A CATEGORICAL MODEL FOR TRAFFIC INCIDENT LIKELIHOOD ESTIMATION

A Thesis

by

SHAMANTH KUCHANGI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2006

Major Subject: Civil Engineering

A CATEGORICAL MODEL FOR TRAFFIC INCIDENT LIKELIHOOD ESTIMATION

A Thesis

by

SHAMANTH KUCHANGI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Paul Nelson
Committee Members,	Yunlong Zhang
	Michael Sherman
Head of Department,	David Rosowsky

December 2006

Major Subject: Civil Engineering

ABSTRACT

A Categorical Model for Traffic Incident Likelihood Estimation.

(December 2006)

Shamanth Kuchangi, B.Tech, Regional Engineering College, Warangal

Chair of Advisory Committee: Dr. Paul Nelson

In this thesis an incident prediction model is formulated and calibrated. The primary idea of the model developed is to correlate the expected number of crashes on any section of a freeway to a set of traffic stream characteristics, so that a reliable estimation of likelihood of crashes can be provided on a real-time basis. Traffic stream variables used as explanatory variables in this model are termed as "incident precursors". The most promising incident precursors for the model formulation for this research were determined by reviewing past research. The statistical model employed is the categorical log-linear model with coefficient of speed variation and occupancy as the precursors. Peak-hour indicators and roadway-type indicators were additional categorical variables used in the model. The model was calibrated using historical loop detector data and crash reports, both of which were available from test beds in Austin, Texas. An examination of the calibrated model indicated that the model distinguished different levels of crash rate for different precursor values and hence could be a useful tool in estimating the likelihood of incidents for real-time freeway incident management systems.

DEDICATION

To My Parents:

Smt Prema and Sri K Prabhu

ACKNOWLEDGMENTS

I would like to thank my advisor and committee chair, Dr. Paul Nelson for his guidance in completing my thesis. I am highly indebted to him for his openness in discussing his thoughts on several topics in traffic and transportation, which has helped me learn and refine my scientific thought process to a great extent. I would like to specifically acknowledge him for inspiring and rekindling my interest in traffic flow theory and for guiding me to a wealth of knowledge. More so, he has personally touched and inspired me by his thoughtfulness and understanding towards his students and by his modesty. I consider it an honor to have been associated with him and having him as my mentor at Texas A&M.

I would like to thank my committee members, Dr. Yunlong Zhang and Dr. Michael Sherman for reviewing my thesis and providing insightful comments, which helped me refine the clarity of my thesis. I would like to thank all my team members at TTI for their technical input and for critiquing my document. I would also like to thank all the faculty members in the Transportation Engineering division at Texas A&M University, specifically Dr. Mark Burris, for his help in academic related advising and the Civil Engineering faculty at the Regional Engineering College, Warangal for the help in building my technical expertise.

I'm also grateful to Dr. Gordon Newell, Dr. Carlos Daganzo, Dr. Joseph Sussman, Dr. John R. Meyer and Dr. S. Raghavachari with whom I never personally interacted or had little interaction, but their books and writings I happened to read have deeply influenced my thoughts and views about traffic theory and several aspects in transportation engineering.

As always, I remain truly indebted to my dear parents, my brother, Jayanth, my family and friends, specifically Ms. Amruthavani, for their love, support and encouragement in all phases of my life.

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION iv

ACKNOWLEDGMENTS v

TABLE OF CONTENTS vii

LIST OF FIGURES ix

LIST OF TABLES x

1 INTRODUCTION 1

1.1 Research Objective 1
1.2 Background and Motivation 2
1.3 Scope of Research and Report Organization 4

2 LITERATURE REVIEW 7

2.1 Freeway Management 7
2.2 Incident Detection 11
2.3 Incident Prediction 16

3 STUDY METHODOLOGY 21

3.1 Precursor Selection 21
3.2 Log-Linear Models 23
3.2.1 Generalized Linear Models 23
3.2.2 Log-Linear Models for Contingency Tables 25
3.2.3 Goodness-of-Fit Tests 26
3.3 Incident Likelihood Model 28
3.4 Model Assumptions 29

4 DATA ANALYSIS AND REDUCTION 30

4.1 Study Area and Data Description 30
4.2 Data Reduction 33
4.2.1 Precursor Calculation from the Data 33
4.2.2 Identification of Incidents 35
4.2.3 Determining the Boundary Values for Precursors 37

Page

4.2.4 Determining Exposure 38

5 RESULTS AND DISCUSSIONS 41

5.1 Model Calibration Results 41
5.2 Discussion of Parameter Estimates 42
5.2.1 Physical Interpretation of Estimated Parameters 43
5.2.2 Statistical Significance of Estimated Parameters 45
5.2.3 Comparison of Results with Canadian Model 47
5.3 Alternate Reduced Category Model 48
5.3.1 Calibration for Reduced Model 48
5.3.2 Discussion of Results for Reduced Model 50
5.4 Limitations 52

6 CONCLUSIONS 55

6.1 Summary 55
6.2 Future Work 58

REFERENCES 60

VITA 64

LIST OF FIGURES

Page

FIGURE 2.1 Causes of Highway Congestion in United States (*12*)..................................9

FIGURE 2.2 Incident Management Processes (*13*) 11

FIGURE 4.1 Two Test Beds for Obtaining Data 31

FIGURE 4.2 Speed Profile at Detector 6006822 (speed in MPH)...................................36

FIGURE 4.3 Speed Profile at Detector 6005721 (speed in MPH)...................................37

FIGURE 5.1 Residual Analysis for Full-Scale Model...51

FIGURE 5.2 Residual Analysis for Reduced Model 52

LIST OF TABLES

Page

TABLE 3.1 Review of Precursors Used for Incident Prediction 23

TABLE 3.2 Two-way Contingency 25

TABLE 4.1 Boundary Values for the Precursors..38

TABLE 4.2 Input Data for Model Calibration..40

TABLE 5.1 Results of Parameter Estimation 42

TABLE 5.2 Boundary Values for 2-Level CVS 49

TABLE 5.3 Parameter Estimation for Reduced Model 49

TABLE 5.4 Goodness of Fit 50

1 INTRODUCTION

1.1 Research Objective

The objective of this thesis was to develop and calibrate a categorical log-linear model to estimate the likelihood of incidents on a real-time scale for freeway sections in the State of Texas. The methodology used here closely followed that suggested by Lee *et al.* (*1, 2*). However the model was slightly modified, in terms of the number of categorical variables, and was calibrated with a dataset that was available for urban freeway traffic in the State of Texas. This work also derived its content in congruence with the requirements of a project at Texas Transportation Institute (TTI), where this model was intended to be used for the development of an Incident Detection and Short-Term Congestion Prediction prototype for Texas Department of Transportation (TxDOT).

This thesis also served the purpose of validating the possible use of the proposed methodology to integrate with the Freeway Management Systems in the United States. However in the scope of this work, the developed model has not been validated against any site data other than that used for calibration, due to the constraints in obtaining data and time constraints within the overall project framework.

This thesis follows the style of *Transportation Research Record*.

1.2 Background and Motivation

Accidents are a common scenario on freeways in the United States. There has been an increase in the number of highway related fatalities over the past 10 years, which is attributed to the increase in population (*3*). National statistics shows that there also has been a reduction in fatality rate per capita (*3*). But still a significant number of fatalities are observed, even after the conventional safety measures have been implemented. Increased use of freeways has not only made our freeways unsafe, but also contributes to the frequent congestion and hence increased delays, adding to the travel time. National Highway Safety Administration has ranked road accidents as the number one cause of death among several age groups (*3*). Studies on incident detection and development of highway safety models have been in vogue before the 1970's (*4*); however, these techniques in philosophy focus on reducing the post crash effects, which can only marginally reduce deaths due to incidents on freeways, rather than on proactively reducing the crash rate. With the development of real-time and intelligent transportation systems, it has been the perspective of researchers and architects of transportation system to focus on crash avoidance, rather than on reducing consequences of crashes (*5*). It is hoped that such a transition would not only save many more lives on freeways, but also avoid frequent congestion due to incidents and hence add value to the overall economy of the country.

Over the past decade fewer research attempts have been made in developing the ability to predict the likelihood of incidents for application in real-time Freeway Management Systems than that is required, to make any conclusive remarks on the possibility of incident prediction. There have been some successful models correlating crash rate with certain precursors of traffic flow. One of the earliest among these, especially for a real-time application was reported by Oh *et al.* (*6*). In their study the probability of disrupted traffic is estimated parametrically using Bayesian methodology with longitudinal coefficient of variation of speed (CVS) as a precursor. In this context, a “precursor” is a traffic stream variable or some combination of the traffic stream variables, the values of which are expected to correlate with incidents. The result showed that the proposed model captures a significant number of accidents (*6*). Other successful and more elaborate studies reported with regard to real-time incident likelihood prediction were by Lee *et al.*, (*1, 2*). These studies used a categorical model to relate crash rate with certain precursors, such as coefficient of variation of speed, occupancy, spatial difference in speed between adjacent detectors along a lane, and other factors such as peak hour factor and roadway geometric factor. Data from a Canadian Expressway with three lanes (and short sections of four lanes) in each direction over a 13 month period were used.

In contrast to the successful models that have been highlighted in the previous paragraph, there have been a few others that have shown less confidence in the possibility to predict likelihood of accidents for applicability in real-time systems. The most striking contrast can be seen in the study reported by Kockelman *et al.* (*7*). This study uses a set of conventional models (non categorical), such as Binomial and others, in attempt to relate potential crash precursors to likelihood of crashes. Speed variation, and average speed along the lane, as well as their section averages were considered as the precursors for this study. Data from a freeway test bed in California has been used for this study. However the researchers in this study concluded that there is no evidence from their data that speed or speed variations correlated with crashes (*7*). A detailed note on these models is provided in the literature review of Section 2 below.

It was observed that all models cited above were sensitive to the selected site conditions and quality of data. Not only the quality of data obtained was of concern, but the different aggregation periods chosen for a specific study could also result in varied correlation between the precursors and the crash rate. This was a possible reason for the varied research results that have been reported in the past. In view of the sensitivity of these real-time incident likelihood models to the site, data quality and aggregation period, it was deemed essential to evaluate the incident likelihood models with specific site conditions to establish their general use on Texas freeways.

1.3 Scope of Research and Report Organization

In this research the potential capabilities of using a categorical modeling technique to estimate the crash rate on freeway sections are demonstrated. In the above mentioned context the effectiveness of selected incident precursors is also examined. The model was validated by examining the statistical significance of the parameters, the overall model fit and the physical interpretation of the parameters. However in this research the model is not validated with historical data or in real-time field conditions due to lack of data for validation and time constraints. The report is organized in the following order.

- In this introductory section the precise objective of the research work is specified. Also, here the general background to the current work and the motivation for taking up this research is provided.
- The introductory section is followed by a literature review section, where a detailed state-of-the-art review on incident management is provided and the development of the transition from incident detection to incident prediction is traced. Here a fit for the current work in the overall practice of incident management is provided.
- A study methodology section contains information on the framework for the developed model. In this section an explanation of the procedure undertaken for incident precursor selection is given. It also contains some background to log-linear models as a first step toward introducing the concepts and terminologies used in the model formulated in this research work. Finally in this section, details of the categorical model formulated for this research in correlating number of crashes to the selected precursors, including the assumptions underlying in the model formulation is provided.
- A section on the data reduction procedures that were employed for model calibration is provided. This section starts with a brief introduction to the study site from which the required data were obtained. Also in this section, details of the data that was available for this study are described. Following this, the data processing that was

carried out on the raw data to obtain the required input for model calibration is elaborated.

