated data for the training phase. The success of such neural network models depends on the validity of the simulator. The Neural network approach precludes the possibility of deducing anything from the final model. The weights obtained from training have very little physical interpretation. Chen et al.[15] propose a method to detect incident by extracting the principal components from sensor data. Since all fundamental traffic variables are related, they extract only the independent variations from the data from two adjacent detector stations. The CUSUM algorithm, another time-series based approach used by Teng and Madanat[19], uses traffic variable measurements to detect deviation from normal conditions in a minimum time for a specified false alarm rate. Data processing algorithms usually do not have methods to explicitly distinguish between incident and recurrent congestion. The lack of physical interpretation is another disadvantage.

Most of the studies on evaluation of incident detection algorithms have been targeted at showing improved performance of the algorithm under scenarios with given system characteristics. Gall et al.[1] propose refinements to distinguish between incident and recurrent congestion, but recognize that the resulting algorithm may not be applicable in a network which is subject to many bottlenecks. Persaud et al.[7] restrict their study to data from the median lane free from trucks. They also restrict the flow rates to scenarios without recurrent bottlenecks. Peeta and Das[39] study the influence of time-dependent demand. However, the test zone has been restricted to a stretch between two ramps, free from effects of merging and exiting traffic. Under such conditions any trivial congestion detection algorithm would have good performance. The emperical study in later sections shows that most of the false alarms can be attributed to traffic conditions downstream of an on-ramp. Ishak et al.[42] study the applicability of Fuzzy ART to incident detection. The scenarios they consider vary only the algorithmic parameters with fixed system characteristics.

This can give a misleading impression of robustness of the incident detection system. As Dudek

et al.[11] indicate, *"the results of an incident detection model have to placed in the proper perspective."*. Stephanedes et al.[43] have also stressed the need for a comprehensive evaluation and identified the limited transferability of algorithms. They have compared the performance of some existing algorithms when the sensor data averaging period is varied and the algorithmic logic is different(single-station or comparative). Abdulhai[4] has outlined the requirements for a universally transferable freeway incident detection framework based on Neural Networks. Most of the simulation variables studied in our work can be directly related to the requirements in Abdulhai[4].

In any network, many of the system characteristics and algorithmic parameters are variable. Thus, in order to demonstrate the applicability of the algorithms, it is imperative to examine the performance under a wide range of scenarios in which these factors are varied. Given the complex interplay between these factors, it would be difficult to expect intuitive performance from the algorithms. It would also be difficult to directly use the MOEs to measure algorithmic performance under the combined influence of the different factors. There have been few studies on the robustness and transferability of the algorithms by varying the different factors. Such a study could lend insight into designing and implementing incident detection systems for an FTMS. A wide range of scenarios are not usually considered because of the lack of data. Simulation would be an ideal testbed to generate data to study the algorithmic performance under numerous scenarios. Moreover, it is difficult to isolate the impacts of the different factors on trade-offs between MOEs when real data is used. Loop detector based detection is subject counting errors(especially due to large trucks and motorcycles).This makes additional data screening necessary. Petty[25] presents ways to fix common problems in loop data.

Although data fusion methods, which use data from different sources in incident detection, have been well studied(Petty[25], Sethi et al.[44], Bhandari et al.[33]) there are no studies on combining different loop-based algorithms. The current work presents some rules to combine loop-based algorithms and evaluates the performance of one combined algorithm over a range of scenarios.

1.2.1 Impact of System Characteristics on Incident Detection

The system characteristics that affect incident detection performance are flow rate(high, medium or low), geometry of the stretch(entry or exit ramps, curvature, add or drop lanes) and properties of the sensor system(loop detector, speed traps, RVC sensor etc.), the sensor data averaging period and the spacing of the sensors. A qualitative discussion of the impact of each of these variables on the performance has been presented in the remainder of this section.

Flow rate: Most algorithms do not detect incident per se but detect congestion. Therefore, when an incident does not cause significant congestion, these algorithms have poor performance in terms of the DTR. At high demand levels, the congestion due to incident is clearly visible, which makes it easier to detect incidents. However, improved detection may be offset by an increase in the FAS. Consider a hypothetical three lane section and without loss of any generality assume that the capacity per lane is 2000 veh per hour. If the flow across the section is 4000 vph, a single lane blockage would not cause noticeable congestion. The section would still be able to serve the demand with the other two lanes. In such a scenario, any algorithm detecting congestion may not be able to detect the incident. However, if the flow were around 6000 vph, any trivial algorithm would be able to detect the congestion. Therefore, it is important to examine the performance of algorithms under different flow conditions.

Complex Geometry: Complex geometrical configurations give rise to additional complications. Recurrent bottlenecks due to entry ramps, weaving sections or drop lanes create *incident-like* situations. Presence of grade or curvature also produces patterns similar to those recognized by most algorithms as incidents. Thus, it may be easier to detect incidents at some sections than others. A robust algorithm would be expected to be relatively insensitive to the geometry. Consider the case of a simple corridor with no ramps. If we assume relatively uniform grade and curvature, the problem of incident detection reduces to that of congestion detection. Again any trivial congestion detection algorithm would perform well. On the other hand, if there are several on- and off-ramps which produce severe recurrent congestion, then it will be difficult to isolate the effect of an incident.

Gall et al.[1] discuss a logic to distinguish between incident and other forms of congestion but limit the applicability of their logic to cases where nonincident congestion is rare. Peeta and Das[39] consider a stretch of with several entry and exit ramps. However, they applied their algorithm only to a section where the incident was located. Moreover, this test section had no off and on ramps. This precludes the possibility of testing the algorithm at stations upstream of an entry ramp particularly susceptible to FAS. In a complex system with considerable geometric complexity, it is necessary to examine the performance of IDAs for incidents at different locations. It is expected that complex geometry will increase the number of FAS because of the presence of a great number of incident-like patterns. The TTD would increase because the algorithm would need more time to isolate the impact of an incident. Since it would be easier to miss incident congestion patterns, the DTR can be expected to decrease.

Spacing of the sensors: The spacing between sensors is critical to detect incidents quickly and reliably. Petty[25] has suggested the possibility of using principles from information theory to optimize the sensor spacing in order to obtain maximum information from the network. Intuitively, as the spacing between the sensors decreases, we can expect algorithmic performance to improve in terms of the time taken to detect the incident and the detection rate. But the number of false alarms may increase due to an increase in the number of checks made by the algorithm. Obviously, high spacing reduces the reliability of the data. Different networks have different sensor spacings. Studying the sensitivity of algorithmic performance with respect to sensor spacing would also indicate the degree of transferability of the algorithm between different networks. There have been few studies[25] investigating the effect of sensor spacing on the algorithmic performance. Therefore, it would be useful to examine the effect of sensor spacing on algorithmic performance.

1.2.2 Impact of Algorithmic Parameters on Incident Detection

The algorithmic parameters & thresholds depend mainly on the measures used by the algorithm and to some degree on the system properties. Algorithmic parameters common to all algorithms that use persistency checks are the number of time intervals needed to confirm incidents and the

number of time intervals needed to clear incidents. For example:- If an algorithm uses the difference in average speed(at a sensor station) between two successive time intervals, the performance of the algorithm would directly depend on the sensor data averaging period(greater the interval length, lower would be the drop in speed, which implies that any threshold for the drop in speed would be lower). Speed drops would be detected only during the onset of the incident. There would be very little persistency checking for such an algorithm. On the other hand if an algorithm used spatial differences in speed to detect incidents, persistency checks would be useful because the incident congestion between two sensors would be sustained during the duration of the incident. Flow rate and system factors like sensor spacing also affect the choice of persistency. At high flow rates or closer sensor spacings, it is easy to quickly detect any form of congestion. Therefore higher persistencies are needed to distinguish incident congestion from recurrent congestion. Low flows and high sensor spacings make detection of incidents difficult. Therefore, lower persistencies would have to be chosen. Sensitivity of algorithmic thresholds to the different variables also depend on the type of logic used eg:- single-station logic or comparative logic.

Persistency checking: The persistency of the algorithm directly affects the MOEs. A higher persistency increases the time taken to detect the incident, but, reduces the number of false alarms and may also reduce the detection rate. A lower persistency increases the detection rate and decreases the time to detect. But it also increases the false alarm rate. A majority of the algorithms have thresholds which have to be calibrated. These thresholds also have to be updated or recalibrated before they can be used at other sites. Peeta and Das[39] and Abdulhai[4] has presented schemes to update the algorithmic thresholds.

Incident characteristics: Other factors specific to incidents such as duration of the incident and location of the incident also affect performance of incident detection algorithms. Longer incidents imply greater congestion in the network. This can increase the number of FAS because of the presence of patterns similar to incidents. Incidents that occur in sections where capacity reductions have significant effect can be detected easily. But, incidents that occur in recurrent congested zones of the network such as a weaving section would be more difficult to detect.

Type of Logic: Single-station algorithms have difficulties in distinguishing between incident and recurrent congestion and under non-incident related changes in traffic operation such as weather. Performance of algorithms based on comparative logic is particularly sensitive to incident location and duration of incidents. Persaud et al.[8] discuss the drawbacks of the different types of logic used for incident detection algorithms.

In the next chapter the performance of two representative algorithms, the McMaster algorithm based on catastrophe theory and the APID(All Purpose Incident Detection) algorithm based on comparative occupancy logic, has been examined under a wide range of scenarios. The system factors that have been varied are flow rate, location of incident, incident duration and sensor spacing. The main algorithmic factor that was chosen to be varied was the number of intervals used in persistency checking. Based on evaluation results, we present some new ideas and algorithms for incident detection. Simulation based testing has helped identify conditions under which the algorithms have poor performance. The microscopic simulator MITSIM[37] has been used to generate data. A detailed description of MITSIM can be found in Ben-Akiva et al.[32], Yang et al.[37] and Ahmed[26].

1.3 Thesis Contribution

The thesis makes the following contributions in the area of evaluating incident detection algorithms:-

i. It explores the trade-offs between the different MOEs over a large range of scenarios and to derive insights into designing a network wide incident detection algorithm.

ii. It examines the influence of key system variables and algorithmic parameters(usually fixed in other evaluation studies) like sensor spacing, flow rate, persistency and type of logic on the performance of an incident detection algorithm.

iii. It uses the insights from the comprehensive evaluation tests, proposes enhancements to existing algorithms and makes them more applicable in the field and proposes new ideas for the network being studied. In the direction of enhancing performance of existing algorithms,

the idea of combining algorithms has been explored. Combined algorithms are shown to perform better than the individual component algorithms.

In addition to these new directions in evaluating incident detection algorithms, two new approaches to incident detection have been presented. They are:-

i. A comparative algorithm(SPDDRP), based on detecting rapid speed drops at a station, is developed. Conventional Incident detection algorithms use elaborate measures and conditions to identify incidents. However, SPDDRP assumes that in a network with dense loop detector spacing, incidents can be detected by observing the speed drop at a sensor due to incident. The validity of this assumption has also been tested.

ii. A family of discrete choice models for incident detection is calibrated and evaluated. Performance of these models have been shown to be promising.

1.4 Thesis Outline

Chapter 2 presents the details of calibration of the McMaster and the APID algorithms. Chapter 3 describes the experiments conducted on I90 West and I93 North to evaluate the performance of the algorithms over a wide range of scenarios and also summarizes the findings from the experiments. Chapter 4 describes an algorithm based on speed drops to exploit dense loop detector spacing. Chapter 5 presents a new incident detection algorithm based on different types of discrete choice models. Chapter 6 presents the idea of combining algorithms to break the barriers imposed by the trade-offs between the MOEs of any single algorithm. Two different frameworks to combine algorithms have been presented. Chapter 7 summarizes the findings and discusses the areas for future research.

Chapter 2

Calibration of the McMaster and APID algorithms

Both the McMaster and APID algorithms are fundamentally threshold algorithms. They identify the different traffic states based on pre-determined thresholds. These thresholds are mostly non-transferable. The first step in using these algorithms is the calculation of these thresholds.

2.1 Description of the system

The Network used in the current study is a part of the Central Artery/ Tunnel(CA/T) Project in Boston. It is a 7.5 mile highway, approximately half of which will be built as a tunnel. The two parts of the network used in the current study are I-90 West and I-93 North. This part of I-90 West is also called the Ted Williams Tunnel. It is a two lane 1.6 mile stretch of tunnel with loop detectors at every 200ft. Speed traps are located at every 600 ft. I-90 West is a simple 2-lane tunnel with no geometric change. As discussed in the earlier section, the problem of detecting incidents in this simple network configuration reduces to that of detecting congestion. The results,presented later, support this claim. The I-93 North varies between 3 and 4 lanes. Loop detectors are located at every 600 ft on normal freeways and at every 200ft in the tunnel. There are 6 on-ramps and 5 off-ramps. The locations of these ramps are shown in Figure 2-1. The flow rates on the different links in Figure 2-1 are based on the OD flows estimated for the year 2004 during the PM peak. This

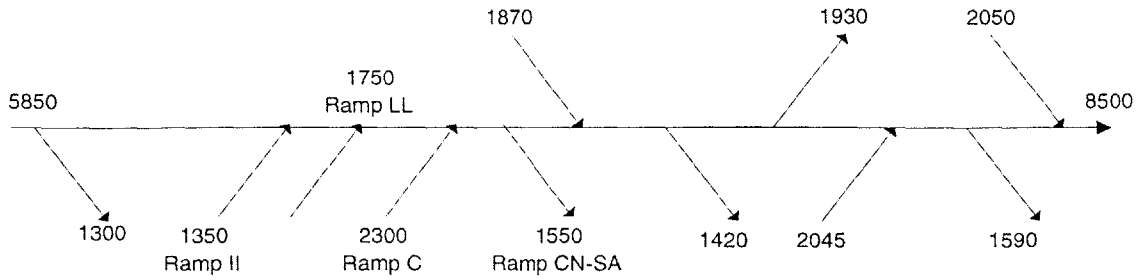


Figure 2-1: I-93 North Bound

demand is provided by the CA/T project and is used as the base demand(100%) for scenario design. Sensor data is generated for all stations on I-93 North and analyzed for presence of incidents. There are significant variations in geometry because of on- and off- ramps, add/drop lanes and curvature.

2.2 The McMaster Algorithm

The McMaster algorithm uses flow-occupancy data to classify the traffic conditions into different states based on pre-determined thresholds. A typical fundamental diagram has been presented in Figure 2-2. The parameters OCMAX, VCRIT and the shape of the curve have to be determined

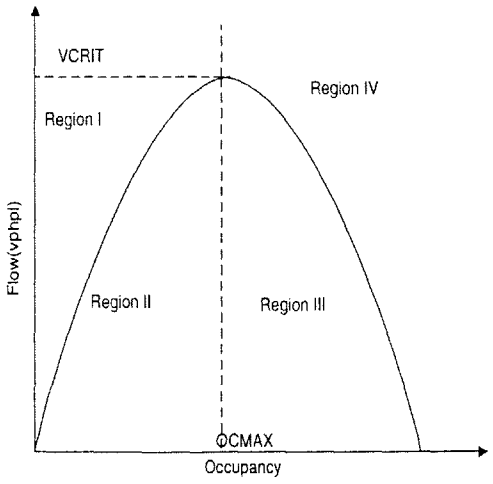


Figure 2-2: Fundamental Diagram of Traffic Flow

before the algorithm can be used. The algorithm uses the flow and occupancy values to classify the current traffic state as one of the four states shown in Figure2-2. Based on this classification, the algorithm decides if incident conditions exist at a particular sensor station. Region I represents a

region of low occupancy and flow varying from low to high. This represents an incident-free condition. Region II is a region of low flow at low occupancies. This typically represents the traffic state at a sensor station downstream of a blocked lane. Region III is the other type of incident state. It represents a high occupancy and low flow state. This is the traffic state upstream of the blocked lane. Region IV is a high occupancy, high flow region. This is typically the state downstream of a permanent bottleneck, e.g. an on-ramp. The parameters of the flow-occupancy plot have to be calibrated.

The McMaster algorithm seeks to ensure that the data in the congested regime can be clearly separated from other random fluctuations. The essence of the McMaster algorithm is that it tries to explain sudden discrete changes in one variable when other related variables are exhibiting smooth and continuous change. Further details can be obtained in Persaud et al.[7]. The Microscopic Traffic Simulator(MITSIM)[37] was used to generate the fundamental diagram by varying the OD flows. Zhou et al.[31], Hall et al.[12] and Hall et al.[13] have dealt extensively with the shape of the flow-occupancy and speed-flow plots. They have fitted different curves to the left portion of the flow-occupancy plot. Particularly, Zhou et al.[31] have estimated a cubic polynomial for the speed-flow plot. In order to better represent the overall flow-occupancy relationship, polynomials of different degrees were fitted to the data consisting of congested and uncongested operation. Two main criteria have been used to select the degree of interpolation.

i. *Better representation of the overall relationship:-* The fit can be used as a criterion to select the best polynomial. Both r-squared and adjusted r-squared values have been used to decide best fit. Polynomials of degree 4 or higher were found to have better fit for most of the stations. However, the r-squared indicated an overfit after degree 4.

ii. *Conservative representation of region II of the fundamental diagram:-* The Region II of the fundamental diagram is the state of traffic at a station downstream of an incident. The parameter OCC_{max} defines region II. It was identified that polynomials of degree 4 or higher conservatively define OCC_{max} making false identifications of incident harder. This reduces the FAS without affecting the DTR or TTD. Therefore these polynomials have been used. Table 2.1 is representative of the trend observed in OCC_{max} for different degrees of interpo-

lation.

Table 2.1: OCC_{max} squared for different degrees

Degree	OCC_{max}
2	23.86
3	23.55
4	21.58
5	21.58

The results from fitting different polynomials to the data for a sample sensor has been produced in Table 2.2. As indicated by the adjusted r-squared values in the table, the model becomes an overfit when the degree of the polynomial is 5. Thus the form of the polynomial used is

$$flow = a_0 + a_1 occ + a_2 occ^2 + a_3 occ^3 + a_4 occ^4$$

The thresholds can be determined from the shape of the left portion of the curve. Some typical parameter values obtained from the simulation have been shown in Table 2.3. The initial calibra-

Table 2.2: R-square and Adjusted R-squared for different degrees

Degree	R-squared	Adjusted R-squared
2	0.7444	0.7420
3	0.7567	0.7557
4	0.7644	0.7607
5	0.7719	0.7605

tion and testing of the algorithms was done for the Ted Williams Tunnel. Projected 2004 PM peak flows were used as the base flows. A typical plot generated from the simulation data has been presented in Figure 2-3. Once an n-degree polynomial is fitted to the data, the threshold curve (curve between 0 and OCC_{max}) is the 97.5% confidence curve. The threshold has been chosen so that 95% of the points fall above the threshold in the no incident case. A sample plot is shown in Figure 2-4.

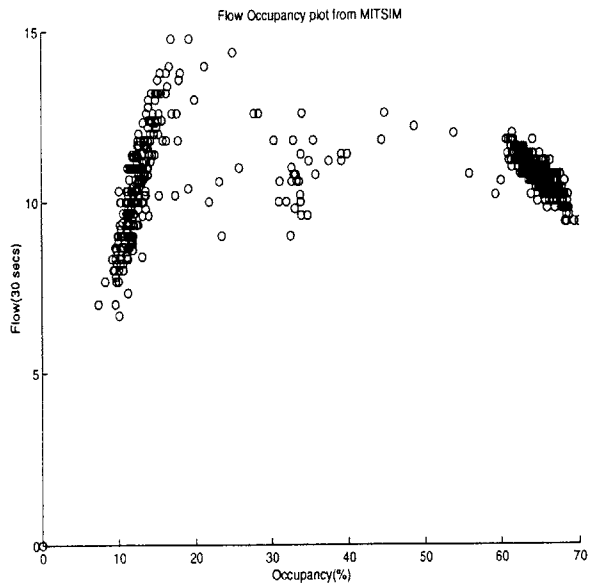


Figure 2-3: Flow Occupancy plot for a sensor

Table 2.3: Typical parameter values for a station for the McMaster algorithm

Parameter	Value
a_0	-2.5
a_1	1.4
a_2	$-6.5x10^{-2}$
a_3	$1.2x10^{-3}$
a_4	$-7x10^{-6}$
$OCC_{max}(in\ \%)$	21.58
$V_{crit}(in\ veh/30\ sec)$	13.5
$Av.Spd.(in\ mph)$	50

2.3 The APID Algorithm

The All Purpose Incident Detection algorithm(APID) uses a combination of different algorithms to identify incident conditions at different levels of flows ranging from low to high. The parameters have to be obtained by observing the values of the different measures used by the algorithm over a wide range of scenarios. Philip et al.[34] have presented the basic framework of the algorithm. However, they have used the same set of parameters for all the stations. They acknowledged that performance could be improved by using different groups of parameters. Their parameters are the initial values for the parameter calibration in our study. Calibration has helped identify different

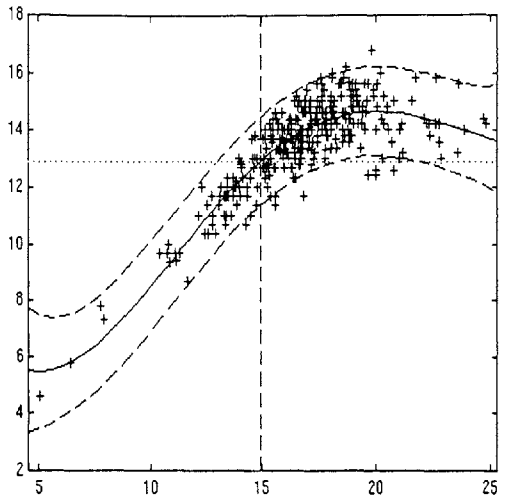
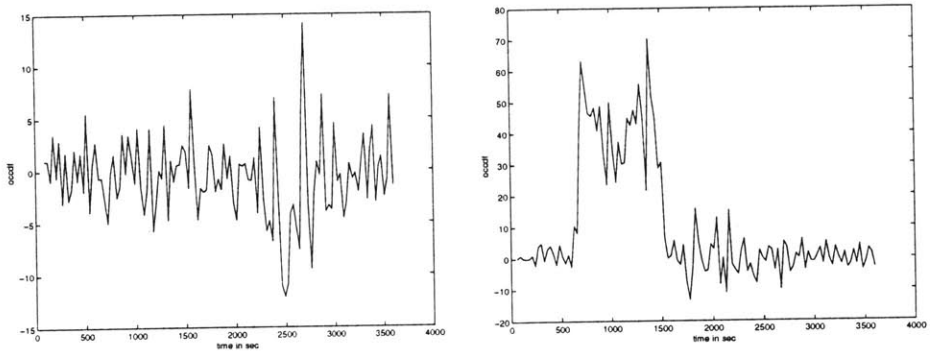


Figure 2-4: Sample plot of McMaster threshold curve with the ±97.5% confidence curves

groups of parameters for different stations. The calibration has been performed for sensors in both the Ted Williams Tunnel and the I-93 North Bound. All stations in the Ted Williams tunnel have used the same set of parameters due its simple geometry. At a spacing of 200 ft, five different groups of parameters have been identified for I-93 North. The algorithm uses three main routines - a low volume routine, a medium volume routine and a high volume routine. The California Incident detection algorithm is used as the high volume routine. The basic framework has been slightly modified for the current study. The basic algorithm uses compression wave testing and suspension of detection when compression waves are detected. Suspension of detection is used to reduce false alarms due to turbulent traffic conditions. Although it decrease FAS appreciably, it also directly affects detection rate. The incidents on I-93 North have been placed in locations close to recurrent bottlenecks which are subject to turbulent traffic conditions. To avoid missed detection, suspension of detection has not been used in the present case. Suspension can always be replaced by higher thresholds and persistency checks for locations of unstable traffic conditions. The persistency of 3 polling periods has been found sufficient to eliminate compression wave testing at a spacing of 200ft. Persistency checking has also been applied to identify the end of the incident and return to normal traffic conditions. Typical parameter thresholds used by the different components of the APID algorithm are presented in Table 2.5. Typical plots of the measures have been produced in Figure 2-5 and Figure 2-6. Based on these plots, the thresholds for the different measures were decided.

