AN INTEGRATED MULTI-LEVEL METHODOLOGY TO MITIGATE
SHORT AND LONG TERM EFFECTS OF INCIDENTS

by

SAMİ DEMİROLUK

A Dissertation submitted to the
Graduate School-New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
Graduate Program in Civil & Environmental Engineering
written under the direction of
Kaan Ozbay
and approved by

________________________________________

________________________________________

________________________________________

________________________________________

New Brunswick, New Jersey
October, 2014
ABSTRACT OF THE DISSERTATION

An Incident Management Methodology to Mitigate Short and Long Term Effects of Incidents

by SAMI DEMIROLUK

Dissertation Director: Kaan Ozbay

Incidents are random, yet disruptive events that occur frequently on our roadways. Incidents also have a major impact on the other motorists, the environment and the economy. For example, as a result of incidents, the motorist can experience delays and delays may cause higher emission rates. Hence, without proper incident management strategies lives can be at risk and the motorist might experience long delays. Therefore, there is a need for better incident management strategies for improving the safety of the motorists, reduce congestion thus improving the productivity of our economy.

In this thesis an integrated approach was used to mitigate the effects of incidents. The proposed had two components: (i) a real-time incident duration model designed to improve decision making during incident management operations, (ii) a novel statistical model for mapping incidents and their severity in space to detect the segments with higher accident risk with the goal of better planning for and responding to future incidents. The novelty of this approach is that it offers solutions for mitigating both short term and long term effects of incidents using a Bayesian framework. This probabilistic
approach enables the representation of the probabilistic range of results rather than point estimates.

Bayesian networks are used for predicting incident duration. The performance of the model was subsequently improved by introducing adaptive learning techniques to the model.

A set of spatial models was estimated and risk maps were developed to investigate the locations with higher crash risk in order to facilitate incident management planning. For purposes of estimating the predictive power of the models, spatial data with different resolutions were used.

The overall findings of this dissertation indicate that the integrated approach can be used for estimating incident duration in the long run as it adapts itself to minor and major changes to the system. It can also be used for effectively pinpointing the locations with higher crash risk since, unlike existing models, the model do not ignore the continuous nature of roadways.
Acknowledgements

I would like to express my deepest gratitude to my advisor Dr. Kaan Ozbay for his support and guidance over the course of my Ph.D. study. I would also like to thank my thesis committee members Dr. Hani Nassif, Dr. Eric Gonzales and Dr. Hediye Tuydes for their valuable comments that helped me shape this dissertation.

I would like to thank my friends and colleagues at Rutgers Intelligent Transportation Systems Laboratory starting with Dr. Bekir Bartin who provided very valuable data for this dissertation. I would also like to thank Eren Erman Ozguven, Anil Yazici, Ender Morgul, Mehmet Yildirimoglu for their support throughout my studies.

I would like to thank my friend and roommate Ilker Dastan for his friendship and support. I would also like to thank my other friends Ufuk Ates, Cem Iyigun, Ozlem Cavus, Kemal Karakayali and Mustafa Karakayali. I have felt so blessed to be surrounded by such great friends who made this journey enjoyable. I would also like to thank Luciana Pereira for her companionship and support.

Finally, I would like to thank to my family from bottom of my heart for their unconditional love and endless support in my studies. If it was not for them, I would be never in the place where I am today.
Dedication

To my parents, Nuran and Alper Demiroluk…
Preface

The work conducted in this dissertation has been presented and published in several conferences and journals. Below is the list of publication derived from this dissertation with corresponding chapter numbers.

Chapter-4

- Demiroluk and Ozbay (2014). Adaptive Learning in Bayesian Networks for Incident Duration Prediction. 93rd Annual Meeting of Transportation Research Board.

Chapter-5

Table of Contents

ABSTRACT OF THE DISSERTATION ................................................................. ii
Acknowledgements ...................................................................................... iv
Dedication ........................................................................................................ v
Preface .............................................................................................................. vi
CHAPTER 1. INTRODUCTION ................................................................. 1
  1.1 Background and Motivation ................................................................. 1
  1.2 Problem Definition .............................................................................. 2
  1.3 Contributions of This Dissertation ...................................................... 8
  1.4 Dissertation Outline ............................................................................. 8
CHAPTER 2. LITERATURE REVIEW .................................................. 11
  2.1 Models for Estimating Incident Duration ............................................. 11
    2.1.1 Regression Models for Incident Duration ...................................... 11
    2.1.2 Hierarchical Models for Incident Duration .................................... 14
    2.1.3 Bayesian Network Models ............................................................. 17
  2.2 Spatial Models for Analyzing and Mapping the Locations with Higher Accident Risk .............................................................................................................. 19
CHAPTER 3. POTENTIAL IMPROVEMENTS IN INCIDENT MANAGEMENT 30
  3.1 Incident Management .......................................................................... 30
  3.2 Need for Better Incident Management Planning and Effective Operations .... 33
  3.3 Conceptual Framework of the Methodology .......................................... 38
  3.4 Proposed Incident Management Improvements ...................................... 47
CHAPTER 4. BAYESIAN NETWORKS AND ADAPTIVE LEARNING FOR PREDICTING INCIDENT DURATION ................................................................. 51
  4.1 Machine Learning Methods ............................................................. 51
    4.1.1 Bayesian Networks ....................................................................... 51
    4.1.2 Structure Learning Algorithms for Automatically Building Bayesian Networks .............................................................................................................. 53
    4.1.3 Connection Strength of Variables in a BN ...................................... 56
    4.1.4 Parameter Estimation for Adaptive Learning in Bayesian Networks .... 57
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.5 Feature Selection</td>
<td>58</td>
</tr>
<tr>
<td>4.1.6 Tools for Building Bayesian Networks</td>
<td>63</td>
</tr>
<tr>
<td>4.2 Methodology</td>
<td>64</td>
</tr>
<tr>
<td>4.2.1 Structure Learning for the Estimation of Incident Duration Prediction Models</td>
<td>64</td>
</tr>
<tr>
<td>4.2.1.1 Description of Data</td>
<td>64</td>
</tr>
<tr>
<td>4.2.1.2 Modeling Results</td>
<td>66</td>
</tr>
<tr>
<td>4.2.1.3 Connection Strength of Variables in the Network</td>
<td>73</td>
</tr>
<tr>
<td>4.2.1.4 Model Performance with Missing Data</td>
<td>74</td>
</tr>
<tr>
<td>4.2.1.5 A Step-by-step Incident Duration Estimation Scenario for Incident Response</td>
<td>77</td>
</tr>
<tr>
<td>4.2.1.6 Feature Selection for Incident Variables</td>
<td>82</td>
</tr>
<tr>
<td>4.2.2 Adaptive Learning in Bayesian Networks for Duration Prediction</td>
<td>86</td>
</tr>
<tr>
<td>4.2.2.1 Data</td>
<td>86</td>
</tr>
<tr>
<td>4.2.2.2 Model Estimation and Validation</td>
<td>87</td>
</tr>
<tr>
<td>4.2.2.3 Real-time Prediction Performance and Adaptive Learning Using 2011 Data</td>
<td>91</td>
</tr>
<tr>
<td>4.2.3 A Complete Adaptive Strategy for Estimating Incident Duration</td>
<td>94</td>
</tr>
<tr>
<td>4.2.3.1 Data</td>
<td>95</td>
</tr>
<tr>
<td>4.2.3.2 Model</td>
<td>97</td>
</tr>
<tr>
<td>5.1 A County Level Model</td>
<td>102</td>
</tr>
<tr>
<td>5.1.1 Data</td>
<td>102</td>
</tr>
<tr>
<td>5.1.2 Model</td>
<td>106</td>
</tr>
<tr>
<td>5.1.3 Discussion of Results</td>
<td>110</td>
</tr>
<tr>
<td>5.2 Bayesian Spatial Modeling and Risk Mapping of Downed Trees along the Roadways Using Data from Hurricanes Irene and Sandy</td>
<td>120</td>
</tr>
<tr>
<td>5.2.1 Data</td>
<td>120</td>
</tr>
<tr>
<td>5.2.2 Zero Inflated Model</td>
<td>124</td>
</tr>
<tr>
<td>5.2.3 Model Selection</td>
<td>127</td>
</tr>
</tbody>
</table>
5.2.4 Results and Discussion .................................................................130
5.3 A Doubly Stochastic Point Process Model for Crashes along a Corridor ....135
  5.3.1 Data ...........................................................................................136
  5.3.2 Modeling .....................................................................................138
  5.3.2.1 Log-Gaussian Cox Model .........................................................138
  5.3.2.2 Alternative Link Level Model: Poisson MCMC ......................142
  5.3.3 Modeling Results and Discussion ...............................................144

CHAPTER 6. CONCLUSIONS AND FUTURE RESEARCH .....................152
  6.1 Conclusions on Bayesian Networks and Adaptive Learning for Predicting
       Incident Duration ............................................................................152
  6.2 Conclusions on Spatial Models for Analyzing and Mapping the Locations with
       Higher Crash Risks for Incident Management ................................155
  6.3 Practical Implementations ...............................................................158
    6.3.1 Incident Duration Estimation and Its Use for Real-world Operations.....158
    6.3.2 Spatial Distribution of Incidents and Allocation of Resources ..........159
    6.3.3 Use of the Proposed Methodology for the Performance Evaluation of
           Incident Management Operations ................................................161
  6.4 Future Research ...........................................................................163

REFERENCES ..........................................................................................165
List of Tables

Table 3.1 Strategies and Technologies for Incident Management (Ozbay and Kachroo, 1999) ................................................................. 34
Table 4.1 Comparison of different machine learning algorithms ......................... 55
Table 4.2 Description of the Variables ................................................................ 65
Table 4.3 BIC scores of Prediction Models .......................................................... 71
Table 4.4 Conditionally Dependent Parameters of Each Parameter in K2 Model ..... 72
Table 4.5 Connection Strength between the variables in K2 model in terms of MI% .... 74
Table 4.6 Performance of the BN model with randomly missing data .................... 75
Table 4.7 Variables in ESP and NJ Turnpike accident databases .......................... 83
Table 4.8 Variable Rankings across Datasets (Top-5 Variables are Marked) ......... 84
Table 4.9 Description of the variables ................................................................. 87
Table 4.10 Summary statistics of the incident durations ....................................... 87
Table 4.11 BIC scores and accuracies of Prediction Models ................................. 89
Table 4.12 Performance of the model for 2011 data with adaptive learning .......... 94
Table 4.13 Summary statistics of the incident data .............................................. 96
Table 4.14 Rank of the features .......................................................................... 97
Table 4.15 Model states (Castillo and Gama, 2009) .............................................. 100
Table 5.1 Descriptive statistics of the variables included in the study .................. 106
Table 5.2 Modeling results of hierarchical model for raw crash counts .............. 112
Table 5.3 Modeling results of hierarchical model for crashes by jurisdiction ........ 114
Table 5.4 Modeling results based on severity weighted crash counts ................. 116
Table 5.5 Summary of the hurricane data ......................................................... 123
Table 5.6 Deviance of models with different combination of the variables .......... 128
Table 5.7 Effect of spatial and temporal terms on model fit .............................. 130
Table 5.8 Estimated posterior density of parameters for the selected model ........ 130
Table 5.9 Descriptive statistics of the variables .................................................. 138
Table 5.10 The results of LGCP model .............................................................. 146
Table 5.11 The results of link level MCMC model ............................................ 148
List of Illustrations

Figure 1.1 A real-time incident management framework with integrated duration prediction model ................................................................................................................. 4

Figure 1.2 A data-driven incident management planning model ........................................ 6

Figure 2.1 Decision tree for incident clearance time prediction (Ozbay and Kachroo, 1999) ................................................................................................................................. 15

Figure 2.2 Classification tree (Smith and Smith, 2001) .................................................... 16

Figure 2.3 BN structure proposed by Ozbay and Noyan (2006) ..................................... 18

Figure 3.1 Steps of incident management ........................................................................ 30

Figure 3.2 New Jersey 2010 accidents based on (a) frequency and (b) severity weights . 37

Figure 3.3 Conceptual structure of the proposed methodology ........................................ 38

Figure 3.4 Example BN .................................................................................................... 52

Figure 4.1 BN structures from (a) Naïve Bayes, (b) TAN and (c) K2 ............................. 68

Figure 4.2 CART Tree for New Jersey data ..................................................................... 76

Figure 4.3 A step-by-step incident duration estimation scenario ..................................... 81

Figure 4.4 CMIM Scores of Variables for (a) NJ Turnpike data, (b) ESP data................. 84

Figure 4.5 Candidate BN structures from (a) Naïve Bayes, (b) TAN and (c) K2 .......... 89

Figure 4.6 Adaptive Learning Mechanism in the BN model ............................................ 93

Figure 4.7 Percent increase in prediction accuracy of the model with adaptive learning for monthly and quarterly 2011 data ...................................................................................... 93

Figure 4.8 Change in the performance of Naïve Bayes by eliminating the features in the model one-by-one ............................................................................................................. 97

Figure 4.9 Adaptive modeling framework for Bayesian Networks ................................... 99

Figure 4.10 Model performance of Adaptive learning framework vs. static models ..... 100

Figure 5.1 (a) Annual crash counts by county in 2010 (b) Crash rates by county in 2010 (per thousand DVMT) ................................................................................................................. 104

Figure 5.2 New Jersey road network (except surface streets) ........................................... 104

Figure 5.3 Graphical representation of the hierarchical Bayesian crash model .......... 109

Figure 5.4 Mean crash rates from (a) raw, (b) from severity weighted crash counts ..... 117
Figure 5.5 Mean crash rates for (a) state (b) authority (six counties are excluded) (c) county (d) municipal roadways by county in 2010 (per thousand DVMT) .......................... 119
Figure 5.6 Municipal locations of downed tree incidents during (a) Hurricane Irene and (b) Hurricane Sandy ........................................................................................................ 122
Figure 5.7 (a) New Jersey roadway network (except surface roads) (b) Roadway density at municipality level ........................................................................................................ 124
Figure 5.8 Predicted mean of (a) the number of downed trees (b) the number of downed trees normalized by roadway density during Hurricane Sandy ................. 133
Figure 5.9 Predicted mean of downed trees normalized by roadway density during Hurricane Sandy in case of (a) ten percent and (b) twenty percent increase in maximum wind speed ...................................................................................................................... 134
Figure 5.10 The marginal posterior distribution of spatial dependence ....................... 147
Figure 5.11 Point level (left) and link level (right) crash risks in 2011 ......................... 150
Figure 5.12 Crash risks from LGCP model (top) and Poisson MCMC model (bottom) between mileposts 0 and 13.5 .................................................................................................................. 151
CHAPTER 1. INTRODUCTION

1.1 Background and Motivation

Incidents, such as accidents, disabled vehicles and other unplanned events, contribute significantly to traffic congestion and delays in urban freeways. If incidents are not responded to and cleared in a timely manner, secondary crashes may also occur. To minimize the effects of incident-induced congestion namely, non-recurring congestion and to prevent secondary accidents, an effective incident management tool which can reliably predict incident durations in real-time is necessary. Real-time duration predictions can provide a better basis for travelers for decision-making (e.g. route diversion) as well as for reducing uncertainty (Khattak et al., 1995). Therefore, the development of an incident duration model which permits users to obtain reliable estimates of incident duration (even if partial information is available about the incident) is of utmost importance. Having an accurate idea of incident duration will facilitate better management of traffic conditions during the incident. However, reducing incident duration also requires a different and more sophisticated long term approach that requires a better understanding of the spatial and temporal characteristics of incidents. This was the second major goal of this dissertation which will attempt to propose spatio-temporal accident frequency and severity models that can be employed to improve incident management operations in a network setting.
1.2 Problem Definition

Incidents are random, yet disruptive events on our roadways. Every year many lives are lost as a result of accidents. According to the Center for Disease Control and Prevention, motor vehicle crashes are the leading cause of death among young people age 5 to 34 in the United States (CDC, 2012). Moreover, in 2010, 2,764,122 people were injured as a result of motor vehicles crashes in 2010 (CDC, 2012). Incidents also have major impact on other motorists, the environment and the economy. For example, as a result of incidents, the motorist can experience delays and delays may cause higher emission rates. Hence, without proper incident management strategies lives can be at risk and the motorist might experience long delays. Therefore, there is a need for better incident management strategies for improving the safety of the motorists, reducing congestion and improving the productivity of our economy.

The incident management methodology employed in this study consisted of two components:

1. An incident duration model to improve decision making during incident management operations,

2. A novel statistical model for mapping incidents and their severity in space in order to detect segments with higher accident risk with the goal of better able to plan for and respond to future incidents.
**Real-Time Duration Prediction**

To undertake the first part of the study methodology, it was essential to develop a practical model capable of automatically predicting incident durations in real-time which might be incorporated into incident management tools of Traffic Management Centers (TMC) to improve decision-making process as regards more effective incident response. Figure 1.1 shows the flow of the real-time incident management framework that was used. In this framework, first, an incident occurs and it is either detected by sensors or reported by travelers after its occurrence. At this stage, information about the details of the incident may be very limited and might only include information on location and number of vehicles involved in the incident. Even though the information may be limited, a Bayesian network-based incident duration prediction model can be utilized to provide predictions about expected duration with this limited information. Based on this first prediction based on limited information, an initial response strategy can be determined and emergency vehicles can be dispatched. Later, when additional information is received from emergency teams and/or additional real-time reports are forthcoming, the inputs to the prediction model can be updated to obtain a new and more accurate prediction of outcomes. A more appropriate response strategy can then be evaluated and necessary changes can be made.
One of the major challenges in making real-time estimations of incident durations is that the initial information acquired about an incident is generally very limited. Moreover, there is no guarantee that the information about different incident parameters will be received in a particular sequence, which is the basic assumption of most of traditional prediction models such as decision trees and linear regression models. These models also try to predict the exact duration of incidents and are insensitive to the stochastic nature of the problem; hence, they do not provide an opportunity to understand the time dependent variation of these predictions.

Since no assumption on the sequence of information flow is required, situations in which the data is incomplete can be handled effectively by Bayesian Networks (BNs). Moreover, BNs can be used to compute incident duration predictions as possible outcomes instead of fixed results. This way, variation in the results can be observed and
more informed decisions can be made for incident response. Unlike Artificial Neural Network (ANN) models that emerge as a black box, all the parameters in BNs have an understandable representation. Therefore, BNs can be constructed directly using expert knowledge, without any time-consuming learning process. Alternatively, machine learning techniques can be utilized to construct BNs from sample data if relationships between the parameters are not well understood or an expert knowledge is not available.

Incident duration prediction models are often developed using a limited dataset and stationary. It is not practical to use such models for predicting future incidents in the long run, however, since traffic networks are dynamic and many features such as demand and roadway conditions are subject to unforeseen changes. Hence, the proposed duration model also required a mechanism for purposes of being able to adapt to future conditions.

**Data-Driven Short and Mid-term Planning of Efficient Incident Management Strategies**

For the second part of the dissertation, a statistical model to represent the spatial distribution of the frequency and severity of incidents along the roadway was developed. There is little research in the literature to estimate models for “spatial prediction of incidents”. However, these models are necessary for the development of long term incident management strategies such as re-allocation of resources for effective incident detection and response. If such strategies can be implemented successfully, it is possible to decrease both the incident detection and incident response times, which are two critical stages contributing to the overall incident duration. Figure 1.2 shows the proposed planning model for the development of better incident management strategies. In this
model, first, the statistical spatial incident prediction model was fed by the incident data. As an output of the spatial model, predictive incident heat maps which show the distribution incident risk along the roadway were generated. Based on these maps, the locations that had higher incident risks could be identified and then these locations could be given priority in terms of resources. For example, incident response teams can be repositioned to the location and/or the area might be continuously monitored.

Figure 1.2 A data-driven incident management planning model

Although, there is some research on resource allocation for incident response, in these studies, optimization techniques are used for minimizing incident response time by re-allocating resources at different locations prone to incidents. The roadway is either segmented (Zografos et al., 2002) or pre-determined fleet depot locations are used for this purpose (Ozbay et al., 2004; Pal and Bose, 2009; Ozbay et al., 2012). Hence, they cannot provide a robust spatial representation of higher accident risk areas and their impact zones as proposed in this study.
Mapping is a very popular research technique that is widely used in examining disease patterns and others because “mapping transforms spatial data into a visual form, enhancing the ability of users to observe, conceptualize, validate, and communicate information” (Miaou et al., 2003). After the introduction of Markov Chain Monte Carlo (MCMC) methods, hierarchical Bayesian models are employed for developing risk maps of diseases in the context of public health studies (Waller and Gotway, 2004). Additionally, these models are used for ranking disease risks across regions based on spatial and temporal variations. It is possible to make an analogy between diseases and crashes since crash risks vary in different regions and there is need for ranking the vulnerability of these regions, as well. Therefore, mapping spatial crash data has the potential to provide a better understanding of crash frequency by modeling and visualizing crash frequency estimates and their contributing factors.

However, traditional crash frequency modeling methods cannot effectively handle spatial crash data. Hierarchical Bayesian models in contrast are proven to handle well global heterogeneity as well as spatial correlation which exist in area-wide crashes (Miaou et al, 2003; MacNab, 2004, Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Huang et al., 2010). Currently, there are few documented studies that use this technique to estimate spatial crash models, even though there are several benefits to utilizing these methods in the transportation field. However, these models generally deal with crash frequency (mapping) at the aggregate level such as county level. There are limited examples which try to model the crashes at intersection or link level (point level). Since crashes continuously occur on roadways, there is also a need for spatial models that can analyze crashes continuously using disaggregate crash data.
1.3 Contributions of This Dissertation

The major contributions of the dissertation can be summarized as follows:

- Compare different structure learning algorithms for selecting the best BN for incident duration prediction. Then use the concept of link strength to analyze the impacts of incident variables in BN-based duration prediction model.

- Propose an adaptive learning scheme by updating model parameters for adapting the BN model for minor changes in the roadways system. Apply the adaptive learning framework proposed by Castillo and Gama (2009), which can adapt both minor and major changes in a system, for BN-based incident prediction.

- Apply severity weights to the crash counts and estimate spatial crash frequency models for different road types using (aggregate) county level data.

- Analyze downed tree incidents during hurricanes using municipality level and estimate spatial incident risk for these incidents.

- Estimate a continuous network level spatial crash risk model using Log-Gaussian Cox (LGCP) model (Moller et al., 1998) and point level data.

1.4 Dissertation Outline

Chapter 2, provides a focus on the current literature detailing both incident duration prediction models and spatial models for mapping crash risk. The most
important studies that have shaped the literature are reviewed in this chapter. The models in these studies are discussed in detail and their advantages and weaknesses are explained.

Chapter 3 discusses the concept of incident management. The strategies and technologies that can help to improve incident management are presented. Then, how use of the proposed methodology would impact the incident management planning and operations is discussed.

Chapter 4 presents first part of the methodology that proposes novel BN models for predicting the incident duration. First, different structure learning algorithms were used for automatically building BNs and the best one was chosen based on the performance metrics. Then, an adaptive learning scheme was introduced in the model in which the parameters to adapt minor changes in the system. Finally, the use of a complete adaptive framework for BNs, that proposed by Castillo and Gama (2009), is highlighted as a method for accurately predicting incident duration.

Chapter 5 focuses on the development of spatial models and risk maps for identification of the locations with higher crash risk. Hierarchical Bayesian modeling is used in this chapter for its flexibility in modeling spatial effects. Moreover, Bayesian modeling also makes it easy to perform a best/worst case scenario analysis. Crash data with different spatial resolutions are used in this chapter. First, a county level model, which is the common practice, and risk maps are developed. Then, a spatial model for the incidents during special events is developed. Downed tree incidents on the roadways during two hurricanes, Irene and Sandy, are considered for this purpose with finer
resolution spatial data (municipality level). Finally, a point level model spatial model is developed for a roadway in New Jersey using Cox Model.