- A results and discussion section contains information on the calibration procedure and presents the model calibration results. A detailed discussion on the significance of the model parameters is provided, based on the calibration results and relating those results to the general observation of traffic flow characteristics. Finally, the limitations of this modeling approach are discussed.
- This thesis concludes with a brief section related to conclusions. In this section a summary of the research problem, research process and research findings are provided. Also some pointers to future work as an extension to this thesis are listed.

2 LITERATURE REVIEW

This section is a brief review of freeway management, from a broad perspective. The literature review presented here builds the motivation for research needs in developing efficient freeway management strategies. A top-down approach is adopted to provide a discussion on the general aspects of freeway management in Subsection 2.1; Subsection 2.2 more specifically treats the incident management process. In Subsection 2.3, a review of incident detection algorithms is presented. Finally in Subsection 2.4, a detailed review on selected state-of-art models related to incident forecasting and prediction is presented.

2.1 Freeway Management

As civilization grows, there is always a growing need for transportation. Mobility of goods, safe and reliable passenger travel, accessibility and security are essential contributing factors for economic development of any society. The contribution of transportation systems to the societal development is as important in Texas and the United States as it is elsewhere. This is manifested by the vast surface transportation system, such as the interstate highway and freeway networks that have grown to now being approximately 55,000 centerline miles (*8*). Freeways are less than 2.4% of the total road network yet carry about 20% of the traffic throughout the United States (*8*).

By any measure the freeway network forms the backbone for the transportation needs in the United States.

However with the increasing population and economic prosperity in the United States, there has also been an increasing demand in freeway usage over the years. It was reported that from 1980 to 1999 there was 76% increase in vehicle miles of travel, while during the same period the increase in highway miles constructed was 1.5% (*8*). Another report indicated that between years 1993 and 2000 there was an anticipated increase of 50% in vehicular traffic (*9*). In the current day situation the traffic on most of the freeways has reached to its maximum capacity limits, specifically at peak hours. Because less can be done in expanding the freeway system, due to lack of space, the focus now has turned to effective management of the freeway system, for efficient operation in order to provide a better quality of travel and reliable travel time for road users.

Ever increasing demand on US freeways has posed a real challenge in freeway management. Travel congestion studies conducted by the Texas Transportation Institute estimated that in the year 2000, 75 metropolitan areas experienced a travel delay of 3.6 billion vehicle-hours, 5.7 billion gallons wastage in fuel and around $67.5 billion loss in productivity (*10*). In addition to the severe congestion and loss in productivity, the National Highway Traffic Safety Administration has reported that more than 42,000 people died on highways and 3 million people are injured due to traffic related crashes in the year 2002 (*11*). The economic cost of these crashes is estimated to be more than $230 billion per year (*9*).

To combat congestion, many freeway management programs have been put in place. These management strategies can be broadly classified as

- Travel demand management
- Traffic responsive operations
- Freeway incident management

Figure 2.1 shows the contributing factors for freeway congestion (*12*). As can seen from Figure 2.1, incidents are the second major cause for highway congestion, and incidents are accountable for 25% of the time congestion is observed on highways. This emphasizes the important role of freeway incident management for reducing highway congestion by reducing highway crashes.

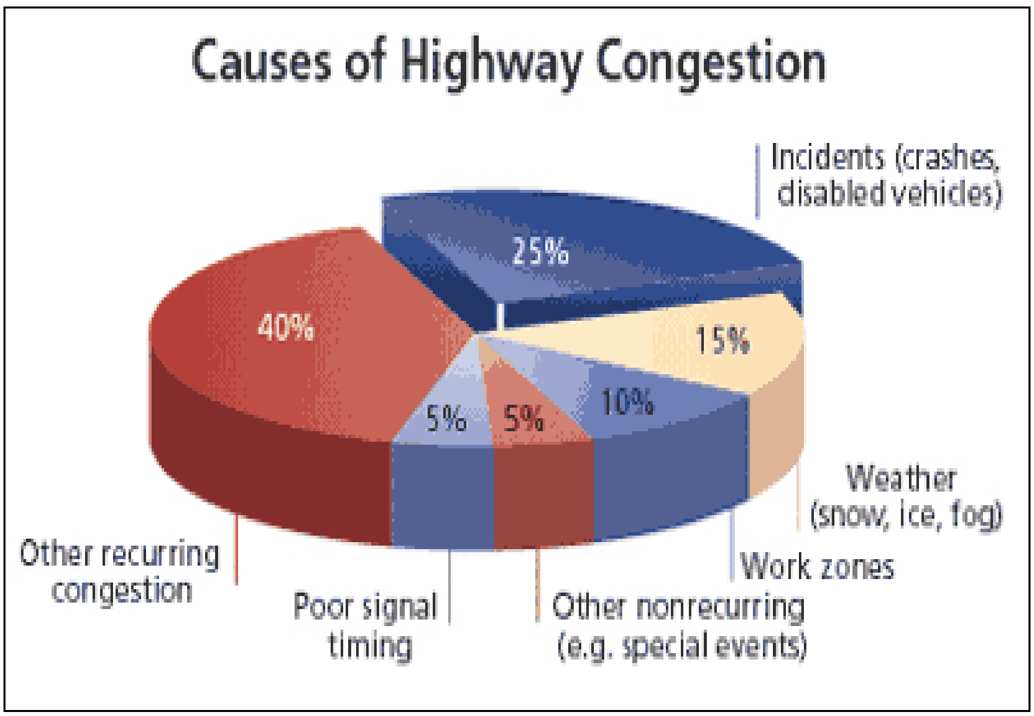

FIGURE 2.1 Causes of Highway Congestion in United States (*12*)

Congestion due to incidents can be minimized by diverting traffic to reduce the propagation of shock waves, and by clearing the incident as quickly as possible. Such measures can avoid secondary crashes and also save human life. But the length of time required in reporting an incident and the resulting response time is high in manual surveillance and traditional police reporting. Such delayed responses may substantially compound the problem of congestion and lead to fatalities. As such, in recent years many major metropolitan cities throughout the country have been establishing efficient incident management programs to reduce congestion and manage freeway incidents (*9*). These incident management programs generally involve the following four stages of management (*9*):

- Incident detection to reduce the time it takes to detect and verify incidents
- Incident response to identify the nature of an incident and initiate appropriate response
- Incident clearance to clear an incident quickly completely from the roadway
- Traffic management and motorist information to minimize the traffic disruption on the highway

The complete process involved in incident management is shown in Figure 2.2. The primary purpose of efficient incident management program is to minimize the total incident duration.

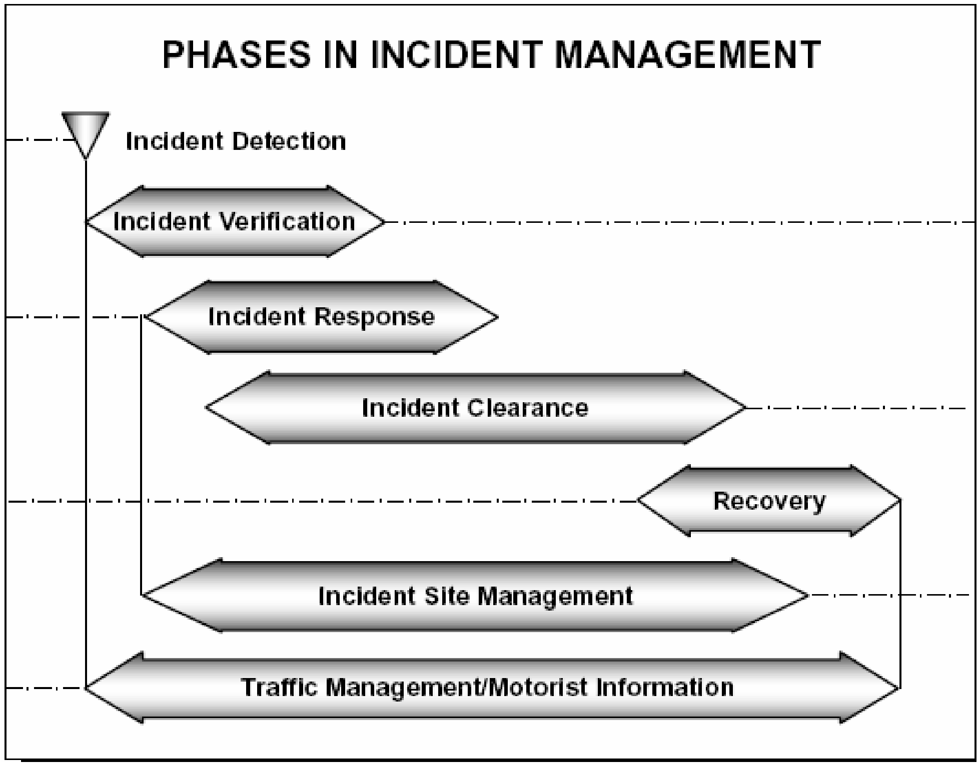

FIGURE 2.2 Incident Management Processes (*13*)

2.2 Incident Detection

Incident detection is an important component of freeway incident management, and also a challenging task for efficient incident management. Many of the techniques used for incident detection can also be useful for incident prediction, which is the focus of this thesis. Hence a brief review of some incident detection techniques is presented here. Several researchers have developed incident detection algorithms to for efficient detection of incidents. Efficiency in detection is determined by factors such as percentage of time incident is detected, false alarm rate, and time to detect incidents. A brief review of some incident detection algorithms is provided here. Incident detection algorithms described here are classified based on their underlying techniques as (*14*)

- Statistics-based algorithms
- Smoothing-based algorithms
- Artificial-intelligence-based algorithms, and
- Probe-based methods

Statistical algorithms in principle determine the deviance of the observed traffic data from the predicted data using some standard statistical techniques. One of the earliest notable statistical methods was derived by Dudek *et al.* at Texas Transportation Institute in 1974 called the standard normal deviate algorithm (SND) (*15*). The SND algorithm is based on the principle that incidents trigger a sudden change in the traffic stream variables. The SND algorithm computes the number of deviations in the 1-minute occupancy from the detectors with the mean value of the 1-minute occupancy for historical data at that location. A threshold is defined in the SND for the allowable deviance. When the measured SND exceeds some critical value, algorithm indicates the presence of an incident. Two successive 1-minute intervals are used to make a consistency test. Other statistical incident detection algorithms are based on Bayesian techniques (*16, 17*). In Bayesian algorithm, frequency distributions for the upstream and downstream occupancies during incident and incident-free conditions are developed, and the likelihood of incidents is estimated by computing conditional probability using Bayesian techniques. Time-series algorithms are another statistical related technique where time-series models are employed to predict normal traffic conditions and detect incidents when detector measurements deviate significantly from model outputs. The autoregressive integrated moving-average (ARIMA) model is one of the popular time series model. ARIMA model assumes that differences in a traffic variable measured in the current time slice (t) and the same traffic variable in the previous time slice (t-1) can be predicted by averaging the errors between the predicted and observed traffic variable from the past three time slices (*18*). These errors are expected to follow a normal pattern under incident-free conditions, and any deviation from the normal distribution of errors indicates a possible incident occurrence.

Smoothing techniques in principle filter short-term noises or non-homogenous conditions from traffic data that cause false alarms, and allow traffic patterns to be more clearly visible in order to detect true incidents (*19*). Some of the smoothing/filtering algorithms employed are double exponential smoothing (DES) algorithm, low-pass filter (LPF) algorithms and the discrete wavelet transform and linear discriminant analysis (DWT-LDA) algorithm (*14*). The DES algorithm weights the preceding and present traffic stream variables obtained from the detector data for forecasting short-term traffic conditions that are expected to reflect actual traffic conditions. This algorithm is expressed as a double exponential smoothing function, with a constant, which weights all the observations over the time window considered for smoothing. Incidents are detected using a tracking signal, which is the algebraic sum of errors between the predicted and observed traffic variable over a 12 minute window. Under incident-free conditions, the tracking signal should dwell around zero since predicted and observed traffic conditions should be similar. LPF algorithms are also known as "Minnesota algorithms" (*14*). This algorithm distinguishes noise as high frequency fluctuations in the data and the low frequency fluctuations that are typically characterized as incident conditions. Three types of smoothing techniques were used to distinguish incidents and recurring congestion to reduce false alarm rate (*14*). The DWT-LDA algorithm consists of two components and uses the techniques of signal processing. The DWT component is used to filter raw traffic data, and the random fluctuations of traffic are discarded. Then, the LDA component is used on the filtered traffic data for feature extraction to identify incidents.