Table 2.4: Measures used by the APID algorithm

Description	Measure
Spatial difference in occ(OCCDF)	occdf(i,t)=occ(i,t)-occ(i+1,t)
Relative spatial difference in occ(OCCRDF)	occdf(i,t)/occ(i,t)
Relative temporal difference in d/s occ(DOCCTD)	docctd(i,t)=(occ(i+1,t-2)-occ(i+1,t))/occ(i+1,t-2)
Downstream Occupancy(DOCC)	docc(i,t)=occ(i+1,t)
Relative temporal diff in speed(SPDTDF)	spdtdf(i,t)=(spd(i,t-2)-spd(i,t))/spd(i,t-2)
Persistency threshold(PERSTH)	threshold to be exceeded in persistency check
Medium Traffic Threshold(MEDTRFTH)	threshold to be exceeded to use med flow alg
Clearence Threshold(CLRTH)	threshold to be exceeded to clear incident



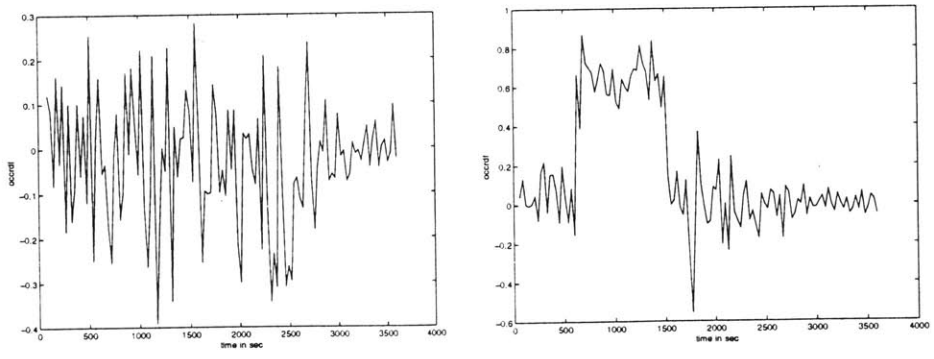
(a) Incident-free

(b) Incident

Figure 2-5: OCCDF plots for incident and non-incident cases from simulation

Table 2.5: Typical APID parameters

	Calif Inc. Det.			**Med Vol Inc. Det**		**Other prms**		
Prm Group	DOCC	OCCDF	OCCRDF	DOCCTD	SPDTDF	PERSTH	MEDTRFTH	CLRTH
1	20	25	0.35	0.40	0.10	0.35	15	0.31
2	17	15	0.65	0.40	0.10	0.65	15	0.31
3	17	55	0.80	0.40	0.10	0.80	15	0.31
4	17	30	0.50	0.40	0.10	0.50	15	0.31
5	17	20	0.50	0.40	0.10	0.50	15	0.31



(a) Incident-free

(b) Incident

Figure 2-6: OCCRDF plots for incident and non-incident cases from simulation

Chapter 3

Evaluation of the McMaster and APID algorithms

3.1 Refinements to the McMaster algorithm

Refinements for I-90

Initial tests were conducted in the Ted William tunnel. The results of the tests indicated that algorithms were not directly transferable to networks with dense loop detector spacing. Incidents and other incident-like traffic states were easily detected which resulted in high number of false alarms. In order to clearly distinguish between incidents and other patterns similar to incidents, refinements were designed. These refinements were also tested in the second series of tests on I-93 North.

Memory of true alarm to simulate actual behavior

The McMaster algorithm primarily uses the traffic state at a station to identify incidents. For a dense spacing of loop detectors, it is possible to identify the incident state due to an incident at several stations in the vicinity of the incident. This implies that it is possible to declare the same incident twice. Subsequent alarms from the same incident would have to be suppressed or recognized to be secondary alarms. Gall et al.[1] indicate that secondary alarms should be identified but do not present any strategy. To prevent this, we use the strategy of declaring an

incident only if it has not been declared before. We do this by checking to see if an incident state exists in the vicinity of the current station. If there is a confirmed incident state, we check if this is a true incident. An alarm has been identified to be a true alarm if it is declared a maximum time t_{max} after the actual occurrence. The value of t_{max} has been established emperically to be 5,7.5 and 10 minutes for spacing of 200ft, 600ft and 1200ft respectively. In real situations, this would correspond to the following. An operator is notified of an incident at a location. He/She responds to the alarm. If the alarm is true, an incident management plan is executed and any further alarms in the vicinity would be secondary. But if the alarm is a false alarm, the state of the station is restored to no incident. So, any further alarms in the vicinity would be(if they are not genuine) treated as false alarms. This strategy is one of the additions to the basic McMaster algorithm.

State Constraint

The original McMaster algorithm suggests that either Region I or Region II can be treated as incident state. For the close detector spacing in the tunnel, this was found to be inappropriate. This condition could be useful for larger sensor spacings. In order to reduce false alarms from this state, it was constrained that during the persistency check interval, the state 3 should be observed upstream of the incident(high density state) at least once. Incident should not be declared if over the entire persistency interval, only state 2 was observed at the station. It should be noted that a condition with state 3 upstream and state 2 downstream may occur for only one set of upstream and downstream stations. However, state 2 can be observed for many sensor station pairs both upstream and downstream of the incident. Gall et al.[1] proposed the original logic for a portion of the Queen Elizabeth Freeway where the spacing of loops is approximately 2200 ft. However, when the spacing is around 200-600 ft, incidents are hardly missed and numerous stations in the vicinity of the incident declare the incident if the state constraint is not imposed.

Count Constraint

The original algorithm does not use any basic reduction in flow. From preliminary investigations, it was found that for 30 sec polling period, there was a reduction in the count at the downstream station compared to the upstream station, at the end of the first polling period during which the

incident occurs. Therefore, it was constrained that for the first persistency check interval, the upstream count should be greater than the downstream count. This condition basically exploits the simple geometry of the Ted Williams Tunnel. For a 2 lane freeway stretch, flow discontinuities can easily be identified for small sensor data averaging periods and dense loop spacings.

Refinements to the McMaster algorithm for I-93 North

The refinements used in the tunnel case were also applied to the I-93 North. But it was found that the count constraint was too tight for the 3 lane and 4 lane case. Therefore this constraint was relaxed. However, the state constraint and the false alarm memory conditions were retained.

3.2 Refinements to the APID algorithm

The algorithm uses three main routines - a low volume routine, a medium volume routine and a high volume routine. The California Incident detection algorithm[11] is used as the high volume routine. The basic framework has been modified for the current study. The basic algorithm uses compression wave testing and suspension of detection when compression waves are detected. Suspension of detection has not been used in the present case. The persistency of 3 polling periods has been found sufficient to eliminate compression wave testing at a spacing of 200ft. Incident state at any station was reset only when the traffic variable values returned to normal for a specified number of consecutive time intervals. Thus, persistency checking was used to declare and clear the incidents.

3.3 Evaluation Results

3.3.1 Evaluation on I-90 West

Experiments on I-90 West

The network in this case consists of one OD pair. Flow rates were increased from 900 vphpl to 2000 vphpl in steps of 100 vphpl. The incident duration considered were 3 , 6 and 10 min. Thus, a total of 12*3 = 36 experiments were conducted. A total of 20 replications were performed per

experiment. For the different experiments, the incident location was fixed. The simple network geometry eliminates the need for studying different incident locations. The incident resulted in a one lane closure. A total of one hour of simulation was performed and the incident occurred at 10 minutes from the start of simulation. For all experiments, the polling period was 30 seconds for both the algorithms. There were a total of 56 stations with 2 sensors each in this network. Therefore the total number of checks made by the algorithm in 20 hours of experiments with a polling period of 30 seconds was 120x20x56x2 = 268800. This number was used in computing the false alarm rate(= (FAS / number of checks)*100). When the number of checks made by the algorithm becomes very high, the percentage of checks resulting in incident declarations become a very small fraction of the total. Expressing the false alarms as FAR therefore does not convey a correct idea of the actual number of false alarms. Therefore, both FAR and FAS(number of false alarms) have been presented. The results are presented in Table 3.2 and Table 3.3.

Analysis of tests on I-90 West

Table 3.1: Original McMaster Algorithm, Persistency=3, Spacing=200ft

Inc. Duration	Parameter	900	1000	1100	1200	1300	1400	1500	1600
	TTD(min)	2.03	2.15	2.11	2.12	2.09	1.77	1.75	1.60
3 min	DTR(%)	70	75	95	100	100	100	100	100
	FAS	2	3	3	5	11	12	14	18
	TTD(min)	2.23	2.21	2.24	2.02	2.10	1.85	1.73	1.58
6 min	DTR(%)	65	80	90	95	100	100	100	100
	FAS	5	4	8	10	18	19	22	24
	TTD(min)	2.23	2.14	2.09	2.04	1.87	1.81	1.77	1.72
10 min	DTR(%)	35	75	95	100	100	100	100	100
	FAS	8	6	11	14	24	27	31	34

i. As depicted in Figure 3-1 and Figure 3-2, APID outperforms the McMaster algorithm. The components of APID are algorithms based on comparative logic. Such algorithms are known to have problems in presence of significant changes in geometry[7]. Since I-90 West is a fairly straight tunnel stretch with almost no variation in geometry, APID has good performance. In the second set of tests on I-93, it will be seen that the APID algorithm has poor performance at greater sensor spacings when there are substantial changes in geometry.

Table 3.2: Refined McMaster Algorithm, Persistency=3, Spacing=200ft

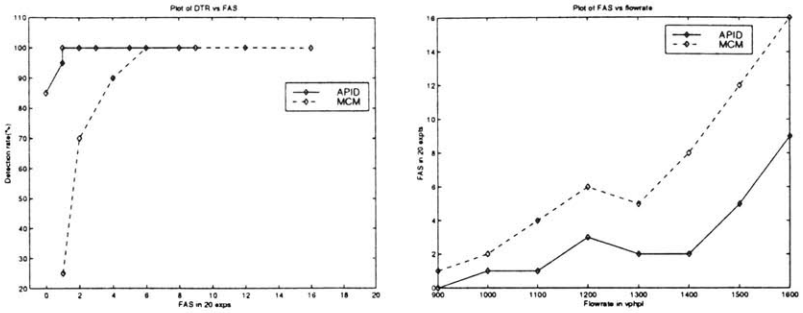
Inc. Duration	Parameter	900	1000	1100	1200	1300	1400	1500	1600
	TTD(min)	2.09	2.18	2	2.05	2.16	1.8	1.8	1.75
3 min	DTR(%)	55	70	95	95	95	100	100	100
	FAS	0	1	2	1	1	2	2	3
	TTD(min)	2.83	2.53	2.24	2	2.08	1.95	1.92	1.85
6 min	DTR(%)	45	70	85	95	95	100	100	100
	FAS	0	2	1	2	3	4	4	4
	TTD(min)	2.2	2.25	2.23	2.2	2.18	1.98	1.95	1.8
10 min	DTR(%)	25	70	90	100	100	100	100	100
	FAS	1	2	4	6	5	8	12	16

Table 3.3: APID algorithm, Persistency=3, Spacing=200ft

Inc. Duration	Parameter	900	1000	1100	1200	1300	1400	1500	1600
	TTD(min)	2.00	2.37	2.08	2.03	1.90	1.63	1.58	1.55
3 min	DTR(%)	30	75	100	100	100	100	100	100
	FAS	0	1	0	0	0	1	0	0
	TTD(min)	3.50	2.81	2.13	1.94	1.78	1.60	1.60	1.58
6 min	DTR(%)	45	90	100	100	100	100	100	100
	FAS	0	1	0	0	1	0	0	2
	TTD(min)	2.91	2.00	1.71	1.65	1.60	1.50	1.50	1.50
10 min	DTR(%)	85	95	100	100	100	100	100	100
	FAS	0	1	1	3	2	2	5	9

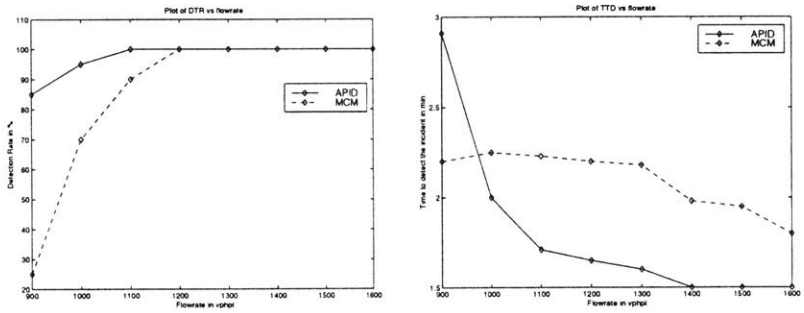
ii. The McMaster in its original form has a high number of false alarms. This can be attributed to the fact that the algorithm in its original form does not effectively distinguish between incident and other forms of congestion when the spacing is 200 ft. The heuristics help identify incidents clearly. Some scenarios even have a FAS of 0 at low flow rates. These results are very similar to the findings of Peeta and Das[39].

iii. Both algorithms have extremely good performance with respect to false alarms. However, as mentioned in earlier sections, these results are *misleading* and do not represent the performance under a general set of conditions. In the second set of experiments on I-93 North in the next section, it will be shown that the performance deteriorates drastically in a more complex network.



(a) DTR vs FAS (b) FAS vs Flow rate

Figure 3-1: Plot of MOEs



(a) DTR vs Flow rate (b) TTD vs Flow rate

Figure 3-2: Plot of MOEs

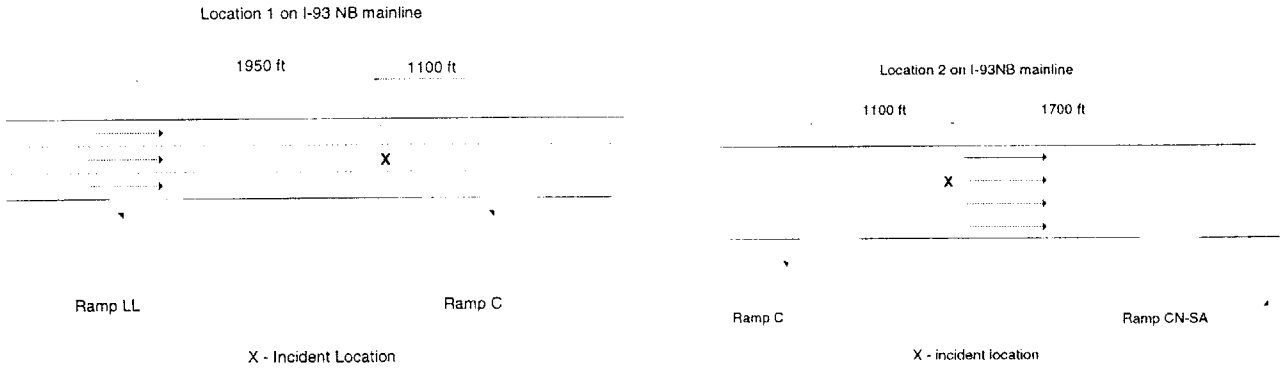
3.3.2 Evaluation on I-93 North

Experiments on I-93 North

Incidents were placed at 2 locations. A total of 20 replications, each of 1.5 hrs were simulated. The OD flow of 2004PM was used as the base flow. The Incidents were simulated to occur 10 minutes after the beginning of the simulation and lasted 15 minutes. A complete lane blockage of the center lane in a 3 lane section was considered for location 1 and a complete blockage of the second lane from the left of a 4 lane section. The total number of stations in the present case is 74. The stations consist of both 3 and 4 lane sections (correspondingly 3 and 4 sensors). Therefore the total number of checks made by the algorithm in 20 experiments of 1.5 hr each with a polling

period of 30 seconds is 180x20x300(total number of sensors) = 1080000 at 200ft spacing, 360000 at 600ft spacing and 180000 at 1200ft spacing. The refinements used for the I-90 West case were also applied to the I-93 North. But it was found that the count constraint was too tight for the 3 lane and 4 lane case. This resulted in many missed detections. Therefore this constraint was relaxed. However, the state constraint and the false alarm memory conditions were retained. Unlike the Ted Williams tunnel, incident alarms were observed before the incidents in some replications. The results have been presented in Table 3.4 to Table 3.1.

i. *Location 1* - South of I-90, downstream of ramp LL and upstream of ramp C.



(a) Location 1

(b) Location 2

Figure 3-3: Incident Locations

ii. *Location 2* - In the weaving section, downstream of ramp C and upstream of ramp CN-SA. This congestion in this section builds up at a much faster rate than location 1.

Analysis of tests on I-93 North

Table 3.4: Basic McMaster Algorithm Persistency=2

		200 ft				600 ft				1200 ft			
Loc	**Demand**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**
Loc 1	**TTD(min)**	1.83	2.56	2.05	1.95	2.19	4.13	5.6	6.82	3.64	4.75	8.03	7.96
	DTR(%)	45	80	95	90	65	95	100	100	70	80	75	70
	FAS	341	251	161	158	202	105	82	54	149	94	118	119
	FAR(%)	0.032	0.023	0.015	0.015	0.056	0.029	0.023	0.015	0.083	0.052	0.066	0.066
Loc 2	**TTD(min)**	2.18	1.27	1.07	1.06	2.81	1.62	1.52	1.43	2.95	2.73	6.17	5.11
	DTR(%)	55	65	70	84	65	85	100	100	100	100	90	95
	FAS	327	229	197	183	168	119	114	121	130	132	109	127
	FAR(%)	0.030	0.021	0.018	0.017	0.016	0.033	0.032	0.034	0.072	0.073	0.061	0.071

Table 3.5: Refined McMaster Algorithm Persistency=2

		200 ft				600 ft				1200 ft			
Loc	**Demand**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**
Loc1	**TTD(min)**	2.58	3.65	1.68	1.92	4.00	4.38	5.28	5.3	-	6.43	8.03	7.96
	DTR(%)	30	85	100	90	5	90	100	100	0	35	75	70
	FAS	110	59	94	76	108	16	35	27	51	26	49	46
	FAR(%)	0.010	0.006	0.009	0.007	0.030	0.004	0.009	0.008	-	0.014	0.027	0.026
Loc2	**TTD(min)**	2.67	1.62	1.47	1.65	3.79	1.82	1.82	2.02	7.08	3.48	6.66	5.63
	DTR(%)	90	100	95	89	70	95	100	30	100	100	95	95
	FAS	228	196	180	166	128	80	71	79	85	59	66	73
	FAR(%)	0.021	0.018	0.017	0.015	0.036	0.022	0.019	0.022	0.047	0.033	0.037	0.041

Table 3.6: APID Algorithm Persistency=2

		200 ft				600 ft				1200 ft			
Loc	**Demand**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**
Loc 1	**TTD(min)**	1.96	2.18	1.82	1.65	2.53	2.55	3.58	3.85	2.75	3.92	7.04	6.73
	DTR(%)	65	95	100	85	75	100	100	100	10	90	65	75
	FAS	93	95	89	90	273	144	119	125	50	32	66	59
	FAR(%)	0.009	0.009	0.008	0.008	0.025	0.040	0.033	0.035	0.028	0.018	0.037	0.033
Loc 2	**TTD(min)**	1.88	1.82	1.53	2.03	3.07	1.93	2.10	2.00	4.00	3.88	3.87	4.10
	DTR(%)	75	100	80	90	75	100	100	30	100	100	95	100
	FAS	173	134	120	111	177	167	148	130	60	74	91	72
	FAR(%)	0.016	0.012	0.011	0.010	0.049	0.046	0.041	0.036	0.033	0.041	0.051	0.040

Table 3.7: Basic McMaster Algorithm Persistency=3

		200 ft				600 ft				1200 ft			
Loc	**Demand**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**
Loc 1	**TTD(min)**	2.75	3.45	2.53	2.83	2.92	6.00	6.31	6.82	3.75	6.75	7.8	8.5
	DTR(%)	50	83	90	90	30	45	65	55	30	20	25	35
	FAS	160	86	133	101	67	38	55	54	56	24	38	41
	FAR(%)	0.015	0.008	0.012	0.009	0.019	0.011	0.015	0.015	0.031	0.013	0.021	0.023
Loc 2	**TTD(min)**	2.41	1.58	1.55	1.58	2.50	2.33	2.05	2.38	3.50	3.30	4.78	4.16
	DTR(%)	80	95	100	95	70	100	100	100	95	100	100	95
	FAS	167	163	189	161	42	73	69	72	44	49	50	61
	FAR(%)	0.015	0.015	0.018	0.015	0.012	0.020	0.019	0.020	0.024	0.027	0.028	0.034

Table 3.8: Refined McMaster Algorithm Persistency=3

		200 ft				600 ft				1200 ft			
Loc	**Demand**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**
Loc 1	**TTD(min)**	2.75	4.08	2.67	2.92	4.50	5.69	6.31	6.82	-	7.25	7.8	8.5
	DTR(%)	10	43	90	90	5	40	65	55	0	10	25	35
	FAS	52	24	48	51	28	5	19	12	-	14	16	18
	FAR(%)	0.005	0.002	0.004	0.005	0.008	0.001	0.005	0.003	-	0.008	0.009	0.010
Loc 2	**TTD(min)**	3.00	2.25	2.12	2.56	3.79	3.10	2.33	2.92	8.00	4.17	5.42	4.87
	DTR(%)	65	100	100	95	60	100	100	5	100	100	100	95
	FAS	139	128	135	114	61	56	52	49	37	41	43	46
	FAR(%)	0.013	0.012	0.013	0.011	0.017	0.016	0.014	0.014	0.021	0.023	0.024	0.026

Table 3.9: APID Algorithm Persistency=3

		200 ft				600 ft				1200 ft			
Loc	**Demand**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**
Loc 1	**TTD(min)**	2.50	2.47	2.34	2.24	3.72	3.32	3.89	4.36	-	4.65	7.35	6.42
	DTR(%)	55	85	95	85	45	95	95	90	-	85	50	30
	FAS	53	33	54	51	49	69	86	79	3	24	58	48
	FAR(%)	0.005	0.003	0.005	0.005	0.014	0.019	0.024	0.022	0.002	0.013	0.032	0.027
Loc 2	**TTD(min)**	2.61	2.33	2.07	2.28	3.92	2.50	2.62	2.60	8.00	5.15	4.85	5.34
	DTR(%)	45	100	75	80	65	100	100	100	5	85	85	95
	FAS	49	53	61	58	90	81	96	93	60	44	58	45
	FAR(%)	0.005	0.005	0.006	0.005	0.025	0.023	0.027	0.026	0.033	0.024	0.032	0.025

Table 3.10: Basic McMaster Algorithm Persistency=4

		200 ft				600 ft				1200 ft			
Loc	**Demand**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**
Loc 1	**TTD(min)**	3.30	3.07	3.26	3.57	3.00	5.00	6.8	6.5	-	7.5	7.67	8.00
	DTR(%)	25	35	85	70	5	15	25	25	0	5	15	5
	FAS	66	100	77	62	25	14	53	29	31	8	18	22
	FAR(%)	0.0061	0.009	0.007	0.006	0.007	0.004	0.015	0.008	0.017	0.004	0.010	0.012
Loc 2	**TTD(min)**	2.32	2.23	2.12	2.28	3.57	3.45	2.83	3.03	4.50	3.95	4.05	4.78
	DTR(%)	70	100	100	100	70	100	100	95	90	100	95	100
	FAS	74	130	148	111	28	48	51	48	25	31	32	37
	FAR(%)	0.007	0.012	0.014	0.010	0.008	0.013	0.014	0.013	0.014	0.017	0.018	0.021

Table 3.11: Refined McMaster Algorithm Persistency=4

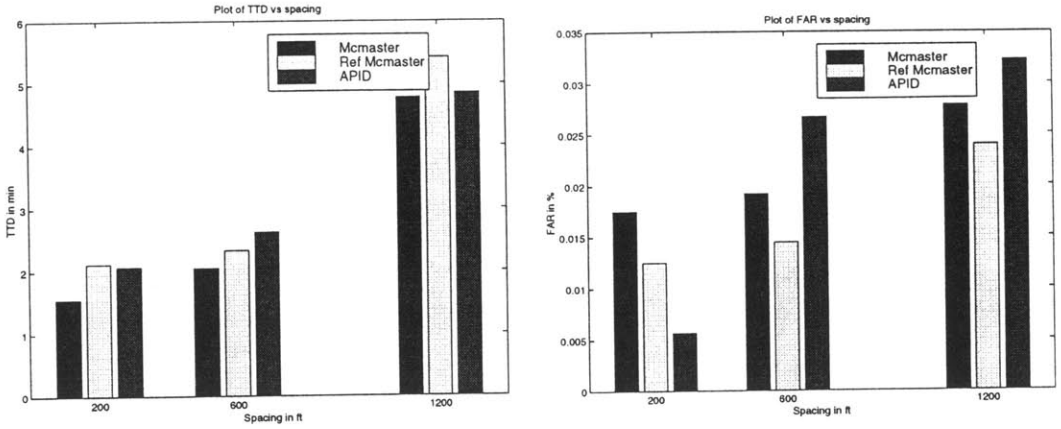
		200 ft				600 ft				1200 ft			
Loc	**Demand**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**
Loc 1	**TTD(min)**	2.00	2.75	3.11	3.46	-	9.5	6.80	6.50	-	-	7.7	8.0
	DTR(%)	5	10	70	60	0	5	25	25	0	0	15	5
	FAS	12	15	39	33	5	4	10	6	-	-	9	7
	FAR(%)	0.001	0.001	0.004	0.003	0.001	0.001	0.003	0.002	-	-	0.005	0.004
Loc 2	**TTD(min)**	3.50	2.85	2.67	3.14	4.33	3.73	3.05	3.44	-	4.70	4.75	5.25
	DTR(%)	45	100	100	95	45	100	100	90	0	100	100	100
	FAS	72	87	83	74	27	40	38	34	14	28	30	31
	FAR(%)	0.007	0.008	0.008	0.007	0.008	0.011	0.009	0.008	0.008	0.016	0.017	0.017

Table 3.12: APID Algorithm Persistency=4

		200 ft				600 ft				1200 ft			
Loc	**Demand**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**	**60%**	**80%**	**100%**	**120%**
Loc 1	**TTD(min)**	2.40	2.97	2.75	2.90	3.80	3.82	4.56	4.89	-	4.87	7.42	6.92
	DTR(%)	25	85	90	75	25	85	90	90	-	25	80	80
	FAS	32	18	21	22	15	42	65	47	-	25	58	43
	FAR(%)	0.003	0.002	0.002	0.002	0.004	0.012	0.018	0.013	-	0.014	0.032	0.024
Loc 2	**TTD(min)**	2.94	2.81	2.60	2.73	4.19	3.00	3.12	3.25	-	5.85	5.57	6.19
	DTR(%)	40	90	75	75	40	100	100	100	-	65	70	65
	FAS	28	38	28	34	56	61	71	68	9	30	51	40
	FAR(%)	0.003	0.004	0.003	0.003	0.016	0.017	0.019	0.019	0.005	0.017	0.028	0.022

Findings

I. *Impact of spacing*



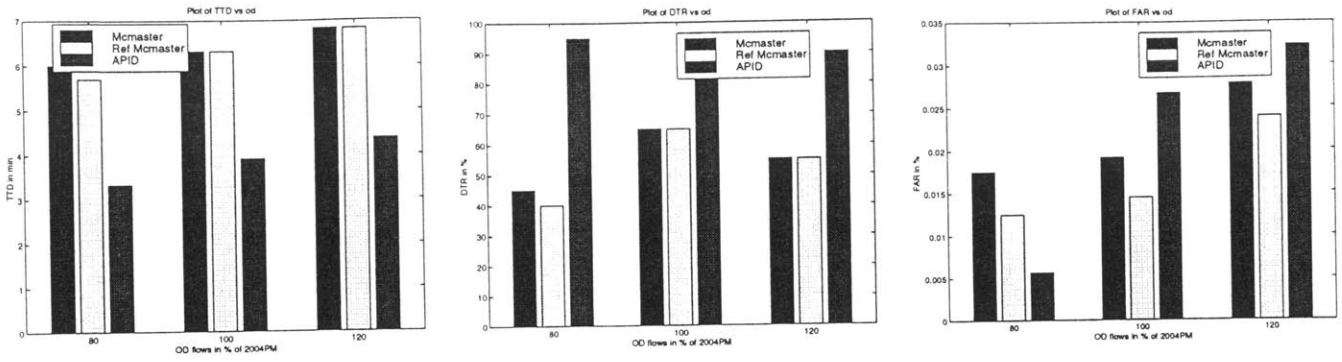
(a) TTD (b) FAR

Figure 3-4: TTD and FAR vs spacing at 100% flowrate with pers=3 at location 2

i. At a spacing of 200ft, the performance at less than 80% flows is worse than that at 100% or 120% flows. Although the FAS at 200 ft is greater than that at 600 ft or 1200 ft, the FAR is greater at higher spacings, indicating that there are greater false alarm confirmations for the same amount of checks made by the algorithm. This could be attributed to large variation in sensor data between 2 adjacent sensors at higher spacings and loss of information about geometry.