Chapter 6 presents the study’s conclusions, and contributions and future tasks to advance the field are discussed.
CHAPTER 2. LITERATURE REVIEW

2.1 Models for Estimating Incident Duration

2.1.1 Regression Models for Incident Duration

One of the earliest models that recognize the dynamic nature of incident prediction was proposed by Khattak et al. (1994). They developed a series of truncated regression models to predict incident duration based on the sequential acquisition of incident information and their decay over the life of an incident. The model when the incident information is first acquired is called a Stage model:

\[ Y = X \beta + \varepsilon \]  
(2.1)

\( Y \) = Vector of \( n \) dependent variable observations on incident duration

\( X \) = Matrix of \( p \) independent variables available at Stage 0

\( \beta \) = Vector of parameters at Stage 0

\( \varepsilon \) = The error term with expected value 0 and variance \( \sigma^2 \)

The above model is valid until some predetermined truncation point, \( \tau_0 \) is reached. Due to the incomplete nature of the information, bias may exist in the model. However, the exclusion of unavailable variables may also reduce multicollinearity in the data.

At Stage 1, more information is received during \( \tau_1 \) or some incidents would cease to exist. The incidents that no longer exist can be assumed to be unobserved and the truncation point can be moved to \( \tau_0 + \tau_1 \). At this stage, the number of observations will
be decreased from \( n \) to \( n-v_1 \) as \( v_1 \) incidents have ceased to exist during the interval \( \tau_1 \).

The model at this point can be estimated from the following equations:

\[
y_i = \beta'x_i + \varepsilon_i \quad > \quad \tau_0 + \tau_1 \quad \text{are included and}
\]

\[
y_i = \beta'x_i + \varepsilon_i \quad \leq \quad \tau_0 + \tau_1 \quad \text{are excluded.}
\]

\( Y = \) Vector of \( n-v_1 \) incident duration observations

\( X = \) Matrix of \( p_1 \) independent variables at Stage 1, \( p_1 = (p_0 - d_1) + a_1 \), where \( a_1 \) denotes the new variables acquired during \( \tau_1 \); \( d_1 \) stands for the decayed information during \( \tau_1 \)

\( \beta = \) Vector of parameters at Stage 1

\( \varepsilon = \) The error term

At Stage \( w \), the model can be estimated as:

\[
y_i = \beta'x_i + \varepsilon_i \quad > \quad \tau_0 + \tau_1 + \ldots + \tau_w \quad \text{are included and}
\]

\[
y_i = \beta'x_i + \varepsilon_i \quad \leq \quad \tau_0 + \tau_1 + \ldots + \tau_w \quad \text{are excluded.}
\]

\( Y = \) Vector of \( n-v_w \) incident duration observations

\( X = \) Matrix of \( p_w \) independent variables at Stage 1, \( p_w = (p_{w-1} - d_w) + a_w \), where \( a_w \) denotes the new variables acquired during \( \tau_w \); \( d_w \) stands for the decayed information during \( \tau_w \)

\( \beta = \) Vector of parameters at Stage 1

\( \varepsilon = \) The error term

These models can predict incident duration at subsequent stages of an incident and can update its value by the addition and/or deletion of explanatory variables of the
incident. By updating the model based on significant factors, they were able to make more accurate incident durations predictions as an incident progressed.

Later, Garib et al. (1997) also employed regression models in their stepwise methodology to estimate incident delays and incident duration. Their five-step modeling approach can be summarized as follows:

- To regress the cumulative incident delay (dependent variable) with all explanatory variables to examine their individual effects
- To employ a step-wise procedure to investigate significant variables
- To regress the dependent variable with all possible combinations of independent variables to find the best functional form
- To introduce each two-factor interaction term
- To employ the step-wise procedure using variables from steps 2-4.

Finally, two models were proposed in the paper as a result undertaking the above steps. Their first model showed that 74% of the variations in incident delay can be estimated using number of lanes affected, number of vehicles involved, incident duration, and traffic demand. The second model described 85% of the delays with the same variables. They also proposed an incident duration model using the above steps, which yielded the following equation:

\[
\log(\text{Duration}) = 0.87 + 0.027X_1X_2 + 0.2X_5 - 0.17X_6 + 0.68X_7 - 0.24X_8
\]  

(2.2)

where

\(\text{Duration} = \) incident duration in minutes

\(X_i = \) number of lanes affected by the incident
\[ X_2 = \text{number of vehicles involved in the incident} \]

\[ X_5 = \text{dummy variable for truck involvement} \]

\[ X_6 = \text{dummy variable for time of day (0 if morning peak (6:30AM-9:30AM) and 1 if afternoon peak (3:30PM-6:30PM))} \]

\[ X_7 = \text{natural logarithm of response time in minutes} \]

\[ X_8 = \text{dummy variable for weather conditions (0 if no rain and 1 otherwise)} \]

The model predicted 81% of variation in incident prediction as a function of six independent variables.

### 2.1.2 Hierarchical Models for Incident Duration

Other studies (Ozbay and Kachroo, 1999; Smith and Smith, 2001) developed incident duration estimation methods based on classification and regression trees (CART) proposed by Breiman et al. (1984). CART is a non-parametric decision tree learning method which produces a classification tree if the decision variable is a categorical variable or a regression tree if the decision variable is a numeric variable. The idea behind the CART is relatively simple. The goal is to predict a class (or response) variable from a set of input variables. This can be done by building a decision tree. At each node of the tree, a test is applied to one of the variables and depending on the result, the left or right branch of the tree is followed. Eventually, each path on the tree is finalized with a prediction of the decision variable.
Ozbay and Kachroo (1999) developed decision trees to predict incident clearance times in Northern Virginia. First, they discovered the duration values for the data does not follow a lognormal or logistic distribution. Before estimating the CART, the significant independent variables were identified using ANOVA tests and other variables are excluded from the model. Then, they built series of decision trees using significant variables as shown in Figure 2.1. They built more detailed decision trees for property damage and personal injury accidents details as explained in Ozbay and Kachroo (1999).

Figure 2.1 Decision tree for incident clearance time prediction (Ozbay and Kachroo, 1999)
Smith and Smith (2001) investigated three models, namely a stochastic model, nonparametric regression model, and classification tree model, for forecasting clearance time of a freeway accident. Due to lack of probabilistic distributions, they did not use the stochastic model for estimating future incident clearance times. They found that the nonparametric regression model suffers from poor performance in predicting the clearance time of future accidents. They built a simpler classification tree than Ozbay and Kachroo (1999) as seen in Figure 2.2. The classification tree has only five nodes and three clearance time class: short, medium and long. They concluded that “the classification tree model is well suited for forecasting the phases of incident duration given a database of incidents with reliable and informative characteristics”.

Figure 2.2 Classification tree (Smith and Smith, 2001)
Kim et al. (2008) claimed that although estimated results from decision trees are quite acceptable, supplemental models can improve the prediction accuracy of incident duration. First, they used CART in their preliminary modeling then redesigned the model by incorporating branch splitting rules to increase the accuracy of it.

In the literature, there is also an attempt to estimate incident durations using Artificial Neural Networks (ANNs). Wei and Lee (2007) developed two adaptive ANN-based models and also used data fusion techniques to forecast incident duration. Their first model is used for forecasting the duration time at the time of incident notification, while the second one provides multi-period updates of duration time after the incident notification.

### 2.1.3 Bayesian Network Models

A BN is a graphical model which is similar to CART. However, unlike CART, an edge in a BN represents probabilistic relationship between nodes. BNs are very flexible tools as it is possible to get a distribution of values for a decision variable instead of a fixed value. Moreover, BNs can still be used for prediction even if the values of some the variables are missing. In the literature there are not too many papers dealing with BNs in the context for incident duration prediction problem. Recently, some researchers have attempted to develop incident duration prediction BNs using a single learning algorithm. Ozbay and Noyan (2006) used K2 algorithm searching best possible graph structure and they also incorporated the expert knowledge for the development of a BN by defining a search order for the variables under consideration. In the study, the incident data
collected in Northern Virginia in 1994 used and the BN structure in Figure 2.3 was estimated. They claimed that the prediction methodology employed by BNs is fully capable of representing the stochastic nature of incidents. Moreover, they demonstrated that dynamic BN estimation trees can be used to generate look up tables even with missing values.

Boyles et al. (2007) developed an incident duration prediction model based on a Naïve Bayesian classifier. They utilized incident records from the Georgia Department of Transportation collected in January and February 2004. They compared their model with linear regression model and found out that Naïve Bayesian classifier performs slightly better than the linear regression model. Furthermore, they analyzed the model performance with missing data to show the robustness of their model.
2.2 Spatial Models for Analyzing and Mapping the Locations with Higher Accident Risk

Previous research on the spatial analysis of crashes utilized data aggregated at different level. For example, Noland (2004) analyzed the effects of infrastructure improvements on fatalities and injuries. Data from 50 states was used to develop a negative binomial regression model which included infrastructure and socio-economic variables. Additionally, Negative binomial models are used by researchers to analyze the crashes at county level (Fridstrom and Ingebrigtsen, 1991; Karlaftis and Tarko, 1998; Amoros et al., 2003; Noland and Oh, 2004). While others utilized such models for modeling crashes for road sections (Abdel-Aty and Radwan, 2000; Lee and Mannering, 2002).

However, Negative binomial regression models suffer from the fact that they cannot handle spatial and temporal correlation (MacNab, 2004; Aguero-Valverde and Jovanis, 2006; Lord and Mannering, 2010). Therefore, more sophisticated methods are proposed by researchers such as hierarchical Bayesian models (Miaou et al, 2003; MacNab, 2004, Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Huang et al., 2010).

Miaou et al. (2003) developed a Poisson hierarchical Bayes model for traffic crash risk mapping at the county level for rural, two-lane, low volume roads in Texas. Their model included six components: a covariate with a fixed regression coefficient equal to 1 for the amount of travel occurring on these roads, a fixed district effect, a fixed or random covariate for spatial crash risk variation, a random spatial effect component for determining spatial correlations based on distance between county
centroids; a fixed or random time effect for year-to-year changes; and an exchangeable random effect term, for pure independent random local space-time variation.

At the first level of hierarchy, conditional on mean $\mu_i$, $Y_{it}$ values are assumed to be mutually independent and Poisson distributed as

$$Y_{it} \sim Po(\mu_i)$$

where indices $i$, $j$, and $t$ represent county, TxDOT district, and time period. The mean of the Poisson is further modeled as

$$\mu_i = \nu_i \lambda_i$$

where total VMT $\nu_i$ is treated as an offset and $\lambda_i$ is the crash rate. The crash rate is also structured as:

$$\log(\lambda_i) = \sum_{j=1}^{J} \sum_{i=1}^{I} \alpha_{ji} I(i \in D_j) + \sum_k \beta_k x_{ik} + \delta_i + \phi_i + e_{it}$$

where $I(i \in D_j)$ is binary indicator function showing if county $i$ is belong to district $D_j$ or not; $\delta_i$ represents year-to-year temporal effects, $\phi_i$ is a random spatial effect; $e_{it}$ is an unstructured space-time random effect; and $\alpha_{ji}$ and $\beta_k$ are regression parameters.

Different variations of the crash rate equation with different degrees of complexity were considered by the authors. For each component that was assumed to have a fixed effect, the second level of hierarchy was chosen to be an appropriate noninformative prior. However, for the components that have random effect, the second level of hierarchy was a prior with unknown parameters. Hyperpriors of these parameters formed the third level of the hierarchy.
For the spatial correlation, Besag’s conditional autoregressive (CAR) model (Besag, 1974) was adopted. The Gaussian CAR models considered in the study is the following general form:

\[
p(\phi_i | \phi_{-i}) \propto \eta^{1/2} \exp \left( -\frac{\eta}{2} \sum_{j \in C_i} w_{ij} (\phi_j - \phi_i)^2 \right)
\]

(2.6)

where \( p(\phi_i | \phi_{-i}) \) is the conditional probability of \( \phi_i \) given \( \phi_{-i} \);

\( \phi_{-i} \) represents all \( \phi \) except \( \phi_i \),

\( \propto \) stands for “proportional to”

\( C_i \) is a set of counties representing “neighbors” of county \( i \)

\( \eta \) is a fixed-effect parameter across all \( I \) and

\( w_{ij} \) is a positive weighting factor associated with the county pair \((i, i^*)\).

They utilized Markov chain Monte Carlo (MCMC) method to estimate the posterior means of model parameters. In their study, they recognized that most of the disease mapping was done for area-based data and thus they used county level data for mapping traffic crashes available in the form of network data. Their work can be extended to develop risk maps for traffic crashes on road networks.

MacNab (2004) developed a Bayesian model to analyze variation of accident risk factors at the regional level. They considered ecological (regional) analysis of accident and injury variations, covariate effects, random spatial effects and age effects simultaneously. Hospital separation data for 83 local health areas in British Columbia (BC), Canada was used to investigate ecological/contextual factors of accident injury among males between 0 to 24 ages. Eighteen local health area characteristics (spatial
covariates) including socio-economic indicators, residential environment indicators (roads and parks), medical services availability and utilization, etc were also studied.

First, they analyzed the regional variations in injury rates and explained such variations by potential risk factors using following model:

\[ \log(\mu_{ij}^h) = \log(n_{ij}) + a_0 + S_0(i) + b_j \]  

(2.7)

where \( \mu_{ij}^h \) represents the expectation of \( Y_{ij} \) conditioning on random spatial effects \( b \). The term \( \log(n_{ij}) \) is an offset; \( a_0 \) is a fixed effect, \( \exp(a_0) \) index the mean injury rate over all local health areas; \( S_0(i) \) is a representing age effect. The \( \exp(b_j) \) s represents local health area age-adjusted injury ratios. On the above model, injury hospitalization count \( y_{ij} \) is assumed to follow a Poisson distribution and the random spatial effects \( b \) are assumed to have spatially structured prior distribution, given as:

\[ y_{ij} | b \sim \text{Poisson}(\mu_{ij}^h) \]  

(2.8)

\[ (b_1, ..., b_J) \sim \text{MVN}(0, \Sigma(\sigma^2, \lambda)) \]

where \( \sum(\sigma^2, \lambda) = \sigma^2 D^{-1}, \quad D = \lambda Q + (1 - \lambda) I_J \)

with \( \sigma^2 \) representing Poisson overdispersion and \( \lambda \) the spatial autocorrelation \( 0 \leq \lambda \leq 1 \); \( I_J \) is an identity matrix of dimension \( J \); the neighborhood matrix \( Q \) has \( j \)th diagonal element equal to the number of neighbors of the corresponding area.

Later, they also investigated the age effect and the effect of age-specific injury separately by extending their model by adding the regression parameters for those effects to their model.
Aguero-Valverde and Jovanis (2006) estimated full Bayes hierarchical models with spatial and temporal effects and space–time interactions by using injury and fatality data for Pennsylvania at the county level. Covariates included socio-demographics, weather conditions, transportation infrastructure, and amount of travel. First, fatal crashes are assumed to follow a Poisson distribution:

$$y_i \sim \text{Poisson}(e_i \theta_i)$$  \hspace{1cm} (2.9)

where $y_i$ is the number of fatal crashes in county $i$, $\theta_i$ the risk in county $i$ and $e_i$ the exposure (daily vehicle-miles traveled) in the county $i$. At the second level of the hierarchy the log risk is modeled as:

$$\log(\theta_i) = \alpha + x_i' \beta + \nu_i + u_i$$  \hspace{1cm} (2.10)

where $x_i$ represents a vector of covariates, $\beta$ a vector of fixed effect variables, $\nu_i$ the uncorrelated heterogeneity and $u_i$ the spatial correlation. At the third level, a uniform prior distribution is assigned to $\alpha$ and a highly non-informative normal distribution is assigned to the $\beta$’s with mean 0 and variance 1000. The prior distribution for the uncorrelated heterogeneity is defined by $N(0, \tau^2_{\nu})$ where $\tau^2_{\nu}$ with a prior gamma distribution. For spatial correlation, CAR is used:

$$[u_i \mid u_j, i \neq j, \tau^2_i] \sim N(\bar{u}_i, \tau^2_i)$$  \hspace{1cm} (2.11)

where

$$\bar{u}_i = \frac{1}{\sum_j w_{ij}} \sum_j u_j w_{ij}$$

and
Finally, they extended their model to include a time-space interaction term:

\[ y_{ij} \sim \text{Poisson}(e_{ij} \theta_{ij}) \]  

(2.12)

\[ \log(\theta_{ij}) = \alpha + \sum_{k} \beta_k x_{ijk} + \nu_i + u_i + (\varphi + \delta_i) t_j \]

where \( y_{ij} \) is the observed number of crashes for the \( i \)th area, in the \( j \)th time interval, \( \beta_k \) the regression coefficient, \( \nu_i \) the uncorrelated heterogeneity, \( u_i \) the spatial correlation, \( \varphi \) the mean linear time trend over all areas, \( t_j \) the time interval \( j \) and \( \delta_i \) the interaction between time and area effect.

Quddus (2008) developed a set of negative binomial regression, econometric, and Bayesian hierarchical models for examining London crash data. Census wards in the Greater London metropolitan area are used as the spatial unit. Variables in different categories such as traffic characteristics, road characteristics and socio-economic factors were considered as the explanatory variables. The relationships between these variables and crash casualties were analyzed in the study using both non-spatial and spatial techniques. There were two spatial techniques in spatial econometrics employed in the study: a spatial autoregressive model (SAM) and a spatial error model (SEM). The SAR model was employed as follows:

\[ Z_i = \rho WZ_i + \beta X_i + \varepsilon_i \]  

(2.13)
where $Z$ was the dependent variable, $WZ$ the spatially lagged dependent variable for spatial weight matrix $W$, $\rho$ was the spatial lag coefficient, $\beta$ the vector of the parameters, $X$ the matrix of explanatory variables, and $\varepsilon$ the vector of normally distributed error terms. The spatial lag was considered as the spatially weighted average of the dependent variable in the neighboring regions. This term was assumed to be correlated with the error terms.

The SEM model was described as follows:

$$Z_i = \beta X_i + u_i$$
$$u_i = \lambda Wu_i + \varepsilon_i$$

where $u$ is the error term for spatial dependence and $\lambda$ is the spatial autoregressive coefficient.

The hierarchical Bayesian model used in the study was similar to those used in previous studies:

$$Y_i \sim \text{Pois}(\theta_i)$$
$$\ln(\theta_i) = \ln(\text{EV}_i) + (\beta_0 + \beta X_i) + \text{SC}_i + \text{UH}_i$$

where $\text{SC}_i$ is the spatial correlation and $\text{UH}_i$ is the uncorrelated heterogeneity. In the study, it was reported that “Bayesian hierarchical models are more appropriate in developing a relationship between area wide traffic crashes and the contributing factors associated with the road infrastructure, socioeconomic and traffic conditions of the area”.

Huang et al. (2010) analyzed crashes at the county level in Florida using hierarchical Bayesian models. The existence of spatial correlation between the counties was checked by using Moran’s $I$. Moran’s $I$ is a metric that range from -1 to 1 and a positive value indicates the presence of spatial correlation in the region under
investigation whereas negative value indicates dispersion (Moran, 1950). Moran’s $I$ is defined as:

$$I = \frac{n \sum_i \sum_j \omega_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_i (Y_i - \bar{Y})^2}$$  \hfill (2.16)

where $n$ is total number of observations, $Y_i$ and $Y_j$ are crash rates in counties $i$ and $j$, $\bar{Y}$ is average crash rate of all observations and $\omega_{ij}$ are the terms of proximity matrix. It was found that the data was spatially correlated. They used set of road and traffic related variables, and demographic and socio-economic variables in model development. The hierarchical model in the study has the following form:

$$Y_{it} \sim \text{Pois}(\mu_{it})$$

$$\mu_{it} = E_{it} R_{it}$$

$$\log(R_{it}) = \beta_0 + \beta X_{it} + \theta_i + \phi_i$$  \hfill (2.17)

where $\mu_{it}$ is the parameter of the model, $R_{it}$ is the relative crash risk at county $i$ in year $t$, $\theta_i$ is unobserved heterogeneity, $\phi_i$ is spatial correlation, and $E_{it}$ is the expected number of crashes at county $i$ in year $t$ and it is estimated from the following equation:

$$E_{it} = e_{it} \frac{\sum_{i,t} Y_{it}}{\sum_{i,t} e_{it}}$$  \hfill (2.18)

where $e_{it}$ is the traffic exposure.

Ossenbruggen et al. (2010) developed tools and performance measures to support decision making to mitigate crash risk. A detection scheme was proposed in the study where classical and spatial statistics that had been developed to identify roadways with the most severe safety needs in New Hampshire. This was based on the null hypothesis
that all roadways have the same level of crash risk. Fatal and nonfatal crash rates, denoted with subscripts $c$ and $f$, are assumed to be randomly distributed as Poisson processes:

$$X_{f,k} \sim \text{Poisson}(\mu_{f,k} = \lambda_f \cdot e_k)$$

$$X_{c,k} \sim \text{Poisson}(\mu_{c,k} = \lambda_c \cdot e_k)$$

(2.19)

where random variables $X_{c,k}$ and $X_{f,k}$ are equal to number of crashes at one-square mile grid location $k$; $e_k$ traffic exposure at location $k$; $\lambda_f$ and $\lambda_c$ statewide fatality and nonfatality crash rates; and $\mu_{f,k}$ and $\mu_{c,k}$ expected number of fatality and nonfatal crashes at $k$.

In this study, the crash rates were estimated in two steps. In the first step, traffic exposure was predicted. To do this, traffic exposure was assumed to be a two-dimensional continuum where exposure varied continuously over space. Further, exposures $e$ at locations, $(s_x, s_y)$ were assumed to have a spatially correlated lognormal distribution. Due to the north to south gradient in population and exposure, a linear trend in log traffic exposure was assumed:

$$\mu(s) = \beta_0 + \beta_1 s_x + \beta_2 s_y$$

(2.20)

For spatial autocorrelation, the Matérn correlation function $R(\phi, \kappa)$ was used (Diggle and Riberio, 1980). In this function, $\phi$ defines the range of the correlation and $\kappa$ describes the smoothness of the spatial random field. The resulting model is called a Gaussian random field shown as:

$$Y \sim N(\mu(s), \sigma^2 R(\phi, \kappa) + \tau^2 I)$$

(2.21)
where $\tau^2$ is the error variance parameter. Locations, where the null hypothesis is rejected, are considered as safety treatment candidates. $P$-value risk rankings are used to identify locations with the most severe safety needs.

Aguero-Valverde and Jovanis (2008) examined the effect of spatial correlation in models of road crash frequency at the link level. They proposed different link neighboring schemes to determine the best one for modeling crash frequency in road networks. They used a hierarchical Bayesian approach with conditional autoregressive effects for the spatial correlation terms. Their analysis of rural roadways in Pennsylvania indicated the importance of including spatial correlation in highway crash models. In this study, the models with spatial correlation showed a significantly better fit with the data than the Poisson lognormal model that incorporated only heterogeneity.

In the doctoral dissertation by Noyan (2007), the use of Cox process for crash modeling was proposed for the first time. In the study, a multi-level approach was used to model risk values from the severities of each individual involved in the crashes on Route 70, Ohio. A total risk value for each crash was calculated by giving different weights to different severities. Then, a hierarchical Bayesian model was developed for the risk estimation. Next, the risk values obtained for the risk estimation model was associated to the crash locations as categorical marks. Finally, the intensity functions were estimated for different mark levels using Cox process. Crash risk profile map along the roadway were generated using these intensity functions to demonstrate the variation in the crash risk along the roadway. It is proposed that these maps give traffic engineers and planners a powerful analysis tool for examining the patterns in crashes of interest.