Many artificial intelligence techniques also have been applied in incident detection, including neural networks, fuzzy logic and their combinations. The commonly used neural network algorithms for incident detection include multi-layer fee-forward neural networks (MLF) and probabilistic neural networks (PNN) (*14*). Neural network algorithms require substantial training through trial-and-error to optimize weights in order to distinguish free flow and bottleneck traffic conditions. To improve incident detection efficiency, neural networks have been combined with other techniques such as wavelet transform (*20*) or probe vehicle data (*21*), to improve incident detection efficiency. As traffic data from loop detectors are usually not of good quality, fuzzy logic is a useful tool for applications involving imperfect data. In fuzzy logic a set of rules are determined to process imperfect data and determine thresholds. Decisions on incident and non-incident states can be determined with varying confidence levels. Fuzzy logic has also been combined with neural networks (*22*) to improve the performance of incident detection. Image processing techniques have been applied to video images for automatic incident detection. The autoscope incident detection algorithm (AIDA) is one such notable algorithm using image processing techniques (*23*).

Most of the above mentioned incident detection algorithms are based on loop-detector data, and are prone to a high false-alarm rate. To overcome this probe-based incident detection systems have been tried. Probes, such as toll transponders and GPS receivers mounted on vehicles, are being tested, as they are increasingly available because of the popularity of electronic toll collection deployments throughout United Sates. Probes provide better information than loop detectors on traffic conditions, including travel times and other spatial traffic measures. TTI developed a probe-based incident detection system using cellular phone and automated vehicle identification (AVI) system installed on freeway facilities in Houston (*24*). Deviation in actual travel time from mean travel time between two destinations was used to identify traffic conditions. This is based on the premise that incidents cause travel time to increase significantly over the normal travel time under incident-free conditions at the same time of day and day of week. Several other probe-based systems have been tried, such as the E-ZPass electronic toll tags used in the TRANSMIT algorithm (*25*), vehicle-equipped radio transponders used in a algorithm developed by University of California at Berkeley (*26*), and GPS data in the ADVANCE algorithm (*27*). Details on the above mentioned algorithms and other probe based incident detection systems can be found on other references (*14, 19*).

2.3 Incident Prediction

Most of the research related to freeway incident management in the past decade has focused on developing efficient incident detection algorithms. However, incident detection algorithms have had a setback from lack of confidence among freeway operators due to high false alarm rate. In addition to that, the focus in incident management has transitioned from reducing crash effects to crash avoidance, to make our highways safer (*5*). In this regard there has been some limited research effort to develop crash forecasting or prediction models. Some of the significant models are reviewed in this section.

Before we review some models related to incident prediction, it is essential to distinguish two kinds of data aggregation that would be involved in most of these models. Data can be aggregated at the controller unit of the detector system, and this will be termed as "system aggregation". Data is also commonly further aggregated for study purposes, which is termed in this report as "study aggregation". A few of the models reviewed here have already been introduced to a brief extent in Section 1.2. Here a more detailed review of previously introduced models is presented, and a broader spectrum of incident prediction models is covered, as this topic is the central theme of this thesis. One of the notably early efforts regarding real-time incident prediction modeling was reported by Oh *et al.* (*6*). Their study estimates parametrically the probability of disrupted traffic using Bayesian methodology with coefficient of variation of speed (CVS) along the lane as a precursor. In this context, a "precursor" is a traffic stream variable or some combination of the traffic stream variables, which is expected to precede incidents. The freeway section that was considered in (*6*) was a 4-lane directional freeway, where probe vehicles were used to record the incident cases. The incident data and the traffic stream data from double loop detectors were collected over a one month period, at morning and evening peak hours only. The detector data available for the study were 10-second system-aggregated data in every 5-minute period, and averaged across the lanes. The result shows that the proposed model captures significant number of accidents (*6*). Other successful and more elaborate studies reported with regard to real-time incident likelihood prediction were by Lee *et al.* (*1, 2*). These studies use a categorical model to relate crash rate with certain precursors, such as coefficient of variation of speed over the study aggregation time, density, spatial difference in speed between adjacent detectors along a lane, and other factors such as peak hour factor and roadway geometric factor. An earlier study by Lee *et al.* (*1*) also used coefficient of variation in speed across the lanes, but was later dropped as it was found to be insignificant. Data from a Canadian Expressway with three lanes (and short sections of four lanes) in each direction over a 13 month period were used. Data were obtained using double-loop detectors along the section in intervals of 20 seconds; study-aggregation used were 5 minutes (*1*), 8 minutes, 3 minutes, and 2 minutes (*2*). In calculating the precursors, the traffic stream data were averaged across all the lanes to obtain the station average. In this model different possible categorical values were defined, and boundary values for each of these categories were found from the detector data. The model parameters were then estimated. The best-fit model was chosen by comparing the model goodness of fit parameters such as likelihood ratio and p-value.

Another technique that has been used by at least a couple of researchers for real-time incident prediction is the spatio-temporal analysis preceding crashes (*28, 29*). Pande *et al.* (*28*) have used Log CVS, average values of speed, occupancy and volume, standard deviation of speed, volume and occupancy as precursors, with a study aggregation of 5 minutes. Efforts were made to individually correlate each of these putative precursors to crash risk. The crash risk was here represented as hazard ratio, defined as the resultant change in the log odds for observing a crash by changing the precursor by one unit. The hazard contour is then plotted to obtain spatio-temporal patterns from which high crash risk situations are identified in real-time. The results of the study have shown that Log CVS, and standard deviation of volume and average occupancy are significantly correlated to crash occurrence (*28*). Ishak *et al.*, (*29*) attempted to use second-order statistical measures derived from spatio-temporal speed contour maps to investigate the characteristics of pre-incident, post-incident and non-incident conditions. The second order statistical measures are Angular Second Momentum (measure of smoothness), Contrast (measure of local variance), Inverse Difference Moment and Entropy (measure of uncertainty) (*29*). The data used for the study came from loop detectors, with a system-aggregation period of 30 seconds; a 5-minute interval was used as the study-aggregation period. These data were collected for 10 minutes prior to crashes in pre-incident analysis, 10 minutes after the crashes for post-incident analysis. The results for this model failed to establish confidence in predicting incidents, as the spatio-temporal patterns were not consistently discernable between pre-incident, post-incident and non-incident cases (*29*).

Traffic volume has been a widely used variable in crash models related to safety research. Though most of models using AADT (average annual daily traffic) are conventionally developed for intersection or road segment safety improvement studies, a few models that use shorter time frame of volume counts, such as hourly volume, are reviewed here for possible use in real-time crash prediction. Persaud *et al.* (*30*) have developed two models, regression and Bayesian models using both macroscopic (AADT) and microscopic volume (hourly volume) variables. The models have been applied to different roads such as collector roads and expressways, and crashes have been distinguished as severe and non-severe. Validation results for the microscopic model, which is of some interest for real-time application, show promise, with the regression model being close to reality than Bayesian model (*30*). Another successful model has been demonstrated by Cedar *et al.* (*31*), using hourly flow. Hourly flow could be a useful precursor, if prediction is required over larger prediction time window.

The most striking contrast to some of the successful models described above can be seen in the study reported by Kockelman *et al.* (*7*). This study uses a set of conventional models (non categorical), such as Binomial and others, to relate potential crash precursors to likelihood of crashes. The precursor used in this approach is coefficient of speed variation. Data from a freeway test bed in California have been used for this study. Traffic stream data other than speed were obtained over 30-second system-aggregation intervals. Since the data were obtained with single-loop detectors, the speed component for obtaining coefficient of speed variation has been derived using standard methods with 150 sec study-aggregation (*7*). However the conclusion of this study says that “...there is no evidence in or across these crash data sets and observations of their corresponding series of 30-second traffic conditions that speeds or their variation trigger crashes” (*7*).

This review of literature has traced the trends in freeway and incident management systems. It has shown the importance of incident prediction capabilities in managing freeway congestion and reducing fatalities on highways. This review gives the state-of-art in real-time crash prediction models, with specifics of some critical components of such models. It also gives a sense of current research needs for developing efficient crash prediction capabilities.

3 STUDY METHODOLOGY

In this thesis an attempt has been made to develop an incident prediction model using the historical traffic and incident data that were obtained from two freeway test beds in Austin, Texas. In this approach a categorical log-linear model was formulated and was calibrated to be able to estimate incident likelihood. A detailed account of the methodology involved in developing the model is presented in this section. In Subsection 3.1 details of the literature survey that was carried out to select incident precursors to be included in the model is given. A brief introduction to the concepts underlying log-linear models is given in Subsection 3.2. This section also serves the purpose of establishing terminology and notation that will be used in discussing incident prediction models. Following this, a description of the model formulated for this research is provided in Subsection 3.3. Finally model assumptions are discussed in Subsection 3.4.

3.1 Precursor Selection

A detailed survey of literature on previous studies on incident detection and forecasting was conducted to identify possible incident precursors in correlating traffic stream variables with incident occurrence. While reviewing several traffic flow variables (and their combinations) as a potential precursors for accident prediction, every study in this process was reviewed for the extent of predictability, and was accordingly classified as positive (‘+’) if results showed satisfactory correlation between incident precursors and occurrence of incidents, negative (‘-’) if no correlation was found and zero (‘0’) for indeterminate cases. It was observed that there was a greater amount of confidence among researchers with coefficient of speed variation along a lane as a precursor than any other precursor. Table 3.1 gives the summary of the review that was conducted. The most common precursors seen in incident-prediction models were speed variation, occupancy, volume and hourly flow (*1, 2, 6, 7, 30, 31*). Among these, volume and hourly flow have been traditional factors in incident correlation for long-term safety studies. Hourly flow is measured as the number of vehicles per hour at a section on a roadways and volume is the average annual daily traffic. These factors are less suitable for real-time incident prediction, as they are based on measurement over longer intervals. They become insensitive in prediction on a real-time scale. It can be seen from the table that speed variation is a widely used precursor and has shown promising correlations in many cases in the past. Density (or occupancy) has also been used as precursor, and has shown good correlations in incident predictions, when used along with coefficient of speed variation. For these reasons, coefficient of speed variation along the lanes (CVS) and occupancy have been chosen as the potential precursors along with other factors such as roadway type and time of the day for this study. (Occupancy was preferred to density, because the former is directly available from our selected dataset.). It should be noted that in some studies time of day indicator and roadway-type could be referred to as precursor. However in this they are referred to as indicators, though they are treated similar to the precursors such as CVS and Occ in the model.

TABLE 3.1 Review of Precursors Used for Incident Prediction

Precursors	Number of studies reviewed	Positive Results	Negative Results	Neutral/Weak Results
Speed Variation along the lane	8	6	2	-
Speed Variation across the lane	1	-	-	1
Occupancy or Density	2	2	-	-
Volume	2	2	-	-
Hourly Flow	2	2	-	-
Headway	1	1	-	-

3.2 Log-Linear Models

Log-linear models are a class of Generalized Linear Models. These models are used for Poisson distributed data, and describe the association and interaction patterns between a set of variables (*32*). In this section some details are provided regarding log-linear models to establish the terminology that will be used further in describing the incident prediction model developed in this thesis. Much of the discussion in this section is adapted from Agresti (*32*)

3.2.1 Generalized Linear Models

A generalized linear model is usually described using three components, a random component, a systematic component and a function. (*32*)

The random component of a generalized linear model consists of the response variable and information on the distribution of the response variable. The response variable could have any kind of distribution depending on the application. If the response variable is a non-negative number, a Poisson distribution might be appropriate for the random component. It is a common practice to use a Poisson distribution for modeling vehicle crash counts.