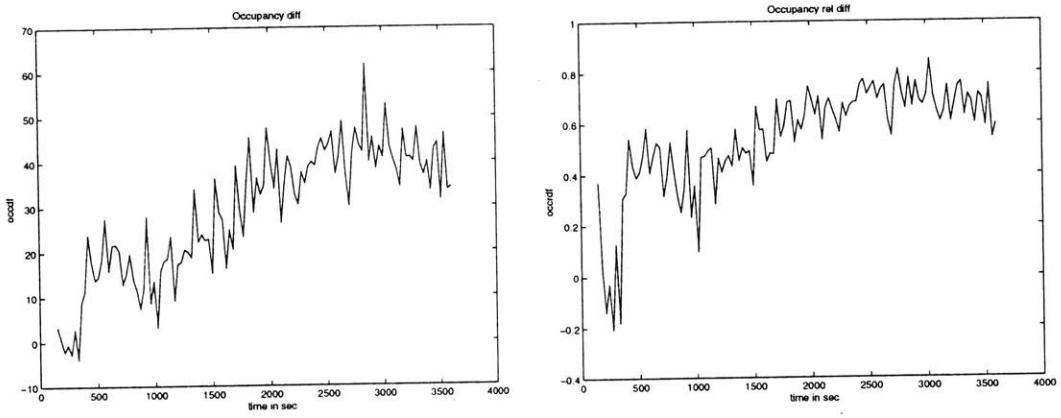
ii. The APID algorithm exhibits poor performance in terms of FAS when the spacing between the sensors becomes high. But the performance in terms of DTR and TTD is much better. This is because of comparative measures used by APID to detect incidents, difference in occupancy(OCCDF) and the relative difference in occupancy(OCCRDF) are difficult to use when there is considerable difference in geometry between two adjacent sensor stations. This has been identified by Persaud et al.[7] and is also seen in the plots of OCCDF and OCCRDF for 2 stations on either side of an on-ramp.

II. *Impact of Flow*



(a) TTD (b) DTR (c) FAR

Figure 3-5: TTD,DTR and FAR vs Flow rate at 600ft spacing with pers=3 at location 1



(a) OCCDF (b) OCCRDF

Figure 3-6: OCCDF and OCCRDF for 2 stations on either side of an on-ramp

i. The level of flow in the network clearly affects the performance of the algorithms. In almost **all** scenarios, the algorithms have very poor performance under conditions of low flow, eg:-Table 3.8, Table 3.11 and Table 3.12 show very low DTR at a dense spacing of 200ft for 60% and 80% flows. This is intuitive because it is not easy to detect the effect of an incident in an uncongested condition. However, when the flow rate is high, the effect of the incident is very noticeable. Therefore the number of false alarms increase.

ii. In some scenarios, it can be seen that a large number of false alarms are observed even at low flows of 60%. Most of the false alarms generated at low flows do not persist

for long. They are cleared immediately after they are declared as the congestion does not build up sufficiently. The APID algorithm has compression wave testing for such turbulent traffic conditions. Therefore, it performs much better than the Conventional McMaster in terms of FAR. But Refined McMaster has better performance than APID even in these scenarios. This indicates that clever heuristics could be used in place the of persistency checking or suspension of detection for better results.

iii. The refined algorithm maintains, in most cases, the performance in terms of TTD & DTR while considerably lowering the FAS. At lower demands, the TTD for the refined algorithm is slightly higher than the original algorithm.

III. *Impact of Persistency*

i. The persistency of the algorithm directly affects the MOEs. A higher persistency increases the time taken to detect the incident, but reduces the number of false alarms and may also reduce the detection rate. A lower persistency increases the detection rate and decreases the time to detect, but increases the false alarm rate. A lower persistency should not be used always because the congestion has not developed completely during the short persistency check period and random fluctuations can be easily confused for incidents. Table 3.8, Table 3.11, Table 3.9 and Table 3.12 clearly show the effect of persistency. The difference between the average TTD for the same scenarios with different persistencies is approximately 0.5 min (30 sec = 1 persistency time interval).

ii. Different algorithms have different performance with respect to persistency tests. When the persistency is increased to 4, the Refined McMaster algorithm fails completely but the APID algorithm continues to detect incidents with high DTR and controlled FAR. When a persistency of 2 is used, the Refined McMaster algorithm has much better performance at lower persistencies with low FAR.

IV. *Impact of Interaction*

i. The choice of persistency depends on the algorithm and the sensor spacing. In order to make the refined McMaster algorithm operational, it would be necessary to use persistency of 2 at high spacings (1200 ft) and high persistency at spacing $\leq$ 600 ft. APID

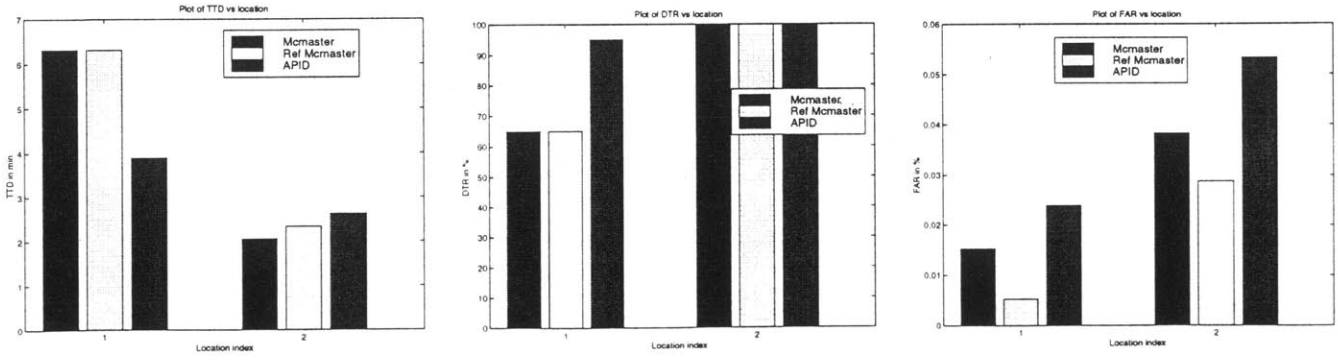
however has very poor performance at low persistencies and all spacings. This implies a higher persistency should be chosen for APID.

ii. The trade-off between the detection rate and the false alarm rate is very clear if the performance of the APID and the refined McMaster algorithms are compared(Table 3.8 compared with Table 3.9 & Table 3.11 compared with Table 3.12). The APID algorithm almost always has a detection rate exceeding 80%. But the false alarm rates are mostly higher than that of the refined McMaster algorithm. The detection times at spacings of 200ft and 600ft are comparable. But at a spacing of 1200 ft, the APID algorithm outperforms the McMaster algorithm in terms of the time to detect and the Detection rate.

V. *Impact of Location*

i. Most of the false alarms were observed either at locations close to the incident during the post-incident phase or at stations immediately upstream of on- and off-ramps. Both algorithms are relatively sensitive to such geometric non-uniformities.

ii. The location of the incident is also critical. Location 2 is in a weaving section where the congestion develops rapidly. Therefore, in all the scenarios it can be seen that TTD and DTR are much better for location 2. However, low TTD and high DTR is accompanied by high FAS.



(a) TTD (b) DTR (c) FAR

Figure 3-7: TTD,DTR and FAR vs Location at 200ft spacing,pers=3,od=100%

Chapter 4

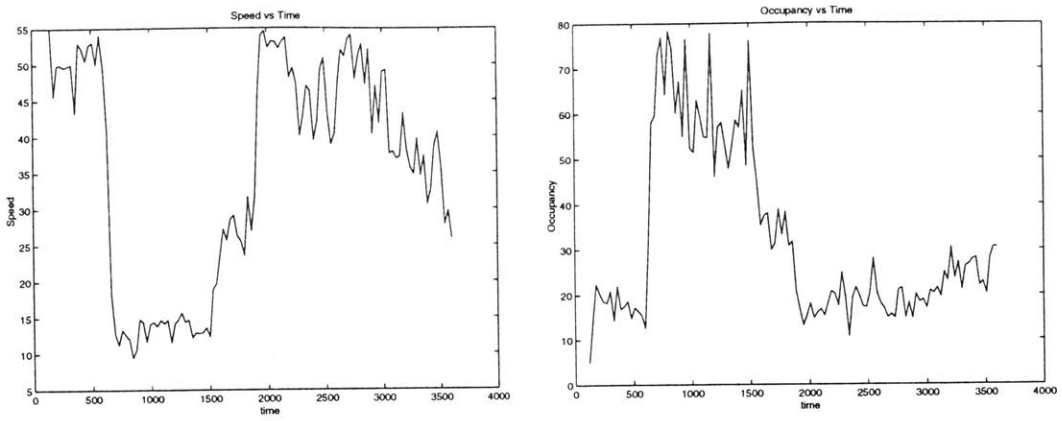
A Heuristic algorithm for dense loop detector spacing

4.1 Introduction

Conventional incident detection algorithms use elaborate criteria based on speed, occupancy and flow to identify congestion and differentiate incident congestion from recurrent congestion. However, at a close spacing of sensors such as that in I-93 North, incidents can be quickly and reliably identified by observing significant changes in speed or occupancy at any sensor station. Occupancy or speed should have been used over flow as indicators because flow is not a clear indicator of the level of congestion as identified by our preliminary experiments and by Lin and Danganzo[45].

4.2 The Proposed Method

Figure 4-1 shows typical change in speed and occupancy at the sensor upstream of the incident lane. Such sharp changes are outside the range of random fluctuations. By detecting such a sharp drop in speed or rise in occupancy, we can detect the incident quickly and reliably. The traffic variables and parameters used by the method are discussed in this section. The method uses the current value of the average speed at the sensor - s_t, the value of average speed in the last interval, s_{t-1} and the value of average speed in the interval before the last, s_{t-2}. The method has been found



(a) Speed Drop

(b) Occupancy Rise

Figure 4-1: Speed Drop and Occupancy Rise at station upstream of incident

to be sensitive to traffic conditions. Under heavy congestion, the method may have difficulties in identifying incidents. But under capacity or close to capacity flows, the method is very effective in declaring incidents. The magnitude of the drop depends on the traffic conditions. The proposed method does not use persistency tests in its current version. This reduces the time to detect the incident considerably when the performance is compared to that of other conventional algorithms. The values of the parameters have been presented in Table 4.1. The framework used to detect incidents and the conditions necessary to characterize incidents have been identified in Figure 4-2 below. The different parameters and the conditions have been summarized below.

4.3 The Framework

The method uses the following variable values to detect incidents.

i. The current value of the average speed at the sensor, s_t

ii. The value of average speed in the last interval, s_{t-1}

iii. The value of average speed in the interval before the last, s_{t-2}

The input required to the algorithm is the same as that for the APID algorithm. Therefore, it is easy to implement. Although speed has been used as the indicator measure, occupancy could also be used. The framework used to detect incidents and the conditions necessary to characterize incidents have been identified in Figure 4-2 below. The values of the parameters have been presented in Table 4.1. The core of the framework has been presented in pseudocode form below:-

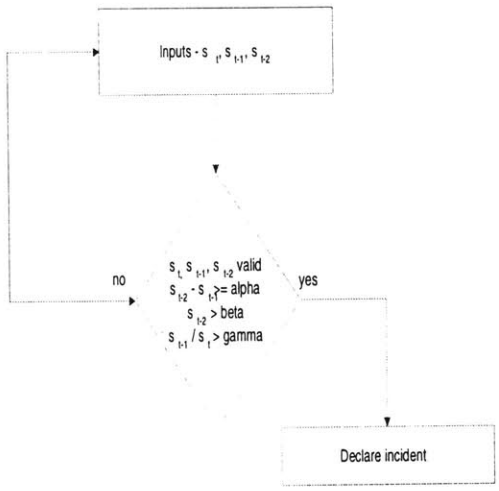


Figure 4-2: Framework to declare incidents

$$\text{if}((s_t, s_{t-1}, s_{t-2} \geq 0.0)\&(s_{t-2} - st - 1 > \alpha)\&$$
$$(s_{t-2} \geq \beta)\&(s_{t-1}/s_t \geq \gamma))$$
$$\textit{DECLARE INCIDENT}$$

The different parameters and the conditions have been summarized below:-

i. $s_{t-2}, s_{t-1}, s_t \geq 0.0$ This condition is essential to eliminate spurious values and to account for sensor malfunctioning. This establishes validity of the data which is part of the test as indicated in the framework in Figure 4-2.

ii. $\alpha = min(s_{t-2} - s_{t-1})$ This is the parameter used primarily to detect incidents. $s_{t-2} - s_{t-1} \geq \alpha$ for incident to be detected.

iii. $\beta =$ minimum average speed.

Table 4.1: Results

Parameter	200 ft	600 ft
α(mph)	25	20
β(mph)	30	30
γ	0.77	0.77

iv. $\gamma = min(s_{t-1}/s_t)$ This is needed to check for random fluctuations. $s_{t-1}/s_t \geq \gamma$ for declaring incidents. In case of incident, $s_{t-1} \geq s_t$. But, under conditions of low flow which implies that s_t can increase compared to s_{t-1}. Its assumed that s_{t-1} can not be lesser than γ % of s_t under any flow conditions for an incident.

4.4 Experiments on I-93 North

Experiments were conducted on the I-93 North to evaluate the performance of the Speed Drop algorithm. All alarms in the first five minutes have been rejected because the network loading has not been completed in that time and the flows have not stabilized. A complete blockage of the

Table 4.2: Scenario properties

Description	Range
Spacing of the sensors(ft)	200, 600
Flow rate(vphpl)	80%,100%,120% of OD2004 PM
Incident Duration(min)	15
Replications per experiment	20
Duration of Simulation(min)	90
Polling Period of algorithm(min)	0.5
Incident Start	10 min after start of simulation

second lane from the left, of a 4 lane section, was simulated. Incident was placed Location 2. The results have been presented in Table 4.3.

Table 4.3: Results at 200 ft and 600 ft sensor spacing

200 ft sensor spacing									
	80%			100%			120%		
Demand	Ref MCM	APID	SD	Ref MCM	APID	SD	Ref MCM	APID	SD
TTD(min)	2.25	2.76	1.22	2.12	2.01	1.03	2.56	2.02	1.27
DTR(%)	100	95	80	100	100	80	95	95	85
FAS	128	39	37	135	59	17	114	61	17
600 ft sensor spacing									
	80%			100%			120%		
Demand	Ref MCM	APID	SD	Ref MCM	APID	SD	Ref MCM	APID	SD
TTD(min)	5.69	3.32	1.62	6.31	3.89	1.82	6.82	4.36	2.00
DTR(%)	40	95	40	65	95	70	55	90	60
FAS	5	69	47	19	86	38	12	79	44

4.5 Summary and Findings

i. The time to detect the incidents is less than half the time taken by the other algorithms both at 200 ft and 600 ft spacing.

ii. At 200 ft spacing, the number of false alarms in 20 replications is less than half those generated by APID at 100% and 120% flow rates. The number of false alarms is significantly smaller than that of McMaster even in its refined form. However at 600 ft spacing, the SPDDRP generates more FAS than the Refined McMaster algorithm but less FAS than APID.

iii. The Detection rate of the SPDDRP algorithm is less than the other algorithms. The advantage to be gained from the SPDDRP algorithm is primarily the TTD. In order to make the SPDDRP algorithm operational, it would be necessary to combine it with another algorithm having complementary characteristics (poor TTD, low FAS and high DTR).

iv. The algorithm is found to be sensitive to traffic conditions. If the conditions without incident are highly congested, the algorithm does not have good performance.

Chapter 5

Discrete choice models for incident detection algorithms

5.1 Introduction

Fundamentally, incident detection algorithms (IDAs) can be probabilistic or deterministic. Probabilistic algorithms are different forms of the Neyman-Pearson (or) the Bayesian formulation[35, 36]. Deterministic algorithms use sensor data to identify incidents. The California algorithms[17, 18] and the McMaster algorithm[20, 7] are two main types of deterministic algorithms. Calibration of these algorithms is based on studying the range of variations in fundamental traffic variables as reported by the sensor data. Most of the calibration efforts use simulated data as a starting point. The advantage of simulation is that a large amount of incident data can be generated. Different scenarios can be studied. The disadvantage of simulation data is that it is subject to different modeling inaccuracies. Factors that affect the performance of an incident detection algorithm are sensor spacing, geometric configuration of the freeway network, location of the incident between two sensors, the duration of persistency checks made by the algorithm etc. All these factors can be varied using simulation. In spite of such extensive calibration, the parameters have to be tuned using the real data once it is available. Thus, using real data in the calibration phase would yield more realistic parameters. But the disadvantages are that it is not very easily available and is often subject to different types of confounding effects, sensor malfunctioning etc. which have to be fil-

tered. A mechanism is necessary to combine the information contained in real and simulated data. This would improve the parameters while, at least partially, eliminating the drawbacks of either data source. Although, we could use information from simulation to get more significant models, we would have to correct the differences in scale between the data sources. The problem of estimation using multiple data sources has been studied in detail Ben-Akiva et al.[32]. They have presented a method to combine revealed preference(RP) and stated preference(SP) data. SP data have complimentary characteristics to RP data. Combining RP and SP data uses the information from SP data on trade-offs among attributes to help identify different parameters while correcting biases contained in the SP data. Real incident data can be treated as RP data and simulated data is equivalent to SP data. Combined estimation using both real and simulated data strengthens the estimation by using the information in the simulated data while correcting for biases in the simulated data using real data. A close analogy exists between that study and the current study as seen above. Therefore, a similar approach is used in this chapter to combine real and simulated incident data.

5.2 Model Specification and Estimation

Incidents are detected by analyzing the values of the fundamental traffic variables: speed, occupancy and flow, measured using loop detectors. Spatial and temporal differences between the values of these variables can be constructed. Spatial differences between values of traffic variables of two adjacent stations and temporal differences in the values of traffic variables at a station can also be constructed. Given the values of these variables, we can compute the probability of an incident. Different specifications have been studied in the following sections.

Traffic state is identified to be in one of the following three states - free flow, congestion due to incident and congestion due to bottlenecks. Incident data is generated using the simulator for different scenarios. Data from one scenario, reasonably free from confounding effects of different types of congestion, is selected. Sensors on either side of the incident were chosen to sample the incident data. During the incident, the state was classified to be incident. Classification of the state as incident-free or recurrent congested was based on visual inspection of the values of the traffic

variables (speed, occupancy) at different times. Incident data from all replications of one scenario were also chosen in order to the enrich the sample with incident congested alternative. In this way, choice-based sampling has been used to obtain the data for estimation.

5.2.1 Binary Logit model to identify congestion

A probabilistic incident detection model based on a discrete choice model is used to evaluate the probability of an incident, given the traffic variables at that instant and at that section. The initial binary choice model is used to identify congestion. For the present case two alternatives exist, congested or non-congested flow. Accordingly, a binary logit model has been used. The model specification is as specified in Table 5.1. V_{cong} is the systematic utility associated with the congestion alternative.

$$V_{cong} = \beta + SPD * speed + OCC * occupancy \tag{5.1}$$

The sensor used for estimation is the first sensor upstream of the incident in the affected lane. The

Table 5.1: Initial choice model specification

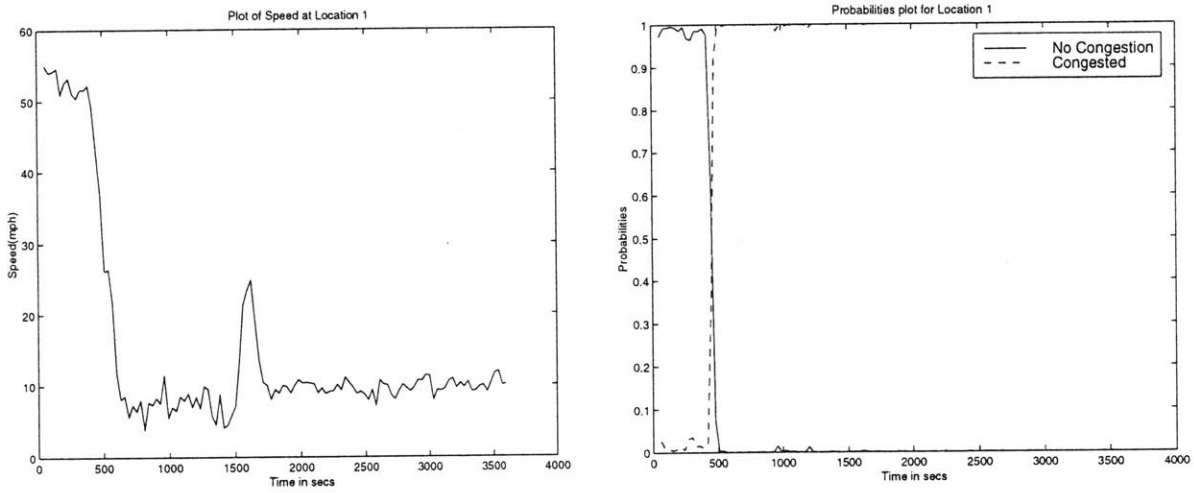
Variable Name	Description
beta	alternative specific constant for no-congestion
SPD	alternative specific for congestion
OCC	alternative specific for congestion

results of the estimation are in Table 5.2 The data for calibration was chosen from Location 2. Demand level of 100% was used for calibration. An incident of 15 min duration was simulated. The incident started 10 min after start of simulation. The model fit is good (0.91) and the log-likelihood is substantially different from that of the trivial model. All the coefficients are significant. The sign of the speed coefficient is intuitive. As the speed increases, the probability of detecting congestion decreases. The sign of the occupancy coefficient is not straightforward because of its correlation with speed measurements. Using the above coefficients, predictions were made for probabilities of of the two states. A plot of the probabilities has been presented in Figure 5-1 and Figure 5-2. This is an extremely simplified model which would be able to detect congestion but not distinguish

Table 5.2: Summary of estimation

Variable Name	Coefficient	Std. Error	t-statistic
beta	-26.85	5.227	-5.137
SPD	-0.1832	0.06197	-2.956
OCC	-0.5468	0.09131	-5.988
Log-likelihood of 0 coeffs model	-471.34		
Log-likelihood of MNL model	-42.471		
ρ^2	0.9099		
$\bar{\rho}^2$	0.9035		

between recurrent and incident congestion. In order to distinguish between recurrent and incident congestion we would have to use a richer specification recognizing the fact that conditions downstream of an incident are different from that downstream of a bottleneck. This is supported by the observation of Hall et al.[1]. In case of location 1 in Figure 5-1, congestion is caused by incident

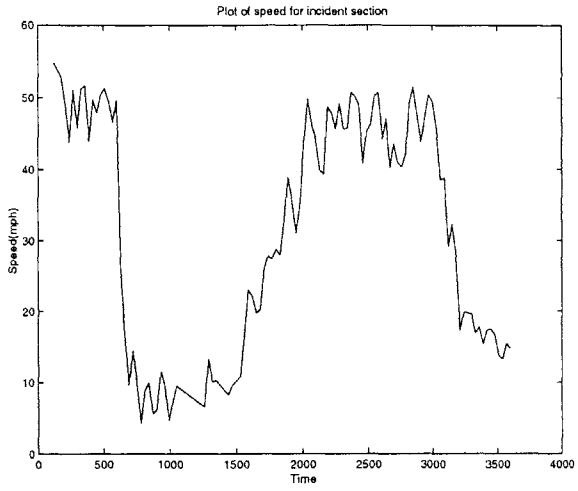


(a) Speed Vs time

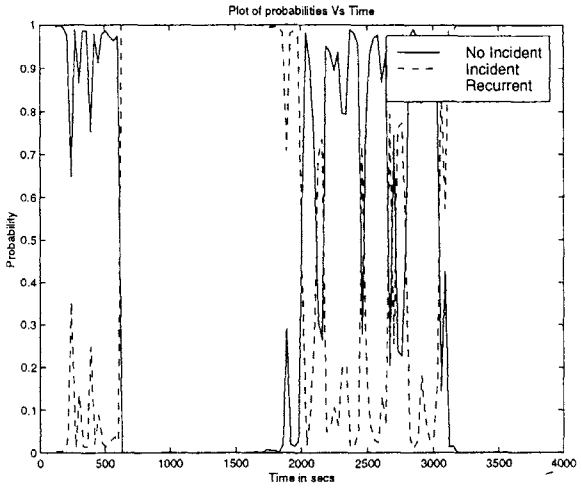
(b) Probabilities Vs time

Figure 5-1: Predicted probabilities for location 1

as well as shock wave propagation from downstream on-ramp. The model predicts the congestion probability accurately. As indicated by the speeds, the congestion starts before the incident. Therefore, the probabilities become one *before* the start of the incident. The probabilities remain at one until the end of the simulation. Clearly, this model is effective in detecting congestion but not incident. It can be seen from Figure 5-2 that the speed drops during the incident for location 2.