Table 2.1 show the summary of the studies reviewed here.
Table 2.1 Summary of Spatial Models Estimated for Accident Analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>Location</th>
<th>Level of Detail</th>
<th>Model</th>
<th>Accident Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huang et al. (2010)</td>
<td>Florida</td>
<td>County</td>
<td>Hierarchical Bayes</td>
<td>All crashes</td>
</tr>
<tr>
<td>Aguero-Valverde and Jovanis (2008)</td>
<td>Pennsylvania</td>
<td>Link</td>
<td>Hierarchical Bayes</td>
<td>All crashes</td>
</tr>
<tr>
<td>Noyan (2007)</td>
<td>Ohio</td>
<td>Point</td>
<td>Cox Model</td>
<td>All crashes</td>
</tr>
<tr>
<td>MacNab (2004)</td>
<td>British Columbia</td>
<td>Regional (83 local health areas)</td>
<td>Bayesian regression</td>
<td>Injury</td>
</tr>
<tr>
<td>Miaou et al. (2003)</td>
<td>Texas</td>
<td>County</td>
<td>Hierarchical Bayes</td>
<td>Fatality/ Injury</td>
</tr>
</tbody>
</table>

It can be observed from Table 3.1 that the researchers chosen to analyze the crash data at different spatial resolutions. Hierarchical Bayes model is used in many studies due to its flexibility in modeling spatial effects. While some studies select only include fatality/injury crashes in the model, others include all crashes in their model.
CHAPTER 3. POTENTIAL IMPROVEMENTS IN INCIDENT MANAGEMENT

3.1 Incident Management

Incident management is aimed at the minimization of impacts of an incident (such as lack of safety and delay). The four stages of an incident management approach are commonly defined as: i) incident detection, ii) incident response, iii) incident clearance, and iv) incident recovery which are summarized in Figure 3.1 and discussed in further detail below.

![Diagram showing steps of incident management]

Figure 3.1 Steps of incident management

At first, these four stages can be seen only sequential but they, in most cases, may be carried out at the same time. For example, while an incident response team is dispatched; alternative routes can be offered to motorists.
Incident Detection/Verification

This phase is the starting point of incident management activities. To generate a response strategy, first the incident should be detected. Detection can be automatized by using relevant traffic data processed by computer-based algorithms. Such algorithms mostly use detector data from freeways comparing traffic conditions between upstream and downstream flows to decide about a potential incident at a given location. Alternatively, detection can be made by other drivers passing through or by police patrols. Reports made by other drivers, mostly via cell phones, are not always reliable, as certain specific pieces of information on an incident is necessary for generating quick and proper response. On the other hand, whenever available, closed circuit cameras might be very useful for verification of incident information on the exact location, severity, type, existing traffic conditions.

One major aspect in this phase is the congestion effecting and this is affected by the incident detection. Most major incidents are detected within 5-15 minutes; however, minor incidents may go unreported for 30 minutes or more (Cambridge Systematics, 1990). On the other hand, early detection of an incident is important for preventing congestion, since early detection allows early response and recovery actions. Otherwise, significant traffic queues build up due to lane blockage and bottlenecking in case of incidents, especially during peak hours.

Incident Response

In Traffic Incident Management Handbook (2000), the term incident response is defined as “dispatching the appropriate personnel and equipment, and activating the
appropriate communication links and motorist information media as soon as there is reasonable certainty that an incident is present”. Since the most important objective is safety, whenever needed, emergency response teams and equipment should be dispatched to an incident scene as soon as incident verified. This may require interagency communication between police, fire department, rescue units, and others.

In current incident management practice, the process of calling in different agencies is carried out by a dispatcher at traffic operations center (Ozbay and Kachroo, 1999). Meanwhile, incident-related information can be disseminated by means of commercial radio broadcasts, variable message signs, route guidance systems, etc. Moreover, the decision should be made as to whether any traffic flow should be diverted, based on estimated incident delay, number of lanes blocked, and type and severity of the incident. During off-peak hours and minor incidents, demand may not exceed the capacity and the impact would be less. In such cases, although the flow is disrupted, it may not be necessary to divert traffic if the remaining capacity is sufficient for traffic operations to continue. In the case of major incidents, diverting traffic flow can be important for network efficiency and public safety, as well as to protect the incident scene and provide for rapid and safe clearance (Sawaya et al., 2005). If the flow is to be diverted, it is important to choose alternative routes with enough capacity in order not to create excessive congestion on the alternative routes.

**Incident Clearance**

Incident clearance involves timely handling of incident scene. This operation may include tow truck operations, the removal of wreckage, and the cleanup of material spills
and debris. In case of some special incidents, like hazardous material (HAZMAT) spills, incident clearance takes more than usual cases as HAZMAT teams are required for response.

**Incident Recovery**

Recovery consists of three tasks: a) restoring traffic flow at the site of the incident; b) preventing more traffic from flowing into the area and c) preventing congestion from spilling across the traffic network (PB Farradyne, 2000). Diverting upstream traffic to an alternative route(s) in the response step can decrease the recovery period. However, the capacity of an alternative route may not be sufficient; or there may not be an alternative route at all. At this point, other steps may become more vital, as decreasing the recovery period depends totally on the duration of other steps.

### 3.2 Need for Better Incident Management Planning and Effective Operations

In the previous section, the four stages of incident management were briefly explained. The overall incident duration depends on the duration of each stage. Clearly, any strategy that can reduce the time spend at any stage can decrease incident duration. Ozbay and Kachroo (1999) summarized the strategies used in the different stages of incident management as shown in Table 3.1.
Table 3.1 Strategies and Technologies for Incident Management (Ozbay and Kachroo, 1999)

<table>
<thead>
<tr>
<th>Goals</th>
<th>Examples of Strategies / Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduce the incident detection time</td>
<td>Traffic Surveillance and Detection System, Closed Circuit TV (CCTV) Cameras, Incident Hot Lines, Highway Patrol, Traffic Management Centers</td>
</tr>
<tr>
<td>Reduce the incident response time</td>
<td>Improved Inter-agency Communication, Increased Number of Service Patrols, Increased Number of Tow Truck Operators, Improved Incident Response Procedures, Computer Based Decision Support Systems</td>
</tr>
<tr>
<td>Reduce the incident clearance time</td>
<td>Improved Incident Response Procedures, Increased Number of Incident Response Teams, Up-to-Date Incident Response Plans, Training of Incident Response Teams</td>
</tr>
<tr>
<td>Reduce the impact of incidents on peak-period traffic</td>
<td>Efficient Traffic Management Strategies, Efficient Diversion Plans, Minimal Capacity Reduction of the Freeway, Reduced Demand due to Improved Traveler Information</td>
</tr>
<tr>
<td>Reduce the impacts of constructions, maintenance, and special events on traffic</td>
<td>Improved procedures, Effective Public Information, Effective Incident/Traffic Management Measures, Reliable and Accurate Forecasts / Traffic Information</td>
</tr>
<tr>
<td>Provide accurate, timely and useful traffic information to motorist</td>
<td>Highway Advisory Radio, Variable Message Signs &amp; Improved Procedures for their use in IM, Improved Communication with Media</td>
</tr>
</tbody>
</table>

As seen in Table 3.1, different stages of incident management require different strategies. For example, to reduce detection and response time, it is important to optimally allocate the available resources such as by deploying traffic surveillance and detection systems or increasing the number of service patrols on a highway. On the other
hand, to reduce clearance time or the effects of the incidents to other motorists, new traffic management procedures are needed and to be implemented, these require multiple agency involvement or information dissemination. Although strategies may differ for different stages of the incident, the main goal of any incident management program is to clear the incidents as quickly as possible from the roadway.

For purposes of strategic incident management planning, the focus should be placed on minimizing the response time by optimizing resource allocation. According to Zagrofos et al. (1993) incident response consists of four components:

- The incident detection and verification time \( (T_1) \): the time between the occurrence of an incident and its detection/verification.
- The dispatch time of incident response teams \( (T_2) \): identification of incident location and severity and the dispatch of first incident response team.
- The travel time of incident response team to the incident location \( (T_3) \): the time elapsed until the first response team arrives on the incident scene.
- The incident clearance time \( (T_4) \): the time to clear the incident scene.

Although these components seem distinct, they are in fact related to a degree. For instance, the travel time of incident response teams gets longer if incident detection and verification time is longer due to possible congestion on the roadway due to the incident. Location of the resources is very critical especially for \( T_1 \) and \( T_3 \). In spite of the fact that the incidents are random events in space and time, there are cases that follow a spatial pattern. The roadway sections which are more prone the accidents called “black spots”. The frequency and severity of incidents at these locations differs significantly from the
rest of the highway because of a spatial characteristic change such as the existence of a ramp, a geometrical defect or annual average daily traffic (AADT).

While Figure 3.2 (a) shows the map of accidents counts (equal weights) in New Jersey by municipality in 2010, Figure 3.2 (b) shows the map of the accidents based on severity weights. For this simple example, different types of accidents given different weights based on their severity. The weights that were assumed were 1, 3, and 5 for property damage, injury and fatality accidents, respectively. From the Figure 3.2 (a), it is easy to observe that accidents are not uniformly distributed in New Jersey. Moreover, when the severity of these accidents is taken into account, accident risks are higher in some areas than the case where only accidents counts are considered. Even this most simple example can shed some light on the spatial distribution of the accidents and might provide some idea about the allocation of resources for incident management. However, in order to optimize resource allocation for incident management, one needs to capture the stochastic nature of incidents and relate them to spatial features of the transportation network. Therefore, robust statistical techniques should be employed to achieve this goal.
Figure 3.2 New Jersey 2010 accidents based on (a) frequency and (b) severity weights
3.3 Conceptual Framework of the Methodology

Figure 3.3 shows the conceptual structure of the proposed methodology. The methodology proposed in this dissertation has two components. In Chapter 4, to improve incident management operations, incident duration prediction models are estimated using BNs and adaptive learning methods. In Chapter 5, for better planning for future incidents, frequency estimation methods are used for identifying the locations more prone to incidents.

![Conceptual Structure of Methodology]

Figure 3.3 Conceptual structure of the proposed methodology

In Section 4.2, a real-time BN-based incident duration prediction model is estimated. The major objectives of this part were:
1. Identify the best ways to estimate Bayesian Network (BN) for more accurate prediction of incident durations given the known difficulties for using traditional parametric prediction methods. This problem was recognized by previous researchers (Khattak et al., 1995) (Garib et al., 1997) (Ozbay and Kachroo, 1999) (Smith and Smith, 2001) (Kim et al., 2008). However, there is also a need for improving the prediction capabilities of BN's by determining the best tree structure to represent the process. Thus, one of the major challenges in estimating BNs for incident duration prediction is the determination of the best tree structure. Previous studies Ozbay and Noyan (2006) and Boyles et al. (2007) did not address this problem. In those studies, only a single structure learning algorithm was chosen by authors. Hence, a single BN was estimated for modeling incident duration. Although the merits and reasons for using the chosen algorithm were explained, they did not mention the possibility of improving their model by using other structure learning algorithms. In this study, different structure learning algorithms were tested to construct the BN structure for incident duration model using 2005 New Jersey Turnpike accident data. The fact that all the incidents were obtained from this major freeway eliminates the need for considering highway specific factors allowing the research team to focus on the task of comparing the effectiveness and practicality of using BNs estimated using three different machine learning techniques.

2. Demonstrate the inherent advantages of BNs over traditional hierarchical models such as linear regression and Classification and Regression Trees (CART) in the case of the absence of all input data. This required acquisition of new data that was not previously used to estimate these models.
3. Finally, once the best BN was obtained, employ novel estimation techniques such as link strength etc. to better understand the relationships among factors contributing to incident durations. This was an important goal of this study because; proposed models can be used to improve incident management procedures if all contributing factors are well understood.

In Section 4.3, an adaptive learning scheme was devised by updating the parameters in the duration prediction model. While the duration prediction model is estimated in this study, a major weakness of BNs is identified. The model is provided with all the training data at the development stage and no future revisions are made to the model. Hence, the model, as in the existing models in the literature (Ozbay and Noyan, 2006; Boyles et al., 2007), lacks an appropriate mechanism for adapting to the changing future conditions. The absence of a learning mechanism can result in a decrease in performance over time. Real-time information used for the prediction may not always be complete; hence, it is not plausible to adapt the model based on this data. In practice, the transportation agencies obtain the data after it is verified by the law enforcement agencies. Hence, there is a need to introduce an adaptive learning mechanism to BNs to update them based on the verified new data with the goal of maintaining their prediction performance in the future.

In this section, the adaptive learning performance of Bayesian model for future data is also demonstrated. First, 2011 incident durations are predicted based solely on the model estimated using 2005 New Jersey incident data. Then, the model is updated by supplying it with 2011 New Jersey incident data for the same locations provided on a
monthly and quarterly basis. The six year difference between training and prediction data can demonstrate how the model performance changes over time and also indicates if the model is still valid after a substantial period of time after its initial estimation.

In Section 4.4, the adaptive learning framework proposed by Castillo and Gama (2009), which is capable of updating the BN structure as well as its parameters, is used for incident duration prediction. Updating the model based on the new data might be sufficient for capturing minor changes in the system. However, it is possible that some of the data might be irrelevant to the duration model in time. Furthermore, the abrupt changes in the system such as lane closure due to long-term work zones, and the addition of new lanes and interchanges to the roadway might require one to develop a completely different model. Hence, a robust adaptive mechanism, which can handle minor as well as major changes to the system, should be embedded into the incident duration prediction model.

In this section, the applicability of the adaptive framework for Bayesian networks proposed by Castillo and Gama (2009) for estimating incident duration is analyzed. This framework is capable of both updating the model and its parameters based on the time sequential data. The performance of this method was investigated using 2011-2012 New Jersey incident data and this was compared to the static BN algorithms considered previously.

In Section 5.1, similar to previous research which used aggregated data at the county level, hierarchical Bayesian models were used to analyze the factors contributing to crash risk and spatial variation among the different regions. Since crashes are rare
events, there is almost always high variance in crash data between different regions. This is called small area problem (Ghosh and Rao, 1994). One way to overcome this problem is to use hierarchical models that factor in spatial correlation. The hierarchical models are also very effective in developing models in the presence of multiple sources of data and uncertainty in parameterizations (Gelfand et al., 2010).

County level crash data used in this section contains data on all crashes that occurred from 2001 to 2010 in New Jersey and these were used in the model development. In addition to developing a general model for raw crash counts, we attempted to develop different models that could be used for different road types. Furthermore, since different crashes have varying impacts, we proposed the use of severity weights to capture the effects of the crashes more realistically.

The main objective of this section is to compare the traditional hierarchical model to a weighted model in order to help us to understand and to demonstrate the importance of representing the spatial variation of crashes as well as their severity. In this dissertation, spatial covariates related to the roadway conditions were used. Due to the limitations of the data, surrogate variables were employed. First, the model was estimated using raw crash counts. Most studies lack an analysis of crash rates by road type using hierarchical Bayesian models that can capture the spatio-temporal characteristics of crashes. To address this gap in the literature, additional models were estimated for those roadways that were under different jurisdictions such as the department of transportation, or other authorities, counties, and municipalities. Finally, crash rate maps were developed based on the modeling results to enable the ability to better visualize the effects of spatial
covariates. These maps were designed to help users to better understand the modeling results without requiring prior knowledge of the hierarchical models.

In Section 5.2, hierarchical Bayesian models are used to model the frequency of downed tree incidents on roadways using factors such as precipitation, wind speed, roadway density, etc. Real variables are used in this section instead of the surrogate variables approach reported in Section 5.1. The effect of spatial correlation and spatial heterogeneity was also considered in the model development. Municipality level counts of the downed trees that contained all events in New Jersey during hurricanes Irene and Sandy were used in the model estimation. Since, the data contained numerous zero counts at the municipality level, a zero-inflated Poisson model was adopted. Until recently, hurricanes in the USA were known to mainly affect the south eastern part of the country (Changnon, 2009). However, two recent hurricanes namely, Irene and Sandy, proved that this assumption was not always true. The most severe damage caused by Irene occurred in Northeastern States of New Jersey, Massachusetts and Vermont (Avila and Cangialosi, 2011). Sandy made landfall near Brigantine, New Jersey with 70 knots maximum sustained winds (Blake et al., 2013). Both hurricanes resulted in a catastrophic storm surge, wide spread property damage and loss of life. Moreover, the infrastructure was affected due to many fallen trees as a result of strong winds. The New Jersey Public Service Electric and Gas Company reported that 48,000 trees had to be removed or trimmed to restore power. Roadway network was also heavily crippled due to storm flooding and downed trees. Roadways are critical for the operations of the agencies responding to hurricanes as well as for the evacuation of the affected people. Therefore,
there is a need to improve roadway safety and accessibility in the aftermath of the hurricanes. By better handling hurricane related incidents on the roadways such as downed trees, the identification of high risk locations and contributing factors, roadway safety and accessibility can be improved.

Currently, there are, to the authors’ knowledge, no studies for modeling hurricane related roadway incidents of downed trees. However, there are some studies that have developed regression models for predicting hurricane related power outages. Although, those studies analyze different impacts of hurricanes, there is, at least, some shared perspective in terms of hurricane related explanatory variables between them and this study. For example, Davidson et al. (2005) conducted an analysis of spatial variation of power outages in Carolinas from five hurricanes and found a significant relationship between wind speed and number of outages. Later, Liu et al. (2005) developed Poisson and negative binomial generalized linear models using the same data. Liu et al. (2008) estimated spatial generalized linear mixed models for power outages using the data from six hurricanes and eight ice storms in East coast. Han et al. (2009) developed negative binomial GLMs to model the spatial distribution of power outages in the Gulf Coast region based on data from ten hurricanes. They also included time since the duration of the hurricane, and the radius of maximum wind variables to measure the vulnerability the power system to the storm damage.

In Section 5.3, a doubly stochastic model was proposed for purposes of point level modeling of the crashes. In this model, the roadway was represented as a continuous
entity without predefined links. This type of model captures the continuous nature of the roadways better than any existing model.

The doubly stochastic model, which is also often referred to as Cox process (1955), might offer a solution to the existing problem in crash risk modeling. Cox process is an extension of Poisson process model where the intensity function of the Poisson process is modeled as a realization of a random field (Moller and Waagepetersen, 2004). One popular method used in the spatial modeling literature is the use of log-Gaussian random field and consequently this model is called log-Gaussian Cox process (LGCP) (Moller et al., 1998). There are several articles in other fields such as epidemiology (Diggle et al., 2005), and ecology (Serra et al., 2013) that use LGCP. For example, Diggle et al. (2005) developed a model for examining the predictive probability that relative risk of a disease at a location exceeds a threshold set by the experts. In the model, Cox model was used for representing spatial variation in risk as a result of known risk factors and unexplained spatial variation. So the question arises, as to whether, although they are very useful, why such models are not considered for mapping of crashes. First of all, the spatial point process models, so far, has been used for modeling spatially continuous events such as contamination in a lake, the disease modeling, etc. In the case of crashes, we cannot directly threat them as spatially continuous events as they are bounded by the roadways. Moreover, spatial covariates are bounded by the roadways as well. The model used in this study was also fairly new and until recently there were limited tools for efficiently implementing the model. Having many point locations in the data also introduces another problem, if Bayesian statistical modeling is used with point
models, as MCMC simulations will be too complex. The simulations of such models
would take hours or even days for a single run. Recently, a tool developed by Rue et al.
(2013) saves the researches from this struggle by using a method called R-INLA
(Integrated Nested Laplace Approximation). Unlike, MCMC which gives exact results,
R-INLA uses Integrated Nested Laplace Approximation (INLA) for the results which
significantly lower the time and computing power needed for MCMC simulations. As the
name suggests, this approximation comes with a cost of lower accuracy than MCMC
approach. But the research shows that in most cases the difference between MCMC and
INLA is negligible. A comparison of these methods is the out of scope of this study and
readers are advised to review the paper by Taylor and Diggle (2014).

In this regard, first, the roadway was divided to form grid cells where the crash
points become the crash counts in the grid cells. Then, these cells were aligned in 1-d so
that the effect of spatial variation is only in the direction of the roadway. This also
enabled one to limit the occurrence of the crashes only on the roadway. At the first level
of the stochasticity, the crash counts in the cells were assumed to follow a Poisson
distribution. In the second level, given the crash counts, the spatial variation on the
roadways was modeled as a log-Gaussian Random field. Using Bayesian modeling
approach and INLA, the model was then developed in R, and the modeling results and
Bayesian inference were used to develop useful maps for determining the crash risks
along the roadway.
3.4 Proposed Incident Management Improvements

This dissertation aimed to develop a comprehensive incident management framework to assist transportation professionals to clear/respond to incidents / accidents in a timely and in an effective manner. This can be achieved by better incident management planning and effective real-time incident management operations. Hence, the improvements to incident management due to this study were described in terms of these two categories. The proposed improvements can be summarized as follows:

Operational:

- Proposed incident duration model enables the researcher and others to predict incident durations even if the partial information is available about an incident. In Bayesian networks, a sequence or completeness of information is not required which makes them very flexible and useful especially at the initial stages of an incident where complete information about the incident might not be available.

- Having an estimate of incident duration at the initial stage of an incident enables transportation professionals to make the best possible estimate with available information and then revise their incident management strategies such as route diversion, lane closures etc. in real time. Therefore, the effects of an incident on other motorists such as delays and secondary incidents can be minimized by being able to make timely decisions even with missing information.
• Unlike traditional models, the BN based duration prediction model can be used to predict other variables of the incident such as number of vehicles involved in the incident or number of injuries. In Bayesian networks, probability distribution for every variable can be estimated using the values of other variables such as type of the incident.

• The concept of link strength was used to analyze the impacts of incident variables on the duration, numerically. By ranking the incident variables based on their impact on the duration, the most influential variables can be identified. This ranking can be used to reduce the number of variables on the model so that it can be easily applied to another roadway where only a subset of incident variables is recorded.

• The idea of on-line learning based on the real-time performance of the model introduced in this dissertation can be a useful property to adapt the model to ever-changing roadway conditions and response strategies. It was observed that adaptive learning techniques can improve the model performance in the short-term (by parameter adaptation) and in the long-term (by combining parameter and structural adaptation).

**Short and Mid-term Planning:**

• Mapping of crashes through spatial representation may help transportation professionals to better understand and analyze incident data and thus plan incident response operations. Mathematical models, generally, provide
numerical results and sometimes it is hard to interpret them, especially if the underlying model is complex. Even though the mathematical background behind the spatial models in this study are rather complex, the output provided from them can be easily understood by decision makers.

- The predictive heat maps developed using the above spatial models can help transportation professionals to develop long term incident management strategies to better allocate resources to problematic areas. These maps do not only display the high risk incident point locations but also spatial distribution of incident risks over a road stretch which can help determining the area of impact.

- Accounting for crash severity in crash counts may help professionals to detect clustering of high risk crashes among the regions. Then, these regions where higher risk crashes are clustered based on severity weights can be further researched. It is necessary to investigate the factors of the likelihood of occurrence of more severe crashes at these locations and the cause behind neighboring locations (e.g. counties) to have similar trends. Using this approach, it is also possible to go beyond predetermined geographical entities such as counties and develop clusters that represent crash distributions over the various road stretches.

- Different crash models based on roadway functional class may serve as a performance measurement tool for helping to analyze the performance of different agencies for different roadway under their jurisdictions.
• Crash databases contain a very large number of variables. It may not be practical/necessary to incorporate all variables in a crash model. It is thus important to identify the most relevant factors. Feature selection methods can help the process of reducing the number of variables so that only a subset of variables most relevant to a target variable can be chosen in the future crash studies.

• Zero inflated hierarchical model has been shown to better represent the data in the case of special incidents such as hurricane related downed tree incidents along roadways.

• The effects of explanatory variables on downed trees as well as the spatial correlation and heterogeneity on the model fit was investigated and the most influential factors were identified.