The systematic component of a generalized linear model is denoted by a set of explanatory variables. A linear combination of the explanatory variables leads to a linear model, or a second order combination of variables leads to a canonical model.

The final component of a generalized linear model is the function that relates the random component to the systematic component. This component is also referred to as a link, as it links the expected value of the response variable (mean of the probability distribution of the response variable) to the explanatory predictor variables. The link function can take any form; the simplest link is the one which relates the explanatory variables to the mean of the response variable as given in the Equation 3.1.

$$\mu = C + \beta_1 X_1 + \beta_2 X_2 ... + \beta_i X_i \tag{3.1}$$

Where:

X’s : Explanatory variables

β's : Coefficients for the explanatory variables

C : Constant

μ : Mean response variable

The link used in Equation 3.1 is called as an identity link. Similarly a log-link is defined when the response variable is taken as log (μ), where μ is the mean of a response distribution, and μ is a positive number. Therefore a log-linear model in general is represented as given in Equation 3.2

$$\log(\mu) = C + \beta_1 X_1 + \beta_2 X_2 ... + \beta_i X_i \tag{3.2}$$

3.2.2 Log-Linear Models for Contingency Tables

The most common use of log-linear models is in the modeling of cell counts in a contingency table. The fundamentals of log-linear model can be easily understood using two-way contingency tables. A contingency table of two categorical variables, each having two categories, is shown in Table 3.2. In the table, π_{i+} and π_{+j} are the conditional probabilities of the respective row (i) and column (j).

TABLE 3.2 Two-way Contingency

		X		
		1	2	
Y	1			π_{i+}
	2			
		π_{+j}		

If the two variables X and Y are statistically independent then the joint probability π_{ij} for the cells in the Table 3.2 is

$$\pi_{ij} = \pi_{i+} * \pi_{+j} \tag{3.3}$$

So if ”*n*” is the total sample size, then the expected cell frequency is obtained as

$$\mu_{ij} = n * \pi_{ij} = n * \pi_{i+} * \pi_{+j} \tag{3.4}$$

Taking logs on both sides of Equation 3.4 yields

$$\log(\mu_{ij}) = C + \lambda_{x(i)} + \lambda_{y(j)} \tag{3.5}$$

This is called an independence log-linear model. In the Equation 3.5, λ_X is the effect parameter for the variable X and λ_Y is the effect parameter for variable Y. As an alternative to the independence model, log-linear models can also contain interaction parameters. The interaction parameter gives the combined effect of the variables used. When all possible combinations of the interaction parameters are used, the model is termed as a "saturated log-linear model". For a two-way contingency table the saturated model is given by

$$\log(\mu_{ij}) = C + \lambda_{x(i)} + \lambda_{y(j)} + \lambda_{xy(k)} \tag{3.6}$$

These concepts can be extended to three-way, four-way or higher-order contingency tables. As the order of the contingency table increases, the model becomes more complex due to the possibilities of multiple combinations of interaction parameters.

3.2.3 Goodness-of-Fit Tests

The statistical models estimated are assessed for significance of the estimated model parameters and for the overall fit of the model. The significance of the model parameters are assessed by indicators "Z" or "sig.". Z is defined as the ratio of the estimate to the standard error for the estimate. For a large sample, Maximum Likelihood estimate, Z is assumed to follow a standard normal distribution. Hence for any given confidence level, if the absolute value of Z calculated is greater than critical Z value form standard normal table, for the chosen confidence level then the estimated parameter is considered to be significant. For a 95% confidence level, critical value of Z is 1.96. “Sig.”, also referred as p-value should be greater than 0.05 at 95% confidence level for estimated parameter to be significant.

Two commonly used overall goodness-of-fit tests for log-linear models are the likelihood ratio and the Pearson statistics. The goodness of fit tests assesses the model by comparing the cell fitted values to the observed counts. These test statistics are useful in comparing best fit models for the given data. Usually the model that has lowest value for the test statistics is considered to be best fit model. The test statistics for the likelihood ratio (G^2) and Pearson (X^2) are given as (*32*)

$$G^2 = 2\sum n_{ij} \log\left(\frac{n_{ij}}{\mu_{ij}}\right) \tag{3.7}$$

and

$$\chi^2 = \sum \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}} . \tag{3.8}$$

Where:

n_{ij} : Observed cell count

μ_{ij} : Expected value for cell count

3.3 Incident Likelihood Model

A categorical log-linear model was chosen to predict the likelihood of crash rate using the precursors selected from the previous task. As indicated in the beginning of this thesis, the model very closely follows the categorical model that was suggested by Lee *et al.* (*2*); however, one of the additional precursors used by Lee et al. (*2*), the speed difference between the two adjacent detectors along a lane at a given point in time, was not considered in this study due to the difficulty in the computation of this precursor from the data that was available for calibration. The functional form of the model considered for this study is.

$$\frac{N}{EXP^{\beta}} = f\left(C * \lambda_{CVS(i)} * \lambda_{Occ(j)} * \lambda_{R(k)} * \lambda_{P(l)}\right) \tag{3.9}$$

Here:

N : Expected number of crashes over the two year analysis time frame;

EXP : Exposure in vehicle-miles of travel over two year time period;

C : Constant, which in this model represents the highest possible crash rate;

$\lambda_{CVS(i)}$: Effect of the crash precursor variable CVS having i levels;

$\lambda_{Occ(j)}$: Effect of the crash precursor variable Occ having j levels;

$\lambda_{R(k)}$: Effect of road geometry (control factor) having k levels;

$\lambda_{P(l)}$: Effect of time of day (control factor) having l levels;

β : Coefficient for exposure

f : Mathematical function

Using a logarithmic function, Equation 3.7 can be expressed statistically as

$$N = \exp(C + \lambda_{CVS(i)} + \lambda_{Occ(j)} + \lambda_{R(k)} + \lambda_{P(l)} + \beta * \ln(EXP)) + \varepsilon_{ijkl} \quad (3.10)$$

Here:

ε_{ijkl} : Random error term

In the above model, $\lambda_{CVS(i)}, \lambda_{Occ(j)}, \lambda_{R(k)}, \lambda_{P(l)}, \beta$ are the parameters to be fitted using the data. N and EXP are inputs to the model estimation process.

3.4 Model Assumptions

Some assumptions made in the model described in the previous section are listed here. Firstly it was assumed that *N*, number of accidents in each category follows a Poisson distribution. This is a reasonable assumption, knowing that number of accidents is a non-negative count. Also crashes in roadway safety are typically modeled using Poisson distribution. A second assumption that is made in the model is independence. That is, none of the interaction parameters are considered in the model. Two reasons for the use of independence model assumption were to follow closely the model form suggested by Lee *et al.*, and more importantly inclusion of interaction parameters makes it difficult to interpret the model in terms of observable traffic behavior. It was deemed important to develop a model that can be simple to interpret in-terms of practical observation, as can be seen from the interpretation given for the developed model in Subsection 5.2.

4 DATA ANALYSIS AND REDUCTION

In the previous section the parameters were identified and model parameters were defined with their categories. In this section, description of the study site characteristics and related data are given. The data that were available from the study site are described in Subsection 4.1. In Subsection 4.2 detailed process for the data reduction that was carried out in this study is explained. This explanation details the process that was adopted in reducing the raw detector data and the incident log data to obtain the input data that were used for the model calibration.

4.1 Study Area and Data Description

Data for the model development were obtained from two selected study sites in Austin, Texas. The two study sites were the freeway sections of US 183 and Loop 1, which connects US 183 to US 290. The map of the two freeway sections that were selected as study sites for this work is shown in Figure 4.1. The alignment indicated as a dotted line in the figure is Loop 1 and the alignment indicated as dashed line is the section of US 183 under study. The entire length of Loop 1 was considered as a study section, traversing a length of 9.7 miles, with 40 detector stations and overall 149 freeway detectors. The section of US 183 spanned 9.2 miles, starting from the junction of IH 35. This section contained 54 detector stations, with overall 174 detectors along the freeway, and stations spaced approximately 0.35 – 0.5 miles. In both of the sections, detectors on the on-ramps, off-ramps and frontage road were not considered in the study.

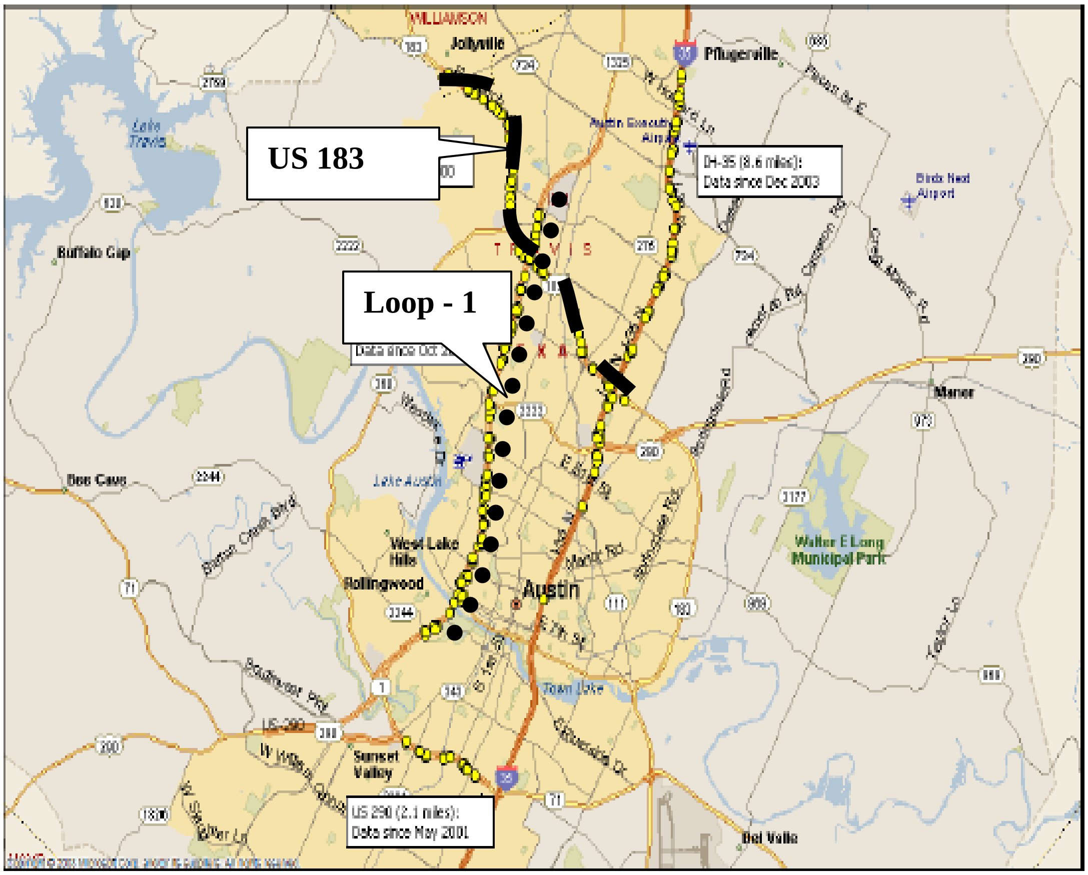

FIGURE 4.1 Two Test Beds for Obtaining Data

Detector data from the two study sites, the section of US 183 and Loop 1, were obtained from TxDOT. The traffic flow data available from the detectors were volume (in vehicles per minute), occupancy (in percent of time), speed (in miles per hour) and percentage of trucks. Although controllers use an aggregation period of 20 seconds, the historical data were aggregated (system-aggregated) over a period of one minute. These system-aggregated data were rounded off to the nearest integer value and presented against a time stamp for every one minute interval. The one-minute system-aggregated data are archived by TxDOT before being fed to the ATMS system. In this study one minute system-aggregated data on all the detectors along the section of US 183 and Loop 1 over a two-year period were considered, specifically for the years 2003 and 2004, from January through December for both years. The detector data were available in comma separated text files (CSV) format.