(a) Speed Vs time



(b) Probabilities Vs time

Figure 5-2: Predicted probabilities for location 2

The probability of congestion during the incident becomes one. Thus, the model is clearly able to detect congestion during the incident. After that the speeds briefly increase back to normal before dropping again due to bottleneck. The same behavior is reflected in the probabilities. The probabilities are especially turbulent between the end of the incident and the onset of the subsequent shock wave. In the next section, we examine a model identify incident congestion.

5.2.2 Binary Logit model to identify incidents

Since our purpose is to detect incidents, the model presented in the previous section is only partially relevant. In this section, we present a binary logit model in which alternatives are incident and no incident. Thus the base alternative (no incident) in this case includes both free flow and recurrent congestion. In order to distinguish between the two alternatives in this case, the specification has to be improved. A good choice of explanatory variables includes spatial difference in flow, occupancy or speed. Flow(counts) was found to be too noisy with substantial variation because of its discrete nature. Speed has been used for the present case. Relative spatial differences(difference in speed between two adjoining detectors normalized with respect to speed of the upstream detector) are found to be smoother than spatial differences because they represent variation normalized with respect to the upstream variable value. Sample plots of spatial speed differences and relative

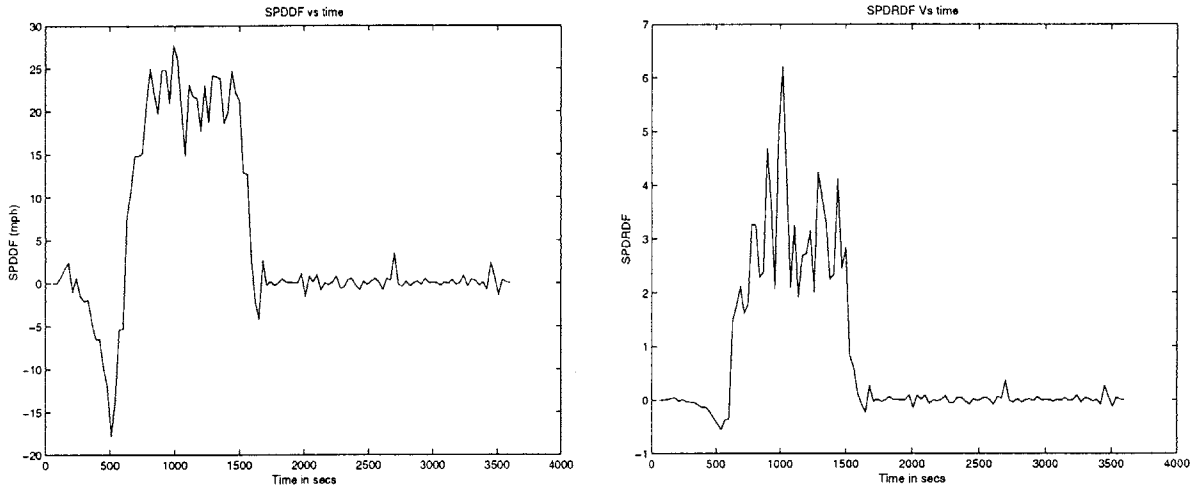
spatial speed differences have been produced in Figure 5-3. V_{inc} is the systematic utility associated with the incident alternative. *spdrdf* and *occrdf* refer to the relative speed difference and relative occupancy difference respectively. They are defined in the equations below.

$$spddf(i,t) = speed(i,t) - speed(i+1,t) \tag{5.2}$$

$$spdrdf(i,t) = (spddf(i,t))/speed(i,t) \tag{5.3}$$

$$occdf(i,t) = occupancy(i,t) - occupancy(i+1,t) \tag{5.4}$$

$$occrdf(i,t) = (occdf(i,t))/occupancy(i,t) \tag{5.5}$$



(a) SPDDF Vs time

(b) SPDRDF Vs time

Figure 5-3: Comparison of SPDDF with SPDRDF

$$V_{inc} = \beta + SPD * speed + OCC * occupancy + FLW * flow + SPDRDF * (spdrdf) \tag{5.6}$$

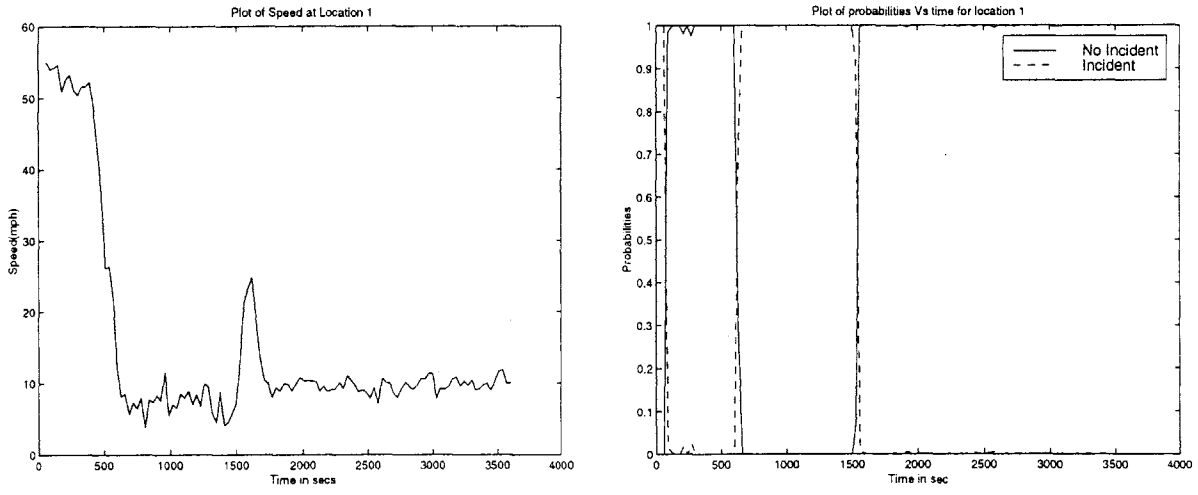
All coefficients from the estimation are seen to be significant. There is also a good fit and the log-likelihood is significantly different from that of the trivial model.

Table 5.3: Initial model specification

Variable Name	Description
beta	alternative specific constant for no-incident
SPD	alternative specific for incident congestion
OCC	alternative specific for incident congestion
FLW	alternative specific for incident congestion
SPDRDF	alternative specific for incident congestion

Table 5.4: Summary of estimation

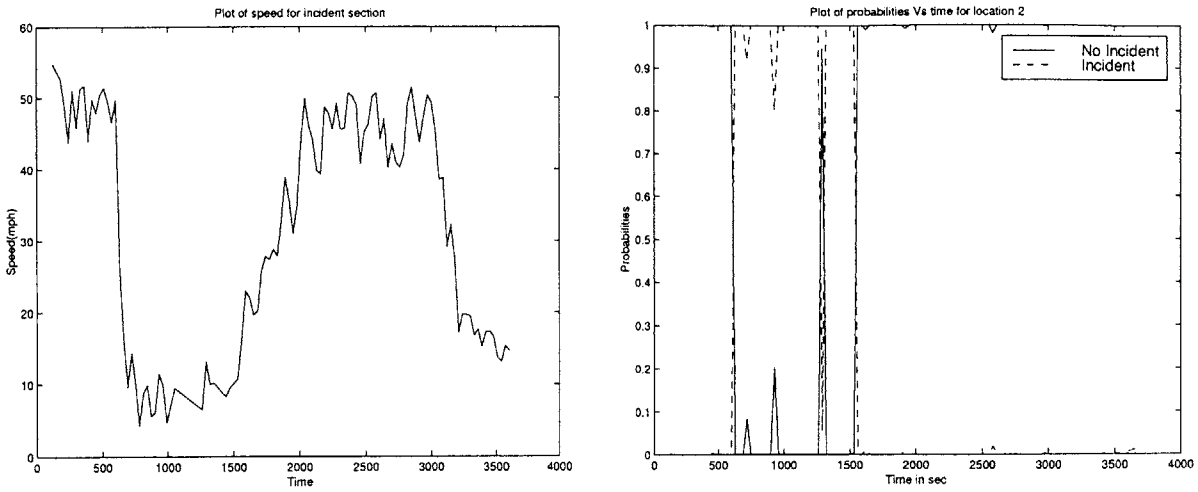
Variable Name	Coefficient	Std. Error	t-statistic
beta	-29.28	4.41	-6.64
SPD	-0.2366	0.1196	-1.978
OCC	-0.2829	0.09681	-2.922
FLW	-1.6150	0.2829	-5.708
SPDRDF	-13.03	2.79	-4.669
Log-likelihood of 0 coeffs model	-471.34		
Log-likelihood of MNL model	-4.4453		
ρ^2	0.9810		
$\bar{\rho}^2$	0.9596		



(a) Speed Vs time

(b) Probabilities Vs time

Figure 5-4: Predicted probabilities for location 1



(a) Speed Vs time

(b) Probabilities Vs time

Figure 5-5: Predicted probabilities for location 2

The probability predictions for location 1, Figure 5-4, are accurate. But for location 2, Figure 5-5, although the trend of the probabilities is accurate most of the time, there is some oscillation in the magnitude of the probability over the incident duration. Such variations is not preferable because this would result in declaring and clearing incidents several times, leading to multiple alarms during the incident.

5.2.3 Binary Logit model to distinguish between incident and recurrent congestion

Many incident detection algorithms proceed in two steps. First, they detect congestion and then distinguish between recurrent and incident congestion. It is easy to detect congestion but the difficulty lies in distinguishing between incident and recurrent congestion. It would be useful to examine the specification necessary to effectively distinguish between incident and recurrent congestion. The model specification is as follows:-

$$V_{inc} = \beta + SPDRDF * spdrdf + OCCRDF * occrdf \tag{5.7}$$

All coefficients from the estimation are seen to be significant. There is also a good fit and the log-

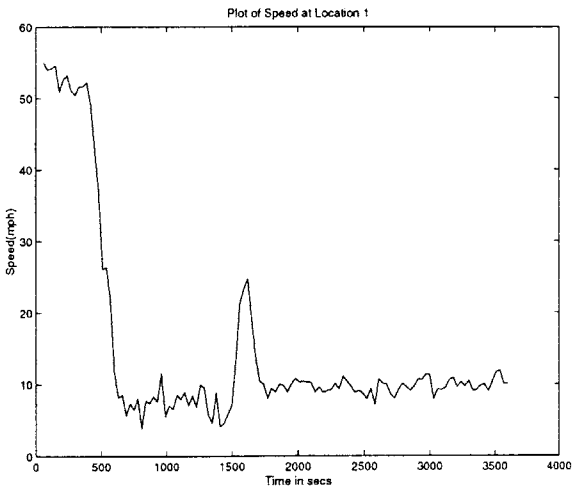
Table 5.5: Initial model specification

Variable Name	Description
beta	alternative specific constant for incident congestion
SPDRDF	alternative specific variable for recurrent congestion
OCCRDF	alternative specific variable for recurrent congestion

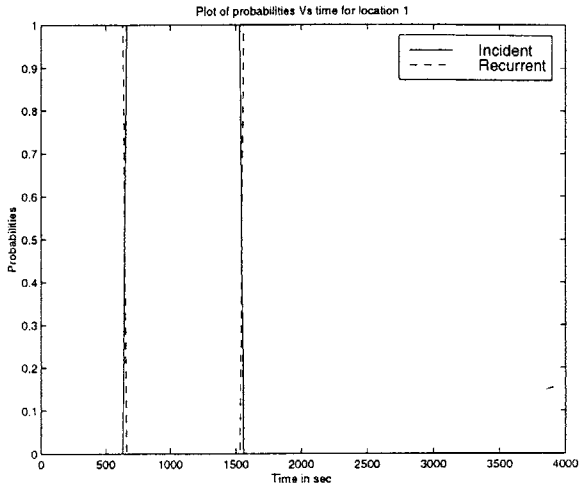
Table 5.6: Summary of estimation

Variable Name	Coefficient	Std. Error	t-statistic
beta	-27.47	9.125	-3.01
SPDRDF	-11.71	5.398	2.169
OCCRDF	-59.35	2.334	-2.543
Log-likelihood of 0 coeffs model	-471.34		
Log-likelihood of MNL model	-2.1519		
ρ^2	0.9894		
$\bar{\rho}^2$	0.9747		

likelihood is significantly different from that of the trivial model. Although, the probabilities are good, the variation in probability over the duration of the incident is still present. The predicted probabilities can be seen to be reasonably accurate for both locations from Figure 5-6 and Figure 5-7.

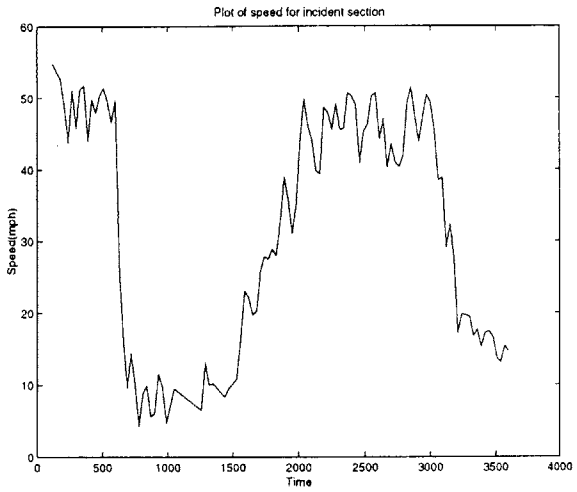


(a) Speed Vs time

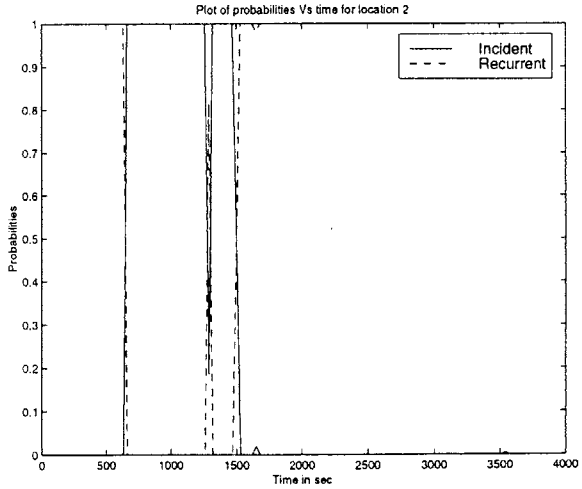


(b) Probabilities Vs time

Figure 5-6: Predicted probabilities for location 1



(a) Speed Vs time



(b) Probabilities Vs time

Figure 5-7: Predicted probabilities for location 2

5.2.4 Multinomial Logit model

A multinomial logit specification can be used to distinguish between the three different traffic states. Each traffic state is treated as a separate alternative.

Table 5.7: Initial model specification

Variable Name	Description
β_{inc}	alternative specific constant for incident congestion alternative
β_{rec}	alternative specific constant for recurrent congestion alternative
OCC_{no}	alternative specific for no congestion congestion alternative
OCC_{inc}	alternative specific for incident congestion alternative
$OCCDF_{inc}$	alternative specific for incident congestion alternative
$OCCDF_{rec}$	alternative specific for recurrent congestion alternative

$$V_{no} = \beta_{inc} + OCC_{no} * occupancy \tag{5.8}$$

$$V_{rec} = \beta_{rec} + OCCRDF_{rec} * occrdf \tag{5.9}$$

$$V_{inc} = OCCRDF_{inc} * occrdf + OCC_{inc} * occupancy \tag{5.10}$$

All coefficients from the estimation are seen to be significant. There is also a good fit and the

Table 5.8: Summary of estimation

Variable Name	Coefficient	Std. Error	t-statistic
β_{inc}	-9.5020	2.6090	-3.6420
β_{rec}	-10.8100	1.3170	-8.2060
OCC_{no}	-0.4112	0.05701	-7.2140
OCC_{inc}	-0.3016	0.0926	-3.2570
$OCCDF_{inc}$	0.3098	0.1983	2.0070
$OCCDF_{rec}$	-0.4107	0.08717	-4.7120
Log-likelihood of 0 coeffs model	-747.056		
Log-likelihood of MNL model	-64.7035		
ρ^2	0.9134		
$\overline{\rho}^2$	0.9054		

log-likelihood is significantly different from that of the trivial model. The signs of the coefficients

are intuitive. The positive sign of OCCDF_INC shows that for a greater occupancy difference, there is a greater probability of incident state than free flow. There is no clear interpretation for the sign of OCCDF_REC. This can be explained by the plot of the Occupancy Difference in Figure 5-8. From the figure, it can be interpreted that the role of OCCDF in distinguishing between no incident and bottleneck is not very clear. OCCDF varies around a mean of 0 for both the no incident and recurrent congested scenarios. It can also be seen in Figure 5-9 and Figure 5-10 that the model is

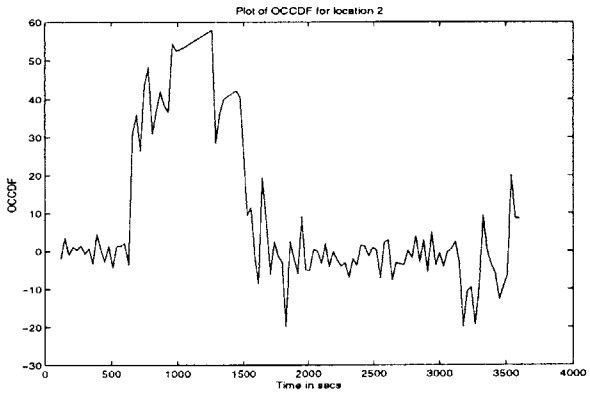
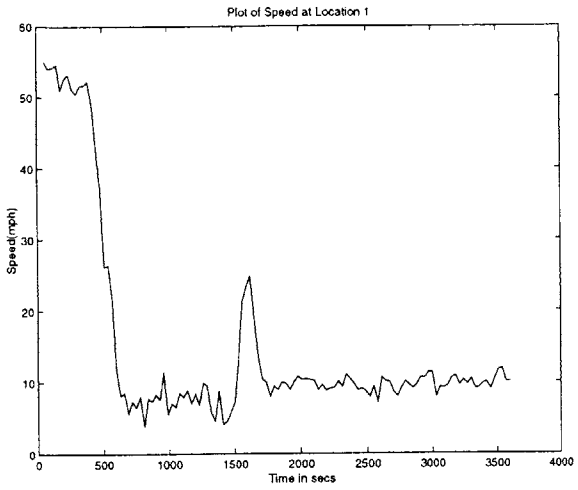
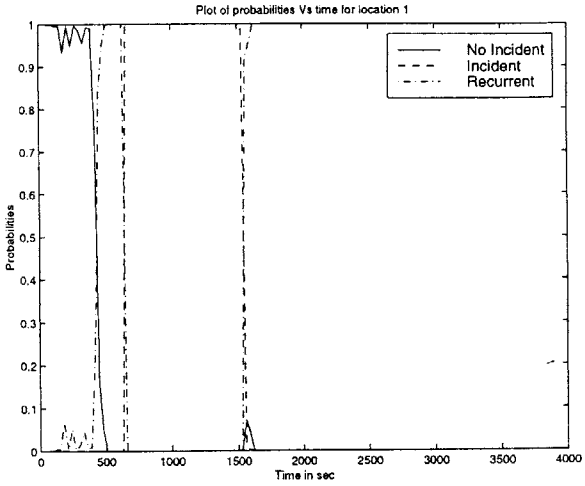


Figure 5-8: Occupancy Difference at Location 2

able to predict the traffic state accurately. Particularly in Figure 5-9, the model is able to identify the propagation of the shock wave at the station. It can be seen from the speed plot that speeds drop much before the occurrence of the incident at t=600 seconds. Thus the model is able to identify and isolate the effect of the incident from that of the bottleneck. In Figure 5-10, the model is able to clearly predict the onset of the incident, the recovery phase, the subsequent bottleneck and the queue close to end of the simulation. The probabilities predicted by the model for location 2 are much more accurate than the previous models. The variation in probabilities is in agreement with the turbulent traffic state after the incident. But the model is able to clearly differentiate between the different states.