• The point level risk maps can provide better insights than maps based on more aggregate level crash data since these point level maps enable one to pinpoint the problematic locations instead of averaging out the crashes over a predefined link length.
CHAPTER 4. BAYESIAN NETWORKS AND 
ADAPTIVE LEARNING FOR PREDICTING 
INCIDENT DURATION

4.1 Machine Learning Methods

4.1.1 Bayesian Networks

A BN is also known as a directed acyclic graph (DAG). If an arc is not present
between two nodes in a DAG, then those two nodes are said to be conditionally
independent. Figure 2 shows an example of a simple BN with 5 nodes. While nodes
represent variables which can be either discrete or continuous, arcs between the nodes
stand for conditional probability of child node given its parent node. Each variable \( X_i \) is
independent of its child nodes given the values of its parents in the graph. The joint
probability distribution \( P \) of the variables in the example DAG is equal to product of its
conditional distributions of all nodes given values of their parents and it is as follows:

\[
P(X_1, X_2, X_3, X_4, X_5) = P(X_5 | X_3, X_2)P(X_4 | X_2)P(X_3 | X_1)P(X_2 | X_1)P(X_1)
\]  

(4.1)

If the conditional probabilities on the right side of the equation exist (i.e.
conditional dependencies between nodes) then the above equation can also be written in
the factorized form as:

\[
P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i | pa(X_i))
\]  

(4.2)

where \( pa(X_i) \) is the set of parents of \( X_i \) in the DAG. The factorized presentation of the
joint probability distribution of variables enables easy computation of the probability of a
variable, given observed values. The assigned value of a variable is propagated through the network and will update the marginal posterior distributions of all the nodes. The propagation will occur in a top-down as well as a bottom-up manner along the arcs in the network, irrespective of the position of the node in the network, and this involves repeated application of Bayes' theorem and the use of the conditional independencies encoded in the network structure (Korb and Nicholson, 2005). The details concerning the belief propagation in BNs can be found in Pearl (1988). For example, the conditional probability distribution of $X_4$ in the network represented by Figure 4.1 is only dependent on $X_1$ and $X_2$. However, suppose the observed value of one or both of these are unknown, then the observed values of $X_3$ and $X_5$ will provide us additional information for computation of the marginal probability distribution of $X_1$ and $X_2$, hence, the marginal probability distribution of $X_4$. A more concrete example to illustrate this concept can be given by borrowing some variables from medical diagnosis problem of Lauritzen and Spiegelhalter (1988) in place of the random variables in Figure 4.1. Say $X_1$ is smoking which can cause lung cancer, $X_2$, or bronchitis, $X_3$. Lung cancer can be detected by an X-ray result, $X_4$. Shortness of breath, $X_5$, may be caused by lung cancer or bronchitis. Evidence of bronchitis or shortness of breath variables will provide information on unknown variables which are whether a patient is smoking or has lung cancer.

Figure 4.1 Example BN
4.1.2 Structure Learning Algorithms for Automatically Building Bayesian Networks

When expert knowledge is available, a BN structure can be easily created by defining the dependencies between the variables. However, if the network structure is not provided by an expert, it is still possible to learn the network structure automatically from the data. “Either the constraint-based approach in which the model satisfies all the conditional independencies observed in the data or the optimization-based approach in which some scoring function is maximized such as marginal likelihood or penalized likelihood (e.g. MDL/BIC) can be utilized to determine the network structure” (Murphy, 2001). Note that, Bayesian Information Criterion (BIC) is a widely used measure for model selection in structure learning which includes a goodness of fit term and a penalty term to account for model complexity (Schwarz, 1978):

\[
BIC = \log L - \frac{d}{2} \log N
\]  

(4.3)

where \(L\) is the sum of likelihood of parameters, \(d\) is the dimension of BN and \(N\) is the sample size. To evaluate BIC score of a BN, first, the maximum likelihood of the parameters of the model is estimated. If the data is complete, this part can be reduced to frequency counting. Then, the probability for each case in the data can be computed using the estimates. “Given a score function, the task is to find the highest-scoring Bayesian network structure among the set of all possible network structures. That is, the task of structural learning has been reduced to a search problem” (Jensen and Nielsen, 2007).

The problem of finding the best scoring model is known to “be NP-hard due to complexity of search space of all possible graphs” (Chikering, 1996). Hence, structure
learning algorithms use search heuristics to limit the size of the search space. They use operators such as arc-insertion or arc-deletion to explore the graph space and local score to minimize the computational effort to the score variation between two neighboring graphs. Now, the structure learning algorithms that were evaluated in this study will be discussed.

Naïve Bayes classifier is a simple probabilistic classifier where all the parameters are assumed independent given the class variable (node), which is the parent of all other variables in a Naïve Bayes classifier. Even though the independence assumption is violated in most cases, Naïve Bayes classifiers are shown to work well in solving many classification problems. Their success is generally attributed to their simplicity as the small number of parameters needed to be estimated (Jing et al., 2008).

To learn and to use the dependencies in the data, the independence assumption in Naïve Bayes can be overridden using an augmented Naïve Bayes classifier (TAN) (Keogh and Pazzani, 1999). Geiger (1992) used an augmented-tree structure for this purpose in which conditional relationships are recovered by using maximum weight spanning tree algorithm.

Chow and Liu (1968) proposed a method derived from the maximum weight spanning tree algorithm (MWST). This method associates a weight to each edge which can be either the mutual information between two variables (Chow and Liu, 1968) or the score variation when one node becomes a parent of the other (Heckerman, 1994). When the weight matrix is created, a usual the MWST algorithm gives an undirected tree that can be oriented given a chosen root (Leray and Francois, 2006).
K2 algorithm, proposed by Cooper and Herskovits (1992), limit the search space to a subset of all possible networks, based on the ordering of the variables. Since the run of the K2 algorithm depends on the given enumeration order, this order is needed to obtained if it is not defined by experts. As Andrieu et al. (2003) propose, the oriented tree obtained with the MWST algorithm can be used to generate this order. We just have to initialize the MWST algorithm with a root node, which can either be the class node or a randomly chosen one. Then the topological order of the tree obtained from the MWST can be used in order to initialize K2 algorithm. Table 4.1 summarizes the initialization requirements of the other algorithms as well as search space and scoring methods employed by the algorithms.

Table 4.1 Comparison of different machine learning algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Search Space</th>
<th>Scoring Method</th>
<th>Initialization Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td>Fixed</td>
<td>-</td>
<td>Class node</td>
</tr>
<tr>
<td>MWST</td>
<td>Tree Space</td>
<td>Penalized Likelihood</td>
<td>Root node</td>
</tr>
<tr>
<td>TAN</td>
<td>Tree Space*</td>
<td>Penalized Likelihood</td>
<td>Root node, class node</td>
</tr>
<tr>
<td>K2</td>
<td>DAG space</td>
<td>Penalized Likelihood</td>
<td>Order of nodes</td>
</tr>
</tbody>
</table>

* limited by variable order

Sahami (1996) introduced a method called k-dependence Bayesian classifiers which includes a Naïve Bayes structure and can also identify k number of dependencies that exist between variables. This is a very flexible algorithm for experimental purposes as it enables to search the space of possible structures by increasing value of k. In the most restrictive case (k=0), the algorithm returns to a Naïve Bayes structure. 1-dependence Bayesian classifier is a TAN structure and it is possible to increase k up to N-1, where N is number of variables in a model.
4.1.3 Connection Strength of Variables in a BN

The connection strength between two variables, such as X and Y, measures how strongly information on the state of X affects the state of Y (and vice versa). Mutual information is the most common implementation of this idea proposed by Pearl (1988). Mutual information is simply the difference between the uncertainty (entropy) of a discrete random variable, Y, and the uncertainty of Y given X and can be defined by:

\[
\text{MI}(X, Y) = U(Y) - U(Y | X)
\]  

(4.4)

where uncertainty \( U(Y) \) is

\[
U(Y) = \sum_{y_i} P(y_i) \log_2 \frac{1}{P(y_i)}
\]

(4.5)

and where \( U(Y | X) \) is calculated by averaging \( U(Y | x_i) \) over all possible states \( x_i \) of X.

\[
U(Y | X) = \sum_{x_i} P(x_i) U(Y | x_i)
\]

Transformation of the above equations yields:

\[
\text{MI}(X, Y) = \sum_{x,y} P(x,y) \log_2 \left( \frac{P(x,y)}{P(x)P(y)} \right)
\]

(4.6)

The absolute degree of uncertainty reduction in a variable may provide less insight than the percentage of the original uncertainty that was removed. Therefore a Mutual Information Percentage metric is proposed by Ebert-Uphoff (2009) to be used in combination with Mutual Information. Mutual Information Percentage can be defined as:
\[ MI\%(X,Y) = \frac{MI(X,Y)}{U(Y)}.100 \]

\[ = \frac{U(Y) - U(Y|X)}{U(Y)}.100 \text{ where } U(Y) \neq 0 \] (4.7)

Note that Mutual Information is not symmetric in X and Y, which means \( MI\%(X,Y) \neq MI\%(Y,X) \) and is undefined if \( U(Y) = 0 \) (Ebert-Uphoff, 2009).

4.1.4 Parameter Estimation for Adaptive Learning in Bayesian Networks

Provided that the data is complete (a value provided for each of the variables for each record), the parameters in a BN can be estimated using batch learning. For each record \( d \) in the data \( D \), the probability \( P(d|M) \) is called the likelihood of model \( M \) given \( d \). For a DAG, the conditional probability distribution can be estimated using either a maximum likelihood or maximum a posteriori approach. Then, the log-likelihood of \( M \) given \( d \) is

\[ L(M \mid D) = \sum_{d \in D} \log_2 P(d \mid M) \] (4.8)

To estimate the conditional probabilities, among the possible models \( M_\theta \), which have the same structure but different parameters, \( \theta \), one must select a parameter estimate \( \tilde{\theta} \) that maximizes the likelihood:

\[ \tilde{\theta} = \arg \max_\theta L(M_\theta \mid D) \] (4.9)

However, when new data is acquired sequentially, it is necessary for the model to adapt to new cases. Hence, the conditional probabilities need to be updated since the new
records will increase the uncertainty in the accuracy of the conditional probabilities. This can only be achieved by modifying certain parameters in a BN structure. As the new records are acquired, the goal is to learn from the new records. Although, we are certain of the network structure, the conditional probabilities vary depending on each new case and it is essential to develop a model which can automatically adapt to new cases. For a BN structure such as the one shown in Figure 1, the variable $X_5$ is influenced by $X_2$ and $X_3$, and the conditional probability can be modeled as $P(X_5 | X_2, X_3)$. The uncertainty in the conditional probability can be explicitly modeled by introducing a new parent $Y$ for $X_5$. To represent each possible result of $X_5$, a prior distribution, $P(Y)$, can be defined for $Y$. When a new case, $r$, is entered to the BN, the propagation will result in a new distribution $P*(Y) = P(Y|r)$ and this way the change in the distribution will reflect what is learned from the case. For the additional cases $P*(Y)$ can be used in the same way and when $i^{th}$ case is acquired, the corresponding variables will be instantiated and $P*(Y) = P(Y|r_1,...,r_{i-1})$ will be updated to $P(Y|r_1,...,r_i)$ (Jensen and Nielsen, 2007).

4.1.5 Feature Selection

Feature selection is the process of reducing the number of features so that only a subset of variables most relevant to target variable can be picked up. There are several methods for selecting/ranking features based on their predictive power, such as mutual information, Fisher score or Pearson correlation. However, these methods do not always guarantee orthogonality between features, and may result in selection of redundant features (Fleuret, 2004). Due its robustness, feature selection methods are widely used in
areas such as text categorization (Forman, 2003), medical imaging (Tourassi et al., 2001) and visual speech recognition (Scanlon, 2004). As a result feature selection process, model complexity can be decreased, the prediction accuracy can be improved, and the underlying process that generated the data can be better understood (Guyon and Elisseeff, 2003).

In order to select an optimal subset of features, an evaluation function is employed by feature selection methods. Evaluation functions measure the discriminating ability of a feature using different metrics. Dash and Liu (1997) grouped the evaluation functions into five categories: distance, information, dependence, consistency, classifier error rate. In this study, information measures were considered. Generally, these measures determine information gain from a feature. Therefore, in this study, an approach based on mutual information (Shannon, 1948) developed by Fleuret (2004) was used for feature selection/ranking. Their approach, called Conditional Mutual Information Maximization (CMIM), consisted of picking features which maximize their mutual information with the class (target variable) to predict, conditionally to the response of any feature already chosen. Selection based on this criterion discards the features similar to those already selected, even if they are individually powerful, as they do not carry additional information about the class to predict.

Entropy (Shannon, 1948) is one of the most fundamental measures in information theory used for determining the uncertainty of a random variable. Let $X$ be a random variables with discrete values, entropy, $H(X)$, can be defined as:
\[ H(X) = -\sum_x p(x) \log_2 p(x) \]  
(4.10)

where \( p(x) = \Pr(X = x) \) is the probability density function of random variable \( X \). The above expression simply quantifies the number of units, generally bits, required to describe the value of \( X \). Based on Shannon’s entropy, a deterministic variable has no entropy as nothing is required to describe it. Suppose we have two random variables \( X \) and \( Y \). The joint entropy of \( H(X, Y) \) of \( X \) and \( Y \) is
\[
H(X, Y) = -\sum_y \sum_x p(x, y) \log_2 p(x, y)
\]
(4.11)

The conditional entropy describes the reduction in uncertainty when one of the variables is known. Assume \( Y \) is given, and then the conditional entropy \( H(X \mid Y) \) of \( X \) with respect to \( Y \) is
\[
H(X \mid Y) = H(X, Y) - H(Y)
\]
(4.12)

If \( X \) and \( Y \) are independent then, \( H(X, Y) = H(X) + H(Y) \). Therefore, knowing the state of \( Y \) does not help describing \( X \). On the other extreme if \( X \) is a deterministic function of \( Y \), then, the conditional entropy will be zero as no further information is necessary to describe \( X \).

Finally, the mutual information \( I(X; Y) \) can be written as:
\[
I(X; Y) = H(Y) - H(Y \mid X) = H(X) - H(X \mid Y)
\]
(4.13)

the mutual information \( I(X; Y) \) measures how much uncertainty of \( X \) is removed when \( Y \) is observed. If the two random variables are independent, then simply their mutual information is zero. Mutual information is also known as “a generalization of the linear
correlation coefficient” since no assumptions are made about the relationships between the variables (Tourassi et al., 2001).

In order to better understand the method proposed in this study, the conditional mutual information concept should be introduced. Suppose we have another variable, $Z$, in addition to $X$ and $Y$. Using above equation, mutual information between $X$ and $Y$ conditionally on $Z$ can be written as:

$$I(X; Y|Z) = H(X|Y) - H(X|Z;Y)$$

This equation computes how much information is shared between $X$ and $Y$, given the value of $Z$.

In other words, it is the difference in the entropy of $X$ given $Y$, and the conditional entropy of $X$ given both $Y$ and $Z$. If $Y$ and $Z$ carry the identical information on $X$, then the value of the conditional mutual information will be zero. The above equation becomes intractable as the number of random variables (or features) in a dataset increases. Fleuret (2004) proposed an estimation method to maximize conditional mutual information by dealing with the tradeoff between individual power and independence by comparing each new feature with the ones already picked. In their approach, a new feature, $X_n$ is good only if $I(X_n, Y|X)$ is high for every feature, $X$, already selected. Therefore a new feature is included in the model if it carries information about $Y$ not yet identified by any of the features already picked. This requires the computation of the following iterative scheme:

$$\nu(1) = \arg\max_{n} I(Y; X_n).$$

$$\forall k, 1 \leq k \leq K \quad \nu (k + 1) = \arg\max_{n} \{\min_{l \leq k} I(Y; X_n | X_{\nu(l)})\}$$
In the above scheme, the algorithm is initialized by picking the feature which maximizes mutual information $I(Y, X_n)$ as no feature selected yet to be conditioned on $Y$. Then, by taking the minimum of the conditional mutual information on all features already picked, one can be sure that a new feature is both informative and different than the preceding ones.

To implement this method, we needed to define a scoring vector $s$ size $N$. Initial values for the vector $s(n)$, were be calculated from the mutual information $I(n)$ values between the target variable and each candidate factor $n$. Then, for the first iteration, the factor with the highest score was be picked by the algorithm. For the following iterations, the scores were updated by taking the minimum of $s(n)$ and conditional mutual information $CMI(n, k)$ between target variable and the factor, $n$, for each of the selected factors, $k$. Finally, the feature with maximum updated $s(n)$ was picked at each cycle. The pseudocode for the algorithm is shown below (Fleuret, 2004):

For each candidate feature $1<n<N$
Calculate $s(n) = I(X_n, Y)$
For each iteration $1<i<N$
If First Iteration Then
Feature ($i$) = argmax($s$)
Else
For each candidate feature $1<n<N$
For All Existing Features $1<k<K$
Calculate $x(k) = CMI(X_n, Y, Feature(k))$
Update $s(n) = \min(\min(x), s(n))$
End If
Feature($i$) = argmax ($s$)
4.1.6 Tools for Building Bayesian Networks

Bayes Net Toolbox (BNT) is an open-source Matlab package for directed graphical models (Murphy, 2001). BNT supports many kinds of conditional probability distributions (CPD) for directed models such as Dirichlet priors for multinomial parameters, Wishart priors for variance/precision parameters, as well as Gaussian priors for weight matrices. Moreover, it is easy to implement user-defined CPDs. Unlike other available software for BNs, this software enables user to choose many algorithm for exact inference such as jtree, Pearl and Hugin/JLO. When exact solution is intractable, approximate inference can be drawn using methods such as sampling (Monte Carlo), variational methods or belief propagation in the software. The parameters can be also estimated (parameter learning) by software using the maximum likelihood (fully observed data) and maximum of posteriori (partially observed data) approaches. For structure learning, exhaustive search and K2 algorithm are implemented in the software. However, BNT Structure Learning Package (SLP), developed by Leray and Francois (2008) can be used together with BNT so as to make use of additional structure learning algorithms such as MWST, Naïve Bayes and augmented naïve Bayes, etc. Another important contribution to BNT, the Link Strength Package, is made by Ebert-Uphoff (2009). This package offers measures for link strength in Discrete Bayesian Networks, such as entropy, the strength of connection along a specific edge, true link strength and blind link strength.
4.2 Methodology

4.2.1 Structure Learning for the Estimation of Incident Duration Prediction Models

In this section, three structure learning algorithms were tested to discover the best BN for incident duration model for the integrated incident dataset. BN structures in this section are learned from Naïve Bayesian classifier, Tree Augmented Naïve Bayes (TAN) and K2 algorithm.

4.2.1.1 Description of Data

The data used in this study was created using two different data sources which were individual accident logs and New Jersey accident records. Accident logs were used to obtain the duration and number of vehicles involved in an incident. On the other hand, the second dataset also included same accidents with different set of parameters such direction, weather conditions, vehicle type, etc. Since these two datasets originated from different sources, even for the same incident, there was a small discrepancy between them, which required utilization of a matching procedure. Therefore, the data were combined using a simple matching algorithm. The steps are the following:

For Each Record in Incident Database(I)
Search for Matching Record in Accident Database(II) with following conditions
Date(I) = Date (II)
|MILEPOST(I) – MILEPOST(II)| <= 0.2 miles
|STARTTIME(I) – STARTTIME(II)| <= 15 minutes
Then match RECORD(I) to RECORD(II)
As a result of the above process, 1531 records from incident logs were matched with the accidents in the second database. Then, 10% of the data was set aside using random sampling prior to model development for validation. The description of the parameters in the combined dataset is given in Table 4.2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>State description</th>
</tr>
</thead>
<tbody>
<tr>
<td>VehNo</td>
<td>Number of vehicles involved</td>
<td>1-6</td>
</tr>
<tr>
<td>VehType</td>
<td>Type of vehicles involved</td>
<td>(1) single car; (2) single truck; (3) single bus; (4) single motorcycle; (5) car vs. car; (6) car vs. truck; (7) car vs. bus; (8) truck vs. truck; (9) truck vs. bus; (10) bus vs. bus; (11) multiple; (12) motorcycle vs. car; (13) motorcycle vs. truck; (14) motorcycle vs. bus; (15) cycle vs. cycle</td>
</tr>
<tr>
<td>Location</td>
<td>Location of incident</td>
<td>(1) open highway; (2) interchange; (3) deceleration lane; (4) acceleration lane; (5) median crossing; (6) eliz dual-dual; (7) Passaic bridge; (8) Ladderman bridge; (9) Hackensack East; (10) Hackensack West; (11) Susquehanna viaduct; (12) service area</td>
</tr>
<tr>
<td>Weather</td>
<td>Weather conditions</td>
<td>(1) clean; (2) rain; (3) snow; (4) cloudy; (5) fog, (6) other; (7):unknown</td>
</tr>
<tr>
<td>Pavement</td>
<td>Pavement conditions</td>
<td>(1) clear; (2) wet; (3) ice; (4) snow; (5) unknown</td>
</tr>
<tr>
<td>Light</td>
<td>Lighting conditions</td>
<td>(1) daylight; (2) dusk; (3) dark/lights; (4) dark/no lights</td>
</tr>
<tr>
<td>AccType</td>
<td>Type of incident</td>
<td>(1) fatal; (2) personal injury; (3) property damage</td>
</tr>
<tr>
<td>Roadway Damage</td>
<td>Presence of roadway damage</td>
<td>0,1 (binary)</td>
</tr>
<tr>
<td>Disabled Vehicle</td>
<td>Involvement of disabled vehicle</td>
<td>0,1 (binary)</td>
</tr>
<tr>
<td>NumInj</td>
<td>Number of injuries</td>
<td>0-7</td>
</tr>
<tr>
<td>NumFat</td>
<td>Number of fatalities</td>
<td>0, 1, 3</td>
</tr>
<tr>
<td>TimeofDay</td>
<td>Duration of incident</td>
<td>(1)Peak hours (6-9 a.m. and 4-7 p.m); (2) Off-peak hours (time of day except peak hours) 3-651 (minutes)</td>
</tr>
</tbody>
</table>
Since BNs work with tabular nodes, prior to model development, incident durations were discretized into 22 categories from 0-30 to 630-660 minutes. These data included incident specific parameters such as incident duration, number of injuries and incident type as well as the parameters related to environmental and roadway conditions. One advantage of predicting incident duration using this data might be the ease of initial estimation as some of these parameters can be identified such as weather conditions, lighting conditions, etc. as soon as the incident is reported. However, the parameters such as number of injuries, number of fatalities or roadway damage are uncertain and the information about these parameters is usually obtained after some time of the occurrence of an incident.

4.2.1.2 Modeling Results

The development of the BNs in this study required use of multiple software packages, hence, BNT and SLP software packages were employed since SLP for Bayes Net complement BNT by offering additional structure learning algorithms while using the basic functions of BNT (Leray and Francois, 2008). The combination of these software packages was able to handle the algorithms used in the present study as they had already been implemented in the packages as built-in functions.

Using the built-in functions of the software packages mentioned above, the following BNs (see Figure 4.2) were estimated:
Figure 4.2 BN structures from (a) Naïve Bayes, (b) TAN and (c) K2

- **Naïve Bayes Model**: This model has a fixed structure around the class node. Therefore, for this algorithm, structure learning is minimal as long as the class node, which is in our case incident duration, is defined. When the only requirement of the algorithm (defining a class node) is supplied, the other nodes are one-to-one connected to the class node. The resulting BN structure is shown in Figure 4.2 (a). According to this structure incident characteristics are conditionally independent of each other given the value of incident duration.
• **TAN (Tree Augmented Naïve Bayes) Model**: The TAN model is simply a modified version of the Naïve Bayes model where the relationships between variables are recovered using a structure learning algorithm. In our case, the mutual relationships between incident characteristics are recovered using the MWST algorithm. The resulting undirected tree can be oriented any chosen root node except class node (incident duration). In Figure 4.2 (b), it can be seen that the tree structure is oriented randomly around VehNo but note that any other node would result in the same structure as the MWST initially produced an undirected tree. As explained before, the relationships between variables recovered using the mutual information (e.g. mutual information between any variable pair, say X and Y, is the measure for the amount of uncertainty removed from X by knowing the state of Y or vice versa). Therefore, the BN structure generated from this algorithm may not only show causality between the variables as the better mutual information between variables does not always necessitate a cause-effect relationship.