Apart from the detector data, incident log data for the two study sites during the year 2003 and 2004 were obtained. This incident log primarily consisted of the following information:

- Approximate location of the incident
- Date incident occurred
- Time of incident
- Time incident was cleared
- Lanes affected
- Direction affected
- Type of incident (collision or stalled or congestion)

The location of the incident in the log was indicated by the nearest cross road to the incident spot. The log also indicated if the incident was upstream or downstream of the cross road or exactly at the cross road ("Cross roads" are the arterial roads perpendicular to the freeway, but essentially outside the freeway system. Normally these cross roads are taken as a reference point when logging freeway incident location in incident reports.). However there was no information available as to how far upstream or downstream of the cross road was the exact incident location along the freeway section. The reasons for the incident were mainly classified as "collision" or "congestion" or "stalled vehicles". The "lanes affected" column in the incident log contains indications if the affected lanes were freeway or cross street or ramps or frontage roads. The affected direction gave information as to whether the incident had occurred on the southbound lanes or the northbound lanes. The incident log was available as an Excel spread sheet.

4.2 Data Reduction

This section contains a description of the process that was involved in reducing the base data that was obtained from TxDOT system to a form that can be used as an input for the calibration of the model. Some of the process involved semi-automation, but a few processes had to be manually carried out. The Statistical Analysis Software package (SAS, *34*) and Excel macro were used for most of the automated processes.

4.2.1 Precursor Calculation from the Data

The raw data from the detectors were sorted based on the detector number and time stamp. The 24-hour data for each detector were stored as a single file, in Excel data sheet (.xls) format. Therefore each file had information on a particular detector for a particular day. In each of these files for every minute, a moving average coefficient of variation of speed and average occupancy was calculated and stored in different columns.

The moving average was calculated over a 5-minute period window, starting from the time interval against which the precursors were reported and the preceding four intervals. Coefficient of variation of speed (CVS) was calculated as

$$CVS = \frac{S.D(\overline{S})}{\overline{S}} = \frac{\sqrt{\dfrac{n\sum_{1}^{n}(S_i)^2 - \left(\sum_{1}^{n} S_i\right)^2}{n(n-1)}}}{\dfrac{\sum_{1}^{n} S_i}{n}} \tag{4.1}$$

Here

$\overline{S}$: Mean speed,

S.D : Represent standard deviation,

S_i : Speed in MPH at time i, and

n : Number of time intervals

CVS defined as in Equation 4.1 gives the fluctuation of speed over 5 minutes with respect to the mean speed at a given location for that time period. Speed fluctuation could occur due to acceleration or deceleration of vehicles. Deceleration can be expected at stations with increasing occupancy, for example at a shock wave. Hence situations of high speed fluctuations with high occupancy could possibly indicate a situation highly prone to crashes.

The precursors obtained were tagged with either “peak” or “non-peak” for time-of-day consideration. Precursors occurring anytime between 6:30 AM to 9:30 AM and 4:00 PM to 7:00 PM were considered as peak, and precursors obtained at other times were considered as non-peak. The roadway-type indicator does not vary with each and every recording of data, but is fixed for every detector location. The roadway type was classified as either "straight" or "other" against each detector considered for this study. This classification was based on the horizontal alignment of the roadway section and the presence of ramps nearby to the detector station. A detector station on a straight alignment and far from the influence area of the ramps was tagged as "straight", and otherwise was tagged as "other".

4.2.2 Identification of Incidents

From the information obtained through the incident log files, it was necessary to verify two aspects of the logged incidents, to the best attainable accuracy. The two important aspects were firstly the exact location of the incident in terms of the detectors that could have first identified the incident and secondly the exact time of incident. It was necessary to verify the manually recorded incident logs for location and time, as the location description provided in the incident log was not accurate enough to pin-point the exact location (in terms of the detector). In order to exactly locate the incidents from the rough identification obtained from the logs, all detector stations near to the location of the reported incident were noted. For these possibly affected detectors, speed profiles were generated from the detector data for the incident day. The speed profiles were then examined for any drop in speed during the time around the logged incident time. The detector station where speed drop was observed approximately near to the incident log time was identified as the affected detector station. From the speed profiles, not only the incident location were identified, but also the time of incident were verified in few cases by looking at the time when the speed drop occurred on a particular affected detector station. Figure 4.2 shows a typical speed profile for incident conditions, where incident details are easily verifiable. In a few other cases, details of the incident were difficult to identify or distinguish from congestion. A typical speed profile where distinguishing incident from congestion is difficult is shown in Figure 4.3. In a strict sense, the time it takes for speed drop to propagate to the nearest upstream detector from the incident location should be deducted from the time the speed drop was noticed in the speed profile plot.

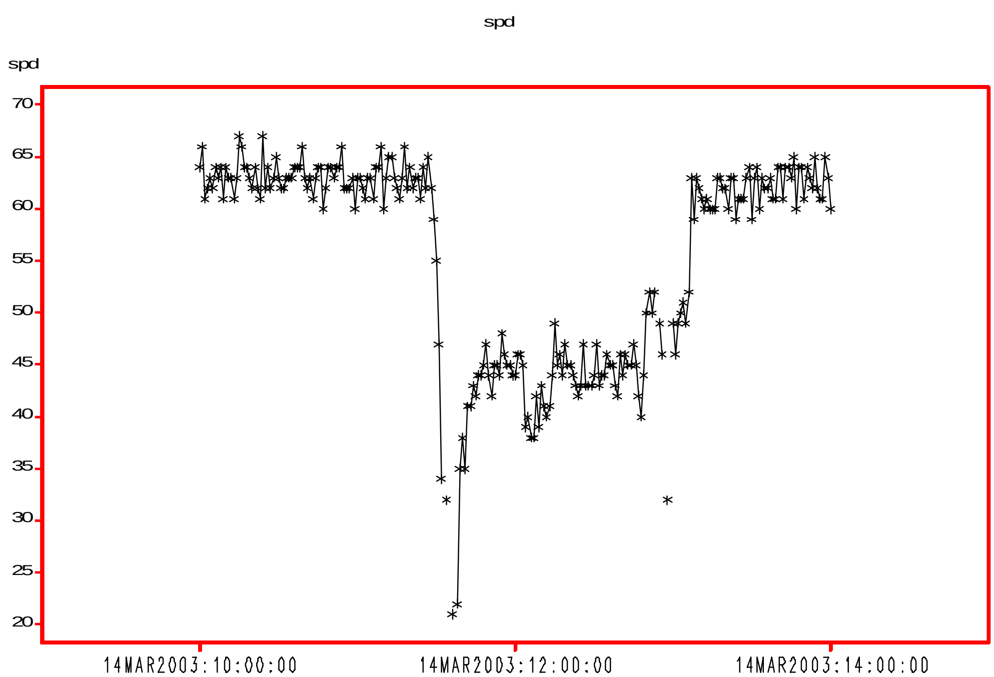

FIGURE 4.2 Speed Profile at Detector 6006822 (speed in MPH)

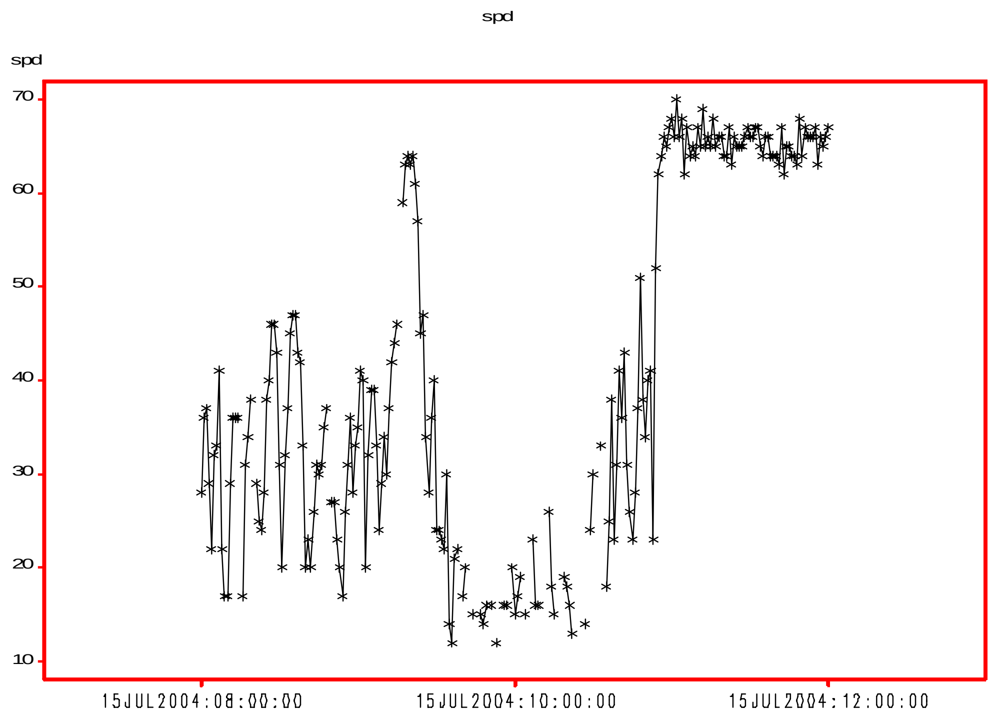

FIGURE 4.3 Speed Profile at Detector 6005721 (speed in MPH)

However this procedure of deducting the time taken by the shock wave to reach nearest upstream detector from the incident log time was not carried out, due to the uncertainty in determining that time period.

4.2.3 Determining the Boundary Values for Precursors

Lee *et al.* (*2*) found that a proportion of 50:30:20, low, medium and high values of precursors respectively, gave the best fit for the categorical model that was developed in their study. Carrying forward with this suggestion, for each of the precursors, coefficient of variation of speed and occupancy, the boundary values for the lowest 50% of the precursor values and next 30% of the precursor values were determined from the processed detector data. In order to minimize the computation time in determining the boundary values using a very large dataset, 24-hour data on a random sample of 10 detectors were taken. Using this sample, representative boundary values were obtained for both of the precursors CVS and Occ, as shown in Table 4.1.

TABLE 4.1 Boundary values for the Precursors

Category	CVS	Occ (%)
L (50) or 1	<= 0.043	<= 3.6
M (30) or 2	> 0.043 & <= 0.227	> 3.6 & <= 5.8
H (20) or 3	> 0.227	> 5.8

Note: Proportion of 50:30:20 (L:M:H) was recommended as the best fit (Chris Lee *et al.*, (*2*))

4.2.4 Determining Exposure

As described in the earlier sections, the suggested categorical model has a dependent variable crash rate, which is defined as the number of crashes over a certain exposure. Exposure is defined for a category ‘i’ as the product of the proportion of time the category existed during the total time of study, the number of vehicles recorded during that proportion of time and the aggregate distance traveled by the vehicles in that category. Boundary values for the precursors, CVS and occupancy were determined as explained in the previous section, each of these having three categories (Low, Medium and High). Apart from that two indicators, to incorporate the effect of roadway geometry and peak traffic conditions as defined earlier, each have two categories. In all any instant of the exposure could lie in one of the 36 possible categories.