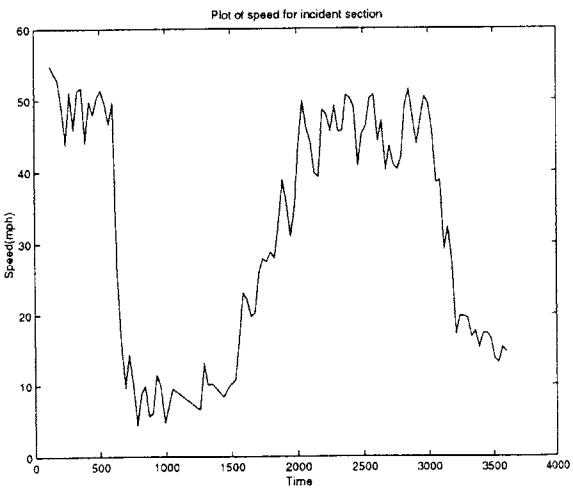


(a) Speed Vs time

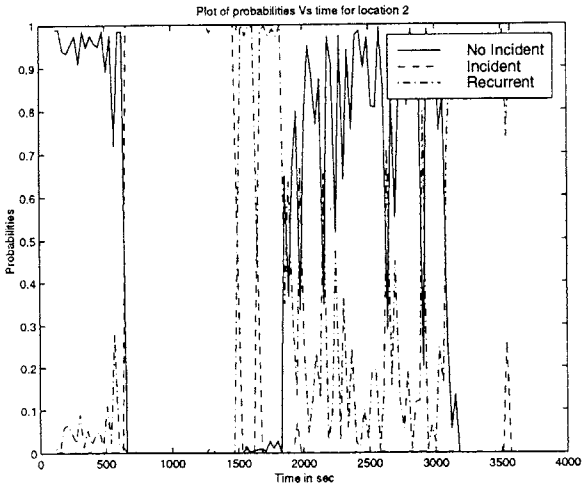


(b) Probabilities Vs time

Figure 5-9: Predicted probabilities for location 1



(a) Speed Vs time



(b) Probabilities Vs time

Figure 5-10: Predicted probabilities for location 2

5.2.5 Nested Logit model

The framework of almost all incident detection algorithms have two main components - a congestion detection routine and a routine to distinguish between incident and recurrent congestion. This essentially means that given the congestion, the algorithm attempts to check if the congestion is due to incident. This suggests a natural extension of the multinomial logit model, discussed earlier, to a nested logit model structure. The a priori hypothesis is that speed/occupancy can be used to identify the presence of congestion and flow/spatial differences can be used to distinguish between the different types of congestion. This is intuitive, because any type of congestion is accompanied by a drop in speed and increase in occupancy. As seen earlier the flows under incident congestion are lesser than recurrent congestion and the spatial differences would be greater. The initial model was estimated using the same data set used to estimate the binary logit and multinomial logit models in the earlier sections. The nested logit model has been estimated using Hielow[29]. The model structure was developed using the model structure presented earlier to identify congestion and the model to distinguish between incident and recurrent congestion. The NL model structure is shown in Figure 5-11

Table 5.9: Model specification

Variable Name	Description
β_{no}	No Congestion alternative specific constant
β_{inc}	Incident alternative specific constant
OCC_{cong}	occupancy alternative specific variable for congestion nest
FLW_{inc}	alternative specific variable for incident congestion alternative
FLW_{rec}	alternative specific variable for recurrent congestion alternative
$OCCRDF_{inc}$	alternative specific variable for incident congestion alternative
$OCCRDF_{rec}$	alternative specific variable for recurrent congestion alternative

Nested Logit Model probability expressions

$$V_{no} = \beta_{no} \tag{5.11}$$

$$V_{cong} = OCC_{cong} * occupancy + V'_{cong} \tag{5.12}$$

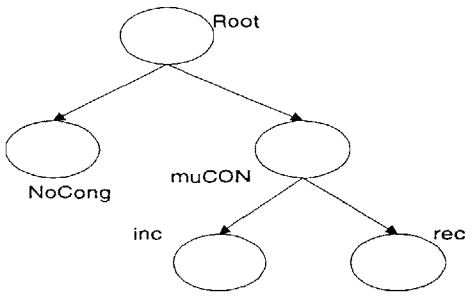


Figure 5-11: Nested logit model structure

$$V'_{cong} = (1/\mu_{cong}) ln(e^{\mu_{cong} V_{inc}} + e^{\mu_{cong} V_{rec}}) \tag{5.13}$$

$$V_{inc} = \beta_{inc} + FLW_{inc} * flow + OCCRDF_{inc} * occrdf \tag{5.14}$$

$$V_{rec} = FLW_{rec} * flow + OCCRDF_{rec} * occrdf \tag{5.15}$$

$$P_{no} = 1/(1 + e^{V_{cong} - V_{no}}) \tag{5.16}$$

$$P_{inc} = (1 - P_{no})(1/(1 + e^{V_{rec} - V_{inc}})) \tag{5.17}$$

$$P_{rec} = (1 - P_{no})(1/(1 + e^{V_{inc} - V_{rec}})) \tag{5.18}$$

V_{no} is the systematic utility associated with the no congestion alternative

V_{inc} is the systematic utility associated with the incident alternative

V_{rec} is the systematic utility associated with the recurrent alternative

P_{no} is the probability of no congestion

P_{inc} is the probability of incident

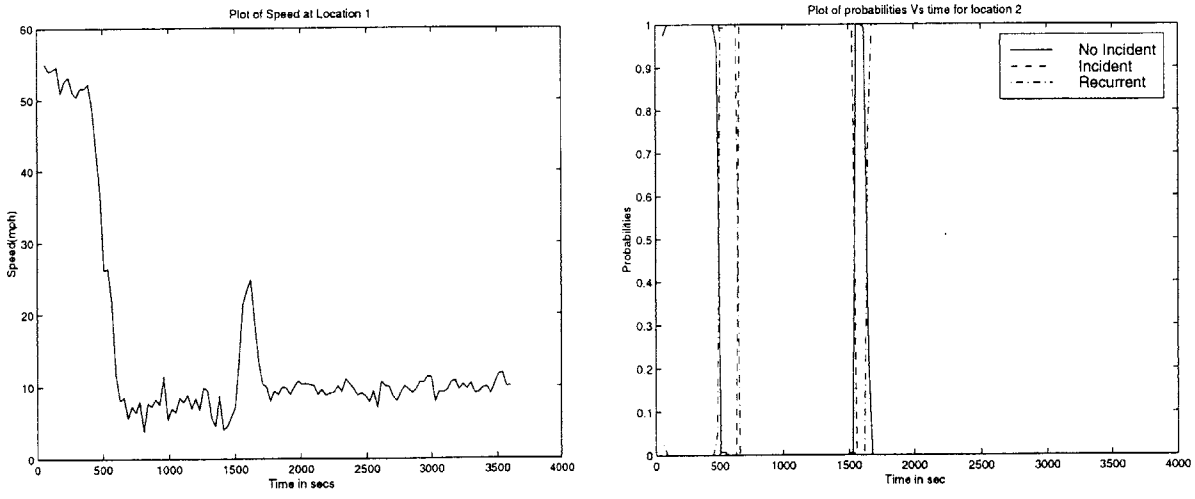
P_{rec} is the probability of recurrent congestion

μ_{cong} is the scale associated with the congestion nest

The results of the estimation are presented in Table 5.10. The signs of the coefficients are intuitive. The coefficient β_{no} is greater than 0, which means that all being equal, there is a greater probability of no congestion. the coefficient OCC_{cong} is positive, which indicates that as the occupancy increases, probability of congestion increases. The results in Table 5.10 are for simulated data. All the coefficients except the alternative specific constants are significant at the 95% significance level. The predicted probabilities are presented in Figure 5-12 and Figure 5-13.

Table 5.10: Summary of estimation

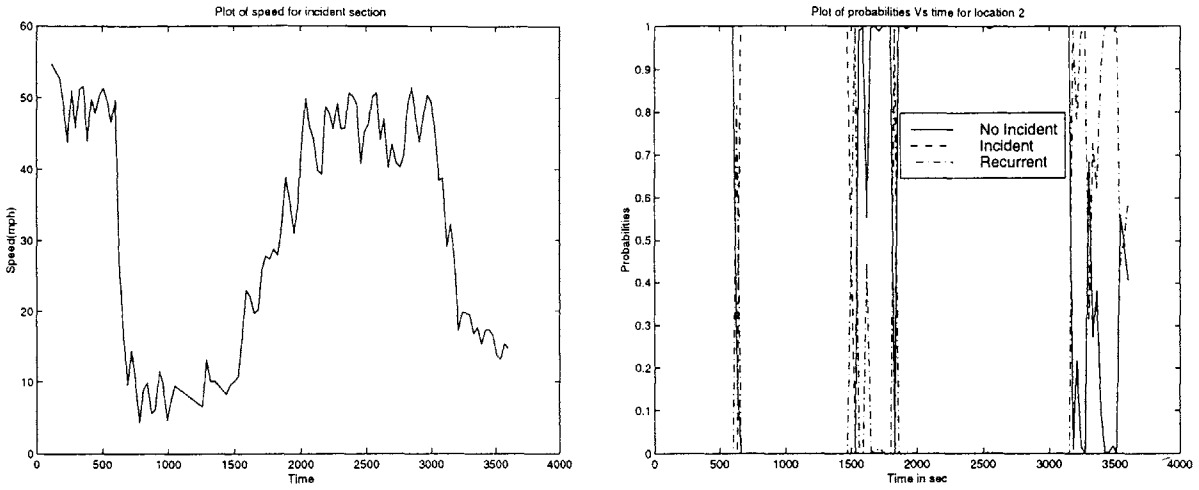
Variable Name	Coefficient	Std. Error	t-statistic
β_{no}	4.063	2.898	1.402
β_{rec}	2.795	5.305	0.5268
OCC_{cong}	-0.4914	0.08209	5.986
$OCCRDF_{inc}$	67.80	19.98	3.394
$OCCRDF_{rec}$	-22.43	10.83	-2.071
FLW_{inc}	-4.582	1.34	-3.419
FLW_{rec}	-1.691	0.9121	-1.854
θ	0.3792	5.305	0.5268
$L_{nl}(L)$	-50.1444(-747.056)		
$\bar{\rho}^2(\rho^2)$	0.9077(0.9204)		



(a) Speed Vs time

(b) Probabilities Vs time

Figure 5-12: Predicted probabilities for location 1



(a) Speed Vs time

(b) Probabilities Vs time

Figure 5-13: Predicted probabilities for location 2

It can be seen from the prediction of the probabilities that the nested logit model captures all the phenomena most accurately in both locations. In case of Location 1, the model identifies the incident, the recovery phase and the subsequent recurrent congestion better than the MNL model. In location 2, the model identifies all phenomena predicted by the MNL model. In addition, the model also captures the brief recovery wave between t=1600 to t=1700, after the incident is cleared. This can be observed by a short increase in speeds at end of the incident. The predictions are also much less ambiguous than the MNL predictions.

5.3 Model Evaluation using MITSIM

Using the coefficients obtained from calibration and the framework developed earlier, the performance of the model has been tested over the same set of data and scenarios used to test the APID and McMaster algorithms. The validity of the simulation with respect to real data has been discussed in the later sections. The algorithm was tested for 20 replications with one incident being simulated in each replication. Each replication was one hour long. These results have been presented in Table 5.11 and Table 5.12.

Table 5.11: MNL model results Persistency=3

		200 ft			600 ft		
Location	**Demand**	**80%**	**100%**	**120%**	**80%**	**100%**	**120%**
Location 1	**TTD(min)**	2.63	1.95	2.00	4.06	3.35	4.10
	DTR(%)	75	100	90	90	100	100
	FAS	29	54	45	28	62	57
Location 2	**TTD(min)**	2.02	1.84	2.16	2.45	2.48	2.67
	DTR(%)	100	80	90	100	100	100
	FAS	52	47	56	41	52	53

Table 5.12: NL model results Persistency=3

		200 ft			600 ft		
Location	**Demand**	**80%**	**100%**	**120%**	**80%**	**100%**	**120%**
Location 1	**TTD(min)**	2.46	1.70	2.00	4.00	2.75	2.62
	DTR(%)	70	100	75	90	100	100
	FAS	42	50	41	26	47	27
Location 2	**TTD(min)**	2.18	2.00	2.14	2.60	2.52	2.62
	DTR(%)	95	80	90	100	100	100
	FAS	64	67	78	52	55	50

5.3.1 Limitations of Simulation

The limitations of the model calibrated with only simulated data are:-

i. Simulation data tends to have lesser variation than field data. But it is also free from confounding effects, invalid observations (due to sensor malfunctioning etc.) or loss of data (failure of loops).

ii. The choice model parameters depend heavily on the accuracy of simulation namely the lane changing, gap acceptance and merging models. They also depend on the accuracy of modeling of incidents (visibility of incidents, presence of rubber necking).

iii. Inherently, the simulation cannot match the real network exactly because of boundary effects and limitations in network data availability.

5.3.2 Summary of Results

i. The probabilistic approach helps collapse all thresholds into a single threshold. This is significant as the number of threshold to be calibrated is just one. This precludes the need for an elaborate calibration process. Thresholds can be established by directly setting costs of declaring incidents and false alarms. Section 5.5.5 illustrates this procedure.

ii. The NL model has better performance than the MNL model. This was expected because the NL model has better power to distinguish between recurrent and incident congestion better by accounting for the correlation between the two congested states.

iii. Both models maintain a detection rate above 70% in all scenarios and have quick detection times of 2-3 min in almost all scenarios.

iv. These choice models have been found to have problems when there are large lanewise differentials in the values of traffic variable.

v. The model was calibrated using data from location 2. The same parameters were used to make predictions for location 1. This indicates that the model parameters are transferable in this case.

5.4 Comparison of performance of different algorithms

The results from the Nested Logit model have been compared with the results from the McMaster and APID incident detection algorithms in Table 5.13.

i. It can be seen that over almost all scenarios, the NL model has smaller TTD than either the Refined McMaster or the APID algorithms.

ii. The FAS generated by the NL model lies between the FAS generated by APID and the refined McMaster algorithms.

iii. The performance of the NL model based algorithms is intuitive because they use both single-station and comparative variables in calibration.

Table 5.13: NL vs others at 200 ft and 600 ft sensor spacing

Results at 200 ft										
Location		**80%**			**100%**			**120%**		
	Demand	**MCM**	**APID**	**NL**	**MCM**	**APID**	**NL**	**MCM**	**APID**	**NL**
Location 1	**TTD(min)**	4.08	2.47	2.46	2.67	2.34	1.70	2.92	2.24	2.00
	DTR(%)	43	85	70	90	95	100	90	85	75
	FAS	24	33	42	48	54	50	51	51	41
Location 2	**TTD(min)**	2.25	2.33	2.18	2.12	2.07	2.00	2.56	2.28	2.14
	DTR(%)	100	100	95	100	75	80	95	80	90
	FAS	128	53	64	135	61	67	114	58	78
Results at 600 ft										
Location		**80%**			**100%**			**120%**		
	Demand	**MCM**	**APID**	**NL**	**MCM**	**APID**	**NL**	**MCM**	**APID**	**NL**
Location 1	**TTD(min)**	5.69	3.32	4.00	6.31	3.89	2.75	6.82	4.36	2.62
	DTR(%)	40	95	90	65	95	100	55	90	100
	FAS	5	69	26	19	86	47	12	79	27
Location 2	**TTD(min)**	3.10	2.50	2.60	2.33	2.62	2.52	2.92	2.60	2.62
	DTR(%)	100	100	100	100	100	100	100	100	100
	FAS	56	81	52	52	96	55	49	93	50

5.5 Combined Estimation using real and simulated data

5.5.1 Introduction

Using both real and simulated data in the calibration phase can yield more realistic parameters. Since simulation can be used to obtain data from different scenarios, the robustness of the model can be improved. However, simulated data is subject to different types of errors and modeling inaccuracies. These errors can be reduced by using real data in conjunction with simulated data during the model estimation phase. As discussed earlier, this approach is similar to that of combining revealed preference and stated preference data which has received much attention in recent years. In addition, all traffic states may not be observed at any given sensor station in the real data. This disadvantage can be overcome to an extent using simulation. The estimation of logit models using simulated data was studied in the previous section. In this section, combined estimation using real and simulated data has been presented. Data from I-880 near Hayward, California has been used.

A brief description of the real data is followed by the description of simulation parameters. Factors that affect modeling of incidents in a simulator have been briefly studied. Results of combined estimation have been discussed followed by results from testing of the model prediction.

5.5.2 Real data

Description of the real incident data used in current study

For the current study, real incident data was obtained from the Internet for I-880 near Hayward, California[24]. The data was collected as part of the Freeway Service Patrol (FSP) study at the University of California at Berkeley. The study section has been presented in appendix C-1. Also produced in appendix B-1 is the geometric configuration of the network (to scale) where the actual location of the exit and entry ramps have been marked. The study section was 9.2 miles long and varied between 3-5 lanes. An HOV lane covered approximately 3.5 miles of the study section. In addition to loop detector data, probe vehicle data was also collected in the study. The data is also available in the processed format, both at one minute intervals and at five minute intervals. The one minute interval data was chosen and some sample plots have been made (shown in appendix D-1).

Format of the Loop data

Loop detectors in the network were placed approximately 1/3 of a mile (approximately 1800 ft). There were a total of 322 mainline detectors, 18 on-ramp detectors and 14 off-ramp detectors. The loop data was collected in 2 different periods. The first study took place from February 16 through March 19, 1993 and the second study took place from September 27 through October 29, 1993. On each day, the loop data is available for both the a.m. and the p.m. peaks (i.e. 6:30-9:30 a.m. and 3:30 - 6:30 p.m.). The raw loop data is in the binary format and has to be converted to the standard form with measurements of speed, occupancy and counts from each sensor.

Format of the Incident data

Detailed data is also available for the incidents that were observed in the study. Some of the details available are time at which the incident was first observed (which is offset from the time of

occurrence by the time taken for the congestion to reach the upstream or downstream station), the location of the incident, the severity of the incident (single vehicle / multiple vehicles), end of the incident and the incident management strategy used to clear the incident.

5.5.3 Simulation data

Predictions under free flow

In order to obtain realistic results, the parameters of the simulator have to be calibrated. The network used for the current study has also been studied by Ashok and Ben-Akiva[23]. The simulator MITSIM needs as input, time dependent OD trip tables. The estimates of dynamic OD flows using the sensor data has been have been obtained by Ashok and Ben-Akiva[23]. Some calibration has also been done for the same network by Ahmed[26]. The number of replications (to obtain a statistically significant estimate) required for all the speed and flow estimates to be within 5 percent accuracy was determined to be 10. In order to measure the accuracy of the model, the root mean square error(*rmse* and root mean square percent error(*rmsep*) were used. This can be defined as

$$rmse = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\{y_i^s - y_i\}^2} \tag{5.19}$$

$$rmsep = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\{(y_i^s - y_i)/yi\}^2} \tag{5.20}$$

Based on the above measure, further tuning of the model produced results which have been produced in Table 5.14. The tuning of the model mainly involved the setting of desirable speed and maximum allowable speed for specific portions of the network. The network in the simulator is subject to boundary effects. It does not accurately represent all field conditions. Based on the maximum speed observed in different parts of the network obtained from Petty[24], this value for reset for different parts of the network. There is a significant drop in difference between the predicted and the actual speeds. The flow predictions are practically constant.

Table 5.14: Error estimates

Measure	Old(flow)	New(flow)	Old(speed)	New(Speed)
RMSEP(%)	2.57	2.39	24.33	17.70
RMSE	27.26	24.74	11.47	6.31

Predictions under incident conditions

In order that the simulator replicates actual behavior to a reasonable extent the manner in which an incident is simulated has to be carefully studied. In order to simulate an incident some parameters have to be specified prior to the simulation. Some of these parameters that could be important under incident conditions are

i. Visibility of the incident

ii. Rubbernecking

iii. Nosing and Yielding parameters

In order to calibrate the simulator under incident conditions, the incident scenario simulated is the one for which the plots have been produced in appendix D-1. There were two simulated incidents in the same time interval. Both were major incidents lasting over 30 minutes each. In both cases one lane was blocked. The first incident started at 7:40 am and lasted for about 34 minutes. The second incident started at 8:42 am and lasted for 47 minutes. Both incidents were simulated with rubber necking, with reduced speeds of 20mph in adjacent lanes. Again, 10 replications were determined to be sufficient to ensure 95% accuracy. The *Root Mean Square*

Table 5.15: Error estimates under incident conditions

Measure	Speed	Flow
RMSEP(%)	22.73	3.06
RMSE	10.59	32.51

Error Percent(RMSEP) & Root Mean Square Error(RMSE) values for speed and flow are close to the magnitude of the corresponding values for non-incident cases. Most of the calibration was

done by varying the visibility distance and the rubbernecking speed. Further improvement could be achieved by varying parameters of the lane changing model.

Simulation details

Using the information about the layout of the network and the map of the network showing the geometry, it was created in MITSIM[37]. For the initial experiments, the section of the network between Hesperian Blvd. and WA-Street is to be used. This region of the network is the longest section in the network between any 2-exit/entry ramps. Hence, it would be a section where the congestion induced due to incident (capacity reduction) can be easily detected and differentiated from the congestion due to bottlenecks (excessive demand). The details of the simulation have been presented in Table 5.16. It has to be noted that the real data contained, in addition to incidents

Table 5.16: Simulation Setup

Simulation Between	07:30:00 to 10:30:00
Incident Between	08:00:00 to 08:30:00
Location of Incident	at midway of section between detectors 1 and 7
Type of Incident	2 lane block with simulated rubber-necking

similar to the simulated incident, other minor incidents that reduced capacity during the periods of data collection. Model calibration using real data should correct these differences.

5.5.4 Model structure for Combined Estimation - NLRPSP

The model estimated using both simulation data(SP) and real data(RP) is referred to as NLRPSP. The model simulated using only simulation data(SP) is referred to as NLSP.

$$V_{no} = \beta_{no} \tag{5.21}$$

$$V_{cong} = OCC_{cong} * occupancy + V'_{cong} \tag{5.22}$$

$$V'_{cong} = (1/\mu_{cong}) ln(e^{\mu_{cong} * V_{inc}} + e^{\mu_{cong} * V_{rec}}) \tag{5.23}$$

$$V_{inc} = \beta_{inc} + FLW_{inc} * flow + OCCRDF_{inc} * occrdf + SPD_{inc} * speed \tag{5.24}$$

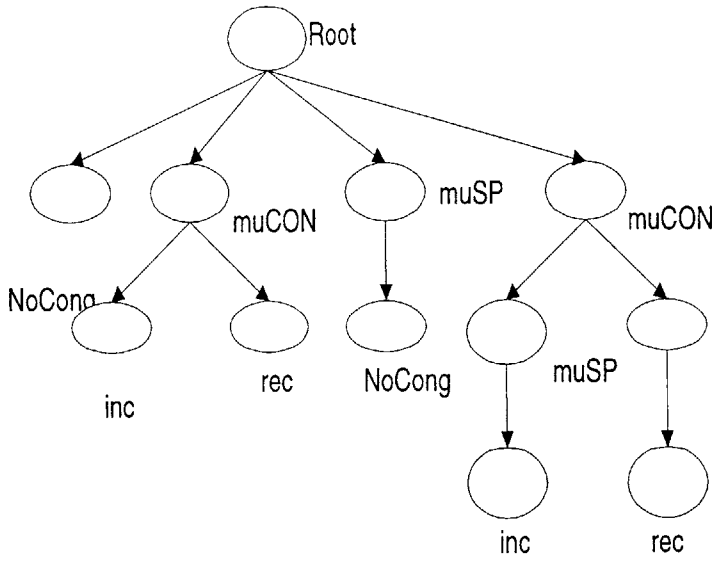


Figure 5-14: Nested logit model structure for RP-SP Combination

$$V_{rec} = FLW_{rec} * flow + OCCRDF_{rec} * occrdf + SPD_{rec} * speed \tag{5.25}$$

$$P_{no} = 1/(1 + e^{V_{cong} - V_{no}}) \tag{5.26}$$

$$P_{inc} = (1 - P_{no})(1/(1 + e^{V_{rec} - V_{inc}})) \tag{5.27}$$

$$P_{rec} = (1 - P_{no})(1/(1 + e^{V_{inc} - V_{rec}})) \tag{5.28}$$

V_{no} is the systematic utility associated with the no congestion alternative

V_{inc} is the systematic utility associated with the incident congestion alternative

V_{rec} is the systematic utility associated with the recurrent congestion alternative

P_{no} is the probability of no congestion

P_{inc} is the probability of incident congestion alternative

P_{rec} is the probability of recurrent congestion alternative

The model for combined estimation has two different types of scale differences. Since we believe that the simulation data(analogous to SP, referred to as SP) is different from the real data(analogous to RP, therefore referred to as RP), we use the scale parameter μ_{SP}. The second type of scale parameter because of the congestion nest is μ_{CON}. In order to estimate both the scale parameters, the model structure considered is presented in Figure 5-14. The μ_{SP} for the SP data is accounted for

Table 5.17: Model specification

Variable Name	Description
β_{no}	No Congestion alternative specific constant
β_{inc}	Incident alternative specific constant
OCC_{cong}	occupancy alternative specific variable for congestion nest
FLW_{inc}	flow alternative specific variable for incident alternative
FLW_{rec}	flow alternative specific variable for recurrent alternative
$OCCRDF_{inc}$	occupancy relative difference alternative specific variable for incident
$OCCRDF_{rec}$	occupancy relative difference alternative specific variable for recurrent
SPD_{inc}	speed alternative specific variable for incident
SPD_{rec}	speed alternative specific variable for recurrent
μ_{SP}	scale parameter introduced to correct scale difference between RP and SP data
μ_{con}	scale parameter of congestion nest

before the μ_{CON} for the congestion dummy node is considered.

5.5.5 Results of Combined Estimation

For the present case, most of the real sensor data did not contain any traffic state III. Therefore the state III points are only from the simulation data. The results of the combined estimation have been presented in Table 5.18.

5.5.6 Using the RP-SP for prediction

The combined model can be used for detecting incident conditions by calculating the probability of an incident, given the traffic state at any particular sensor. There are several advantages in using a choice model to compute this probability. Some of them are summarized below.

i. In addition to predicting probability of an incident, we could also predict the probability of other states. The equivalent procedure for a deterministic algorithm would be to calculate the relative distance of a traffic point to the different incident states.

ii. We have an effective method to improve our model using real data. By combining real data with simulated data, we can increase the reliability of our model and reduce the bias due to

Table 5.18: RP-SP Estimation results

Coef	Value(t-stat)
β_{no}	2.161(8.1)
β_{rec}	-2.325(-4.01)
FLW_{inc}	-1.381(-6.72)
FLW_{rec}	-1.159(-5.82)
$OCCRDF_{inc}$	-4.866(1.465)
$OCCRDF_{rec}$	-5.555(-3.71)
OCC_{cong}	0.5706(11.37)
SPD_{inc}	-0.2577(-4.27)
SPD_{rec}	-0.2381(-4.08)
μ_{con}	0.3350(-11)
μ_{SP}	0.7139(-8.57)
Log-likelihood(trivial model)	-1707.55
Log-likelihood	-683.05
Fit	0.60

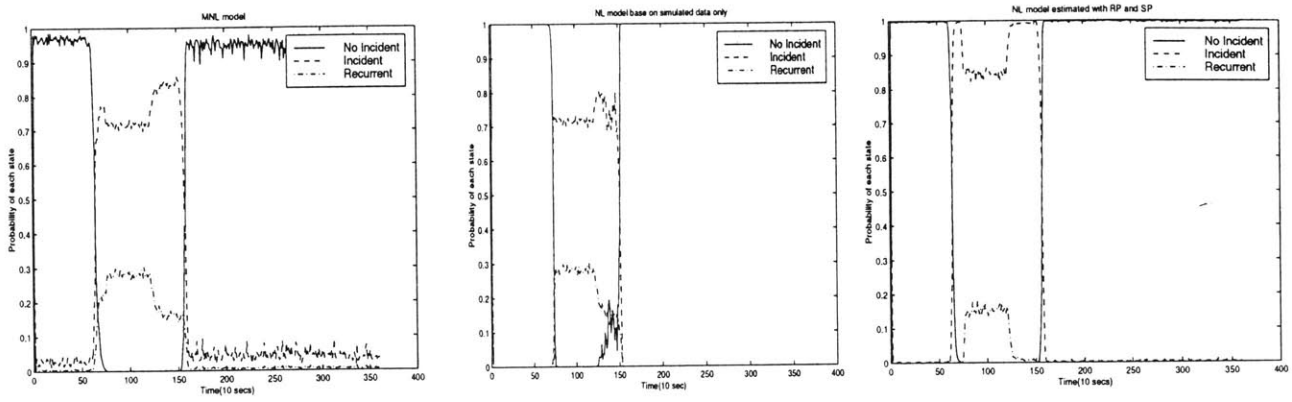
either data source. We also reduce the effort spent in tuning the model to real data when the model is used in the field.

iii. We get new coefficients from our SP data(simulation in this case). This was also demonstrated in the current study because we did not have state III in the real data, we can use the coefficient from the simulation.