• **K2 Model**: The K2 algorithm limits the search space to a subset of all possible networks based on the ordering of the variables. Since no expert opinion is assumed in this study, the initial search ordering is obtained by using the MWST algorithm. According to output of the MWST, the initial ordering of variables was: Duration > NumInj > Disabled Vehicle > Acc Type > NumFat > VehType > NJTPK damage? > Location > Light > TimeofDay > Pavement > Weather > VehNo. Where Duration variable accepts the most and VehNo accepts the least
number of variables as parents. This algorithm searches the entire DAG space (limited by variable order) and also gives room for additional input, mainly expert knowledge for variable ordering. Therefore, it may perform very well especially if expert knowledge is available. Figure 4.2 (c) show the BN created by the K2 algorithm. It can be observed from the figure that this BN is less complex than the previous BNs (less number of edges) and two variables were excluded from the other variables. Light depends on TimeofDay and these two variables are independent of the other variables in the BN.

To determine the best BN among the three BN structures, their BIC scores were calculated and these are shown in Table 4.3 (see section 4.1.2 for definition of BIC). As a result, K2 algorithm is selected to derive the BN for the incident durations. This algorithm resulted in a network with the highest score compared to the other for the structure learning algorithms, which were Naïve Bayes and TAN. The BIC scores also suggested that:

- the independence assumption for Naïve Bayes model was violated since the model scored the last among the three algorithm.
- by relaxing the independence assumption of Naïve Bayes, a BN, better representing the data, can be derived (TAN model)
- K2 achieved the best representation of the data even when employing a simpler structure as it employs a larger search space.
Table 4.3 BIC scores of Prediction Models

<table>
<thead>
<tr>
<th>Structure Learning Algorithm</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td>-15146</td>
</tr>
<tr>
<td>TAN</td>
<td>-13663</td>
</tr>
<tr>
<td>K2</td>
<td>-11861</td>
</tr>
</tbody>
</table>

The selected BN model for the incident duration prediction can be very effective for illustrating causal relationships between nodes which is neglected by classification trees. This structure makes it possible to predict incident duration even in those cases where a single parameter is known since any evidence provided in a node will result in updating the probability distributions of connected nodes. On the other hand, traditional methods such as classification trees require the a-priori knowledge of sequential information. Moreover, missing information in the sequence would result in the failure of being able to estimate incident duration due to the hierarchy involved in their structure.

CART are constructed in such a way that only one variable of interest can be estimated from the tree as opposed to flexible dynamic representation capability of BNs. As the probabilistic relationships between multiple variables are captured by BNs, there is no need to define a classification variable (Cooper and Herkovits, 1992). For example, K2 model can be used to predict the conditional probability distribution of the number of injuries in an incident providing the values of the parameters which are conditionally dependent on number of injuries (see Table 4.4).

At this point it is important to explain the conditional independency concept more thoroughly. Suppose that a probability distribution for the variable AccType is under
consideration in Table 4.4. AccType variable has three distinct parents which are NumFat, VehNo and Roadway Damage. This means that the probability distribution of AccType variable is “conditionally independent” of the variables other than NumInj, VehType and VehNo. Therefore, the probability distribution of this variable can be computed using the observed values of these three variables only. In the absence of the observed values of these variables that are connected to AccType variable, using the observed values of other variables, the probability distribution of these three variables can be constructed since the relationships in the BN can be reversed. This property also enables us to choose any of the parameters in the structure as decision variable as long as it is connected to the other parameters.

Table 4.4 Conditionally Dependent Parameters of Each Parameter in K2 Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Conditionally Dependent Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>VehNo</td>
<td>Roadway Damage</td>
</tr>
<tr>
<td>VehType</td>
<td>VehNo, Roadway Damage</td>
</tr>
<tr>
<td>Location</td>
<td>Roadway Damage</td>
</tr>
<tr>
<td>Weather</td>
<td>Roadway Damage</td>
</tr>
<tr>
<td>Pavement</td>
<td>Weather, Roadway Damage</td>
</tr>
<tr>
<td>Light</td>
<td>TimeofDay</td>
</tr>
<tr>
<td>AccType</td>
<td>NumFat, VehNo, Roadway Damage</td>
</tr>
<tr>
<td>Roadway Damage</td>
<td>-</td>
</tr>
<tr>
<td>Disabled Vehicle</td>
<td>-</td>
</tr>
<tr>
<td>NumInj</td>
<td>Disabled Vehicle, Acc Type, NumFat, VehNo, Roadway Damage</td>
</tr>
<tr>
<td>NumFat</td>
<td>VehNo, Roadway Damage</td>
</tr>
<tr>
<td>TimeofDay</td>
<td>-</td>
</tr>
<tr>
<td>Duration</td>
<td>NumFat, VehNo, Roadway Damage</td>
</tr>
</tbody>
</table>
4.2.1.3 Connection Strength of Variables in the Network

Using the link strength package and the BN learned from the K2 algorithm, the connection strengths between all variables in the network were calculated (see Table 3.5). The last column in Table 3.5 shows the connection strength of the incident characteristics to incident duration. For example, number of fatalities removes 0.42% uncertainty in the incident duration or number of vehicles involved in an incident removes 0.12% of the uncertainty in the duration variable. Of course, it is possible to investigate the connection strength of the other variables in the network using Table 3.5. For instance, when the vehicle type variable is known, 27.7% of the uncertainty in roadway damage can be removed.

Connection strength is a very useful property that can help researchers and others to better understand and interpret the BNs. Even if there is no direct edge (connection) between two variables in the network, how much they are affecting each other can be measured by simply calculating the connection strength. For example, although there is no direct edge between Location and Roadway Damage variables, apparently, 1.87% of the uncertainty in Roadway Damage variable can be recovered by identifying the location of an incident.
### Table 4.5 Connection Strength between the variables in K2 model in terms of MI%

<table>
<thead>
<tr>
<th>Parameters</th>
<th>VehNo</th>
<th>VehType</th>
<th>Location</th>
<th>Weather</th>
<th>Pavement</th>
<th>Light</th>
<th>AccType</th>
<th>Roadway Damage</th>
<th>Disabled Vehicle</th>
<th>NumInj</th>
<th>NumFat</th>
<th>TimeofDay</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>VehNo</td>
<td>-</td>
<td>52.2</td>
<td>3.59</td>
<td>3.38</td>
<td>1.95</td>
<td>0</td>
<td>0.19</td>
<td>27.7</td>
<td>0</td>
<td>0</td>
<td>26.6</td>
<td>0</td>
<td>0.12</td>
</tr>
<tr>
<td>VehType</td>
<td>66.6</td>
<td>-</td>
<td>2.6</td>
<td>2.45</td>
<td>1.51</td>
<td>0</td>
<td>0.08</td>
<td>27.2</td>
<td>0</td>
<td>0</td>
<td>8.91</td>
<td>0</td>
<td>0.03</td>
</tr>
<tr>
<td>Location</td>
<td>5.15</td>
<td>2.92</td>
<td>-</td>
<td>0.17</td>
<td>0.1</td>
<td>0</td>
<td>0.01</td>
<td>1.87</td>
<td>0</td>
<td>0</td>
<td>1.87</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Weather</td>
<td>2.5</td>
<td>1.42</td>
<td>0.09</td>
<td>-</td>
<td>60.1</td>
<td>0</td>
<td>0.01</td>
<td>0.98</td>
<td>0</td>
<td>0</td>
<td>1.35</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Pavement</td>
<td>1.64</td>
<td>0.99</td>
<td>0.06</td>
<td>68</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0.69</td>
<td>0</td>
<td>0</td>
<td>0.76</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Light</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>59</td>
<td>0</td>
</tr>
<tr>
<td>AccType</td>
<td>0.13</td>
<td>0.04</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0.02</td>
<td>0.01</td>
<td>55.3</td>
<td>92.8</td>
<td>0.38</td>
</tr>
<tr>
<td>Roadway Damage</td>
<td>18.6</td>
<td>14.3</td>
<td>0.88</td>
<td>0.88</td>
<td>0.55</td>
<td>0</td>
<td>0.02</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>2.3</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Disabled Vehicle</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0.18</td>
<td>0.04</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NumInj</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>94.2</td>
<td>0</td>
<td>8.94</td>
<td>-</td>
<td>6.71</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>NumFat</td>
<td>0.18</td>
<td>0.05</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0</td>
<td>0.92</td>
<td>0.02</td>
<td>0.01</td>
<td>0.04</td>
<td>-</td>
<td>0</td>
<td>0.42</td>
</tr>
<tr>
<td>TimeofDay</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>33.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>Duration</td>
<td>0.15</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0</td>
<td>0.68</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>76.3</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>

**MI% are calculated up to 2 significant figures**

### 4.2.1.4 Model Performance with Missing Data

In this section, we investigated the prediction accuracy of the BN in the case of randomly missing input data. Here, prediction accuracy is defined as the number of correctly predicted incident durations by the model, expressed as a percentage of the total number of samples in the data. First, prediction accuracy of the model is computed, in which incident duration is predicted, for the 10% of the data which is set aside using random sampling prior to model development. For the complete test dataset, K2 model predicted 53.02% of the incidents accurately. For the sake of completeness, the prediction accuracy of Naïve Bayes and TAN model were also calculated as 51.01% and 49.66%, respectively. Although the best performing BN was chosen previously using BIC scores
in this study, these values again confirmed that the K2 model represented the data best among all “structure learning algorithms” tested in this study.

One of the main problems in the real-time incident prediction is that some incident characteristics are unknown most of the time. Hence, to evaluate the performance of the BN model, some of the information is omitted randomly from the dataset. Performance of the model is evaluated by starting from complete data and in each following step, 10% of the information is randomly excluded from the dataset until half of the information is missing (see Table 4.6). The results show that model performance decays as the amount of missing information increases and is almost parallel to percentage of known information. However, the model still has the ability to predict well even with randomly missing information. This is a very important property that other static models lack.

Table 4.6 Performance of the BN model with randomly missing data

<table>
<thead>
<tr>
<th>Percentage of known data</th>
<th>100%</th>
<th>90%</th>
<th>80%</th>
<th>70%</th>
<th>60%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction Accuracy</td>
<td>53.02</td>
<td>44.97</td>
<td>40.94</td>
<td>36.91</td>
<td>30.87</td>
<td>26.17</td>
</tr>
<tr>
<td>Overall Performance</td>
<td>100%</td>
<td>84.81%</td>
<td>77.22%</td>
<td>69.62%</td>
<td>58.23%</td>
<td>49.37%</td>
</tr>
</tbody>
</table>

To compare the model developed in this study with CART, a CART tree was generated using the same data (see Figure 4.3). The ‘classregtree’ function of MATLAB was used to create the tree structure. ‘Prune’ option was used to obtain an optimal sequence of pruned subtrees and the number of observation to impure nodes to split was set to 250. As a result, the tree shown in Figure 4.3 was estimated. The prediction accuracy of the CART tree shown in Figure 4.3 was found to be 20.81%.
The CART tree cannot estimate the incident duration unless information is acquired sequentially. On the other hand, BN models can predict the probability of an accident to be within one of the pre-determined duration intervals with little available information. This is simple to explain because, in BNs, any evidence results in updating the probabilities of outcomes of other parameters that are connected to the variable representing the new evidence. Another advantage of BNs over the example classification tree and other regression based models is that it gives predictions in the form of a probability distribution of possible outcomes instead of a fixed outcome namely a point estimate. The probability of outcomes other than the one predicted can also be evaluated.
and more informed decisions can be made about response strategy based on the variability of outcome.

### 4.2.1.5 A Step-by-step Incident Duration Estimation Scenario for Incident Response

A three step incident scenario was constructed to demonstrate the use of the duration prediction models in incident response. The scenario included the following steps:

- **Step 1:** In this step, an incident is reported to TMC by commuters. Hence, the information is very limited and only location and number of vehicles involved in the incident is known from the verbal reports. Weather conditions are also assumed to be known since there is a high chance that this is monitored along the affected route by the TMC. Hence, at the end of the first step:
  - Location is set to 1 (open highway).
  - Number of vehicles involved is set to 2.
  - Weather parameter is set to 1 (clear).

  The initial prediction from three out of the four models incident duration was estimated as 30-60 minutes (Figure 4.4(a)).

- **Step 2:** After the incident is initially reported by the cars traveling by it, emergency teams are dispatched by response personnel and when they arrive to the incident site, they update response personnel with the following additional information also used by us to develop our duration models:
  - Lighting condition is set to 3 (dark/lights).
Road Damage is set to 0 (no damage).

Pavement is set to 1 (clear).

Number of injuries is set to 2.

As a result of the additional information, variability of incident duration predictions increased. While predicted incident duration remained at 30-60 minutes, its probability decreased slightly (see Figure 4.4(b)).

- Step 3: In the last step, the remaining information is reported to TMC and used to update the models to make the final duration estimation
  - Number of fatalities is set to 0 and since there were injuries type of incident is set to 2 (personal injury).
  - Type of vehicles involved in the incident set to 5 (car vs. car).
  - Involvement of disabled vehicle is set to 0.

After all parameters were entered to the model, the predicted incident duration was 30-60 minutes with more than 50% probability (Figure 4.4(c)). In fact, there were two incidents in the data with the same parameters and their durations 29 and 37 minutes which indicates incident duration interval estimated by the models are accurate.
(a) Step 1
(b) Step 2
In case of an incident scenario such as the one used to demonstrate the real-world application of BN models, a CART tree cannot estimate the incident duration unless information is acquired sequentially. On the other hand, BN models can predict the probability of an accident to be within one of the pre-determined duration intervals with little information availability. This is simple to explain because, in BNs, any evidence
results in updating the probabilities of outcomes of other parameters that are connected to the variable representing the new evidence. Another advantage of BNs over the example classification tree and other regression based models is that it gives predictions in the form of a probability distribution of possible outcomes instead of a fixed outcome namely a point estimate. The probability of outcomes other than the one predicted can also be evaluated and more informed decisions can be made about the response strategy based on the variability of outcome. For example, given the information in Step 1 and Step 2 of this incident scenario, the probabilities for number of fatalities using the BN model were 0.994 for no fatality, 0.006 for 1 fatality and 0 for 3 fatalities.

4.2.1.6 Feature Selection for Incident Variables

In this section, two different datasets are used to compare the effect of different features on different roadways. First, New Jersey Turnpike accident records, which were used in the previous section, were employed. Then, Emergency Service Patrol (ESP) logs were also employed for the major interstate highways including I-78, I-80, I-280 and I-287. The two databases have nine features in common. In this study, we created another feature called “accessibility” which is a criteria based on the distance to the closest exit on the roadway from the accident location. Table 4.7 shows the features of the two databases.
Table 4.7 Variables in ESP and NJ Turnpike accident databases

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>State description</th>
</tr>
</thead>
<tbody>
<tr>
<td>VehNo</td>
<td>Number of vehicles involved</td>
<td>1-6</td>
</tr>
<tr>
<td>Weather</td>
<td>Weather conditions</td>
<td>(1) clean; (2) rain; (3) snow; (4) cloudy; (5) fog; (6) other; (7):unknown</td>
</tr>
<tr>
<td>Pavement</td>
<td>Pavement conditions</td>
<td>(1) clear; (2) wet; (3) ice; (4) snow; (5) unknown</td>
</tr>
<tr>
<td>Light</td>
<td>Lighting conditions</td>
<td>(1) daylight; (2) dusk; (3) dark/lights; (4) dark/no lights</td>
</tr>
<tr>
<td>AccType</td>
<td>Type of incident</td>
<td>(1) fatal; (2) personal injury; (3) property damage</td>
</tr>
<tr>
<td>NumInj</td>
<td>Number of injuries</td>
<td>0-7</td>
</tr>
<tr>
<td>NumFat</td>
<td>Number of fatalities</td>
<td>0-3</td>
</tr>
<tr>
<td>TruckInv</td>
<td>Involment of a truck</td>
<td>(0) no, (1) yes</td>
</tr>
<tr>
<td>TimeofDay</td>
<td></td>
<td>(1)Peak hours (6-9 a.m. and 4-7 p.m); (2) Off-peak hours (time of day except peak hours)</td>
</tr>
<tr>
<td>Accessibility</td>
<td>Distance to the closest exit</td>
<td>0-1 miles: 1, 1-2 miles:2, 2-3 miles:3, 3-4 miles:4, 4-5 miles:5, 5-6 miles:6, 6-7 miles:7</td>
</tr>
<tr>
<td>Duration</td>
<td>Duration of incident</td>
<td>3-651 (minutes)</td>
</tr>
</tbody>
</table>

Figure 4.5 shows the Conditional Mutual Information Maximization (CMIM) values for each parameter across two databases and illustrates how CMIM scores vary across the databases. For each accident dataset, the feature selection algorithm was applied. For example, for ESP data, the selection algorithm starts with 0 variables and calculates the mutual information between decision (target variable) and each candidate variable. Then, mutual information values are ranked and the variable with the largest mutual information, Number of injuries, is chosen. This approach enables the selection of the most informative and relevant variable. Starting with the second variable, the algorithm changes its search strategy according to variables that have already been picked up. For each iteration cycle, the CMIM criterion for each variable in the candidate set and decision is calculated for each existing variable. The lowest CMIM value for the variable is picked and compared based on the calculated mutual information. If the additional information introduced by a variable is already captured or the information is irrelevant,
the CMIM value will be lower than the mutual information. The lowest of the CMIM and the mutual information is then updated as the score of the variable. Finally, the variable with the highest score is picked until all the variables are picked based on their influence on the decision. This process is performed recursively until all the variables are processed.

![CMIM Scores](image)

Based on the results presented in Figure 4.5, factors are ranked based on CMIM values and summarized in Table 4.8. The features that are ranked among the top-5 in both incident dataset were marked along with its ranking at corresponding dataset.

Table 4.8 Variable Rankings across Datasets (Top-5 Variables are Marked)

<table>
<thead>
<tr>
<th>Feature</th>
<th>NJ Turnpike</th>
<th>ESP</th>
</tr>
</thead>
<tbody>
<tr>
<td>VehNo</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Weather</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Pavement</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Light</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>AccType</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>NumInj</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NumFat</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>TruckInv</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>TimeofDay</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Accessibility</td>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>
An immediate finding based on the analysis is that the top-5 factors that affect incident duration shows similarities across datasets, but with varying rankings. The variables that are found to be at top-5 are listed below.

- Number of injuries (2/2)
- Accident type (2/2)
- Number of vehicles (2/2)
- Lighting conditions (2/2)
- Truck involvement (1/2)
- Accessibility (1/2)

One important common feature of these variables is most of them can be acquired at the initial stage of an incident. This fact is very useful for incident management professionals for incident duration estimation for those situations in which no other information is available. Although the set of top-ranking factors include similar variables, the ranking of some parameters shows sharp changes based on the dataset. In particular, accessibility variable was found to be very influential in New Jersey Turnpike data while it is ranked eighth for ESP data. This might stem from the fact that the toll roads such as the one discussed here, are limited access roads. On the other hand, the interstate roadways are more accessible compared to the toll roads. This fact can affect the response time of the emergency teams to who are sent to the incident site. The reason behind why the number of fatalities is ranked the least influential factor could be the scarcity of this type of incident in the data.
4.2.2 Adapting Learning in Bayesian Networks for Incident Duration Prediction

4.2.2.1 Data

In this section, individual incident logs and New Jersey Department of Transportation crash records in 2005 and 2011 were used by means of the matching strategy explained previously. 2221 records in 2005 and 1951 records in 2011 from incident logs were matched with the accidents in the second database. The outliers in the records in 2005 dataset are discarded using the method explained in Ramaswamy et al. (2000). In this method, the distance of a data point from \( k^{th} \) (\( k=5 \)) nearest neighbor is calculated and then the points are ranked by the distance from its \( k^{th} \) nearest neighbor and the top \( n \) points are identified as outliers (\( n = 247, 10\% \) of the data).

To better represent the incident conditions in the model, the list of variables is updated to include time and location related variables as shown in Table 4.9.

Incident durations were discretized into 4 categories: incidents lasting less than 30 minutes: low, incidents lasting between 30 to 60 minutes: medium, incidents lasting between 60 to 90 minutes: high, and incidents lasting more than 90 minutes: very high.

Table 4.10 shows the frequency of the records in those categories and summary statistics for incident duration data. From Table 2, it can be seen that the distribution of the incident durations in 2005 and 2011 are quite similar. The accessibility measure, distance, which indicates the distance of the incident location from the nearest exit in miles was also discretized by rounding up the distance to the nearest integer value.
Table 4.9 Description of the variables

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>State description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month</td>
<td>Month of year</td>
<td>0-12</td>
</tr>
<tr>
<td>DayofWeek</td>
<td>Day of week</td>
<td>(1) weekday; (2) weekend</td>
</tr>
<tr>
<td>TimeofDay</td>
<td>Time of day</td>
<td>(1) morning peak (6-9 a.m.); (2) afternoon off-peak (9 a.m.-4 p.m.); (3) evening peak (4-7 p.m.); (4) evening off-peak (7 p.m. - 6 a.m)</td>
</tr>
<tr>
<td>NumFat</td>
<td>Number of fatalities</td>
<td>0, 1, 3</td>
</tr>
<tr>
<td>NumInj</td>
<td>Number of injuries</td>
<td>0-12</td>
</tr>
<tr>
<td>CrshType</td>
<td>Type of crash</td>
<td>(1) rear end; (2) side swipe (same direction); (3) right angle; (4) head-on; (5) side swipe (opposite direction); (6) struck parked vehicle; (7) left/u turn; (8) backing; (9) encroachment; (10) overturned; (11) fixed-object; (12) animal; (13) pedestrian; (14) pedalcyclist; (15) non-fixed object</td>
</tr>
<tr>
<td>VehNo</td>
<td>Number of vehicles involved</td>
<td>1-6</td>
</tr>
<tr>
<td>Pavement</td>
<td>Pavement conditions</td>
<td>(1) dry; (2) wet; (3) snowy; (4) icy</td>
</tr>
<tr>
<td>Light</td>
<td>Lighting conditions</td>
<td>(1) daylight; (2) dawn; (3) dusk; (4) dark (street lights off); (5) dark (no lights); (6) dark (continuous lighting)</td>
</tr>
<tr>
<td>Weather</td>
<td>Weather conditions</td>
<td>(1) clear; (2) rain; (3) snow; (4) fog/smog/smoke</td>
</tr>
<tr>
<td>Roadway Damage</td>
<td>Presence of roadway damage</td>
<td>0,1 (binary)</td>
</tr>
<tr>
<td>NumTrkInv</td>
<td>Number of trucks involved</td>
<td>0-4</td>
</tr>
<tr>
<td>Location</td>
<td>Link on which incident occurred</td>
<td>1-28</td>
</tr>
<tr>
<td>Distance</td>
<td>Distance from the closest exit (in miles)</td>
<td>0-6.6</td>
</tr>
<tr>
<td>Duration</td>
<td>Duration of incident (in minutes)</td>
<td>3-762</td>
</tr>
</tbody>
</table>

Table 4.10 Summary statistics of the incident durations

<table>
<thead>
<tr>
<th>Year</th>
<th>0-30 min</th>
<th>30-60 min</th>
<th>60-90 min</th>
<th>90+ min</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>13.4%</td>
<td>51.9%</td>
<td>23.0%</td>
<td>11.7%</td>
<td>3</td>
<td>762</td>
<td>62.4</td>
<td>57.5</td>
</tr>
<tr>
<td>2011</td>
<td>13.5%</td>
<td>52.2%</td>
<td>22.7%</td>
<td>11.6%</td>
<td>5</td>
<td>615</td>
<td>60.8</td>
<td>48.7</td>
</tr>
</tbody>
</table>

4.2.2.2 Model Estimation and Validation

Three BNs were estimated using Naïve Bayesian classifier, Tree Augmented Naïve Bayes (TAN) as shown in Figure 4.6
To determine the best BN among these three BN structures, their BIC scores were calculated and shown in Table 4.11. The Bayesian Information Criterion (BIC) assesses the overall fit of a model and facilitates the comparison of BN models. For example, if we have two different models and there is no prior information on which model performs better, BIC identifies the model that is more likely to have generated the observed data. The smaller the value of the statistic (or the more negative the value is) the better the fit of the model.