In this manner, exposure for each of the 36 possible categorical values was determined using the processed detector data. To determine the roadway geometry factor, all the detectors were classified either as a straight or other section by identifying the location of the station from the detector map. A detector station was labeled as straight if it existed on a straight section of the road and not influenced by the on-ramps or off-ramps. Otherwise that station detectors were labeled as "other". For the precursor to incorporate the effect of peak hour on incident, all the data collected between 6:30am to 9:30am and 4:00pm to 7:00pm will be classified as peak time data, while data that were collected other than these periods was classified as "non-peak". With this it was possible to count the proportion of time and the number of vehicles in each of these 36 possible categories. The distance traversed was taken as the average length between the two detector stations. In order to reduce the computation time in calculating the exposure for two years period, exposure was determined over three months period and averaged by the number of days. The daily average was extrapolated for 2 year period for use in the model.

The final data after all the necessary reduction that was used for the model calibration are shown in Table 4.2.

TABLE 4.2 Input Data for Model Calibration

N	CVS	Occ	R	P	ln(EXP) in VMT
5	3	3	0	0	10.089
12	3	3	0	1	8.274
14	3	3	1	0	10.171
27	3	3	1	1	8.605
0	1	1	0	0	14.797
0	1	1	0	1	10.145
1	1	1	1	0	15.623
0	1	1	1	1	11.343
4	2	2	0	0	11.664
3	2	2	0	1	8.436
0	2	2	1	0	12.363
0	2	2	1	1	9.270
1	3	1	0	0	12.708
0	3	1	0	1	4.280
1	3	1	1	0	13.560
1	3	1	1	1	6.359
0	3	2	0	0	7.741
0	3	2	0	1	3.187
3	3	2	1	0	7.939
7	3	2	1	1	4.505
9	2	3	0	0	13.204
11	2	3	0	1	11.282
8	2	3	1	0	13.303
11	2	3	1	1	11.505
2	2	1	0	0	13.441
4	2	1	0	1	7.388
3	2	1	1	0	14.169
1	2	1	1	1	8.975
0	1	3	0	0	15.509
5	1	3	0	1	13.092
4	1	3	1	0	15.691
9	1	3	1	1	13.278
0	1	2	0	0	14.169
1	1	2	0	1	10.779
1	1	2	1	0	14.530
1	1	2	1	1	11.472

5 RESULTS AND DISCUSSIONS

This section contains the results of the calibration for the parameters used in the incident prediction model. It also briefly details the software environment used for calibration of the model. A detailed discussion is provided in Subsection 5.2 on the significance of the parameters used in the model as interpreted from the calibration results. The significance is discussed from a physical standpoint, where the parameters are examined to see how well the model can explain the effects of the parameters in terms of what occurs in traffic stream. Also a statistical discussion of the model parameters is made based on the 'Z' values and "P-value" obtained from the model calibration results. The model calibration result from this research is compared with the results of model calibration obtained by Lee *et al.* (*2*). The result of an alternative reduced category model is presented in Subsection 5.3. The final Subsection 5.4 highlights some of the limitations in the modeling approach undertaken in this research work.

5.1 Model Calibration Results

The input data shown in Table 4.2 were used to calibrate the model parameters. The parameters for all the precursors and the β for exposure given in Equation 2 were determined using the Statistical Package for Social Sciences (SPSS) (*34*). SPSS package is a menu driven software application with several tools for statistical analysis of data.

SPSS also has an analysis tool specifically for general log-linear models, which was used in this research for parameter calibration. In this software, the estimates for the parameters are determined by Maximum Likelihood methodology. The estimated parameters for the proposed model, and the significance of these parameters i.e. values, are presented in the Table 5.1

TABLE 5.1 Results of Parameter Estimation

Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
C	2.693	.832	3.237	.001	1.062	4.324
$\lambda_{CVS=1}$	-1.395	.566	-2.466	.014	-2.504	-.286
$\lambda_{CVS=2}$	-.373	.357	-1.045	.296	-1.071	.326
$\lambda_{CVS=3}$	0(a)	.	.	.	.	.
$\lambda_{Occ=1}$	-2.059	.299	-6.884	.000	-2.646	-1.473
$\lambda_{Occ=2}$	-1.632	.361	-4.522	.000	-2.339	-.924
$\lambda_{Occ=3}$	0(a)	.	.	.	.	.
$\lambda_{P=0}$	-.615	.301	-2.047	.041	-1.205	-.026
$\lambda_{P=1}$	0(a)	.	.	.	.	.
$\lambda_{R=0}$	-.462	.173	-2.670	.008	-.801	-.123
$\lambda_{R=1}$	0(a)	.	.	.	.	.
β	.043	.099	.437	.662	-.151	.237

Note: Estimates marked 'a' are aliased cells

5.2 Discussion of Parameter Estimates

The estimated parameters (λ's) for the model are shown in Table 5.1. CVS and Occ parameters with subscript 1 indicate the lowest level of that precursor, 2 indicates medium level, and 3 indicate highest or most severe level. While subscript '0' for peak-time factor (λ_P) indicates a non peak hour and '1' indicates peak hour. Similarly subscript '0' for roadway type (λ_R) indicates straight section of road without any on-ramps or off-ramps. β is the exponential parameter for exposure, which would have a value zero if the categorical variable selected were perfectly explanatory. C is a constant, which in the present model setup can be interpreted as the maximum crash rate at which incidents are predicted to occur. The information contained in Table 5.1 is useful in analyzing two different aspects of the model parameter estimates. Firstly the effect of different categorical values in a given precursor can be analyzed, and as well the effect of different categorical variables to the extent they influence incident prediction can be analyzed. Secondly the statistical significance of each of the parameters can be assessed. Both of these types of analyses are presented in the following paragraphs of this section.

5.2.1 Physical Interpretation of Estimated Parameters

Any traffic model can be valid, which is to say it can reflect what can be observed in real-time, or which can explain some observed phenomenon. This kind of verification, we term here as physical interpretation of the model. It can be observed in the estimated parameters that one category in each precursor is set to zero value. This category is referred to as aliased cell, and all other estimates will be with reference to the respective precursor's aliased cell. Each precursor is examined here for understanding its physical significance. Parameters for CVS shows that medium and low category have negative sign with decreasing value (real scale) as we move from high to low category. This means that when the traffic state transitions from high category of CVS to medium category, the rate of incident occurrence decreases by an amount. A further transition from medium state to low CVS results in further decrease in the possible rate of incident occurrence. The results very much agree with the general observation one can make out on a highway. Many other studies too have shown results that as CVS increases there is a higher probability of incidents than under lower CVS (*2*). A quantitative measure for the decrease in the crash rate is indicated by the numerical values of the estimated parameters. The model also indicates that the reduction in crash rate is less from high to medium CVS when compared to the reduction in crash rate from medium to low CVS category.

Occupancy also shows a similar trend to CVS. As we move from higher to lower occupancy the rate of crash occurrence decreases. However the calibrated model shows that reduction in probability of crash occurrence is more significant as we move from higher occupancy state to a medium occupancy state than CVS. And the same is true when there is a transition from medium occupancy state to low occupancy state. The results are aligned with expectations. It is more likely we encounter accidents on highways during high occupancy level than during low occupancy, as high occupancy requires high attention from the drivers so as to not involve in accidents. Our model very much reflects this obvious observation. But comparing the values of CVS categories with occupancy categories, it could be noted that occupancy has higher range between maximum and minimum estimates and hence is a stronger precursor to predict incidents than CVS. The range of the estimated λ values between high and low levels is much larger for occupancy than CVS. On the contrast by reducing the occupancy on a section there is a high likelihood that probability in number of crashes can be considerably reduced, more than what we could achieve by controlling speed variation.

From the results we could see that peak-hour indicator and roadway-type indicator have a negative value for the estimated parameters with subscript ‘0’. This means that non-peak hours have less chance of incidents to occur than during peak hours. Also we can state from the results that straight sections of roadway, without on-ramps and off-ramps have lesser likelihood of incident occurrence than curved sections or sections near to ramps. The overall results shows that roadway-type indicator and peak-hour indicator are less sensitive when compared to crash precursors, such as CVS and occupancy. However peak-hour factor and roadway-type are still significant precursors in forecasting incidents, for the developed model.

5.2.2 Statistical Significance of Estimated Parameters

Another useful view of the model results comes from the statistical significance of the parameter estimates. The columns ‘Z’ and ‘Sig.’ in Table 5.1 are the primary indicators to measure the significance of the estimated parameters as was explained in Section 3.2. Looking at the ‘Z’ values every parameter except $\lambda_{CVS\,=\,2}$ and β are statistically significant with ‘Z’ values having magnitude greater than 1.96 at 95% confidence level. Correspondingly the significance (Sig.) also indicates about the same about parameter significance as was seen with the ‘Z’ values. This is based on the criterion that significance value of less than 0.05 declares any parameter used in the model as statistically significant.

In the above context $\lambda_{CVS=2}$ and β are statistically less significant. However $\lambda_{CVS=2}$ is still included in the model, for two reasons. Firstly, reduction in the number of CVS categories would mean lesser number of states to represent traffic flow, and removal of $\lambda_{CVS=2}$ would reduce the total possible states to two-third's of the current possible states. This means that overall model is less sensitive to the change in traffic state to predict the likelihood of accidents, which is not desirable. Secondly, to be consistent with the source model upon which the present model is based. The estimate for β being not significant indicates that β is not different from zero. This is acceptable as the dependent variable; N (number of crashes in two year period) is denominated in terms of exposure of time. In that case it should be noted that the model results cannot be used on any other section of roadway as exposure is not denominated in vehicle miles of travel, but is denominated only by time. Hence the model is not transferable to sections with different traffic flow condition from that of the test bed used for calibration. On the other hand, justification for retaining the exposure parameter β should be made in order for the model to be valid for use in freeway sections other than the test bed. This is because number of accidents by itself gives insufficient information for any logical decision making. As we can see that two roads with similar number of accidents, but first road reported over 2 year period and second road over 1 year period means different in terms of crash rate. In the example second road has higher crash rate or likelihood of incident than the first road. Hence exposure is an important factor in explaining the risk component in incident reporting. As such even though β is statistically not significant, it has a significant role for the physical interpretation of the overall model.

5.2.3 Comparison of Results with Canadian Model

Here comparison in parameter estimates between the model presented above and a similar Canadian model (*2*), with similarly defined categorical values. The model developed for this thesis will be referred to as Texas model. Such a comparison is interesting, as the Texas model is based on similar lines as the Canadian model, but with a little modification and calibrated with completely different data. This could provide insight into the sensitivity of the model to different data conditions. Details of the Canadian model are provided in Section 2.3.

In contrast to the Texas model, Canadian model shows that likelihood of incident occurrence is more sensitive to the level of CVS precursor than to density. The Canadian model indicates that reduction in crash rate between high and medium level of density is minimal, that means to say high and medium level of density pose more or less equal probability of involving in crashes. But low densities pose a considerably smaller crash rate. While the Texas model shows that there is a considerable reduction in crash rate as level of occupancy (substitute for density of Canadian model) reduces from high to medium and also from medium to low compared to Canadian model. Another distinct feature between the two models was that Texas model indicated that crash rate was more sensitive to peak hour indicator than roadway-type indicator, while Canadian model indicated the reverse. Coefficient for the exposure variable has values of the same order (close to zero) in both the models.

The Canadian model uses an additional precursor, the speed difference between longitudinally adjacent detectors (Q), and this precursor is found to be a significant precursor in capturing incident conditions. Lack of this precursor Q in Texas model could have resulted in slightly varied calibration results. Apart from that, data quality can be a significant reason for the observed differences. However, it is difficult to ascertain an exact reason at this level of analysis. But at a broader look, the models seem to be mutually consistent in terms of the effect of the categorical variables on crash rate.