Table 5.19: RP-SP Estimation results

Coef	Value
β_{no}	2.161(8.1)
β_{rec}	-2.325(-4.01)
$FLOW_{inc}$	-1.381(-6.72)
$FLOW_{rec}$	-1.159(-5.82)
$OCCRDF_{inc}$	-4.866(1.465)
$OCCRDF_{rec}$	-5.555(-3.71)
OCC_{cong}	0.5706(11.37)
SPD_{inc}	-0.2577(-4.27)
SPD_{rec}	-0.2381(-4.08)
θ_{cong}	0.3350(-11)

Based on the coefficients obtained above, the probability of an incident was calculated using three different models, an MNL model, a Nested Logit model using only simulated data and a Nested Logit model calibrated using both simulated and real data. The plots of the probabilities has been shown compared in Figure 5-15 On comparison of the three plots, it can be seen that the



(a) MNL model (b) NLSP model (c) NLRPSP model

Figure 5-15: Predicted probabilities

performance of the MNL and the NLSP(Nested Logit model estimated with SP i.e. simulation data only) models, both based on simulated data is almost comparable. The NLSP predictions have much lesser variation. The prediction of the NLRPSP model, which uses both real and simulated data, is much better than either model in the sense that the probability of incident has increased and the probability of recurrent congestion has reduced. This model has not been evaluated because of a lack of accurate real incident data. It would be difficult to observe any improved prediction if real incident information is not very accurate. However, it is clear from Figure 5-15 that the probabilities are less ambiguous when calibrated with real incident data.

Chapter 6

Combining different algorithms for improved performance

6.1 Combining the SPDDRP and McMaster algorithms

6.1.1 Introduction

The results from testing of the McMaster and the APID algorithms have been summarized in Figure 6-1 at a persistency = 3. It can be seen that the McMaster algorithm in its refined form has less FAS than the APID algorithm. But APID is better in terms of the TTD. In order to further improve the TTD of the McMaster Algorithm or the APID algorithm, it would be advantageous to combine it with the Speed Drop algorithm discussed earlier. The Speed Drop algorithm has quick detection properties but poor detection rate which can be compensated by combining the other algorithms in an OR mode so that TTD and DTR over all scenarios improve while not increasing the FAS too much. SPDDRP does not perform well in congested conditions where the McMaster algorithm has good performance. Therefore the combination uses the best of both algorithms.

The algorithm chosen to be combined with SPDDRP was the McMaster algorithm in its refined form. The SPDDRP algorithm has been improved by using it in an OR mode with the McMaster algorithm that improves the DTR without worsening the FAS. Two main observations have been used to select the McMaster algorithm

		persistency = 3								
		200 ft			600 ft			1200 ft		
		80 %	100 %	120 %	80 %	100 %	120 %	80 %	100 %	120 %
loc1	TTD	APID	APID	APID	APID	APID	APID	APID	APID	APID
loc1	DTR	APID	APID	MCM	APID	APID	APID	APID	APID	MCM
loc1	FAS	MCM	MCM	MCM	MCM	MCM	MCM	MCM	MCM	MCM
loc2	TTD	MCM	APID	APID	APID	MCM	APID	MCM	APID	MCM
loc2	DTR	MCM	MCM	MCM	MCM	MCM	MCM	MCM	MCM	MCM
loc2	FAS	APID	APID	APID	MCM	MCM	MCM	MCM	MCM	MCM

Figure 6-1: Summary of Results

6.1.2 Methodology

Th following rules are used for combining algorithms:-

i. If the two algorithms are perfectly complementary to each other they can be combined in an OR fashion for maximum benefits. The speed drop algorithm studied earlier uses a high threshold to obtain quick detection time while maintaining a low false alarm rate. But the high threshold also leads to missed detections. Thus, in order to improve the performance of the SPDDRP algorithm, it would have to be combined with another algorithm that has less rigid thresholds for incident detection but acceptable false alarm and detection properties. Combining in an "OR" mode leads to higher FAS, higher DTR and probably lower TTD since the earliest confirmation detects the incident.

ii. In order to combine algorithms in an AND fashion, the algorithms have to be functionally similar. There should not be considerable differences in the spatial and temporal locations where incidents are detected by the algorithms. If the algorithms detect the same incident at very different locations or at different times, there would be a lot of missed detections.

6.1.3 Framework

Basic Idea

By virtue of the model structure shown below, under uncongested conditions, the SPDDRP declaration would mostly declare incident. However, when speed is less than α, we rely on Refined

McMaster would declare the incident.

(if speed $\geq \alpha$*)*

(if SPDDRP == 1)

confirm incident

else*((if Ref MCM == 1)&(if 0.5*SPDDRP == 1))*

confirm incident

else

(if Ref MCM == 1)

confirm incident

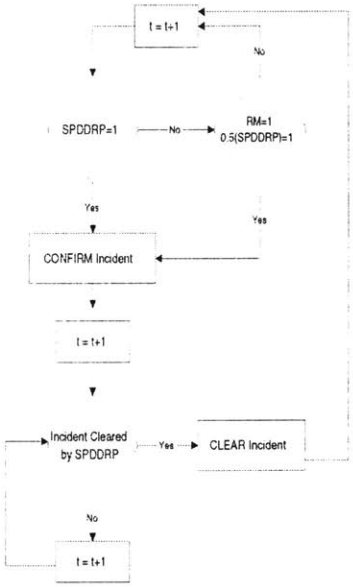


Figure 6-2: Combination of SPDDRP and MCM

Since the SPDDRP algorithm does not use persistency checks, it can be easily combined with a version of the McMaster that does not use these checks. When the number of persistency checks are relaxed, the FAS increase. In order to limit the increase in FAS, McMaster was combined with a relaxed version of the SPDDRP algorithm to confirm detections. We will demonstrate that the FAS do not increase unreasonably because of relaxing persistency checks. The framework used has been presented in Figure 6-2. The check for random fluctuations has been eliminated in the

version of the SPDDRP algorithm that is used in conjunction with the McMaster algorithm. The original version of the speed drop is used as the main indicator of incidents.

6.1.4 Evaluation of the combined algorithm

The combined algorithm was tested over the same range of scenarios, at location 2, used to test the SPDDRP algorithm. The parameters are summarized in Table 6.1. The parameter α_{comb} is the value of α parameter for the SPDDRP module combined with the McMaster algorithm in an AND fashion. The results are presented in Table 6.2 and compared with SPDDRP results for a spacing of 200 ft. The combined algorithm improves the time to detect of the SPDDRP algorithm in addition to the improvement in the detection rate which improves to 100% for all scenarios.

Table 6.1: Parameters used

Parameter	**200 ft**
α(mph)	24
β(mph)	30
γ	0.77
α_{comb}(mph)	12

Table 6.2: Results at 200 ft spacing

	80%		**100%**		**120%**	
Demand	**SD**	**COMB**	**SD**	**COMB**	**SD**	**COMB**
TTD(min)	1.22	1.07	1.03	1.09	1.27	1.06
DTR(%)	80	100	80	100	85	100
FAS	37	44	17	28	17	30

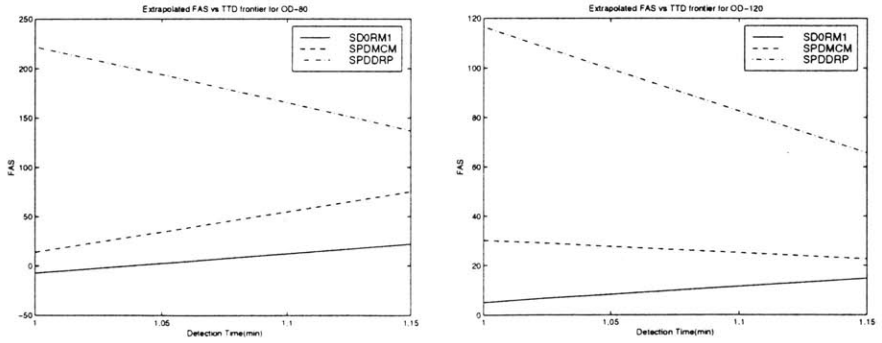
6.1.5 Findings and Summary

i. In order to combine the Refined McMaster algorithm with the SPDDRP algorithm, the persistency is reduced to 1.

ii. As expected, the detection rate has improved and the false alarms have increased because of the ORing of the algorithms. However, the reduction in the time to detect the incident is unexpected. The McMaster algorithm using the modified SPDDRP algorithm for confirmation *detects incidents quicker* than the original SPDDRP algorithm for some scenarios. By increasing the α_{comb} parameter, we can reduce the FAS but the DTR would decrease.

iii. However, when the McMaster algorithm operates with a persistency greater than one, it is not able to confirm the incident over successive time periods.

iv. A performance frontier can be constructed for every incident detection algorithm. This would be the maximum improvement that can be attained in the MOEs for a specific scenario for different values of thresholds. The limit of the maximum attainable MOEs can be obtained by calculating the MOEs for the different values of the input parameters. The parameter α and α_{comb} were varied to obtain the MOE frontiers. Sample plots of the frontiers for the 80% and 120% OD have been presented in Figure 6-3. The frontiers have been presented for all three algorithms - SPDDRP, SDRM(MCM AND 0.5*SPDDRP i.e. relaxed speed drop and refined McMaster) and SPDMCM (SPDDRP OR SDRM). It can be seen that SDRM algorithm has favorable FAS but bad DTR. Also, for a given TTD SPDMCM generates higher FAS than SDRM.

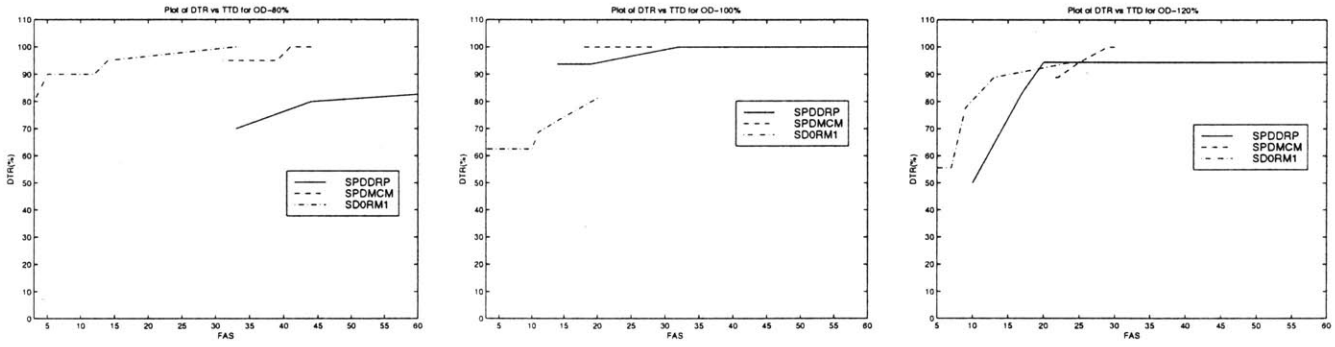


(a) OD 80%

(b) OD 100%

Figure 6-3: FAS vs TTD at 200ft spacing

v. From Figure 6-4, it can be seen that the SPDMCM algorithm has much better detection performance than the other algorithms over all scenarios.



(a) OD 80% (b) OD 100% (c) OD 120%

Figure 6-4: DTR vs FAS at 200ft spacing

6.2 Using a LOGIT model to combine algorithms

6.2.1 Introduction

Two different algorithms may detect the same incident at different sensor stations and at different times. Therefore, it becomes difficult to relate the incident indications from different algorithms. *Incident indication* refers to the detection of incident state by an algorithm at a station. Persistence of this preliminary incident state results in incident confirmation. This is particularly true for networks with dense loop detector spacing and small sensor data averaging periods. The number of incident declarations in a given time period could also be different. For example:- single-station algorithms, such as the McMaster algorithm, observe the traffic state as different regions in the fundamental diagram. Since they do not observe temporal drops in values of traffic variables, they may give more incident indications than other algorithms like the SPDDRP algorithm. Table 6.3 shows

Table 6.3: Comparison of number of incident indications

Algorithm	No of Indications
McMaster	24
APID	10
SPDDRP	2
COMB	7

the number of indications of the different algorithms at a single station upstream of an incident for a single experiment. All other algorithms in Table 6.3 except the McMaster algorithm are based on comparative thresholds. As discussed in chapter 3, algorithms with comparative thresholds give many more indications at sections with complex geometry. In combining algorithms, if we have to wait for confirmation from all algorithms, the time taken to confirm an incident may increase disproportionately, offsetting the advantages of combining them. The fundamental requirement for combining algorithms is a weight function that can be used to indicate the degree to which an incident indication from an algorithm can be trusted. For example:- Consider the combination of the SPDDRP and McMaster. We trust the SPDDRP algorithm completely because of the high threshold level and low FAS. In this case, weight = 1 for SPDDRP and weight = 0 for McMaster under uncongested conditions. By assigning different weights, we may be able to reduce the FAS

of the combined algorithm.

A logit model based on algorithmic decisions would automatically assign weights to the decisions using coefficients. The decisions of different algorithms under different scenarios can be used to calibrate such a model. The resultant decision would be the output of all individual model decisions.

Another way to improve the performance of a logit-based algorithm is to note the *strength* of an incident indication. *Strength* of an incident indication can be defined as the magnitude of the deviation of a measurement from the corresponding threshold. It is obvious that a stronger signal, indicated by greater deviation from the threshold in the correct direction (For example:- incident is indicated by a positive deviation of measured occupancy from the threshold occupancy upstream of the incident), can be trusted more than a weak signal which would correspond to a small deviation. The *Strength of signal* idea can be easily used in a logit model by multiplying the algorithmic decisions with magnitude of the deviation. Table 6.4 presents the deviation from the threshold for

Table 6.4: Comparison drop magnitude for incidents vs False alarms(SPDDRP)

OD(%)	Mean Drop for True alarm	Mean Drop for FAS
80	27.95	27.0611
100	28.34	26.54
120	29.15	26.55

true detections and FAS for the SPDDRP algorithm for one scenario. Since a greater deviation from the threshold implies a stronger signal, it is evident that the incident signal is stronger for true incidents than FAS from Table 6.4. The mean drop for true incidents is greater than that for FAS and the magnitude of this drop for true incidents increases for higher OD. However no such pattern is clear for the drop in case of FAS.

Thus, two types of logit models can be used to combine algorithms:-

I. A multinomial logit model in which the decisions of the Refined McMaster, APID and the SPDDRP algorithms(LOGCMB-I) are used as binary explanatory variables.

II. A multinomial logit model in which the decisions of the Refined McMaster, APID and an interaction term between decision of SPDDRP algorithm and the speed drop magnitude (as an indicator of the strength of the signal) are used to make quick decisions(LOGCMB-II).

Choice based sampling

Table 6.3 shows the number of algorithmic decisions favoring an incident for different algorithms over a single replication. Clearly such decisions can be identified only over the duration of the incident during the experiment. Data from location 2 was used to collect incident and non-incident data. The number of observations for which any one of the component algorithms makes an incident declaration was collected from all replications of one scenario from location 2. This was done to obtain more incident data.

LOGCMB-I

The model specification is presented in Table 6.5. The results of estimation of the LOGCMB-I model is presented in Table 6.6. In addition to algorithm indications, interaction variables are also been used. The interaction variables could be a proxy for measuring the most significant algorithm combinations.

$$V_{no} = \beta_{no} + SPD_{no} * speed \tag{6.1}$$

$$V_{inc} = FLW_{inc} * flow + OCCDF_{inc} * occdf + spd_{inc} * speed + \tag{6.2}$$

$$MCM_{inc} * mcm + APD_{inc} * apd + DRP_{inc} * drp + DRPMCM_{inc} * drp * mcm +$$

$$DRPAPD_{inc} * drp * apd + MCMAPD_{inc} * mcm * apd$$

$$V_{rec} = \beta_{rec} + flw_{rec} * flow + OCCDF_{rec} * occdf \tag{6.3}$$

V_{no} is the systematic utility associated with the no congestion alternative

V_{inc} is the systematic utility associated with the incident congestion alternative

V_{rec} is the systematic utility associated with the recurrent congestion alternative

Table 6.5: Model specification

Var Name	Description
β_{no}	No Congestion alternative specific constant
SPD_{no}	speed alternative specific variable for no congestion
β_{rec}	Incident alternative specific constant
FLW_{rec}	flow alternative specific variable for recurrent congestion alternative
$OCCDF_{rec}$	occupancy difference alternative specific variable for recurrent congestion alternative
FLW_{inc}	flow alternative specific variable for incident congestion alternative
$OCCDF_{inc}$	occupancy difference alternative specific variable for incident congestion alternative
SPD_{inc}	speed alternative specific variable for incident congestion alternative
MCM_{inc}	= 1, if McMaster declares incident; = 0 otherwise
APD_{inc}	= 1, if APID declares incident; = 0 otherwise
DRP_{inc}	= 1, if SPDDRP declares incident; = 0 otherwise
$DRPMCM_{inc}$	$= MCM_{inc} * SPD_{inc}$
$DRPAPD_{inc}$	$= APD_{inc} * SPD_{inc}$
$MCMAPD_{inc}$	$= MCM_{inc} * APD_{inc}$

P_{no} is the probability of no congestion alternative

P_{inc} is the probability of incident congestion

P_{rec} is the probability of recurrent congestion

mcm = 1 if McMaster algorithm gives incident indication; = 0 otherwise

apd = 1 if APID algorithm gives incident indication; = 0 otherwise

drp = 1 if SPDDRP give incident indication; = 0 otherwise

There is a good model fit. The interaction term measuring the effect of using the McMaster and the APID algorithms is not significant at 95% level of significance. $OCCDF_{inc} > OCCDF_{rec}$ which is intuitive because as the occupancy difference increases between the upstream and the downstream stations, the probability of incident increases.

LOGCMB-II

The results of estimation of the LOGCMB-I model have been presented in Table 6.7. The effect of strength of the incident signal has been incorporated using the $SPDDRP_{inc}$ variable.

Table 6.6: Estimation results

Var Name	Coef	Std. Err	t-stat
β_{no}	-537.2	37.02	-14.51
SPD_{no}	11.16	0.7277	15.33
β_{rec}	1.824	0.7445	2.45
FLW_{rec}	-0.7081	0.3504	-2.021
$OCCDF_{rec}$	1.429	0.8851	1.615
FLW_{inc}	-0.4397	0.3587	-1.226
$OCCDF_{inc}$	1.563	0.8852	1.766
SPD_{inc}	-0.3355	0.04629	-7.247
MCM_{inc}	5.713	0.5184	11.02
APD_{inc}	3.345	0.5948	5.624
DRP_{inc}	-0.3355	0.9726	-2.476
$DRPMCM_{inc}$	-6.493	1.216	-5.339
$DRPAPD_{inc}$	-2.284	1.169	-1.953
$MCMAPD_{inc}$	-1.29	1.013	-1.273
$L_{nl}(L)$	-142.902(-1054.37)		
$\bar{\rho}^2(\rho^2)$	0.8511(0.8645)		

$$V_{no} = \beta_{no} + SPD_{no} * speed \tag{6.4}$$

$$V_{inc} = FLW_{inc} * flow + OCCDF_{inc} * occdf + spd_{inc} * speed + \tag{6.5}$$

$$MCM_{inc} * mcm + APD_{inc} * apd + SPDDRP_{inc} * speed_drop + DRPMCM_{inc} * drp * mcm +$$

$$DRPAPD_{inc} * drp * apd + MCMAPD_{inc} * mcm * apd$$

$$V_{rec} = \beta_{rec} + flw_{rec} * flow + OCCDF_{rec} * occdf \tag{6.6}$$

V_{no} is the systematic utility associated with the no congestion alternative

V_{inc} is the systematic utility associated with the incident alternative

V_{rec} is the systematic utility associated with the recurrent alternative

P_{no} is the probability of no congestion

P_{inc} is the probability of incident

P_{rec} is the probability of recurrent congestion

mcm = 1 if McMaster algorithm gives incident indication; = 0 otherwise

Table 6.7: Model specification

Var Name	Description
β_{no}	No Congestion alternative specific constant
SPD_{no}	speed alternative specific variable for no congestion
β_{rec}	Incident alternative specific constant
FLW_{rec}	flow alternative specific variable for recurrent alternative
$OCCDF_{rec}$	occupancy difference alternative specific variable for recurrent alternative
FLW_{inc}	flow alternative specific variable for incident alternative
$OCCDF_{inc}$	occupancy difference alternative specific variable for incident
SPD_{inc}	speed alternative specific variable for incident alternative
MCM_{inc}	= 1, if McMaster declares incident; = 0 otherwise
APD_{inc}	= 1, if APID declares incident; = 0 otherwise
$SPDDRP_{inc}$	=speed drop,if SPDDRP declares incident; = 0 otherwise
$DRPMCM_{inc}$	$= MCM_{inc} * SPD_{inc}$
$DRPAPD_{inc}$	$= APD_{inc} * SPD_{inc}$
$MCMAPD_{inc}$	$= MCM_{inc} * APD_{inc}$

apd = 1 if APID algorithm gives incident indication; = 0 otherwise

drp = 1 if SPDDRP algorithm gives incident indication; = 0 otherwise

$speed_drop$ = drop in speed if SPDDRP give incident indication; = 0 otherwise

The observations made for LOGCMB-I are also true for LOGCMB-II.

6.2.2 Testing of the models

The models have been calibrated using data from location 2. In order to test the models, they were evaluated at different levels of OD flows(80% - 120%), at location 1. Three levels of OD, 80%, 100% and 120%, were used for testing. Table 6.9 and Table 6.10 show the results of testing for LOGCMB-I and LOGCMB-II respectively.

Table 6.8: Estimation Results

Var Name	Coef	Std. Err	t-stat
β_{no}	-639.9	76.08	-8.411
SPD_{no}	13.36	1.585	8.428
β_{rec}	1.833	0.743	2.467
FLW_{rec}	-0.8268	0.7804	-1.059
$OCCDF_{rec}$	-0.4659	0.9196	-0.5066
FLW_{inc}	-0.5554	0.7841	-0.7083
$OCCDF_{inc}$	-0.3298	0.9197	-0.3586
SPD_{inc}	-0.3381	0.04677	-7.229
MCM_{inc}	5.726	0.5237	10.93
APD_{inc}	3.353	0.5994	5.593
$SPDDRP_{inc}$	-0.1058	0.03681	-2.875
$DRPMCM_{inc}$	-6.616	1.17	-5.656
$DRPAPD_{inc}$	-2.435	1.062	-2.293
$MCMAPD_{inc}$	-1.374	1.005	-1.367
$L_{nl}(L)$	-140.02(-1054.37)		
$\overline{\rho}^2(\rho^2)$	0.8539(0.8672)		

Table 6.9: LOGCMB-I model results Persistency=3

		200 ft		
Location	**Demand**	**80%**	**100%**	**120%**
Location 1	**TTD(min)**	3.25	2.32	2.68
	DTR(%)	10	85	85
	FAS	20	10	7

Table 6.10: LOGCMB-II model results Persistency=3

		200 ft		
Location	**Demand**	**80%**	**100%**	**120%**
Location 1	**TTD(min)**	3.33	2.23	2.42
	DTR(%)	15	90	90
	FAS	20	18	9

6.2.3 Discussion

i. The LOGCMB algorithms were calibrated using data from location 2. They were validated at location 1.

ii. The FAS generated by the LOGCMB algorithms is the Minimum for the given scenario compared to any other algorithm. It generates less FAS than the SPDDRP algorithm. The DTR and TTD at high flows is comparable to that of the APID and McMaster algorithms but less than that of SPDMCM presented in the previous section.

iii. The LOGCMB-II algorithm detects incidents quicker than the LOGCMB-I algorithm. This is intuitive because the LOGCMB-II algorithm also uses the strength of incident signal.

iv. The poor DTR of the LOGCMB algorithm at low flows can be improved in the same way discussed earlier for the SPDDRP algorithm.

Chapter 7

Summary, Conclusions and Future directions

7.1 Summary and Conclusions

The purpose of the current study was to explore the influence of different system characteristics and algorithmic parameters on the performance of incident detection algorithms(IDAs)and to propose new directions in the design and evaluation of IDAs. The limitations of many existing studies in evaluating IDAs were identified. The need for a broad-based evaluation was justified with case studies on I-93 North and I-90 West. The insights from the comprehensive evaluation were used to propose new ideas for designing IDAs.

7.1.1 Evaluation of the McMaster and APID algorithms

The performance of the McMaster and APID algorithms was evaluated in two case studies. The algorithms were calibrated prior to evaluation. The first study on I-90 West helped to design refinements to enhance the algorithms. These refinements were tested in the second study on I-93 North.