Table 4.11 BIC scores and accuracies of Prediction Models

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>BIC</th>
<th>Prediction Accuracy (%)</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td>-39037</td>
<td>63.1±3.6</td>
<td>0.134</td>
<td>0.267</td>
</tr>
<tr>
<td>TAN</td>
<td>-37365</td>
<td>61.4±5.6</td>
<td>0.142</td>
<td>0.287</td>
</tr>
<tr>
<td>K2</td>
<td>-34281</td>
<td>60.9±4.9</td>
<td>0.149</td>
<td>0.306</td>
</tr>
</tbody>
</table>
When BIC scores of the models are considered, Naïve Bayes algorithm results in a network with the lowest BIC score compared to the other structure learning algorithms, which are TAN and K2. The BIC scores also suggest that:

- Although complete independence among the variables is assumed, Naïve Bayes model ranks the best among the three algorithms.
- When relaxing the independence assumption of Naïve Bayes, a better representation of the data cannot be achieved (TAN model).
- K2 achieves the worst representation of the data by employing a simpler structure and excluding the many variables which might help better representation of the data.

Table 4.11 gives the prediction accuracy of each model. Here, prediction accuracy is defined as the number of correctly predicted incident durations by the model; expressed as a percentage of the total number of samples in the data. Prediction accuracy of the models was computed using a 10-fold cross-validation method, in which incident durations are randomly partitioned to 10 equal size subsamples and one of the subsamples is kept for validation. Other samples were used for model estimation. This process was then repeated 10 times to produce prediction accuracy for the models in Table 4.11.

For the dataset, Naïve Bayes model predicts 63.1% of the incidents accurately on average. The prediction accuracy of TAN and K2 model were also calculated as 61.4% and 60.9%, respectively. The values of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) among the models also present a similar trend. Although the best performing BN was found previously using BIC scores, these values again confirmed that
the Naïve Bayes model represents the data best among all the models estimated using other “structure learning algorithms” tested in this study.

4.2.2.3 Real-time Prediction Performance and Adaptive Learning Using 2011 Data

One of the main problems in the real-time incident duration prediction models is that most of them are estimated using a particular dataset and the prediction performance of the model is only tested using a subsample of this data. Even if the subsample is not used in estimating the model, the data used in testing still belongs to the same sample and the same timeframe when the data is collected.

Hence, to evaluate the real performance of a BN model for future data, it is important to predict future incident durations using the model. In this study, 2011 NJ incident data was used for the prediction of the incident durations. Since the model was estimated using 2005 NJ incident data, there was a considerable time difference between the training and the prediction data. Therefore, the prediction data may reveal how the model behaves in the future.

Another related problem in the real-time incident duration prediction is that the data acquired in any real-time transportation system evolves over time. However, most models are static and they can only provide estimation based on archival data. Generally, it is not easy to tune model parameters by using the data acquired after model development. This may result in a decrease in the model’s performance over time as the model cannot adapt to the ever changing incident conditions and duration of real life situations.
In this present study, an adaptive learning scheme was introduced into the BN model, to overcome the problems pointed out above. Figure 4.7 demonstrates how adaptive learning occurs in the proposed BN prediction model. At t=0, the adaptive model is initialized by using the parameters of the base model. At the end of the first time interval t, the data is received and the adaptive model learns the new data. The scheme also keeps track of another instance of the model without learning the data for the current time interval t. At the end of next time interval t+1, both instance are used to predict incident durations in time interval t+1. If the learning in the previous time interval helped improve model performance (prediction accuracy), the parameters of the model are updated. The updated model is used for the next time interval and so on. On the other hand, if the learning data in the previous time interval does not increase the prediction accuracy, the parameters of the model are kept the same for the next time interval. Figure 4.8 shows the prediction results for the 2011 data provided by the adaptive learning scheme using the new data over the base model. First, month-to-month prediction accuracy performance of the base model and the model with adaptive learning is compared. For adaptive learning, the model is fed with the monthly data and the model parameters are automatically adjusted as long as the learning criteria are met. For the base model, the predictions are estimated without changing the initial parameters. The same performance analysis is done once again by dividing the data into quarters.
Figure 4.7 Adaptive Learning Mechanism in the BN model

Figure 4.8 Percent increase in prediction accuracy of the model with adaptive learning for monthly and quarterly 2011 data
The results indicated that:

- Adaptive learning can improve the prediction of the model by adjusting the model parameters when the new data is acquired. It can be seen from Figure 4.8 and Table 4.12 that the prediction accuracy of the base model is significantly improved after introducing adaptive learning to the base model.

- Even without adaptive learning and after the passage of 6 years since development of the model, BN models can still produce reasonable predictions (a 15% decrease in model performance).

- If the data is fed to the model monthly, the results appeared to fluctuate more than the case where the data was fed to the model quarterly (see Figure 4.8).

<table>
<thead>
<tr>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>April</td>
<td>May</td>
<td>June</td>
</tr>
<tr>
<td>Monthly</td>
<td>18.9</td>
<td>17.5</td>
</tr>
<tr>
<td>Quarterly</td>
<td>10.1</td>
<td></td>
</tr>
</tbody>
</table>

### 4.2.3 A Complete Adaptive Strategy for Estimating Incident Duration

Models developed using batch learning suffer from not being able to adapt to future conditions as it is not possible to incorporate the new sequential data to those models. One immediate remedy can be the updating of the model parameters as the new data arrives. Adapting the model in this way is quite easy. Demiroluk and Ozbay (2014)
discuss adapting the parameters of an incident duration model by using sequential data. On the other hand, adapting the model structure can be computationally expensive task as it requires the search of possible structures (Castillo and Gama, 2006).

It is not easy to adapt and improve a model to ever changing real life conditions. Due to changing conditions some of the data might be outdated and needs to be discarded. In time, even, the existing model may not represent the system accurately anymore. Widmer and Kubat (1996) named this problem, concept drift. They explain that some concept of interest may depend on a hidden context and changes in the hidden context may result in unexpected changes (gradual or abrupt) in the target concept (Widmar and Kubat, 1996).

While handling concept drift, it is important to distinguish between noise in the data and actual concept drift. Some algorithms might be too sensitive to noise and treat it as a drift, on the other hand, some algorithms are very robust to noise which makes difficult to adapt to changes. An efficient algorithm should resist unnecessary changes in the model due to noise without compromising its flexibility to adapt to drift (Widmer and Kubat, 1996). There are many articles on handling concept drift using different learning algorithms: rule-based learning (Schlimmer and Granger, 1986; Widmer and Kubat 1996), decision trees (Harries et al., 1998), and Bayesian classifiers (Castillo and Gama, 2006).

4.2.3.1 Data

1952 records in 2011 and 1709 record in 2012 New Jersey incident data are used for model development. Table 4.13 shows the summary statistics of the data. Details of
the variables are not discussed here since the same variables as described in Section 4.2.2 were used.

Table 4.13 Summary statistics of the incident data

<table>
<thead>
<tr>
<th>Year</th>
<th>Frequency</th>
<th>Duration (minutes)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0-30 min</td>
<td>30-60 min</td>
<td>60-90 min</td>
</tr>
<tr>
<td>2011</td>
<td>13.5%</td>
<td>52.2%</td>
<td>22.7%</td>
</tr>
<tr>
<td>2012</td>
<td>14.4%</td>
<td>53.2%</td>
<td>20.6%</td>
</tr>
</tbody>
</table>

Prior to model development, the Conditional Mutual Information Maximization approach proposed by Fleuret (2004) was used for finding the best features that represent the data. In order to select an optimal subset of features, an evaluation function is employed by feature selection methods. The approach used in this study which is called Conditional Mutual Information Maximization (CMIM) requires information measures for the evaluation function (Fleuret, 2004). Generally, these measures determine information gain from a feature. CMIM involves selecting features which maximize their mutual information with the class (target variable) to predict, conditionally the response of any feature already selected. This criterion discards the features similar to those already selected. Discarded features may be powerful individually; however they do not carry additional information (relative to the features that are already picked) about the class to predict. To select the optimum number of features, the features are backward eliminated based on Naïve Bayes classifier results. Figure 4.9 shows the performance trend that was observed by eliminating features one-by-one.
Based on the model performance, it is concluded that eliminating three variables (as shown in Table 4.14) with the lowest rank will not degrade the performance of the model.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Feature</th>
<th>Rank</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of Trucks Involved</td>
<td>8</td>
<td>Weather</td>
</tr>
<tr>
<td>2</td>
<td>Location</td>
<td>9</td>
<td>Number of Vehicles Involved</td>
</tr>
<tr>
<td>3</td>
<td>Crash Type</td>
<td>10</td>
<td>Pavement</td>
</tr>
<tr>
<td>4</td>
<td># of Injuries</td>
<td>11</td>
<td>Time of Day</td>
</tr>
<tr>
<td>5</td>
<td># of Fatalities</td>
<td>12</td>
<td>Workzone</td>
</tr>
<tr>
<td>6</td>
<td>Distance</td>
<td>13</td>
<td>Roadway Damage</td>
</tr>
<tr>
<td>7</td>
<td>Light</td>
<td>14</td>
<td>Day of Week</td>
</tr>
</tbody>
</table>

**4.2.3.2 Model**

In this section, an adaptive online learning model which incorporates parameter and structure learning for BNs in a system where the data is received sequentially is
discussed. Castillo and Gama (2009) developed an adaptive framework for BNs for this purpose. Their method has two main components: performance management and handling concept drift.

Their method starts with a simple Naïve Bayes structure since at the beginning there are few data points and it is suggested that the more complex models do not perform better than Naïve Bayes with few training data (Castillo and Gama, 2009). Hence, only parameter learning occurs at this stage. As the size of the training data increases, the algorithm searches for more complex structures to recover the dependencies between the model variables. They use the idea behind k-dependence Bayesian networks (Sahami, 1996) to automatically adapt the model for more or less complex network structures depending on the conditions. Due to computational difficulties, instead of Sahami’s algorithm, they used more simplistic hill-climbing algorithm to recover the dependencies in the network structure. Here, their algorithm was extended by using Sahami’s algorithm.

Figure 4.10 shows the flow of the algorithm. First, the data arrive to a system in batches at each time step. For a defined BN structure and its parameters, the algorithm “sequentially predicts the classes of the next batch”. For each batch of examples, the current model is used for prediction, the correct class is observed and some performance indicators are assessed. Then, the indicator values are used to estimate the actual system’s state. Finally, the model is adapted according to the estimated state. The controlling mechanisms in the model select the best adaptive actions according to the current learning goal. To this end, two performance indicators are monitored over time: batch
error and model error. Batch error is defined as the proportion of misclassified examples in a batch. Model error is the proportion of misclassified examples in the total of the examples that were classified using the same structure. The possible states of the systems are defined in Table 4.15.

Figure 4.10 Adaptive modeling framework for Bayesian Networks
Table 4.15 Model states (Castillo and Gama, 2009)

<table>
<thead>
<tr>
<th>State</th>
<th>Abbreviation</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>is improving</td>
<td>S1</td>
<td>model performance is improving</td>
</tr>
<tr>
<td>stop improving</td>
<td>S2</td>
<td>model performance stops improving</td>
</tr>
<tr>
<td>concept drift alert</td>
<td>S3</td>
<td>possible concept change is detected for the first time</td>
</tr>
<tr>
<td>concept drift</td>
<td>S4</td>
<td>gradual concept change in the model</td>
</tr>
<tr>
<td>concept shift</td>
<td>S5</td>
<td>abrupt concept change in the model</td>
</tr>
<tr>
<td>stable performance</td>
<td>S6</td>
<td>model performance at a plateau</td>
</tr>
</tbody>
</table>

For testing the algorithm, 2011 and 2012 New Jersey incident data divided to monthly sequential batches and the adaptive algorithm run using the data. Figure 4.11 shows the results of adaptive algorithm. The figure also presents the performance of static modes: Naïve bayes and k-dependence Bayesian classifiers (k=1, 2, 3). The static models are built using 2011 incident data and 2012 incidents are predicted using the models.

Figure 4.11 Model performance of Adaptive learning framework vs. static models
From Figure 4.11 Model performance, it can be seen that performance of the BN with integrated adaptive learning improves over time. While the slope, which is the indication of the degree of the improvement in performance, is steep in 2011, the slope becomes flatter as the learning slows down in 2012. This shows that the model stabilizes with increasing number of training samples. This might also be due to the fact that no structure change occurred during the training. It was also observed that the performance of the static models decayed over time. Based on the results, the performance of the static BNs can be improved by implementing the adaptive learning scheme explained in this section.
CHAPTER 5. SPATIAL MODELS FOR ANALYZING AND MAPPING THE LOCATION WITH HIGHER CRASH RISKS FOR INCIDENT MANAGEMENT

5.1 A County Level Model

In this study, a new hierarchical Bayesian model was developed for purposes of spatial crash modeling in which the severity weights were applied to crashes with different severities to obtain a weighted crash count. Our main objective was to compare the traditional hierarchical model to a weighted model to better understand and to demonstrate the importance of representing the spatial variation of crashes as well as their severity. Spatial covariates related to the roadway conditions were used in the model development and due to limitations of the data surrogate variables were employed. First, the model was estimated using raw crash counts. Most studies lack an analysis of crash rates by road type using hierarchical Bayesian models that can capture spatio-temporal characteristics of crashes. To address this gap in the literature, additional models were estimated for roadways under different jurisdictions such as the department of transportation, authority, county, and municipality. Finally, crash rate maps were developed based on modeling results to help visualize the effects of the spatial covariates.

5.1.1 Data

In this study, New Jersey Department of Transportation (NJDOT)’s crash records database were used for developing the crash risk model (NJDOT, 2012). NJDOT keeps
this yearly database containing five different tables: accident, driver, vehicle, pedestrian, and occupant. While accident table includes the general information about these accidents, other tables such as occupant table contains detailed information on each occupant of the vehicle(s) included in the crash. The covariates considered in this study (ratio of crashes related to roadway defects, ratio of crashes occurred on the curves, and ratio of wet weather crashes) are contained in accident and vehicle tables. Severity was factored in as the most severe injury in the crash and was gathered from the physical condition field in the occupant table. In the database, physical condition of the victim is coded as complaint of pain, moderate injury, incapacitated, and killed.

New Jersey has 21 counties and the location of each crash is available at a county level for all crashes in the state. All crash records from 2001 to 2010 were used in the development of the model. Figure 5.1 (a) shows the raw annual crash counts by county in 2010. As seen from the figure, the counties in the central and northern part of New Jersey, close to the New York metropolitan area, have higher crash counts than the rest of the state. This could be a result of the highly developed roadway network in that region (see Figure 5.2). NJDOT publish roadway miles and daily vehicle miles traveled (DVMT) in each county for each year on their website. DVMT is considered as the exposure variable in this study. The crash rate map was developed using crash counts and daily vehicle miles traveled (DVMT) in each county, as shown in Figure 5.1 (b). From this crash rate map, it can be observed that the crash count in a county is highly related to exposure to traffic (DMVT).
Figure 5.1 (a) Annual crash counts by county in 2010 (b) Crash rates by county in 2010 (per thousand DVMT)

Figure 5.2 New Jersey road network (except surface streets)
In this study, the goal was to develop a crash model that only depends on spatially varying roadway characteristics. This task was challenging due to the limitations of the crash data since this did not include any spatial covariates. Only covariate directly available from NJDOT is the roadway mileage in each county. Instead, surrogate variables approach, similar to the approach proposed in (Miaou et al., 2003), is employed to represent the spatial covariates. The proportion of curve crashes were considered as surrogate variables to represent the number of horizontal curves. The detailed roadway data which includes the number of horizontal curves on the roadway was not available. On the other hand, when a crash occurs, the road characteristics were coded as straight or curve. Therefore, the ratio of the crashes that occurred on the curves was included as a surrogate variable. The detailed weather data was also not available; however, when a crash occurs, the road surface condition was recorded. The proportion of crashes that occurred on a wet roadway surface condition was used to indicate the percentage of time that the roadway was wet due to rain or snow, ice, etc. Another covariate considered in the study was the number of roadway hazards in each county. As with the other covariates, this data was not available. However, when a crash occurred the apparent contributing circumstances were recorded in the database in one of four categories: driver actions, vehicle factors, pedestrian factors, and roadway/environmental factors. Roadway factors included obstruction/debris in road, physical obstructions, etc. Surrogate variable proportion of crashes related to roadway defects was devised to indicate the number of roadway defects in a county. The descriptive statistics of the variables included in the study is given in Table 5.1.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>All crashes</td>
<td>Frequency of all crashes</td>
<td>14716.6</td>
<td>1718</td>
<td>57047</td>
<td>11429.8</td>
</tr>
<tr>
<td>Fatal</td>
<td>Frequency of fatal crashes</td>
<td>30.33</td>
<td>5</td>
<td>70</td>
<td>14.96</td>
</tr>
<tr>
<td>Incapacitating</td>
<td>Frequency of incapacitating injury crashes</td>
<td>22.6</td>
<td>4</td>
<td>57</td>
<td>11.2</td>
</tr>
<tr>
<td>Injury</td>
<td>Frequency of other injury crashes</td>
<td>3504.5</td>
<td>424</td>
<td>10041</td>
<td>2276.7</td>
</tr>
<tr>
<td>Property damage</td>
<td>Frequency of property damage crashes</td>
<td>11159.1</td>
<td>1268</td>
<td>27723</td>
<td>7226.15</td>
</tr>
<tr>
<td>DVMT</td>
<td>Daily vehicle miles traveled (in thousands)</td>
<td>9486.7</td>
<td>2160</td>
<td>21587</td>
<td>5367.5</td>
</tr>
<tr>
<td>Roadway mileage</td>
<td>Total length of roadways (in miles)</td>
<td>1823.34</td>
<td>620</td>
<td>3509</td>
<td>758.09</td>
</tr>
<tr>
<td>Wet roadway</td>
<td>Proportion of wet roadway crashes</td>
<td>0.24</td>
<td>0.12</td>
<td>0.38</td>
<td>0.05</td>
</tr>
<tr>
<td>Curve crashes</td>
<td>Proportion of curve crashes</td>
<td>0.13</td>
<td>0.06</td>
<td>0.3</td>
<td>0.05</td>
</tr>
<tr>
<td>Roadway Defects</td>
<td>Proportion of roadway related crashes</td>
<td>0.08</td>
<td>0.02</td>
<td>0.25</td>
<td>0.05</td>
</tr>
</tbody>
</table>

5.1.2 Model

In this study, the hierarchical Bayesian generalized linear model was considered for estimating the model. The full hierarchical Bayesian modeling requires a three step approach. For the first step, conditional on mean, $\mu_i$, weighted crash counts, $Y_i$ was assumed to follow a Poisson distribution:

$$Y_i \sim Po(\mu_i)$$

(5.1)

where $Y_i$ was the observed number of crashes in county $i$ at time $t$ (years), $i=1,...,N$ and $t=1,...,T$; and $\mu_i$ was the mean of the Poisson process for county $i$ at time $t$. The mean was further formulated as:

$$\mu_i = e_i \lambda_i$$

(5.2)
In the above equation, crash risk rate, $\lambda_{it}$, was assumed to be proportional to traffic exposure, $e_{it}$, hence DVMT was considered as an offset. Then, the crash risk rate, $\lambda_{it}$, was modeled as:

$$
\log(\lambda_{it}) = \alpha + \sum_k \beta_k X_{itk} + \theta_i + \phi_i
$$

(5.3)

where $\alpha$ represented the intercept term, $\beta_k$ was the coefficient for spatial covariate $k$, $X_{itk}$ was the observed value of $k$th covariate for county $i$ at time $t$, $\theta_i$ stands for region-wide or statewide global heterogeneity, and $\phi_i$ represented spatially correlated random effects for county $i$.

Next, coefficients were modeled by using non-informative priors. For the random effect terms, second level of hierarchy is defined. The heterogeneity term was modeled using a normal prior:

$$
\theta_i \sim \text{N}
\left(0, \frac{1}{\tau_h}\right)
$$

(5.4)

where $\tau_h$ was a precision parameter that controlled the amount of $\theta_i$ among municipalities. The heterogeneity term enabled us to include extra-Poisson variability due to unobserved variables over the entire state. For modeling the spatially correlated random effects, a CAR prior was adopted, proposed by Besag (1974):

$$
\phi_i \mid \phi_j \sim \text{N}
\left(\sum_{j \neq i} \frac{w_{ij}}{w_c} \phi_j, \frac{1}{\tau_c w_c}\right)
$$

(5.5)
where $\tau_c$ was a precision parameter that controls clustering, $ij$ was a neighbor municipality adjacent to municipality $i$, $w_{ij}$ was the weight of the neighbor $j$, and $w_i$ represented the sum of the weights of the neighbors of municipality $i$. Note that, in this study, it was assumed that all neighbors had equal weight. By including a spatially correlated random effects term, extra-Poisson variability in the log-relative risk which varies from municipality to municipality could be modeled in such a way that nearby municipalities would have more similar rates.

In the next step, the hyperpriors were defined for structured random effect terms which were heterogeneity and clustering precision parameters. Setting the same priors for $\tau_h$ and $\tau_c$ to give same prior emphasis on them was incorrect for two reasons. First, $\tau_h$ uses a marginal specification, whereas, $\tau_c$ is specified conditionally. Second, $\tau_c$ is multiplied by number of neighbors before it is involved in prior specification. Therefore, hyperpriors for the precision parameters, $\tau_h$ and $\tau_c$, were defined based on a fair priors criteria approximated by Bernardinelli et al. (1995) as follows:

$$sd(\theta) = \frac{1}{\tau_h} \approx \frac{1}{0.7\sqrt{\bar{w}\tau_c}} \approx sd(\phi)$$

(5.6)

where $sd(\theta)$ and $sd(\phi)$ are standard deviations of $\theta_i$ and $\phi_i$ respectively, and $\bar{w}$ average number of fair neighbors. The graphical representation of the hierarchical models is shown in Figure 5.3.
The proportion of variability in the random effects is another metric which was included in the modeling to analyze clustering later on and this was defined as:

$$\psi = \frac{sd(\phi_i)}{sd(\phi_i) + sd(\theta_i)}$$  \hfill (5.7)

Several parameterization approaches of the hierarchical model were considered in the study and to evaluate the model fit, the deviance information criterion (DIC) was used. DIC was proposed by Spiegelhalter et al. (2002) to compare the fit and complexity of hierarchical models in which the number of parameters is not clearly defined. DIC is calculated using the posterior distribution of deviance and can be considered as more general form of Akaike’s information criterion (AIC). DIC is defined as in the following equation:
\[ DIC = D(\bar{\theta}) + 2p_D \] (5.8)

where \( D(\bar{\theta}) \) is the classical estimate of fit (posterior mean of deviance) and \( p_D \) is the effective number of parameters. It should be noted that similar to AIC, the models with lower DIC are preferred.

### 5.1.3 Discussion of Results

The hierarchical Bayesian model, which takes into account spatial correlation and unobserved heterogeneity, was estimated using the software package called WinBUGS (2003). WinBUGS uses MCMC methods to sample from the joint posterior probability distribution of the multiple variables. Starting with only an intercept, different subsets of variables were considered in the model development. Multiple chains were simulated for these models. 50,000 iterations were performed for each model and the first 10,000 iterations were considered as burn-ins and were not included in the sampling. The model with the lowest DIC was selected. The convergence of the model was evaluated using the mixing of MCMC chains and Gelman-Rubin statistics in WINBUGS.

Although MCMC methods are very effective in the estimation of hierarchical models, sometimes they can be very slow to converge due to the high correlation between parameters or due to non-informative priors (Gelfand et al., 1995). Several methods are proposed to overcome this issue and increase the efficiency of MCMC estimation of hierarchical models such as hierarchical centering (Gelfand et al., 1995), orthogonal
polynomials (Hills and Smith, 1992) and parameter expansion (Liu et al., 1998). In order to improve convergence of the model in this study, covariates were standardized (i.e. mean 0 and standard deviation 1) and hierarchical centering was used for assessing random effects. After these two procedures, the MCMC chains are mixed well and the convergence is reached rather quickly based on the visual inspection of monitoring plots in WINBUGS.