5.3 Alternate Reduced Category Model

It was observed from the model result that was presented in Table 5.1 that level 2 for CVS is statistically not significant. This means that the numerical estimate for $\lambda_{CVS = 2}$ is not significantly different from zero in a statistical sense. This can be interpreted to say that one cannot be sure a reduction in CVS from high to medium would not decrease the crash rate. Because level 2 and 3 for CVS had same effect on crash rate, an attempt was made to combine these two categories of CVS and estimate model parameters for the reduced model.

5.3.1 Calibration for Reduced Model

On a logical basis, for determining the boundary values for the two categories of CVS a 50:50 (low: high) was considered. As explained in Subsection 4.2 of this report, boundary values for the new proportion of CVS were determined using the same historical data available from study site. The boundary values for CVS in the model are given in Table 5.2. The proportion and boundary values for the remaining precursors remained the same as in the previous model.

TABLE 5.2 Boundary Values for 2-Level CVS

Category	CVS
L (50) or 1	<= 0.043
H (50) or 2	> 0.043

Once the boundary values were determined, input data for the model calibration was prepared in the same procedure as explained in Subsection 4.2. Calibration results for the reduced model with new set of data are presented in Table 5.3.

TABLE 5.3 Parameter Estimation for Reduced Model

Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Constant	3.075	.479	6.417	.000	2.136	4.014
$\lambda_{CVS=1}$	-1.243	.477	-2.606	.009	-2.177	-.308
$\lambda_{CVS=2}$	0(a)	.	.	.	.	.
$\lambda_{Occ=1}$	-2.075	.284	-7.310	.000	-2.631	-1.518
$\lambda_{Occ=2}$	-1.492	.320	-4.661	.000	-2.119	-.864
$\lambda_{Occ=3}$	0(a)	.	.	.	.	.
$\lambda_{P=0}$	-.759	.270	-2.810	.005	-1.288	-.230
$\lambda_{P=1}$	0(a)	.	.	.	.	.
$\lambda_{R=0}$	-.443	.171	-2.592	.010	-.779	-.108
$\lambda_{R=1}$	0(a)	.	.	.	.	.
β	.006	.005	1.209	.227	-.004	.015

Note: Estimates marked ‘a’ are aliased cells

5.3.2 Discussion of Results for Reduced Model

It can be seen in Table 5.3 that the estimate for the constant term in the reduced model is higher than in the "full-scale" model (of Section 5.2). This indicates that the maximum crash rate is comparatively higher in the reduced model than in the full-scale model. The physical interpretation of the parameters that was made for the full-scale model holds good for the reduced model too. All parameters, except β are statistically significant here. That leads to the conclusion that β is not significantly different from zero at 95% confidence level.

The overall model fit was examined using the goodness of fit parameter like likelihood ratio. The definition of likelihood ratio was introduced in Section 3. The likelihood ratio for full-scale model and the reduced model is given in Table 5.4. It can be seen from the table that reduced model has a lesser likelihood ratio than the full-scale model. The lesser the likelihood ratio better is the overall model fit. Hence among the two models the reduced model seems to be a better fit for the data used from the test-bed in Texas.

TABLE 5.4 Goodness of Fit

	Likelihood Ratio
Full-scale model	53.95
Reduced Model	14.15

Goodness of fit for the models developed here can also be compared using residual analysis. Advantage of residual analysis is that the influence of each categorical value on the over all model fit can be assessed. The Normal Q-Q plots of adjusted residuals for full-scale model and the reduced model are shown in Figure 5.1 and Figure 5.2 respectively. The adjusted residual is defined as the difference between the estimated value for a precursor category and the mean value for category divided by standard error for the estimate. In the Figure 5.1 and Figure 5.2, closer the adjusted residuals are to the straight line, better is the model fit. It can be noted from the residual analysis that reduced model has better fit than full-scale model.

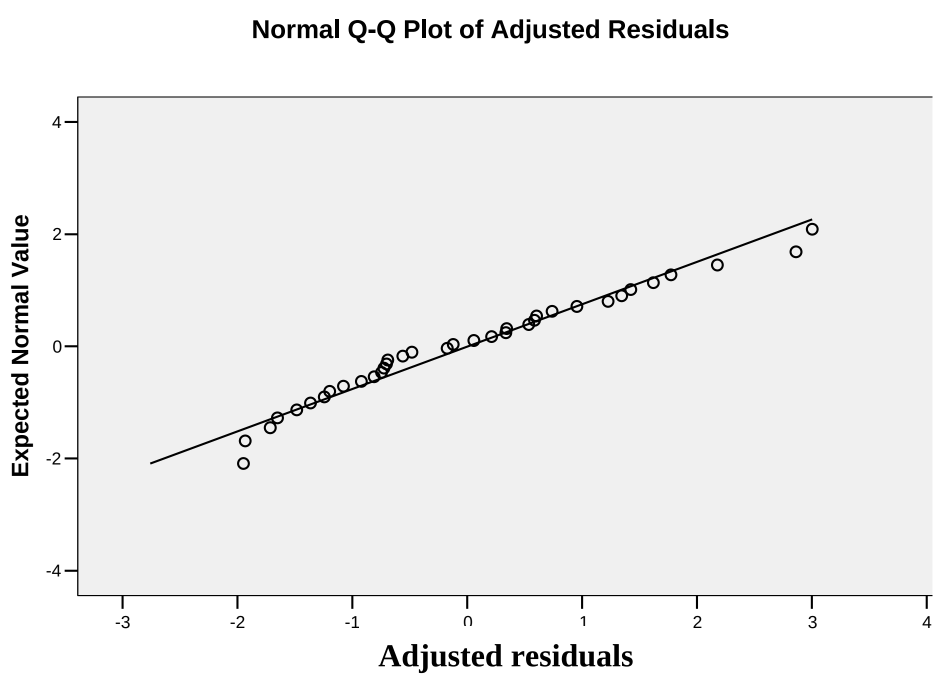

FIGURE 5.1 Residual Analysis for Full-Scale Model

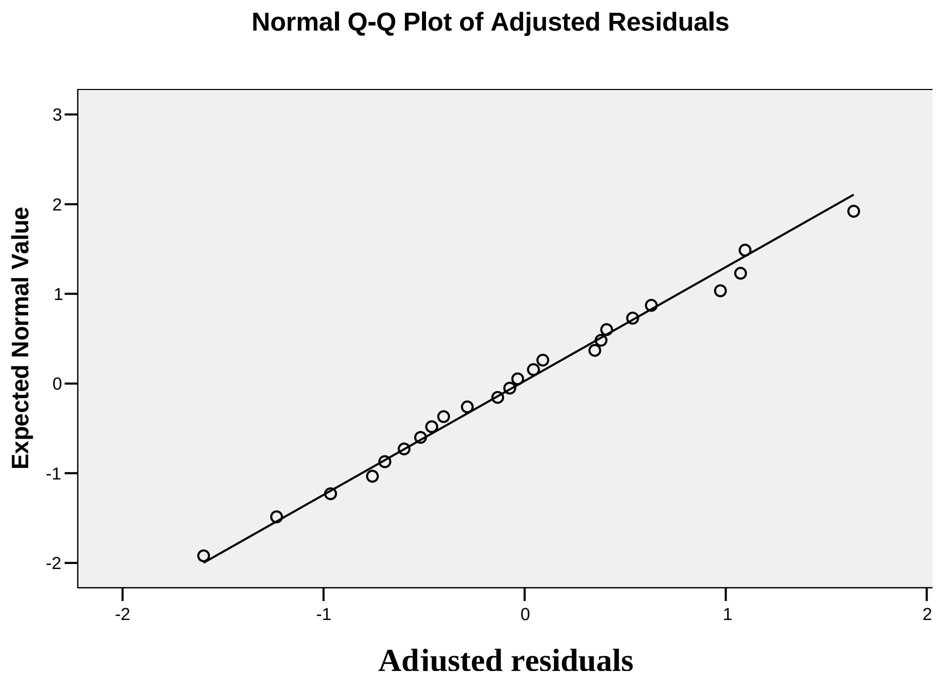

FIGURE 5.2 Residual Analysis for Reduced Model

5.4 Limitations

The model developed in this thesis has some critical components that significantly influence the model parameters. First among those critical components is the study-level aggregation. In the current study, 5 minutes was chosen as the study-aggregation period. However there is no set rule for choosing study-aggregation period. It is usually found that smaller aggregation periods give rise to lot of noise that makes it difficult to distinguish the traffic patterns, while aggregation periods that are too large will be insensitive to changes in traffic within the aggregation period. Hence an optimal study-aggregation period should be selected based on engineering judgment or on an empirical basis. It is also noted that optimal study-aggregation period may be different in different applications. However in the current study there was no systematic effort to identify the optimal study-aggregation period.

A second critical component in the categorical model structure is the definition of appropriate categorical variables and their boundary values. In this thesis, a proportion of 50:30:20 low, medium and high values of precursor were used for the full-scale model. This proportion was based on the recommendation by Lee *et al.*, as this proportion gave the best fir model for their data. However the results of the model in this thesis show that for the data used here, a proportion of 50:30:20 is not necessarily the best choice. However different proportions for category were not experimented with in this thesis, due to time limitations. But certainly experimentation with different categories could yield a better fit model.

It needs no mention that quality of the data is a primary concern in calibrating the incident prediction model used in this thesis. In the process of data reduction, it was noticed that there were several issues with the data obtained from loop detectors. Broadly speaking, some major concerns noted in the loop detector data used for this thesis were missing data, erroneous data and mismatch of data with detectors. Data were missing either for a complete day with no response from a few detectors, due to malfunctioning of detectors, or often some variables were just missing. For example, volumes and occupancy were recorded but there was no speed reported. Erroneous data that were noticed for example were numbers appearing for speed such as ‘88’ or ‘44’ over long periods of time, which appeared erroneous. Other examples of erroneous data were logical errors such as low volumes, low speed, but zero occupancy. Such errors significantly prevailed in the loop-detector data used here. Some of the errors were cleaned in the data reduction, to an extent. For example, detector data files with no response were excluded while processing. Also use of 5-minute study aggregation smoothened the fluctuations in the data values that were noticed due to missing data. Still the quality of data is of importance in order to have confidence in incident prediction models. Dealing with improving the quality of data remains a practical issue in calibrating incident prediction or detection models. Effects of data quality should be borne in mind when interpreting incident prediction models, such as the model described in this thesis, in terms of reliability.

Another limitation of this thesis was that there was no validation done to evaluate the performance of the model. Although it is possible to evaluate the model either with the historical data or in field conditions, neither of these was conducted. Evaluation of the model with historical data was not in scope of this thesis as the model requires at least two years of historical data, which was not available. Field evaluation was deemed impractical within the project time frame and cost constraints. Some possible validations for future study are highlighted in Subsection 6.2.

6 CONCLUSIONS

6.1 Summary

In this thesis a categorical log-linear model was formulated and calibrated to predict likelihood of crashes on freeways. The model formulated here was based similar to model demonstrated by Lee *et al.* (*1, 2*) for Canadian roadway data. Some specific issues that were addressed in this thesis were

- To determine if it was possible to develop satisfactory model for incident prediction
- To examine incident precursors that might be useful for incident prediction
- To calibrate the model with data available on freeway section in Texas
- To interpret the model parameters to understand traffic characteristics during incident and non-incident conditions

A detailed review of literature was conducted to enlist the possible incident precursors that could be used in the model. After a careful examination of the past studies, coefficient of variation in speed (CVS) and density were found to be the most promising precursors, specifically for real-time applications. In the current model CVS and occupancy were used as the precursors considering their capabilities as demonstrated in past studies, and also in view of the ease in which they can be derived from the available loop detector data. Apart from the two precursors, two other indicators were used in the model. These indicators were peak-hour indicator and roadway-type indicator to incorporate the effect of peak hour and roadway geometrics in predicting incidents. Exposure (in vehicle miles of travel) was another explanatory variable used in the model. While CVS, occupancy, peak-hour indicator and roadway-type were categorical variables, exposure was a continuous variable.