Findings

i. In order to identify the applicability and transferability of IDAs, their performance must be evaluated over a broad range of scenarios.

ii. The Refined McMaster and the APID algorithms exhibited excellent performance on I-90 West especially with respect to FAS. The network in this case was devoid of any complex geometry. This study was representative of many existing studies which give a misleading impression of the applicability of the IDA.

iii. The performance of both algorithms deteriorated in the second study on I-93 North. The geometry of the network varied considerably due to on & off ramps, add & drop lanes and curvature. The performance in these scenarios indicates the restricted applicability of these algorithms.

iv. Over all scenarios, the Refined McMaster algorithm generated less FAS than APID, but APID detected incidents quicker than the Refined McMaster algorithm.

7.1.2 Detecting incidents based on speed drop: SPDDRP

Designing new ideas for given conditions may be better than trying to transfer algorithms from different sites to the current site directly. Simple ideas which exploit the best features of the network under study would be good candidates.The SPDDRP algorithm for dense loop detector spacing supports this observation.

Findings

i. The SPDDRP algorithm was a simple and effective idea to detect incidents quickly(less TTD) and reliably(less FAS).

ii. SPDDRP had a poor DTR and it was applicable only to non-congested traffic.

7.1.3 Discrete choice model based Incident Detection

The predictions of algorithms based on discrete choice models were seen to be powerful and elegant. A family of logit models were calibrated and evaluated for I-93 North. The model predictions were tested over the same range of scenarios used to test the McMaster and APID algorithms.

Findings

i. The probabilistic approach helped collapse all thresholds into a single threshold. This eliminated the need for an elaborate calibration process.

ii. The Nested logit model predicted all traffic phenomena correctly and in a much better way than the other models (MNL,binary logit).

iii. Combined estimation of these models using real and simulated data yielded more realistic parameters. The advantage of combined estimation using both real and simulated data was shown in a case study on I-880.

7.1.4 Combining algorithms

The trade-offs between the different MOEs indicated that it was difficult to improve all the MOEs simultaneously as increasing the *type I error* (increase in FAS) decreased the *type II error*(decreased the missed detections) and vice-versa. Each algorithm can be thought of as having a trade-off frontier which represents the maximum attainable performance of the algorithm for a given set of system variables. Combining algorithms was shown to be equivalent to combining different trade-off frontiers, one for each algorithm, to improve the overall performance. Two methods to combine algorithms were presented:-

Findings

i. The combination of the SPDDRP idea and the Refined McMaster algorithm was a simple example to demonstrate the idea of combining complementary algorithms. The combined algorithm(SPDMCM) was shown to have better performance than either component algorithm.

ii. The effectiveness of combining algorithmic decisions using a logit framework was shown using the LOGCM-I and LOGCMB-II models. These models were evaluated in different tests on I-93 North. Calibration of these algorithms was performed using data from location 2 and the algorithms were validated using data from location 1.

iii. The LOGCMB-II model detected incidents quicker than LOGCMB-I because it used the *strength* of the incident signal.

7.2 Future Research

I. Using BOSS quattro to calibrate parameters of McMaster and APID algorithms: BOSS quattro is a software package which is designed to analyze and optimize the influence of parameters on responses yielded by external software applications. A model gives BOSS quattro the required information about the parameters affecting the external software application. Using a driver system, BOSS quattro reads the available parameters from models and collects the analysis responses after the application run. New responses can be defined by combining the responses and parameters using mathematical operators and functions. Parametric study, Monte-Carlo study, Design of experiments, Multiobjective Constrained optimization and Updating are some of the built in analysis and optimization routines. The user can also write his/her own driver to conduct any specialized optimization or analysis routine that he/she may want to perform. Using this software and an incident analyzing tool which generates MOEs for given sets of sensor data and a simple objective function subject to some constraints, the calibration process can be automated.

II. The algorithms were calibrated for I-93 North and I-90 West. Similar calibration is necessary for I-93 South and I-90 East before a network-wide implementation for the CA/T project.

III. The performance of the algorithms has been evaluated for single lane blockages. The performance has to be studied for other types of incidents such as partial lane blockage, multiple incidents etc.

IV. The advantages of combined estimation using real and simulated data for logit models have been demonstrated using data from I-880. Similar estimation would have to be performed for the freeways in the Central Artery Network before the parameters can be used in the field.

V. On-line updating schemes would have to be used to constantly update the parameters to significant network-wide changes like change in weather conditions that affect the overall network state.

VI. Combining algorithms is a difficult problem. In general this problem could be difficult. When persistency checks are used, different algorithms declare incidents at different points in space and time. Combining algorithms becomes a cumbersome task. Combining algorithms using the Logit framework is easier. Decisions of the McMaster, APID and SPDDRP algorithms have been incorporated in the LOGIT framework. Other algorithms could also be used into the model to improve the specification.

VII. The model specification of the logit model can be enriched to increase its applicability. Traffic variables from two adjoining stations are typically used by different algorithms. Chen et al.[15] have studied the use of Principal Component Analysis(PCA) to extract the independent variations from the traffic data. These independent variations could also be used in the model specification to improve the prediction power. PCA ia also the core of the Preprocessor Feature Extraction used by Abdulhai and Ritchie[5]. It has been used in conjunction with Neural Networks. An analogous approach is recommended for use with choice models.

Appendix A

Results of Calibration

Table A.1: McMaster Parameters for the Ted Williams Tunnel

SNo	b1	b2	b3	b4	b5	OCMAX	avSpd	VCRIT	n	i1	i2
1	-1.07e-05	1.90e-03	-0.1202	3.0295	-10.51	22.65	62.06	15.66	2	10	11
2	-1.10e-05	1.92e-03	-0.1200	3.0021	-10.36	22.84	61.85	15.52	2	12	13
3	-1.04e-05	1.88e-03	-0.1202	3.0280	-10.58	22.01	61.65	15.42	2	14	15
4	-1.06e-05	1.85e-03	-0.1167	2.9577	-10.22	22.13	61.36	15.67	2	16	17
5	-1.17e-05	2.01e-03	-0.1247	3.0838	-10.82	22.21	61.20	15.41	2	18	19
6	-1.14e-05	2.01e-03	-0.1260	3.1260	-11.16	21.50	61.33	15.35	2	20	21
7	-1.12e-05	1.97e-03	-0.1237	3.0866	-10.93	22.09	61.03	15.49	2	22	23
8	-1.07e-05	1.87e-03	-0.1179	2.9825	-10.40	22.89	60.92	15.61	2	24	25
9	-1.05e-05	1.83e-03	-0.1148	2.8994	-9.97	22.86	60.64	15.29	2	26	27
10	-1.21e-05	2.05e-03	-0.1241	3.0306	-10.69	22.51	59.99	14.86	2	28	29
11	-1.01e-05	1.78e-03	-0.1129	2.9234	-10.96	23.89	57.43	15.39	2	31	32
12	-1.07e-05	1.79e-03	-0.1109	2.8643	-11.23	24.91	55.05	14.87	2	34	35
13	-9.84e-06	1.72e-03	-0.1091	2.8874	-12.32	26.71	50.13	14.82	2	37	38
14	-2.69e-05	3.79e-03	-0.1953	4.3240	-20.79	24.31	48.09	13.95	2	39	40
15	-9.01e-06	1.66e-03	-0.1049	2.7278	-11.20	27.50	43.09	13.97	2	41	42
16	-4.56e-06	9.65e-04	-0.0710	2.1313	-7.58	29.54	44.70	14.86	2	43	44
17	1.75e-06	1.54e-04	-0.0370	1.6072	-4.80	29.96	46.51	15.71	2	45	46
18	7.82e-06	-4.95e-04	-0.0144	1.3122	-3.35	29.06	47.76	16.14	2	47	48
19	1.40e-05	-1.17e-03	0.0087	1.0337	-2.35	28.00	48.42	16.43	2	49	50
20	1.12e-05	-9.33e-04	0.0024	1.1004	-2.51	28.43	48.88	16.59	2	51	52

Table A.2: McMaster Parameters for the Ted Williams Tunnel(contd.)

SNo	b1	b2	b3	b4	b5	OCMAX	avSpd	VCRIT	n	i1	i2
21	1.08e-05	-9.00e-04	0.0022	1.0985	-2.52	29.14	49.11	16.85	2	53	54
22	8.80e-06	-7.13e-04	-0.0030	1.1439	-2.67	29.60	49.29	16.80	2	55	56
23	1.62e-05	-1.38e-03	0.0146	1.0187	-2.82	28.14	49.31	16.77	2	71	72
24	2.31e-05	-1.90e-03	0.0230	1.0776	-4.14	26.69	49.24	16.63	2	73	74
25	5.55e-05	-4.76e-03	0.1106	-0.0233	0.66	25.44	49.34	16.51	2	75	76
26	2.94e-05	-2.99e-03	0.0882	-0.2851	2.41	33.47	49.26	16.62	2	77	78
27	4.80e-05	-4.29e-03	0.1017	0.0258	0.77	25.84	49.35	16.71	2	79	80
28	5.05e-05	-4.28e-03	0.0974	0.1165	0.21	26.32	49.35	16.95	2	81	82
29	5.12e-05	-4.36e-03	0.0989	0.1186	0.10	26.31	49.35	16.78	2	83	84
30	6.16e-05	-5.08e-03	0.1157	-0.0152	0.22	26.55	49.35	16.92	2	85	86
31	4.76e-05	-4.08e-03	0.0909	0.2289	-0.60	26.48	49.36	16.89	2	87	88
32	2.31e-05	-2.15e-03	0.0416	0.6957	-1.80	27.50	49.31	17.31	2	89	90
33	2.77e-05	-2.55e-03	0.0532	0.5641	-1.37	27.55	49.26	17.30	2	91	92
34	2.10e-05	-1.95e-03	0.0362	0.7410	-1.90	28.29	49.24	17.32	2	93	94
35	1.46e-05	-1.41e-03	0.0219	0.8774	-2.23	29.10	49.15	17.51	2	95	96
36	1.74e-05	-1.62e-03	0.0257	0.8747	-2.44	28.08	49.22	17.38	2	97	98
37	1.58e-05	-1.53e-03	0.0246	0.8719	-2.38	28.83	49.16	17.53	2	99	100
38	1.92e-05	-1.79e-03	0.0302	0.8367	-2.44	28.33	49.14	17.24	2	101	102
39	1.45e-05	-1.40e-03	0.0202	0.9144	-2.39	28.58	49.09	17.31	2	103	104
40	1.49e-05	-1.61e-03	0.0406	0.4271	-1.32	32.71	49.04	16.81	2	105	106
41	1.40e-05	-1.31e-03	0.0179	0.9287	-2.39	29.27	49.08	17.45	2	107	108
42	1.89e-05	-1.77e-03	0.0316	0.7814	-2.09	28.68	49.08	17.30	2	109	110
43	1.08e-05	-1.01e-03	0.0092	1.0134	-2.72	29.88	49.20	17.35	2	111	112
44	1.80e-05	-1.51e-03	0.0178	0.9714	-2.11	27.72	51.25	16.85	2	113	114
45	2.45e-05	-1.95e-03	0.0244	0.9648	-1.99	26.03	53.06	16.52	2	116	117
46	2.01e-05	-1.50e-03	0.0072	1.1893	-2.45	25.03	54.48	16.27	2	119	120
47	5.08e-05	-3.59e-03	0.0529	0.8448	-1.59	23.82	55.90	16.43	2	122	123
48	6.48e-05	-4.44e-03	0.0687	0.7631	-1.41	23.29	57.25	16.59	2	126	127
49	8.54e-05	-5.74e-03	0.0942	0.6050	-1.25	22.74	58.28	16.56	2	130	131
50	9.27e-05	-6.36e-03	0.1097	0.4781	-1.04	22.34	58.93	16.62	2	134	135
51	8.94e-05	-6.23e-03	0.1097	0.4579	-0.89	22.55	59.33	16.91	2	138	139
52	1.17e-04	-8.00e-03	0.1498	0.1012	0.16	22.31	59.66	16.94	2	144	145
53	1.17e-04	-8.16e-03	0.1548	0.0718	0.12	22.23	60.18	17.15	2	146	147

Table A.3: McMaster Parameters for the I-93 North

sno	b1	b2	b3	b4	b5	ocmax	avspd	vcrit	n	i1	i2	i3	i4
1	1.75e-04	-3.55e-03	-4.99e-03	0.87	0.01	19.04	44.10	8.00	4	535	536	537	538
2	-5.35e-05	2.17e-03	-5.99e-02	1.12	-0.72	19.43	45.54	7.81	4	531	532	533	534
3	-3.92e-05	2.66e-03	-7.77e-02	1.23	-1.25	18.46	47.27	7.24	4	12	13	14	15
4	-6.41e-06	1.09e-03	-6.38e-02	1.41	-2.16	18.82	49.90	8.32	3	530	529	528	
5	-6.81e-06	1.16e-03	-6.75e-02	1.49	-2.59	18.70	50.25	8.39	3	26	27	28	
6	-6.85e-06	1.16e-03	-6.73e-02	1.48	-2.81	18.79	49.10	8.07	3	525	526	527	
7	-7.33e-06	1.21e-03	-6.79e-02	1.43	-2.74	18.00	48.30	7.40	3	522	523	524	
8	-7.21e-06	1.19e-03	-6.65e-02	1.40	-2.50	17.83	47.72	7.38	3	385	386	387	
9	-6.31e-06	1.06e-03	-6.05e-02	1.33	-2.49	18.67	46.17	7.33	3	381	382	383	
10	-4.31e-06	7.83e-04	-4.98e-02	1.24	-3.34	21.70	36.31	7.12	4	252	253	254	269
11	-3.90e-06	7.14e-04	-4.63e-02	1.19	-3.21	35.20	36.22	7.78	4	517	518	519	520
12	-4.42e-06	8.00e-04	-5.06e-02	1.25	-3.40	21.36	37.73	7.01	4	378	379	380	521
13	-4.00e-06	7.64e-04	-5.20e-02	1.41	-3.01	24.68	47.42	10.15	3	514	515	516	
14	-4.31e-06	7.99e-04	-5.31e-02	1.41	-3.33	24.45	47.38	9.55	4	375	377	374	376
15	-5.28e-06	9.64e-04	-6.17e-02	1.56	-3.27	22.63	49.39	10.24	3	255	256	257	
16	-4.49e-06	8.47e-04	-5.63e-02	1.48	-3.07	23.73	48.91	10.32	3	371	372	373	
17	-5.86e-06	1.03e-03	-6.39e-02	1.56	-2.95	22.05	49.79	10.09	3	258	259	260	
18	-6.01e-06	1.05e-03	-6.43e-02	1.55	-2.79	21.77	49.69	10.05	3	365	366	367	
19	-5.61e-06	1.01e-03	-6.31e-02	1.57	-3.13	22.32	49.55	10.24	3	368	369	370	
20	-6.02e-06	1.05e-03	-6.39e-02	1.54	-2.75	21.77	49.83	9.98	3	362	363	364	
21	-5.98e-06	1.03e-03	-6.24e-02	1.50	-2.53	21.93	49.67	9.86	3	261	262	263	
22	-6.19e-06	1.06e-03	-6.29e-02	1.48	-2.31	21.40	49.62	9.70	3	359	360	361	
23	-6.24e-06	1.06e-03	-6.26e-02	1.47	-2.29	21.29	49.44	9.63	3	356	357	358	
24	-6.07e-06	1.03e-03	-6.08e-02	1.43	-2.09	21.44	49.00	9.55	3	353	354	355	
25	-6.00e-06	1.01e-03	-6.01e-02	1.42	-2.00	21.24	48.47	9.54	3	264	265	266	
26	-4.46e-06	7.79e-04	-4.89e-02	1.25	-1.77	23.65	46.88	9.51	3	349	350	351	
27	-2.47e-06	5.14e-04	-4.10e-02	1.31	-1.59	28.53	44.30	12.83	3	238	239	240	
28	-4.05e-06	7.85e-04	-5.39e-02	1.45	-2.03	23.61	41.92	11.33	3	728	729	730	
29	-5.89e-06	1.04e-03	-6.50e-02	1.61	-2.05	22.20	44.87	11.55	3	731	732	733	
30	-6.16e-06	1.09e-03	-6.75e-02	1.66	-2.17	21.88	47.06	11.69	3	1111	2222	3333	
31	-6.56e-06	1.15e-03	-7.10e-02	1.72	-2.44	21.63	48.99	11.81	3	244	245	246	
32	-6.81e-06	1.19e-03	-7.25e-02	1.74	-2.52	21.42	49.48	11.80	3	343	344	345	
33	-6.99e-06	1.21e-03	-7.34e-02	1.76	-2.62	21.32	49.63	11.77	3	340	341	342	
34	-6.80e-06	1.18e-03	-7.24e-02	1.75	-2.68	21.60	49.40	11.73	3	247	248	249	
35	-5.69e-06	1.03e-03	-6.69e-02	1.72	-3.11	23.72	46.04	12.21	4	70	71	72	73

Table A.4: McMaster Parameters for the I-93 North(contd.)

36	-6.69e-06	1.15e-03	-7.12e-02	1.77	-3.00	22.87	47.61	12.25	4	339	338	337	336
37	-7.08e-06	1.20e-03	-7.26e-02	1.78	-2.90	22.71	47.66	12.26	4	74	75	76	77
38	-7.07e-06	1.19e-03	-7.19e-02	1.77	-2.76	22.85	47.62	12.32	4	335	334	333	332
39	-7.00e-06	1.17e-03	-7.09e-02	1.75	-2.62	23.16	47.31	12.34	4	331	330	329	328
40	-6.94e-06	1.16e-03	-7.02e-02	1.73	-2.54	23.41	47.12	12.35	4	78	79	80	81
41	-6.92e-06	1.15e-03	-6.96e-02	1.72	-2.52	23.97	46.76	12.33	4	82	83	84	85
42	-6.49e-06	1.09e-03	-6.66e-02	1.68	-2.38	31.18	44.36	12.76	4	86	87	88	89
43	-6.14e-06	1.03e-03	-6.40e-02	1.64	-2.28	31.37	43.68	12.87	4	96	97	98	99
44	-5.97e-06	1.01e-03	-6.29e-02	1.62	-2.25	31.25	43.55	12.92	4	324	325	326	327
45	-5.81e-06	9.87e-04	-6.19e-02	1.62	-2.25	31.41	43.41	12.96	4	320	321	322	323
46	-5.64e-06	9.61e-04	-6.08e-02	1.60	-2.28	31.66	43.09	12.96	4	316	317	318	319
47	-4.90e-06	8.58e-04	-5.63e-02	1.54	-2.19	32.60	42.04	13.02	4	312	313	314	315
48	-4.09e-06	7.43e-04	-5.10e-02	1.46	-2.05	33.20	40.55	12.97	4	106	107	108	109
49	-4.11e-06	7.28e-04	-4.98e-02	1.43	-2.11	32.71	39.62	12.59	4	308	309	310	311
50	4.77e-06	6.73e-05	-3.36e-02	1.24	-2.63	22.61	40.75	10.39	4	304	305	306	307
51	1.84e-05	-7.23e-04	-1.31e-02	1.09	-2.27	26.31	41.56	13.03	3	117	118	119	
52	-3.61e-05	1.62e-03	-3.22e-02	1.00	-1.63	25.02	43.02	14.42	4	300	301	302	303
53	-1.19e-04	4.38e-03	-6.17e-02	1.17	-1.61	23.18	46.71	14.33	4	125	126	127	128
54	-1.09e-04	3.53e-03	-4.53e-02	1.10	-1.51	23.25	47.63	14.50	4	296	297	298	299
55	-9.26e-05	2.98e-03	-3.95e-02	1.08	-1.48	22.84	47.97	14.47	4	292	293	294	295
56	-6.58e-05	1.99e-03	-2.74e-02	1.04	-1.41	22.49	48.52	14.49	4	133	134	135	136
57	-6.30e-05	1.88e-03	-2.54e-02	1.03	-1.34	22.04	48.89	14.47	4	288	289	290	291
58	-9.23e-05	2.97e-03	-3.79e-02	1.08	-1.32	21.17	49.10	14.28	4	144	145	146	147
59	-5.21e-05	1.13e-03	-1.06e-02	0.94	-1.36	21.78	48.61	14.20	4	284	285	286	287
60	3.77e-05	-3.04e-03	5.15e-02	0.62	-1.33	23.07	47.57	13.86	4	280	281	282	283
61	4.60e-05	-3.33e-03	5.28e-02	0.64	-1.46	23.25	47.26	13.50	4	152	153	154	155
62	-7.63e-05	1.76e-03	-2.41e-02	1.02	-1.30	19.61	48.36	12.47	4	158	159	160	161
63	-2.29e-04	5.11e-03	-4.61e-02	1.07	-1.23	19.64	48.87	12.62	4	1723	1724	1725	1726
64	-1.37e-04	3.36e-03	-3.59e-02	1.06	-1.16	19.22	49.20	12.80	4	1727	1728	1729	1730
65	-6.44e-05	1.55e-03	-2.06e-02	1.01	-1.14	19.12	49.51	13.11	4	1731	1732	1733	1734
66	-5.87e-05	1.08e-03	-1.14e-02	0.96	-1.10	19.69	49.80	13.50	4	1735	1736	1737	1738
67	-4.32e-05	9.31e-04	-1.28e-02	0.98	-1.07	18.98	50.01	13.65	4	1739	1740	1741	1742
68	-4.04e-05	1.01e-03	-1.56e-02	1.00	-1.06	18.35	50.12	13.58	4	1743	1744	1745	1746
69	-1.05e-04	3.28e-03	-4.05e-02	1.08	-1.09	18.18	50.19	13.43	4	1747	1748	1749	1750
70	-1.03e-04	3.27e-03	-4.19e-02	1.09	-1.10	18.56	50.17	13.48	4	1751	1752	1753	1754

Table A.5: McMaster Parameters for the I-93 North(contd.)

71	-1.47e-04	4.33e-03	-4.85e-02	1.10	-1.08	18.67	50.11	13.52	4	1755	1756	1757	1758
72	-7.11e-05	1.94e-03	-2.43e-02	1.02	-1.07	18.55	49.99	13.68	4	1759	1760	1761	1762
73	-9.52e-05	2.95e-03	-3.86e-02	1.08	-1.08	18.07	49.81	13.33	4	1763	1764	1765	1766
74	-6.03e-05	1.74e-03	-2.46e-02	1.04	-1.07	18.67	49.67	13.57	4	1767	1768	1769	1770
75	-9.19e-05	2.40e-03	-2.76e-02	1.03	-1.08	18.95	49.44	13.59	4	1771	1772	1773	1774
76	-7.13e-05	1.62e-03	-1.81e-02	0.99	-1.09	18.98	49.25	13.54	4	1775	1776	1777	1778
77	-4.54e-05	1.17e-03	-1.88e-02	1.02	-1.09	19.23	49.11	13.70	4	1779	1780	1781	1782
78	-4.29e-05	8.04e-04	-1.05e-02	0.97	-1.09	19.64	49.02	13.74	4	1783	1784	1785	1786
79	-6.48e-05	1.28e-03	-1.39e-02	0.97	-1.08	19.83	48.52	13.60	4	1789	1790	1791	1792
80	-6.06e-05	1.34e-03	-1.87e-02	0.99	-1.24	20.67	46.75	13.24	4	1793	1794	1795	1796
81	3.39e-05	-2.17e-03	2.09e-02	0.82	-1.58	22.53	44.15	11.73	4	1797	1798	1799	1800
82	-3.36e-05	8.68e-04	-2.20e-02	1.04	-1.37	20.35	47.39	12.27	3	700	701	702	
83	-8.38e-06	-9.05e-05	-8.63e-03	0.99	-1.28	19.12	48.62	12.59	3	3855	3856	3857	
84	-2.46e-05	7.10e-04	-1.76e-02	1.02	-1.24	18.57	49.50	12.95	3	697	698	699	
85	-9.32e-05	2.94e-03	-3.98e-02	1.08	-1.26	18.58	49.08	12.86	3	3852	3853	3854	
86	-1.15e-04	4.01e-03	-5.78e-02	1.14	-1.50	19.91	47.10	12.16	3	686	687	688	
87	-3.46e-06	-2.11e-04	-6.08e-03	0.90	-1.49	27.02	41.42	13.01	4	3848	3849	3850	3851
88	9.84e-06	-5.78e-04	-7.90e-03	0.99	-1.36	27.34	42.78	13.53	4	689	690	691	692
89	7.47e-06	-6.54e-04	-7.13e-03	1.00	-1.28	25.47	44.57	12.77	4	3844	3845	3847	3846
90	9.01e-06	-9.94e-04	2.34e-03	0.96	-1.25	26.91	44.36	13.48	4	693	694	695	696
91	3.33e-05	-1.72e-03	9.52e-03	0.93	-1.24	25.70	44.68	13.56	4	655	656	657	658
92	1.50e-05	-1.32e-03	8.44e-03	0.93	-1.20	24.40	45.32	13.41	4	3840	3841	3842	3843
93	3.45e-05	-2.35e-03	2.55e-02	0.85	-1.18	24.21	46.30	13.42	4	3836	3837	3838	3839
94	8.06e-06	-1.66e-03	2.03e-02	0.86	-1.17	23.77	46.51	13.39	4	3832	3833	3834	3835
95	3.03e-05	-2.22e-03	2.44e-02	0.85	-1.24	23.65	45.97	13.16	4	3828	3829	3830	3831
96	4.74e-05	-2.92e-03	3.39e-02	0.80	-1.32	24.50	45.20	12.93	4	3823	3824	3825	3826
97	4.79e-05	-2.94e-03	2.99e-02	0.85	-1.42	23.02	45.05	12.43	4	3820	3821	3822	3827
98	3.73e-05	-2.27e-03	2.26e-02	0.87	-1.31	25.45	45.66	13.77	3	3817	3818	3819	
99	2.21e-05	-1.74e-03	1.79e-02	0.89	-1.21	22.96	47.01	13.83	3	3814	3815	3816	
100	-6.81e-05	1.96e-03	-2.68e-02	1.06	-1.20	20.41	48.52	14.02	3	3811	3812	3813	
101	-8.33e-05	2.61e-03	-3.47e-02	1.08	-1.18	20.43	48.66	14.22	3	717	718	719	
102	-8.73e-05	2.90e-03	-3.93e-02	1.10	-1.15	20.28	48.73	14.34	3	3808	3809	3810	
103	-1.10e-04	3.71e-03	-4.84e-02	1.14	-1.14	20.16	48.76	14.24	3	714	715	716	
104	-1.25e-04	4.33e-03	-5.70e-02	1.17	-1.15	20.33	48.71	14.25	3	3803	3804	3805	
105	-1.82e-04	6.57e-03	-8.64e-02	1.27	-1.46	19.97	48.01	13.17	3	703	704	705	
106	-4.65e-05	1.47e-03	-2.19e-02	0.98	-1.32	23.73	44.73	14.84	4	3799	3800	3801	3802
107	-3.13e-05	6.28e-04	-8.17e-03	0.94	-1.31	23.14	46.02	14.90	4	706	707	708	709
108	-8.96e-05	3.08e-03	-4.01e-02	1.09	-1.19	21.25	47.60	15.09	4	2799	2800	2801	2802
109	-1.31e-04	4.67e-03	-5.88e-02	1.16	-1.12	20.61	48.23	15.05	4	710	711	712	713