Table 4.2 shows the estimated hierarchical crash risk model for raw county level crash counts. All variables considered in the study were found to be statistically significant. The results indicate that based on the estimates of covariate effects, the “curve crashes” are the most significant variable in explaining crash risk over space. This result confirms the findings of Miaou et al. (2003) in which curve crashes also yielded the most significant covariate effects. “Roadway mileage” was found to be the second most influential variable. This result shows the general trend that as the roadway mileage increases in a county, if all other variables are controlled in the model, the number of crashes increases. Previous research found that the roadway mileage increases crash risk (Aguero-Valverde and Jovanis, 2006; Quddus, 2008). “Roadway defects” are found to be the third most influential variable. This result implies that although a surrogate variable was used in the study to represent the number of roadway defects, roadway defects elevate the crash risk. The least influential variable in the model was found to be wet roadway. This was the case in Aguero-Valverde and Jovanis (2006) as well. However, this might stem from the fact that during the wet weather, traffic volume might be lower.
Furthermore, drivers might be more careful and potentially drive more slowly during the wet conditions.

Table 5.2 Modeling results of hierarchical model for raw crash counts

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D.</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.17730</td>
<td>0.01619</td>
<td>0.00006</td>
<td>0.17730</td>
</tr>
<tr>
<td>Roadway mileage</td>
<td>0.07259</td>
<td>0.00834</td>
<td>0.00037</td>
<td>0.07207</td>
</tr>
<tr>
<td>Curve crashes</td>
<td>0.11170</td>
<td>0.00667</td>
<td>0.00021</td>
<td>0.11170</td>
</tr>
<tr>
<td>Wet roadway</td>
<td>0.02788</td>
<td>0.00195</td>
<td>0.00004</td>
<td>0.02787</td>
</tr>
<tr>
<td>Roadway defects</td>
<td>0.05564</td>
<td>0.00065</td>
<td>0.00001</td>
<td>0.05564</td>
</tr>
<tr>
<td>( \phi )</td>
<td>0.47370</td>
<td>0.01637</td>
<td>0.00077</td>
<td>0.47580</td>
</tr>
<tr>
<td>( \theta )</td>
<td>0.06362</td>
<td>0.02647</td>
<td>0.00129</td>
<td>0.06010</td>
</tr>
<tr>
<td>( \psi )</td>
<td>0.88320</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \bar{D} : 2554.63, D(\bar{\theta}) = 2343.62, p_D = 24.816, DIC = 2765.65 \]

The model also included terms that captured the global heterogeneity and spatial correlation. The spatial correlation was significantly higher than the global heterogeneity. The standard deviation of \( \sigma_\phi \) was 0.4737 while the standard deviation of \( \sigma_\theta \) was 0.06362. Based on the proportion of variability in the spatial random effects, \( \psi \), it can be implied that there was a strong spatial correlation in crash risk between neighboring counties.

In the literature, there is a lack of analysis of crash rates according to road types using this kind of hierarchical Bayesian models. To address this gap in the literature, crashes were classified by the type of roadways under different jurisdictions to re-estimate the model. In New Jersey, there are four levels of jurisdiction for roadways: state (NJDOT), authority (toll roads), county, and municipal. Although the roadway
jurisdiction is provided in the crash records, DVMT for the roadways was not available; therefore, the county DVMT values were used in the each model. On the other hand, the roadway mileage by jurisdiction was available and was used for model estimations. Six counties were excluded from the analysis of authority roadways since the length of authority roadways was less than five miles. Table 5.3 shows the estimated hierarchical crash risk model for roadways by jurisdiction. As expected, the effects of the contributing factors differ by the roadway jurisdiction. While roadway mileage is the most influential factor on state and authority roadways, it is the second most influential factor on county and municipal roadways. Moreover, it is only negatively related to crash risk on state roadways. This result might be related to the low average speed of vehicles due to high congestion on the state roadways. The curve crash was also another significant variable for all roadway jurisdictions. However, it was positively related to the crash risk for state roadways. This might suggest that more safety precautions are needed in the vicinity of sharp curves on state roadways. The effect of wet roadways was found to be significant and positive on authority roadways, and municipal roadways. This can be attributed to higher speed limits on authority roadways. On the other hand, municipal roadways are generally two-lane roadways and there is only a small margin for error for drivers when the roadway is wet. The roadway defect was found to be a factor for all road types except for county roadways. Based on the proportion of variability in the spatial random effects, $\psi$, the spatial correlation was strong for all jurisdictions. The highest value of $\psi = 0.82$ occurred for municipal roadway, which indicate that they are highly clustered in the neighboring regions.
Table 5.3 Modeling results of hierarchical model for crashes by jurisdiction

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D.</th>
<th>2.50%</th>
<th>97.50%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>State</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.9493</td>
<td>0.03721</td>
<td>-1.026</td>
<td>-0.8723</td>
</tr>
<tr>
<td>Roadway mileage</td>
<td>-0.2091</td>
<td>0.02726</td>
<td>-0.2651</td>
<td>-0.1549</td>
</tr>
<tr>
<td>Curve crashes</td>
<td>0.03301</td>
<td>0.006056</td>
<td>0.0212</td>
<td>0.04499</td>
</tr>
<tr>
<td>Wet roadway</td>
<td>0.003339</td>
<td>0.003007</td>
<td>-0.00255</td>
<td>0.009248</td>
</tr>
<tr>
<td>Roadway defects</td>
<td>0.03294</td>
<td>0.006668</td>
<td>0.01983</td>
<td>0.04603</td>
</tr>
<tr>
<td>( \sigma_\phi )</td>
<td>0.24</td>
<td>0.04201</td>
<td>0.1567</td>
<td>0.3136</td>
</tr>
<tr>
<td>( \sigma_\theta )</td>
<td>0.1508</td>
<td>0.05595</td>
<td>0.03403</td>
<td>0.2423</td>
</tr>
<tr>
<td>( \psi )</td>
<td>0.6206</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Authority</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.409</td>
<td>0.06581</td>
<td>-2.554</td>
<td>-2.266</td>
</tr>
<tr>
<td>Roadway mileage</td>
<td>0.2218</td>
<td>0.03592</td>
<td>0.1517</td>
<td>0.2914</td>
</tr>
<tr>
<td>Curve crashes</td>
<td>-0.03364</td>
<td>0.006284</td>
<td>-0.04597</td>
<td>-0.0213</td>
</tr>
<tr>
<td>Wet roadway</td>
<td>0.009369</td>
<td>0.004767</td>
<td>1.28E-05</td>
<td>0.01861</td>
</tr>
<tr>
<td>Roadway defects</td>
<td>0.03012</td>
<td>0.011</td>
<td>0.008366</td>
<td>0.05154</td>
</tr>
<tr>
<td>( \sigma_\phi )</td>
<td>0.5563</td>
<td>0.09232</td>
<td>0.2789</td>
<td>0.6597</td>
</tr>
<tr>
<td>( \sigma_\theta )</td>
<td>0.1935</td>
<td>0.1356</td>
<td>0.02669</td>
<td>0.5039</td>
</tr>
<tr>
<td>( \psi )</td>
<td>0.7539</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>County</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.011</td>
<td>0.03142</td>
<td>-1.079</td>
<td>-0.9433</td>
</tr>
<tr>
<td>Roadway mileage</td>
<td>0.06992</td>
<td>0.003202</td>
<td>0.06366</td>
<td>0.07622</td>
</tr>
<tr>
<td>Curve crashes</td>
<td>-0.1006</td>
<td>0.0129</td>
<td>-0.1261</td>
<td>-0.07604</td>
</tr>
<tr>
<td>Wet roadway</td>
<td>-0.00325</td>
<td>0.003561</td>
<td>-0.01025</td>
<td>0.003712</td>
</tr>
<tr>
<td>Roadway defects</td>
<td>-0.02149</td>
<td>0.01428</td>
<td>-0.04959</td>
<td>0.006426</td>
</tr>
<tr>
<td>( \sigma_\phi )</td>
<td>0.3563</td>
<td>0.04159</td>
<td>0.2459</td>
<td>0.4096</td>
</tr>
<tr>
<td>( \sigma_\theta )</td>
<td>0.1132</td>
<td>0.07493</td>
<td>0.02235</td>
<td>0.2929</td>
</tr>
<tr>
<td>( \psi )</td>
<td>0.7701</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Municipal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.05</td>
<td>0.03888</td>
<td>-1.135</td>
<td>-0.9675</td>
</tr>
<tr>
<td>Roadway mileage</td>
<td>0.2073</td>
<td>0.01293</td>
<td>0.1818</td>
<td>0.2321</td>
</tr>
<tr>
<td>Curve crashes</td>
<td>-0.5059</td>
<td>0.01428</td>
<td>-0.5349</td>
<td>-0.4784</td>
</tr>
<tr>
<td>Wet roadway</td>
<td>0.03914</td>
<td>0.004267</td>
<td>0.03075</td>
<td>0.04746</td>
</tr>
<tr>
<td>Roadway defects</td>
<td>-0.0473</td>
<td>0.009711</td>
<td>-0.06643</td>
<td>-0.02846</td>
</tr>
<tr>
<td>( \sigma_\phi )</td>
<td>0.6068</td>
<td>0.04957</td>
<td>0.4705</td>
<td>0.6754</td>
</tr>
<tr>
<td>( \sigma_\theta )</td>
<td>0.1394</td>
<td>0.09315</td>
<td>0.02308</td>
<td>0.3561</td>
</tr>
<tr>
<td>( \psi )</td>
<td>0.8212</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Not all crashes have the same impact. More severe crashes result in higher costs and potential fatalities. Consequently, instead of only using a frequency based approach, it is possible to assign different weights to crashes with varying severity. FHWA Five Percent Report proposes a severity weighing scheme for crashes in which the weights for fatality are 15, incapacitating injury 7, non-incapacitating injury 4, possible injury 2 and property damage is only 1 (FHWA, 2012). Although the theoretical background behind the weighing scheme is not explained in the report, it suggests a guideline for weighing the crashes according to their severity. The severity weights in this report were adopted to re-estimate the previous model. Since non-incapacitating injury and possible injury crashes were not differentiated in the data, for all injury crashes, except incapacitating injury, 3 was assigned as the weight. Table 5.4 shows the results of modeling when the severity weights are applied to crash counts. The results indicate a similar trend to the model based on raw crash counts. Note that the spatial correlation decreased in this case while global unobserved heterogeneity increased. This might suggest that there are other unobserved underlying factors affecting the severity of the crashes analyzed.

Figure 5.4 shows the crash rate maps for 2010 which were developed based on the model findings of the mean crash rates from raw crash counts (Figure 5.4 (a)) and from the crash counts when the severity weights were applied (Figure 5.4 (b)). In Figure 5.4 (a), the natural breaks in Figure 5.1 (b) are used to make the comparison between the

\[1 \bar{D}: 6219.92, D(\bar{D}) = 6195.14, p_D = 24.783, DIC = 6244.71\]
\[2 \bar{D}: 2715.11, D(\bar{D}) = 2696.15, p_D = 18.954, DIC = 2734.06\]
\[3 \bar{D}: 9977.43, D(\bar{D}) = 9952.36, p_D = 25.066, DIC = 10002.5\]
\[4 \bar{D}: 12066.5, D(\bar{D}) = 12041.4, p_D = 25.079, DIC = 12091.6\]
crash rates from the model and the real data easy to read. It can be observed that the model estimates are proximate to the real crash rates, which confirms the fit of the model to the data.

Table 5.4 Modeling results based on severity weighted crash counts

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D.</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.5666</td>
<td>0.0263</td>
<td>0.5105</td>
<td>0.6229</td>
</tr>
<tr>
<td>Roadway mileage</td>
<td>0.0974</td>
<td>0.0066</td>
<td>0.0840</td>
<td>0.1101</td>
</tr>
<tr>
<td>Curve crashes</td>
<td>0.1210</td>
<td>0.0053</td>
<td>0.1107</td>
<td>0.1316</td>
</tr>
<tr>
<td>Wet roadway</td>
<td>0.0231</td>
<td>0.0016</td>
<td>0.0200</td>
<td>0.0262</td>
</tr>
<tr>
<td>Roadway defects</td>
<td>0.0672</td>
<td>0.0005</td>
<td>0.0661</td>
<td>0.0682</td>
</tr>
<tr>
<td>$\sigma_\phi$</td>
<td>0.4705</td>
<td>0.0299</td>
<td>0.4094</td>
<td>0.5290</td>
</tr>
<tr>
<td>$\sigma_\theta$</td>
<td>0.1030</td>
<td>0.0492</td>
<td>0.0296</td>
<td>0.1878</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.8242</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\bar{D}: 2643.32, D(\bar{\theta}) = 2432.57, p_D = 24.646, DIC = 2854.06$

The comparison of Figure 5.4 (a) and (b) indicates that the crash severity is not uniformly distributed in New Jersey. Figure 5.4 (c) presents the changes in crash rates between both models. Three of the counties (Passaic, Essex and Hudson) have the highest difference when the severity weights are considered. Since these are neighboring counties, the site specific factors due to shared roadways might play a role in the higher rankings. Additionally, the five counties that follow the top three counties displayed a similar trend. While Union county neighbors the three counties mentioned earlier, four southern counties, Ocean, Atlantic, Camden, and Cumberland are also neighbors of each other, this trend was not followed for the remaining counties, except for Warren, Hunterdon, and Morris counties which showed the smallest change between the two models. It is necessary to investigate the factors of the likelihood of occurrence of more
severe crashes at these locations and the cause behind the neighboring counties to have similar trends.

Figure 5.4 Mean crash rates from (a) raw, (b) from severity weighted crash counts
Crash rate maps were also developed for the different road types as shown in Figure 5.5. Since DVMT data were not available at the jurisdiction level, county DVMT miles were used for estimating the crash rates. However, the roadway length by jurisdiction data were available and the model was estimated using this data. The limitations of these data is that they might have a negative effect on the accuracy of the results as the exposure to traffic is different for various road types. The model results revealed lower crash rates than the actual rates for each road type, but the sum of the crash rates for different road types was equal to the crash rate of a county for all crashes. This was due to the fact that the same county level DVMT was used for all models. These results indicate that crash rates vary for different counties in terms of road types. Thus, crash rates are not spatially consistent within each county independent of the road type. This can have major policy implications because one county that appears to have lower overall crash rates can have higher crash rates in terms of its county or municipal roadways. Moreover, it was observed that the state and authority roadways have lower crash rates than the county and municipal roadways. This result can be used to address possible funding and other geometric and operational improvement issues for roads where there is room for improvement compared to other roads located within the same county.
Figure 5.5 Mean crash rates for (a) state (b) authority (six counties are excluded) (c) county (d) municipal roadways by county in 2010 (per thousand DVMT)
5.2 Bayesian Spatial Modeling and Risk Mapping of Downed Trees along the Roadways Using Data from Hurricanes Irene and Sandy

In this study, hierarchical Bayesian models were used to model the frequency of downed tree incidents on roadways using factors such as precipitation, wind speed, roadway density, etc. The effect of spatial correlation and spatial heterogeneity was also considered in the model development. Municipality level counts of the downed trees that contained all events in New Jersey during hurricanes Irene and Sandy were used in the model estimation. Since, the data contains numerous zero counts at the municipality level, the zero-inflated Poisson model was adopted.

5.2.1 Data

Data from multiple sources was used in the study. The data include: the number of downed trees, total precipitation, average wind speed, maximum wind speed, roadway mileage, roadway density, and distance from shore (see Table 5.5). For hurricane related data, a two day timeframe selected where the effects of the hurricane was the most severe. Hence, for hurricane Irene, the data covered the August 28 and August 29 2011 period and for hurricane Sandy, the data covered October 29 and October 30 2012.

The number of downed trees in each municipality was obtained from Transportation Operations Coordinating Committee (TRANSCOM) event data. TRANSCOM is a coalition of 16 transportation and traffic enforcement agencies in the New York, New Jersey, and Connecticut metropolitan area. TRANSCOM provides real-time event and travel time data to its members using the existing traffic and transportation
systems of its member agencies (xcm.org, 2014). TRANSCOM event data reports several types of events that occur on the roadways such as accidents, disabled vehicles, cargo spills, flooding, downed trees, vehicle fires, etc. The event data also contain roadway facility names, event types, the start and end time of the event, the event descriptions, and detailed location information (coordinates, milepost, city, county). The downed trees and their locations were extracted from this database. As a result, 204 incidents during Irene and 224 incidents during Sandy were obtained. New Jersey has 566 municipalities and the location of each downed tree was aggregated at the municipality level for all incidents in the state. Figure 5.6 shows locations of the downed tree incidents in terms of municipalities. The map of the downed trees during both hurricanes demonstrated that many municipalities had zero incidents. Hence, the statistical model used in the study needed to handle the data set with many zeros.
Figure 5.6 Municipal locations of downed tree incidents during (a) Hurricane Irene and (b) Hurricane Sandy

The weather related data was obtained from two sources: National Oceanic and Atmospheric Administration (NOAA) (2014) and New Jersey Weather and Climate Network (2014). There were many weather stations included in the NOAA network, however, only those at the airports (8) had the capability to report wind speed data. Therefore, to derive a better picture of the weather during the hurricanes, the data from 31 New Jersey Weather and Climate Networks’ weather stations was also used for generating the required precipitation and wind speed data. The locations of the weather sensors used in the study are shown in Figure 5.7 (a) with diamond icons.
Roadway mileage data obtained from NJDOT’s straight line diagrams (NJDOT, 2014) are seen in Figure 5.7 (a). To calculate roadway density, in addition to roadway miles, municipality land areas were needed. This data was obtained from Bureau of Geographic Information Systems (NJGIS, 2014). The roadway density is presented in Figure 5.7 (b). As seen from the figure, the central and northern part of New Jersey, close to the New York metropolitan area, has a denser roadway network compared with the rest of the State.

Distance from the shore was defined as the Euclidean distance of the centroid of the municipality from the shore and calculated from the GIS map of the municipalities of New Jersey.

Table 5.5 Summary of the hurricane data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hurricane Irene</th>
<th>Hurricane Sandy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std.Dev.</td>
</tr>
<tr>
<td>Downed trees</td>
<td>0.4</td>
<td>0.99</td>
</tr>
<tr>
<td>Total Precipitation (inches)</td>
<td>6.25</td>
<td>1.21</td>
</tr>
<tr>
<td>Average wind speed (miles/hr)</td>
<td>7.48</td>
<td>2.83</td>
</tr>
<tr>
<td>Maximum wind speed (miles/hr)</td>
<td>44</td>
<td>9.28</td>
</tr>
<tr>
<td>Roadway mileage (miles)</td>
<td>20.36</td>
<td>22</td>
</tr>
<tr>
<td>Roadway density (1/miles)</td>
<td>2.637</td>
<td>1.79</td>
</tr>
<tr>
<td>Distance from the shore (miles)</td>
<td>18.87</td>
<td>14.8</td>
</tr>
</tbody>
</table>
Figure 5.7 (a) New Jersey roadway network (except surface roads) (b) Roadway density at municipality level

5.2.2 Zero Inflated Model

In this study, the hierarchical Bayesian generalized linear model was considered for estimating the model. It is common to fit the count data using Poisson or Negative Binomial models. However, when the count data contains numerous zeros, these distributions may not fit the data well and may underestimate the probability of zero counts. The hurricane data used in this study also contained many zeroes as seen in Figure 3. Thus, it was crucial to adopt a model which could accurately represent this data.
The zero-inflated Poisson (ZIP) model proposed by Lambert (Lambert, 1992) is proven successful in dealing with discrete data with many zeros. Manufacturing (Lambert, 1992), medicine, sports, crime are some of the fields that this model is used in the past. The ZIP model is a modified version Poisson model which allows numerous zero counts in the data. This model assumes that the data is generated by two independent processes for zero counts and for non-zero counts.

The full hierarchical Bayesian modeling requires a three step approach. For the first step, conditional on mean, \( \lambda_{it} \), the number of downed trees, \( Y_{it} \) was assumed to follow a zero inflated Poisson distribution:

\[
\begin{align*}
Y_{it} &\sim 0 \quad \text{with probability } \pi_i \\
Y_{it} &\sim \text{Po}(\lambda_{it}) \quad \text{with probability } 1-\pi_i
\end{align*}
\]  

(5.9)

where \( Y_{it} \) was the observed number of downed trees in municipality \( i \) at time \( t \), \( i=1,\ldots,N \) and \( t=1,\ldots,T \) (1:Irene, 2:Sandy), \( \lambda_{it} \) was the mean of downed trees for municipality \( i \) at time \( t \), and \( \pi_i \) was the proportion of additional zeros for municipality \( i \). Then ZIP model became:

\[
Y_{it} \sim \text{ZIP}(\pi_i, \lambda_{it})
\]  

(5.10)

The excessive zeros \( \pi_i \) were modeled for each municipality independent of covariates with following structure:

\[
\log \left( \frac{\pi_i}{1-\pi_i} \right) = X_{it}^\top \beta^\pi
\]  

(5.11)

Then, the log-link function required by the generalized linear model, \( \lambda_{it} \), was modeled as:
\[
\log(\lambda_{it}) = \alpha + \sum_k \beta_k X_{ik} + \theta_i + \phi_i + \delta_t
\]  
(5.11)

where \(\alpha\) represents the intercept term, \(\beta_k\) is the coefficient for spatial covariate \(k\), \(X_{ik}\) is the observed value of \(k\)th covariate for municipality \(i\) at time \(t\), \(\theta_i\) stands for region-wide or statewide global heterogeneity, \(\phi_i\) represents spatially correlated random effects for county \(i\), and \(\delta_t\) is a term for fixed temporal effects of hurricanes varying by \(t\).

Next, coefficients were modeled by using non-informative priors. For the random effect terms, second level of hierarchy was defined. Following the convention from the literature, we assumed a normal prior for the heterogeneity term:

\[
\theta_i \overset{iid}{\sim} N(0, \frac{1}{\tau_h})
\]  
(5.12)

where \(\tau_h\) is a precision parameter that controls the amount of \(\theta_i\) among municipalities. The heterogeneity term enabled us to include extra-Poisson variability due to unobserved variables over the entire state.

Similarly, the error term for fixed time effects was modeled using a normal prior:

\[
\delta_t \overset{iid}{\sim} N(0, \frac{1}{\tau})
\]  
(5.13)

where \(\tau\) was a precision parameter that controls the amount of \(\delta_t\) among hurricanes.

For modeling the spatially correlated random effects, a CAR prior was adopted as proposed by Besag (1974):

\[
\phi_i | \phi_{-i} \sim N\left(\sum_{j \neq i} \left(\frac{w_{ij}}{w_{ii}}\right) \phi_j, \frac{1}{\tau_\phi w_{ii}}\right)
\]  
(5.14)
where $\tau_c$ is a precision parameter that controls clustering, $ij$ is a neighbor municipality adjacent to municipality $i$, $wij$ is the weight of the neighbor $j$, and $w_{i+}$ represents the sum of the weights of the neighbors of municipality $i$. Note that, in this study, it was assumed that all neighbors had equal weight instead of a decaying weight as a function of distance between the centroids of geographic units in this case municipalities. By including a spatially correlated random effects term, extra-Poisson variability in the log-relative risk which varies from municipality to municipality can be modeled in such a way that nearby municipalities will have more similar rates.

In the next step, the hyperpriors are defined for structured random effect terms which are heterogeneity and clustering precision parameters and for fixed time effects. Same non-informative priors are set for $\tau_l$, $\tau_h$ and $\tau_c$.

5.2.3 Model Selection

Several parameterization approaches of the proposed hierarchical model were considered in the model development. Deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002) is used to compare the fit and complexity of the hierarchical models in which the number of parameters is not clearly defined.