Two sections of freeway in Austin, Texas were used as a test bed for this study. Historical 1-minute traffic data from loop detectors and crash reports were used for calibration of the model. Data for years 2003 and 2004 were specifically used. In all 154 collisions were identified over this period in the study site that were used for model calibration. An aggregation time of 5 minutes was used for deriving the precursors from the traffic data. The precursors CVS and occupancy were categorized using 50:30:20 proportion of low, medium and high values respectively for the full scale model. This proportion was used carrying forward the recommendation of Lee *et al.*, as it gave the best fit model in their case. A second model was calibrated with a different proportion of 50:50 for occupancy only, while other specifications remaining same as the full-scale model. Historical data from study site were used to determine the boundary values for the categorical values. The two indicators, each had two categories, peak and non-peak, straight road or other. So in all the full-scale model was designed to allow 36 possible categorical values to define a traffic state at any instant of time. The reduced model was designed to allow 24 possible categorical values.

The model parameters were estimated using Maximum Likelihood Method and were examined for determining their statistical significance as well as their ability to explain observable reality on freeways. The statistical examination indicated that medium and high level of CVS were not distinguishable with the current categorization criteria. The parameter for exposure showed a value close to zero, and also was statistically insignificant. However it was deemed important to retain coefficient for exposure in the model to make meaningful physical interpretation of the model. A detailed explanation of this is given in Section 5.2.2. An alternative model was also tried with a reduced number of categorical values for CVS in response to the observations from statistical inference for the full-scale model. In both the models calibrated for this thesis, all parameter estimates were found to be meaningful in terms of their physical interpretation. Models indicated that if CVS or occupancy level increases, the rate of crash occurrence also increases. It was also observable from model parameters that peak hours and freeway sections with curvature or ramps were prone to high crash rate. All these observations validate the models and strengthen the confidence in ability of the models to predict incidents realistically.

Some important lessons learned from this modeling exercise were to clean the data, if required to ensure good quality data for calibration. Incident prediction models, specifically categorical models are sensitive to aggregation period and the proportions for categorical variables. Hence sufficient care should be taken after examining the data to identify appropriate aggregation period and category boundaries.

Work done in this thesis has reaffirmed that categorical models are a useful tool for incident prediction. More broadly this thesis demonstrates that it is quite possible to estimate the likelihood of incident on freeway sections, provided sufficient attention is paid to data quality and certain model design parameters such as precursors, study aggregation period and proportions for categorical values.

6.2 Future Work

From the lessons learned in this thesis, possible future work that could be extended are identified here and enlisted below.

- In this research only aggregation periods of 5 minutes were used. However a more appropriate aggregation time could be determined by experimenting with different aggregation times. With such an effort one can examine how aggregation time could actually influence the model results.
- Another controlling element in the developed model is the categorical boundaries. It was seen from the initial model results that 50:30:20 proportions was not the most appropriate one for the data used here. Though a reduced category model was tried as an alternative here, several other alternative categories are possible with three levels or two levels. Also, in the model tried here and both CVS and occupancy are categorized with the same proportions. But it could be possible to obtain optimal results by defining categorical values distinctly for CVS and Occupancy.
- An important observation that was made in this research was that expected crashes and exposure have a nonlinear relationship ($\beta << 1$). This topic, whether expected crashes and exposure follow linear or nonlinear relation is been in debate for while among transportation safety professionals (*35, 36*). Some light can be thrown to the debate by doing a controlled experiment, by keeping β constant at different values and examine the relevance of model parameters.

- Currently with the available two-year data, the model could not be validated by splitting the data in two parts, due to insufficient crash samples. But upon availability of sufficient additional data, the developed model can be validated to see how well the model estimates the actual incident conditions. Model can also be tested in field conditions by integrating the model with TxDOT's Advanced Traffic Management Systems.

REFERENCES

1. Lee, C; Saccomanno, F; Hellinga, B. Analysis of Crash Precursors on Instrumented Freeways. In *Transportation Research Record 1784*, TRB, National Research Council, Washington, D. C., 2002, pp 1-8.

2. Lee, C; Saccomanno, F; Hellinga, B. Real-Time Crash Prediction Model for Application to Crash Prevention in Freeway Traffic. In *Transportation Research Record 1840*, TRB, National Research Council, Washington, D.C., 2003, pp 67-77.

3. Top 10 Causes of Death in United States in 2003. In NHTSA Website http://www.nhtsa.dot.gov/, Accessed on 04/25/2006.

4. Dudek, C. L; Messer, C. J; Nuckles, N. B. Incident Detection on Urban Freeways. In *Transportation Research Record 495*, TRB, National Research Council, Washington, D.C., 1974, pp 12-24.

5. Sussman, J. Transitions in the World of Transportation: A Systems View. In *Transportation Quarterly,* Vol.56, No.1, Eno Transportation Foundation Inc., Washington, D.C., 2002, pp 21-32.

6. Oh, C; Oh, J; Ritchie, S. G; Chang, M. Real-Time Estimation of Freeway Accident Likelihood. Working Paper, UCI-ITS-WP-00-21, Institute of Transportation Studies, Irvine, California, December 2000.

7. Kockelman, K. M; Ma, J. Freeway Speeds and Speed Variations Preceding Crashes, Within and Across Lanes. Presented at 14th Annual Meeting and Exposition, ITS America 2004, San Antonio, April 2004.

8. *Freeway Management and Operation Handbook*. Federal Highway Administration, Washington, D.C., September 2003.

9. Robinson, M; Nowak, P. M. An Overview of Freeway Incident Management in the United States. Michigan Department of Transportation, January 1993, www.michigan.gov/documents/freeway_incident_management_8579_7.pdf, Accessed on 07/1/2006.

10. Schrank, D; Lomax, T. The 2002 Urban Mobility Report. Texas Transportation Institute, College Station, June 2002.

11. National Highway Traffic Safety Administration (NHTSA). Motor Vehicle Traffic Crashes as a Leading Cause of Death in the United States, 2002. Traffic Safety Facts Research Note. NHTSA Washington, D.C., 2005.

12. Helman, D. Traffic Incident Management. *Public Roads,* Vol. 68, No. 3, December 2004, pp 14-21

13. Federal Highway Administration. *Traffic Incident Management Handbook*, FHWA, Washington, D.C., November 2000.

14. Parkany, E; Xie, C. A Complete Review of Incident Detection Algorithms and Their Deployment: What Works and What Doesn't. Research Report NETCR 37, University of Massachusetts, Amherst, February 2005.

15. Dudek, C L; Messer, C. J; Nuckles, N B. Incident Detection on Urban Freeways. In *Transportation Research Record, 495*, TRB, National Research Council, Washington, D. C., 1974, pp 12-24.

16. Levin, M; Krause, G. M. Incident Detection: a Bayesian Approach. In *Transportation Research Record 682*, TRB, National Research Council, Washington, D. C., 1978, pp 52-58.

17. Tsai, J; Case, E. R. Development of Freeway Incident Detection Algorithms by Using Pattern Recognition Techniques. In *Transportation Research Record 722,* TRB, National Research Council, Washington, D. C., 1979, pp 113-116

18. Ahmed, M. S; Cook, A. R. Application of Time-Series Analysis Techniques to Freeway Incident Detection. In *Transportation Research Record 841,* TRB, National Research Council, Washington, D. C., 1982, pp 19-21.

19. Balke, K. An Evaluation of Existing Incident Detection Algorithms. Research Report 1232-20, Texas Transportation Institute, College Station, November 1993.

20. Adeli, H; Samant, A. An Adaptive Conjugate Gradient Neural Network-Wavelet Model for Traffic Incident Detection. In *Computer-Aided Civil and Infrastructure Engineering,* Vol. 15, No. 4, 2000, pp 251-160

21. Ivan, J. N. Neural Network Representations for Arterial Street Incident Detection Data Fusion. *Transportation Research Part C,* Vol. 5, No. 3, 1997, pp 245-254.

22. Hsiao, C. H; Lin, C. T; Cassidy, M. Application of Fuzzy Logic and Neural Networks to Automatically Detect Freeway Incidents. *Journal of Transportation Engineering,* Vol. 120, No. 5, 1994, pp 753-772.

23. Michalopoulos, P G. Vehicle Detection Video through Image Processing: The Autoscope System. *IEEE Transactions on Vehicular Technology,* Vol. 40, No. 1, 1991, pp 21-29.

24. Balke, K; Dudek, C L; Mountain, C E. Using Probe-Measured Travel Time to Detect Major Freeway Incidents in Houston, Texas. In *Transportation Research Record 1554*, TRB, National Research Council, Washington, D.C., 1996, pp 213-220.

25. Niver, E; Mouskos, K. C; Batz, T; Dwyer, P. Evaluation of TRANSCOM's System for Managing Incidents and Traffic (TRANSMIT). *IEEE Transactions on Intelligent Transportation Systems,* Vol. 1, No. 1, 2000, pp 15-31.

26. Petty, K. F; Skabardonis, A; Varaiya, P. P. Incident Detection with Probe Vehicles: Performance, Infrastructure Requirements and Fesibility. *Proceeding of Transportation Systems* 1997, Chania, Greece, 1997, pp 125-130.

27. Sethi, V; Bhandari, N; Koppelman, F. S; Schofer, J. L. Arterial Incident Detection Using Fixed Detector and Probe Vehicle Data. *Transportation Research Part C,* Vol. 3, No. 2, 1995, pp 99-112.

28. Pande, A; Abdel-Aty, M; Hsia, L. Spatio-Temporal Variation of Risk Preceding Crash Occurrence on Freeways. Transportation Research Board 84th Annual Meeting, TRB, Washington, D.C., 2005.

29. Ishak, S; Alecsandru, C. Analysis of Freeway Pre-Incident, Post-Incident and Non-Incident Conditions Using Second Order Spatio-Temporal Traffic Performance Measures. Transportation Research Board 84th Annual Meeting, TRB, Washington, D.C., 2005.

30. Persaud, B; Dzbik, L. Accident Prediction Models for Freeways. In *Transportation Research Record 1401*, National Research Council, Washington D. C, 1993, pp 55-60.

31. Ceder, A; Livneh, M. Relationships Between Road Accidents and Hourly Traffic Volume –I. *Accident Analysis and Prevention, Vol. 14*, 1982, pp 19-34.

32. Agresti, A. *An Introduction to Categorical Data Analysis*. John Wiley & Sons Inc., New York, 1996.

33. *SAS OnlineDoc, Version 8*. SAS Institute Inc., Cary, North Carolina, 1999, http://v8doc.sas.com/sashtml, Accessed on 06/20/2005.

34. *SPSS 13.0 Brief Guide*, SPSS Inc, Chicago, 2004.

35. Lord, D; Manar A; Vizioli A. Modeling Crash-Flow-Density and Crash-Flow-V/C Ratio Relationships for Rural and Urban Freeway Segments. *Accident Analysis and Prevention,* Vol. 37, 2005, pp 185-199.

36. Lord D. Issues Related to the Application of Accident Prediction Models for the Computation of Accident Risk on Transportation Networks. Presented at 81st Annual Meeting of the Transportation Research Board, Washington, D.C., 2001.

VITA

Name:	Shamanth Kuchangi
Address:	Texas Transportation Institute 701 N. Post Oak, Suite 430 Houston, TX 77024-3827
Email Address:	s-kuchangi@ttimail.tamu.edu
Education:	M.S., Civil Engineering Texas A&M University, College Station, Texas, 2006
	B. Tech., Civil Engineering, Regional Engineering College Warangal, India, 2001