Table A.6: APID Parameters for Ted Williams tunnel

Parameter Name	Value
compwavetestperiod	0
compwavetestenabled	0
compwaveth1	100
compwaveth2	100
persistenabled	1
medtrfenabled	1
ligttrfenabled	1
incdetth1	10.2
incdetth2	0.40
incdetth3	28.8
medtrfincdet1	0.40
medtrfincdet2	0.10
persistencyth	0.40
medtrfth	15.00
lowtrfth	0.00
incclrth	0.31

Table A.7: APID Parameters for I-93 North

Parameter Name	Group1	Group2	Group3	Group4	Group5
compwavetestperiod	0	0	0	0	0
compwavetestenabled	0	0	0	0	0
compwaveth1	100	100	100	100	100
compwaveth2	100	100	100	100	100
persistenabled	1	1	1	1	1
medtrfenabled	1	1	1	1	1
ligttrfenabled	1	1	1	1	1
incdetth1	25	30	55	30	20
incdetth2	0.35	0.80	0.40	0.40	0.40
incdetth3	20	10	17	17	17
medtrfincdet1	0.40	0.80	0.40	0.40	0.40
medtrfincdet2	0.10	0.10	0.10	0.10	0.10
persistencyth	0.35	0.80	0.80	0.50	0.50
medtrfth	15.00	15.00	15.00	15.00	15.00
lowtrfth	0.00	0.00	0.00	0.00	0.00
incclrth	0.31	0.31	0.31	0.31	0.31

Appendix B

I880 Geometric Layout

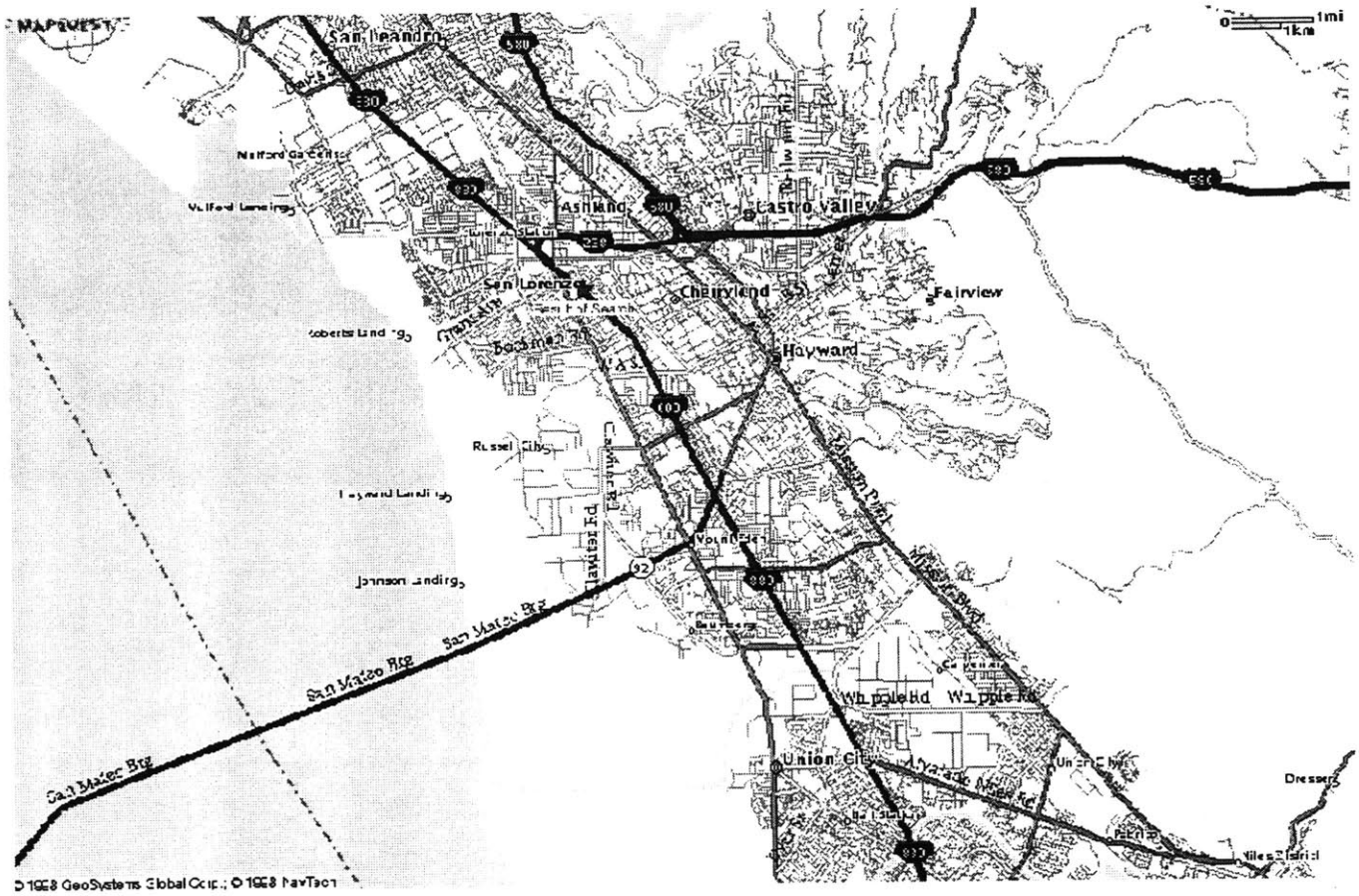
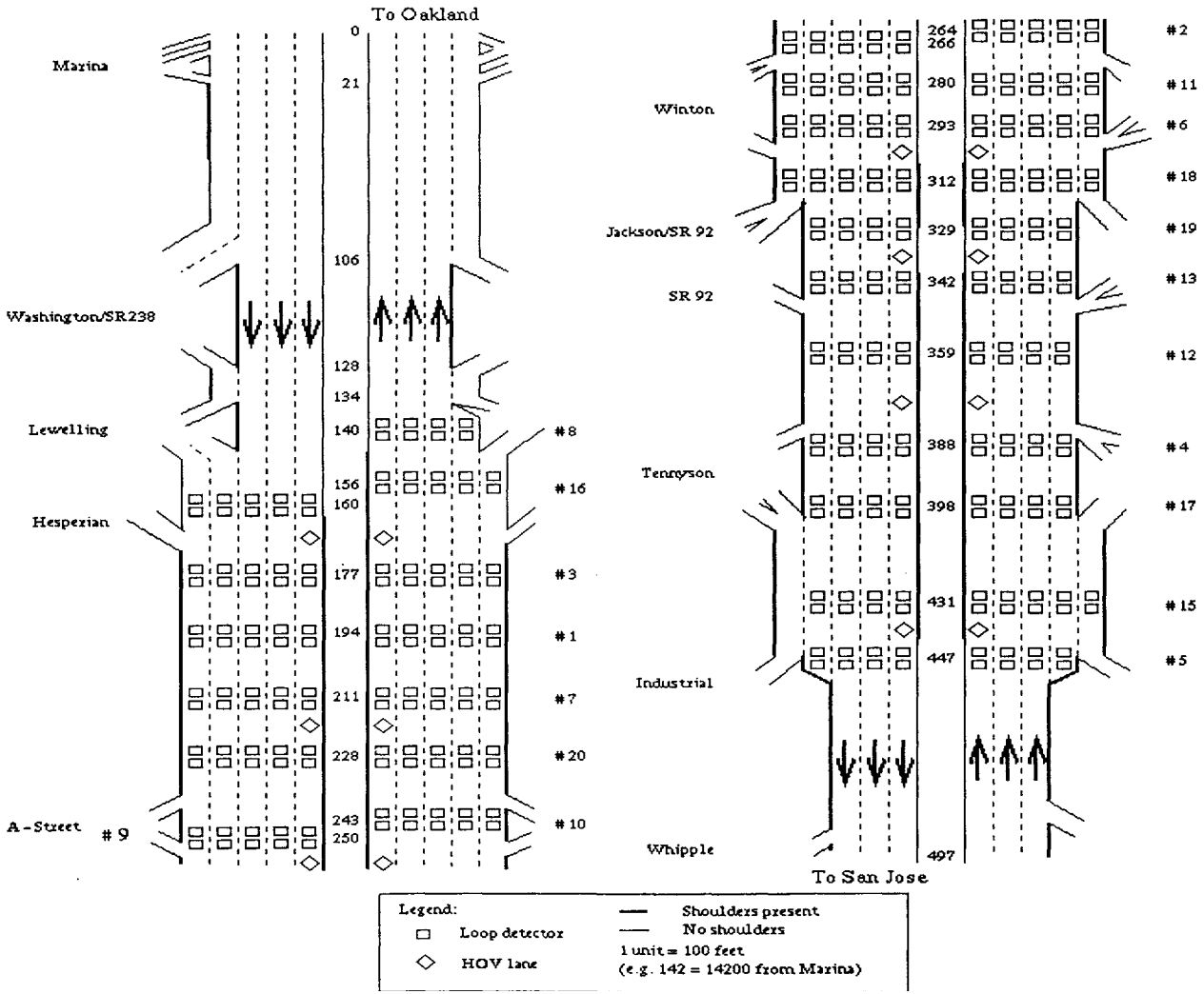


Figure B-1: Layout of I880 from www.mapquest.com

Appendix C

I880 Layout of Sensors



Appendix D

Sample plots from real data

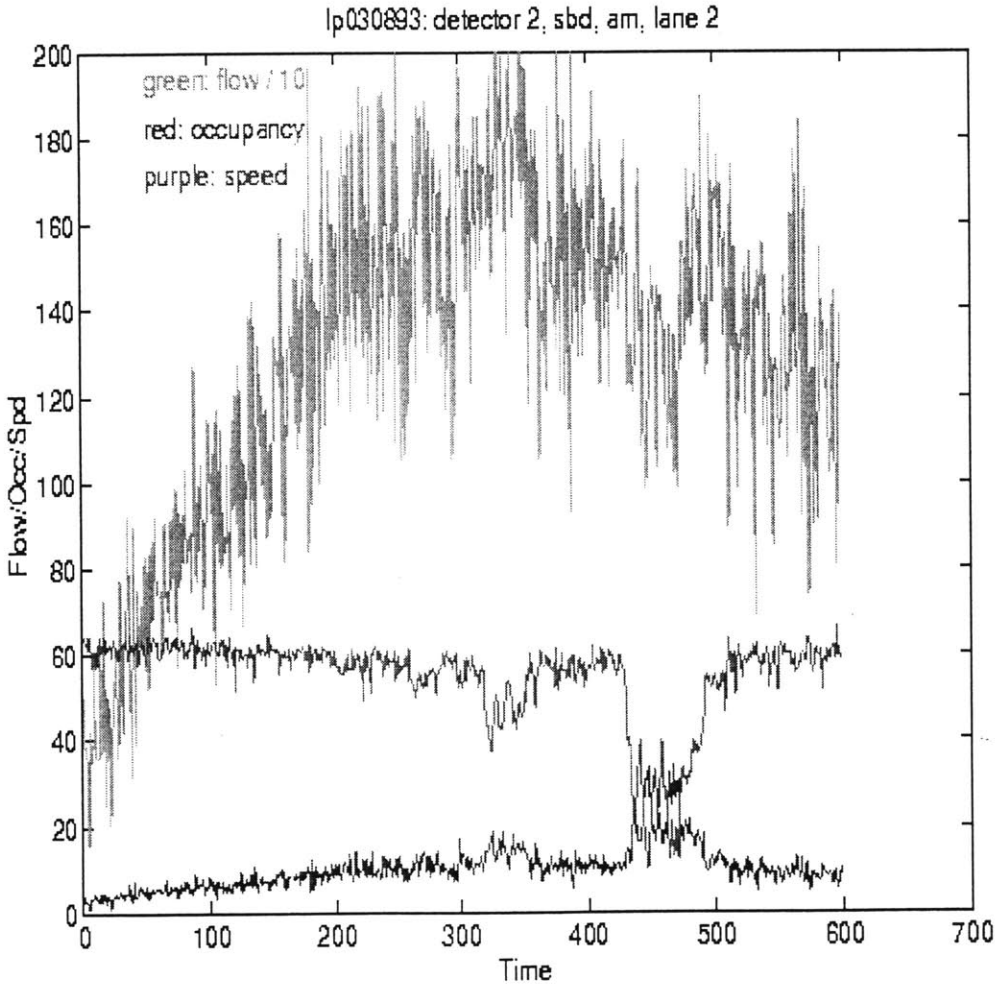


Figure D-1: flow/occ/spd plots

Appendix E

Geometric Layout of the CA/T network

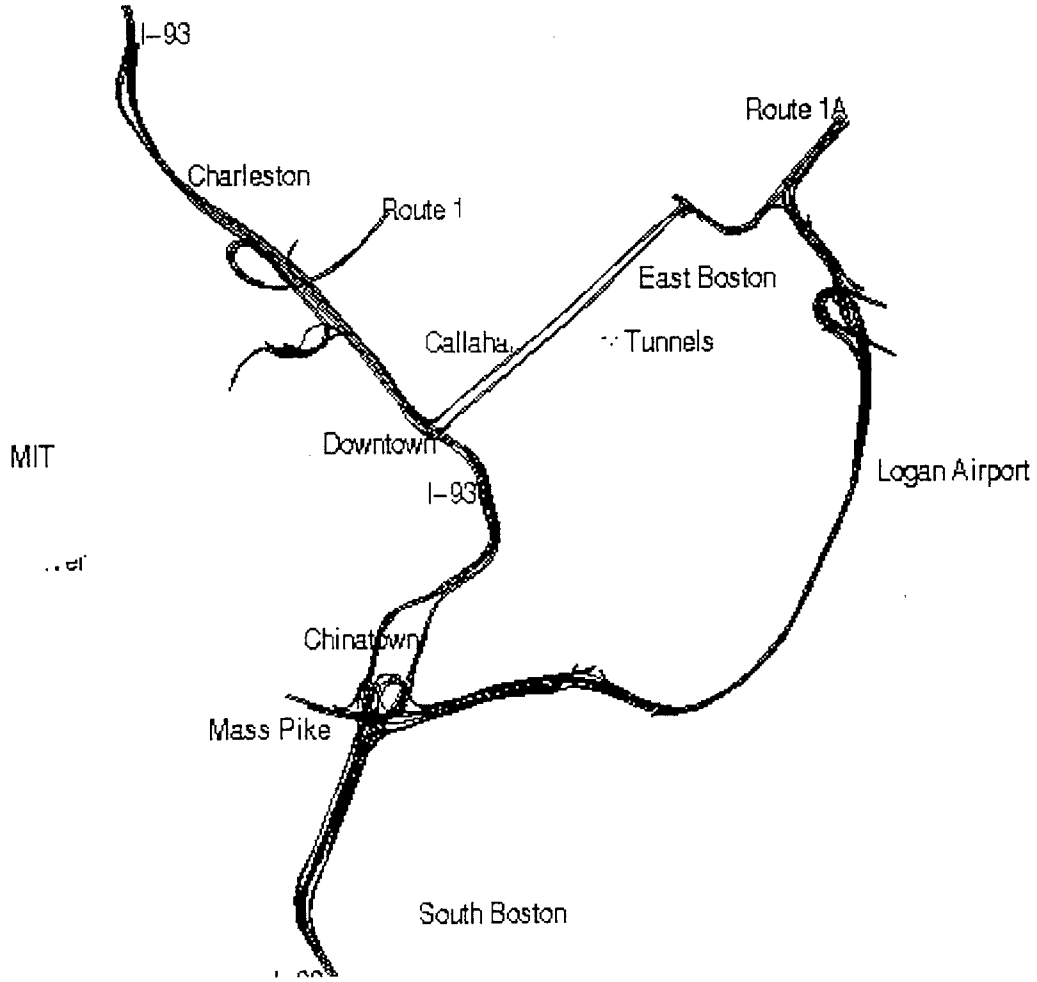


Figure E-1: The CAT network

Bibliography

[1] Gall A.I. and Hall F.L. Distinguishing between incident congestion and recurrent congestion: A proposed logic. *Transportation Research Record*, 1232:1–8, 1988.

[2] Cook A.R. and Cleveland D.E. Detection of freeway capacity reducing incidents by traffic stream measurements. *Transportation Research Record*, 495:1–11, 1974.

[3] Willsky A.S., Chow E.Y., Gershwin S.B., Greene C.S., Houpt P.K., and Kurkjian A.L. Dynamic model-based techniques for the detection of incidents on freeways. *IEEE Transactions on Automatic Control*, AC25:347–360, 1980.

[4] Abdulhai B. *A Neuro-Genetic based Universally Transferable Freeway Incident Detection Framework.* PhD thesis, University of California at Irvine, December 1996.

[5] Abdulhai B. and Ritchie S. A preprocessor feature extractor and a postprocessor probabilistic output interpretor for improved freeway incident detection. *Transportation research record*, 1998.

[6] Persaud B.N. *Study of a freeway bottleneck to explore some unresolved traffic flow issues.* PhD thesis, University of Toronto at Ontario, May 1986.

[7] Persaud B.N. and Hall F.L. Catastrophe theory and patterns in 30-second freeway traffic data - implications for incident detection. *Transportation Research - Part A*, 23(2):103–113, 1989.

[8] Persaud B.N., Hall F.L., and Hall L.M. Congestion identification aspects of the mcmaster incident detection algorithm. *Transportation Research Record*, 1287:167–175, 1990.

[9] Rothrock C.A. and Keefer L.E. Measurement of urban traffic congestion. *Highway Research Board Bulletin*, 156:1–13, 1957.

[10] Wright C.C. Some properties of the fundamental relations of traffic flow. *Proceedings of the 5th International Symposium on the Theory of Traffic Flow*, pages 19–32, 1972.

[11] Dudek C.L. and Messer C.J. Incident detection on urban freeways. *Transportation Research Record*, 495:12–24, 1974.

[12] Hall F.L., Pushkar A., and Shi Y. Some observations on speed-flow and flow-occupancy relationships under congested conditions. *Transportation Research Record*, 1398:24–30, 1993.

[13] Hall F.L., Hurdle V.F., and Banks J.H. Synthesis of recent work on the nature of speed-flow and flow-occupancy(or density) relationships on freeways. *Transportation Research Record*, 1365:12–18, 1992.

[14] Hall F.L., Shi Y., and Atala G. On-line testing of the McMaster incident detection algorithm under recurrent congestion. *Transportation research record*, 1394:1–7, 1993.

[15] Chen H., Boyle R., and Montgomery F. Motorway incident detection using principal component analysis. In Brown R., editor, *Incident Detection and Management, Bloomsbury square training center on Monday, 2 June 1997*, pages 1–7, London, UK, 1997. IEE Road Transport Group.

[16] Dia H. and Rose G. The impact of data quantity on the performance of neural network freeway incident detection models. Technical report, Institute of Transport Studies, The Australian Key Centre in Transport Management, Monash University, Clayton VIC, Australia, 1996.

[17] Payne H.J., Helfenbein E.D., and Knobel H.C. Development and testing of incident-detection algorithms, research methodology and detailed results. *TechnicalReport FHWA-RD-76-12*, 2, April 1976.

[18] Payne H.J. and Tignor S.C. Freeway incident-detection algorithms based on decision trees with states. *Transportation Research Record*, 682(30):37+, 1978.

[19] Teng H.L. and Madanat S. The CUSUM algorithm for freeway incident detection. *submitted to Transportation Research - Part B*, 1997.

[20] Forbes G. J. Identifying incident congestion. *ITE Journal*, 62(6):17–22, June 1992.

[21] Payne H. J. *Traffic flow theory with application to Freeway Ramp Control and Incident Detection*. Technology Service Corporation, 1974.

[22] Collins J.F., Hopkins C.M., and Martin J.A. Automatic incident detection : TRRL algorithms HIOCC and PATREG. Technical report, Transport and Road Research Laboratory, Crowthorne, Berkshire, 1979.

[23] Ashok K. and Ben-Akiva M. Dynamic O-D matrix estimation and prediction for real-time traffic management systems. *Transportation and Traffic flow theory*, pages 465–484, 1993.

[24] Petty K.F. The freeway service patrol homepage. *http://www.path.berkeley.edu/FSP/*, 1995.

[25] Petty K.F. *Incidents on the Freeway: Detection and Management*. PhD thesis, University of California at Berkeley, December 1997.

[26] Ahmed K.I. *Modeling Drivers' accelaration and Lane Changing Behavior*. PhD thesis, Massachusetts Institute of Technology, February 1999.

[27] Sanwal K.K., Petty K.F., and Walrand J. An extended macroscopic model for traffic flow. *Transportation Research*, 30(1):1–9, 1996.

[28] J.A. Lindley. Qualification of urban freeway congestion and analysis of remedial measures. *Technical Report RD/87-052, FHWA, Washington, D.C.*, 1986.

[29] Bierlaire M. HieLoW, a package that estimates hierarchical logit models. Technical report, World Conference on Transportation Research, Sydney, Australia, 1995.

[30] Levin M. and Krause G.M. Incident detection algorithms–part 1: Off-line evaluation, part 2: On-line evaluation. *Transportation Research Record*, 722:49–64, 1979.

[31] Zhou M. and Hall F.L. Investigation of the speed-flow relationship under congested conditions on a freeway. *Presentated at the* 78^{th} *Transportation Research Board Meeting*, 1999.

[32] Ben-Akiva M.E., Koutsopoulos H.N., Mishalani R.G., and Yang Q. Simulation labratory for evaluating dynamic traffic management systems. *ASCE journal of transportation engg.*, 123(4):283–289, 1997.

[33] Bhandari N., Sethi V., Koppelman F.S., and Schofer J.L. Arterial incident detection using fixed vehicle detector and probe vehicle data. *Transportation Research Record*, 1510:60–72, 1995.

[34] Masters P.H., Lam J.K., and Wong Kam. Incident detection algorithms for compass - an advanced traffic management system. *Vehicle Navigation and Information Systems Conference*, pages 295–310, 1991.

[35] Bickel P.J. and Doksum K.A. *Mathematical Statistics*. Holden-Day, Inc., 1977.

[36] Varshney P.K. *Distributed detection and data fusion. Signal processing and digital filtering*. Springer, New York, 1997.

[37] Yang Q. and Koutsopoulos H.N. A microscopic traffic simulator for evaluation of dynamic traffic management systems. *Transportation Research - part C*, 4(3):113–129, 1997.

[38] Blahut R.E. *Principles and Practice of Information Theory*. Addison-Wesley, Reading, Mass., 1987.

[39] Peeta S. and Das D. Continuous learning framework for freeway incident detection. *Transportation Research Record*, 1644:124–144, 1998.

[40] Ahmed S.A. and Cook A.R. Application of time series analysis techniques to freeway incident detection. In *The 61st Annual Meeting of the Transportation Research Board*, Washington, D. C., 1982. Transportation Research Board.

[41] Ritchie S.G. and Cheu R.L. Simulation of freeway incident detection using artificial neural networks. *Transportation Research - Part C*, pages 203–217, 1993.

[42] Ishak S.S. and Al-Deek H.M. Fuzzy art neural network model for automated detection of freeway incidents. *Transportation research record*, 1634(0361):56–72, 1998.

[43] Chassiakos A.P. Stephenedes Y.J. and Michapoulos P.G. Comparative perofrmance evaluation of incident detection algorithms. *Transportation Research Record*, 1360, 1992.

[44] Sethi V., Bhandari N., Koppelman F.S., and Schofer J.L. Arterial incident detection using fixed vehicle detector and probe vehicle data. *Transportation Research - part C*, 3(2):99–112, 1994.

[45] Lin Wei-Hua and Daganzo C.F. A simple detection scheme for delay-inducing freeway incidents. *Transportation Research - Part B*, pages 141–155, 1996.