The hierarchical Bayesian model presented in the previous section is the most general form of this model, which takes into account spatial covariate effects as well as spatial correlation, unobserved heterogeneity, and fixed hurricane effects. To better understand the combination of covariate and spatial effects, a forward selection approach
was followed. First, a simple model with an intercept was estimated. Next, the models for each variable with an intercept were estimated. The best model was selected based on DIC statistics. Then, this process was continued until all of the variables were included in the final model. As a result of this incremental approach where the most significant variable was added to the model in each step, 32 models were estimated. These models were estimated using WinBUGS (Spiegelhalter et al., 2014). Starting with only an intercept, different subsets of variables were considered in the model development. Multiple chains were simulated for these models. 20,000 iterations were performed for each model and the first 10,000 iterations were considered as burn-ins and were not included in the sampling. The convergence of the models was evaluated using the mixing of MCMC chains and Gelman-Rubin statistics in WINBUGS.

Table 5.6 shows the results of model estimation.

<table>
<thead>
<tr>
<th>Model #</th>
<th>$\alpha$</th>
<th>$\beta_{\text{miles}}$</th>
<th>$\beta_{\text{den}}$</th>
<th>$\beta_{\text{maxwind}}$</th>
<th>$\beta_{\text{av.wind}}$</th>
<th>$\beta_{\text{rain}}$</th>
<th>$\beta_{\text{distance}}$</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1895.3</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1790.7</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1701.4</td>
</tr>
<tr>
<td>19</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>1699.6</td>
</tr>
<tr>
<td>24</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td>1699.2</td>
</tr>
<tr>
<td>28</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>1697.6</td>
</tr>
<tr>
<td>32</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1699.8</td>
</tr>
</tbody>
</table>

Based on Table 5.6, roadway miles, roadway density, maximum wind speed, average wind speed, and rain variables were found to be significant factors as they performed best at each step.
in improving the model fit (based on decreasing DIC values). However, as seen in Table 2, distance variable adversely affected the model fit as it had a higher DIC value (1699.8) than the previous steps. Hence, Model 28 was selected.

After selecting the contributing variables, the effects of spatial and temporal terms in the model fit was also considered individually as well as jointly. Table 5.7 shows the variation in the model fit as a result of including the different combinations of these terms in the model. First, they were included in the model individually. Spatial heterogeneity term outperformed the other two with a DIC value of 1486.6. Spatial correlation ranked as a close second with a DIC value of 1488.3, while the fixed time term performed significantly worse, 1697.6, than both models with spatial terms. These terms were then included in the model as pairs. Spatial correlation and spatial heterogeneity performed better than the pairs that included a fixed time term with DIC of 1486.9. Yet, none of the pairs improved the model fit more than the spatial heterogeneity term alone. Finally, all the terms were considered jointly in the model estimation. The model performance was adversely affected by the addition of fixed time effects term to model with the spatial heterogeneity and spatial correlation terms with a DIC value of 1487.4. This might indicate that the source of variation in downed trees incidents among the municipalities is mainly spatial rather than temporal. Based on the estimation results, the spatial heterogeneity is favored over the spatial correlation and fixed temporal effects and was included in the final model.
Table 5.7 Effect of spatial and temporal terms on model fit

<table>
<thead>
<tr>
<th>Term</th>
<th>$\bar{D}$</th>
<th>$D(\bar{\theta})$</th>
<th>$p_D$</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>1300.9</td>
<td>1114.9</td>
<td>186</td>
<td>1486.6</td>
</tr>
<tr>
<td>$\phi$</td>
<td>1614.3</td>
<td>1531</td>
<td>83.3</td>
<td>1697.6</td>
</tr>
<tr>
<td>$\delta$</td>
<td>1300.7</td>
<td>1113</td>
<td>187.7</td>
<td>1488.3</td>
</tr>
<tr>
<td>$\phi, \theta$</td>
<td>1298.8</td>
<td>1111.1</td>
<td>187.8</td>
<td>1486.9</td>
</tr>
<tr>
<td>$\theta, \delta$</td>
<td>1302.4</td>
<td>1118.2</td>
<td>184.2</td>
<td>1487</td>
</tr>
<tr>
<td>$\phi, \delta$</td>
<td>1616.8</td>
<td>1533.8</td>
<td>83</td>
<td>1699.8</td>
</tr>
<tr>
<td>$\phi, \theta, \delta$</td>
<td>1301.6</td>
<td>1115.8</td>
<td>185.8</td>
<td>1487.4</td>
</tr>
</tbody>
</table>

$\theta$: spatial heterogeneity, $\phi$: spatial correlation, $\delta$: fixed time

5.2.4 Results and Discussion

The model was estimated based on the findings discussed in the previous section.

Table 5.8 shows the model parameters for the selected model.

Table 5.8 Estimated posterior density of parameters for the selected model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>MC Error</th>
<th>2.50%</th>
<th>97.50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>-1.298</td>
<td>0.147</td>
<td>0.0104</td>
<td>-1.608</td>
<td>-1.026</td>
</tr>
<tr>
<td>$\beta_{\text{rain}}$</td>
<td>0.075</td>
<td>0.052</td>
<td>0.0019</td>
<td>-0.027</td>
<td>0.180</td>
</tr>
<tr>
<td>$\beta_{\text{avgwind}}$</td>
<td>-0.032</td>
<td>0.027</td>
<td>0.0013</td>
<td>-0.086</td>
<td>0.022</td>
</tr>
<tr>
<td>$\beta_{\text{maxwind}}$</td>
<td>0.021</td>
<td>0.009</td>
<td>0.0004</td>
<td>0.004</td>
<td>0.040</td>
</tr>
<tr>
<td>$\beta_{\text{density}}$</td>
<td>-0.040</td>
<td>0.008</td>
<td>0.0004</td>
<td>-0.058</td>
<td>-0.025</td>
</tr>
<tr>
<td>$\beta_{\text{miles}}$</td>
<td>0.008</td>
<td>0.001</td>
<td>0.0001</td>
<td>0.006</td>
<td>0.011</td>
</tr>
<tr>
<td>$\sigma_\theta$</td>
<td>0.967</td>
<td>0.090</td>
<td>0.0065</td>
<td>0.791</td>
<td>1.144</td>
</tr>
</tbody>
</table>

* For comparison, Poisson model $\bar{D}:1498.9, D(\bar{\theta})=1314.14, p_D=183.7, DIC=1582.7$

Based on Table 5.8:

- The ZIP model can handle zero-inflated outcomes better than the Poisson model due to the fact that ZIP model has a DIC value of 1486.6, while Poisson model has a DIC value of 1582.7.
• The effect of precipitation is found to be insignificant since the confidence interval contains zero. This might suggest that the total precipitation levels did not play an important role in terms of downed trees.

• Average wind speed was also found to be insignificant in the model.

• Maximum wind speed in a municipality during the hurricane was found to be significant. This suggests that while average wind speed does not have great impact on the number of downed trees, high wind speeds can be hypothesized to be responsible for the downed trees.

• Roadway density was found to be significant and had a negative effect on the number of downed trees. If we assume that roadway density is an indication of the degree of urbanization in a municipality, this might be interpreted as it is more likely to have trees next to the roadways.

• Roadway miles was found to be significant with a positive sign. The data only contained information on downed trees along the roadways, hence, we can conclude that the more the roadway miles there are in a municipality, the more chance of a downed tree there will be on a roadway.

• The area-wide spatial heterogeneity term was found to be significant indicating that there are other unobserved region-wide factors that are affecting the number of downed trees. One likely candidate might be the density of trees along the roadways.

Using the model parameters, the predicted mean number of downed trees was plotted for Hurricane Sandy as shown in Figure 5.8 (a). Hurricane Sandy made landfall in
Brigantine, NJ (see Figure 5.8) on October 29, 2012 and then it moved north ashore (Blake et al., 2013). Based on Figure 5.8(a), it can be observed that the predicted mean number of downed trees varies among municipalities without any significant spatial pattern. Later, the mean number of downed trees was normalized by roadway density as shown in Figure 5.8(b). In the normalized map, it is apparent that number of downed trees is clustered around the municipalities along the shoreline in central Jersey. This might indicate that in the less urbanized part of the state close to the shore, the effect of these two hurricanes in terms of downed trees are higher than in other municipalities. To check the validity of this claim, a scenario analysis was also performed for Hurricane Sandy. In this scenario, the effect of a ten percent and twenty percent increase in maximum wind speed to spatial risk pattern was examined (see Figure 5.9). It was observed that an increase in the maximum speed resulted in more clustering on the shore. However, it is important to note that during Irene and Sandy, evacuation orders were mainly issued for these same coastal regions of New Jersey. Thus, there is a higher likelihood that the downed trees at the coastal municipalities might have affected the critical links necessary for evacuation.

The downed tree incidents during both hurricanes were rare; hence, there are many municipalities that had no occurrence of this type of incident. When there are many zero counts in the data, the traditional Poisson model tends to underestimate the probability of zero counts. As a result, modeling the data with a ZIP assumption is proven to be valid since the performance of the Poisson model with the same specifications is found to be less robust than the ZIP model.
Figure 5.8 Predicted mean of (a) the number of downed trees (b) the number of downed trees normalized by roadway density during Hurricane Sandy
Figure 5.9 Predicted mean of downed trees normalized by roadway density during Hurricane Sandy in case of (a) ten percent and (b) twenty percent increase in maximum wind speed.
5.3 A Doubly Stochastic Point Process Model for Modeling Crashes along a Corridor

In this section, a doubly stochastic model for point level modeling of crashes is described. In this model, the roadway is represented as a continuous entity without predefined links. Such a model captures the continuous nature of the roadways better than any existing model.

First, we divided the roadway into grid cells hence the crash points became the crash counts in the grid cells. Then, we aligned these cells in 1-d so that the effect of spatial variation is only in the direction of the roadway. This also enabled us to limit the occurrence of the crashes only on the roadway. At the first level of the stochasticity, the crash counts in the cells were assumed to follow a Poisson distribution. In the second level, given the crash counts, spatial variation on the roadways was modeled as a log-Gaussian Random field. Using Bayesian modeling approach and INLA, the model was developed in R. The modeling results and Bayesian inference was used to develop useful maps for determining the crash risks along the roadway. To our knowledge, this is the first example of this approach for crash modeling. It is believed that the maps generated from the modeling results provides better insights then the maps based on link level crash counts since the point level maps enable one to pinpoint the problematic locations instead of averaging out the crashes over a predefined link length.
5.3.1 Data

New Jersey Department of Transportation (NJDOT)’s crash records database was used for developing the crash risk model (NJDOT, 2014). NJDOT keeps this yearly database of all crashes occurring on all roadways in New Jersey. The database contains the locations of the crashes along with many items of crash specific information. For this study, the spatial patterns of the crashes that occurred on I-287 in 2011 were analyzed. I-287 is a 67.54 miles long interstate highway and it is one of the busiest roadways in the state of New Jersey.

For the modeling, a collection of cells of the region of interest was required. First, the roadway was divided to 0.1x0.1 mile regular grid lattice. As a result, 676 square cells were obtained. Then, the cell crash counts were calculated in ArcGIS using the locations of 2,844 crashes that occurred on the roadway in 2011.

In addition to crash locations, several spatial covariates were considered. These include AADT, skid number, surface distress, rut depth, speed, number of lanes, lane drop, on-ramp, off-ramp, population variables. All covariates except population were extracted from the Roadway Centerlines database provided by NJDOT. Population data is obtained from the 2010 Census Block data (NJGIN, 2014).

Previous research shows that there is a relationship between pavement performance and crash occurrence (Jiang et al., 2013). Hence, the effects of three pavement performance variables are investigated. These are skid number, surface distress, and rut depth. Skid resistance is the force developed when a tire is prevented from rotating slides along the pavement surface (Roads and Transportation Association of Canada, 1977). Skid resistance is an important pavement parameter because
inadequate skid resistance might lead to higher incidences of skid related accidents. Skid resistance is generally quantified using some form of friction measurement such as a friction factor or skid number. Surface distress is another indication of poor or unfavorable pavement performance or signs of impending failure (Highway Research Board, 1970). Surface distress is related to roughness (the more cracks, distortion and disintegration, the rougher the pavement will be) as well as structural integrity (surface distress can be a sign of impending or current structural problems). Rutting is defined as the permanent or unrecoverable traffic-associated deformation within pavement layers that accumulates over time (Paterson, 1986). Rut depth is a measure of permanent deformation due to rutting.

It is safe to assume traffic exposure directly affects the crash counts in each cell. Hence, AADT was considered as the exposure variable in the study. It was also intended to observe the significance of number of lanes and lane drop at a location on spatial crash patterns so they were included as covariates in the model. Another pair of variables (on-ramp and off-ramp) was included to see if these locations were more prone to crashes. Last, the population variable was included to test the effect of the population of the nearest census block on the crash patterns.

The descriptive statistics of the variables included in the study are shown in Table 5.9.
Table 5.9 Descriptive statistics of the variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crashes</td>
<td>4.207101</td>
<td>4.266494</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>AADT</td>
<td>82665.25</td>
<td>19456.63</td>
<td>39118</td>
<td>117671</td>
</tr>
<tr>
<td>Skid Number</td>
<td>45.68402</td>
<td>9.590207</td>
<td>0</td>
<td>63.3</td>
</tr>
<tr>
<td>Surface Distress</td>
<td>3.447988</td>
<td>1.309394</td>
<td>1.27</td>
<td>5</td>
</tr>
<tr>
<td>Rut Depth</td>
<td>0.186095</td>
<td>0.102482</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>Link Speed</td>
<td>63.56509</td>
<td>3.508326</td>
<td>55</td>
<td>65</td>
</tr>
<tr>
<td>Number of Lanes</td>
<td>3.094675</td>
<td>0.66882</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Lane Drop (binary)</td>
<td>0.014793</td>
<td>0.120813</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Population</td>
<td>1917.778</td>
<td>812.1769</td>
<td>551</td>
<td>4392</td>
</tr>
<tr>
<td>On-ramp (binary)</td>
<td>0.136095</td>
<td>0.343143</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Off-ramp (binary)</td>
<td>0.122781</td>
<td>0.328429</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

5.3.2 Modeling

5.3.2.1 Log-Gaussian Cox Model

A Cox process or ‘doubly stochastic’ process is an inhomogeneous Poisson process with a stochastic intensity function $\lambda(x)$. The point process is defined by the following two assumptions:

i) $\Lambda = \{\Lambda(x) : x \in \mathbb{R}^2\}$ is a non-negative valued stochastic process

ii) Conditional on the realization $\Lambda(x) = \lambda(x) : x \in \mathbb{R}^2$, the point process is an inhomogeneous Poisson process with intensity function $\lambda(x)$

In the spatial point process modeling context, intensity stands for the number of events per unit area. Hence, in the case of crashes, intensity represents the number of crashes per unit area.
Møller et al. (1998) developed log-Gaussian Cox processes (LGCPs). LGCP are special case of Cox process with $\Lambda(x) = \exp\{S(x)\}$, where $S$ is a Gaussian process. In this model, spatially varying intensity can be modeled by including one or more spatially indexed explanatory variables $z(x)$. The typical approach would be to retain the stationarity of $S(x)$ but to replace the constant intensity $\lambda$ by a regression model (Diggle et al., 2013) as in the following:

$$\Lambda(x) = \exp\{\beta + S(x)\} \quad (5.15)$$

The resulting Cox process is now an intensity re-weighted stationary point process (Baddeley et al., 2000). This is the analogue for point process data of the linear Gaussian latent model with a spatially varying mean and a stationary residual (Diggle and Riberio, 2007). Like the linear Gaussian model, it is data driven. However, this approach provides a flexible and relatively tractable class of empirical models for describing spatially correlated phenomena. This makes it useful in applications where the scientific focus is on spatial prediction rather than on testing mechanistic hypotheses (Diggle et al., 2013).

Cox processes are very useful models for spatial point patterns (Diggle et al., 2013). For example, Cox process is also a suitable candidate for modeling the spatial distribution of crashes. The observed spatial pattern of crashes results from the spatial variation in the traffic exposure to observed and unobserved factors. On the other hand, these models may not be useful for modeling the spatial distribution of secondary incident sites, since, in this case, the spatial pattern is a result of interactions between the crashes.
In this study, we propose a LGCP model for crash locations with intensity
\[ \Lambda(x) = E(x)R(x), \]
where \( E(x) \) stands for traffic exposure, and \( R(x) \) represents crash risk, \( R(x) = \exp\{S(x)\} \). Conditional on \( R(x) \), crash counts in grid cells \( L_i \) of the corridor of interest are independent and Poisson distributed with means:
\[ \mu_i = \int_{L_i} E(x)R(x)dx \]  
(5.16)

One advantage of LGCP is that it readily allows inclusion of spatial covariates into a model at different spatial resolution. For the crash data, the covariates were incorporated into the model in the following form:
\[ \Lambda(x) = E(x)\exp\{z(x)'\beta + S(x)\} \]  
(5.17)

where \( z(x) \) denotes the covariate surfaces and \( \beta \) is the vector of parameters for covariates. The above equation applies to the continuous case. For computation, a discrete version of this equation is needed.

Assume that crash risk within each cell, \( \Lambda_i, i=1,...,676 \), is constant or has a small spatial variation. Then we can use a GLM structure for the likelihood and estimation of log crash risk (or intensity) in each cell in the following form:
\[ \eta_i = \beta_0 + \log(E_i) + \sum_j \beta_j z_{ij} + S_i \]  
(5.18)

where \( \eta_i \) is log crash risk in cell \( i \), \( \beta_0 \) represent the intercept, \( E_i \) is the exposure, \( z_{ij} \) spatial covariates, \( \beta_j \) the parameters for covariates and \( S_i \) the spatial dependence. As the number of cells tends to go to infinity, this process behaves like its spatially continuous counterpart as shown in equation (5.17).
For spatial dependence, a valid spatial covariance function needs to be defined. For LGCP, a widely used family is the Matern (1960) covariance function:

\[ C(u) = \sigma^2 r(u; \phi, \kappa) \]  

(5.19)

where

\[ r(u; \phi, \kappa) = \left\{ \frac{2^{\kappa-1}}{\Gamma(\kappa)} \right\}^{-1} (u / \phi)\kappa K_{\kappa}(u / \phi) \]

where \( \Gamma(.) \) is the complete Gamma function, \( K_{\kappa}(.) \) is a modified Bessel function of order \( \kappa \), and \( \phi > 0 \) and \( \kappa > 0 \) are parameters. \( \phi \) has units of distance and \( \kappa \) is a dimensionless shape parameter that determines the differentiability of the corresponding Gaussian process. Generally \( \kappa \) is selected from a set of values as it is hard to estimate it empirically. We use values of \( \kappa = 0.5, 1.5, 2.5 \).

Considering a bounded region \( \Omega \subseteq \mathbb{R}^2 \), the likelihood for an LGCP has the following form:

\[ \pi(Y | \lambda) = \exp \left( \int_{\Omega} \lambda(x) dx \right) \prod_{x_i \in Y} \lambda(x_i) \]  

(5.20)

where the integral is complicated by the stochastic nature of \( \lambda(x) \). However, this integral can be computed numerically using traditional methods such as Markov Chain Monte Carlo (MCMC) methods. LGCP fits within the Bayesian hierarchical modeling framework. Moreover, it is a latent Gaussian model; hence, it can be easily embedded within the Integrated Nested Laplace Approximation (INLA) framework. INLA produces results faster and negates the need to assess the convergence and mixing properties as in MCMC algorithm (Illian et al., 2012) because INLA is based on numerical integration rather than simulation (as in MCMC).
5.3.2.2 Alternative Link Level Model: Poisson MCMC

An alternative model to the one estimated in this study would be to estimate the spatial variation of the data using the hierarchical Poisson-Gaussian Markov random field model, proposed by Besag (1974). This model has been used in traffic safety literature for the past decade for modeling crashes at, generally, aggregate level such as county level (Miaou et al., 2003; MacNab, 2004; Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Huang et al., 2010). However, there are also some examples of its use at intersections and at link level (Aguero-Valverde and Jovanis, 2008). For comparison purposes, the model at the link level was estimated for the same corridor. This model used the predefined links on the corridor instead of the point locations. In this case, the crash counts were assigned to the links.

The full hierarchical Bayesian modeling required a three step approach. For the first step, conditional on mean, $\mu_i$, crash counts, $Y_i$ are assumed to follow a Poisson distribution:

$$ Y_i^{ind} \sim Po(\mu_i) $$

(5.21)

where $Y_i$ is the observed number of crashes in link $i$, $i=1,..,N$; and $\mu_i$ is the mean of the Poisson process for link $i$. The mean was further formulated as:

$$ \mu_i = e_i \lambda_i $$

(5.22)

In the above equation, crash risk rate, $\lambda_i$, was assumed to be proportional to traffic exposure, $e_i$, hence AADT was considered as an offset for exposure to traffic. Then, the crash risk rate, $\lambda_i$, was modeled as:
\[ \log(\lambda_i) = \alpha + \sum_k \beta_k X_{ik} + \theta_i + \phi_i \]  

(5.23)

where \( \alpha \) represented the intercept term, \( \beta_k \) was the coefficient for spatial covariate \( k \), \( X_{ik} \) was the observed value of \( k \)th covariate for link \( i \), \( \theta_i \) stands for spatial heterogeneity, \( \phi_i \) represented spatially correlated random effects for link \( i \).

Next, coefficients were modeled by using non-informative priors. For the random effect terms, second level of hierarchy was defined. The heterogeneity term was modeled using a normal prior:

\[ \theta_i \sim \text{iid} N \left( 0, \frac{1}{\tau_h} \right) \]  

(5.24)

where \( \tau_h \) is a precision parameter that controls the amount of \( \theta_i \) among municipalities. The heterogeneity term enabled us to include extra-Poisson variability due to unobserved variables over the entire state. For modeling the spatially correlated random effects, a CAR prior was adopted, as proposed by Besag (1974):

\[ \phi_i | \phi_i \sim N \left( \sum_{j \in i^*} \left( \frac{w_{ij}}{w_i^*} \right) \phi_j, \frac{1}{\tau_c w_i^*} \right) \]  

(5.25)

where \( \tau_c \) is a precision parameter that controls clustering, \( ij \) is a neighbor link adjacent to link \( i \), \( w_{ij} \) is the weight of the neighbor \( j \), and \( w_i^* \) represents the sum of the weights of the neighbors of link \( i \). Note that, in this study, it was assumed that all neighbors had equal weight since this is common practice in the literature. By including a spatially correlated random effects term, extra-Poisson variability in the log-relative risk which varies from municipality to municipality can be modeled in such a way that nearby municipalities
will have more similar rates. In the next step, the same non-informative hyperpriors were defined for heterogeneity and clustering precision parameters.

5.3.3 Modeling Results and Discussion

Based on the model specifications presented in the previous section, LGCP for point level data and Poisson MCMC for link level data were estimated. Model estimation results are presented in Table 5.10 and Table 5.11.

The spatial covariates in the data have different resolutions. The LGCP model is a versatile approach for easily incorporating these covariates. Based on the 95% credible set (which is counterpart of confidence interval in Bayesian statistic) of the posterior distribution of the variables, we can make the following observations (for the variables presented in Table 5.9):

- The skid number has a positive relationship with respect to the crash risk. However, the credible set contains a zero, hence, the skid number is insignificant. This means the skid resistance along the roadways is sufficient; hence, it does not impact crash risk significantly.
- The surface distress index was significant and negatively correlated with the crash risk. Before interpreting this value, it is necessary to give some background information on the surface distress index. Surface distress index is a value given to a roadway section on a scale of 0-100. While the index value of 100 would indicate a perfect road with no measurable distress or rough ride, an index value of 60 is considered as terminable serviceability
and the road is considered failed (FHWA, 2009). Hence, this result indicates that the rough roadway section increases the crash risk.

- Rut depth was significant and had a positive relationship with the crash risk. Therefore, the higher degree of permanent traffic-related deformation of the pavement, the higher the crash risk on the roadway.

- Speed limit was found to be significant and negatively correlated with the crash risk. Although this sounds counter intuitive, this might point to the fact that the drivers might not be following lower speed limits along the roadway and consequently experience a greater crash risk at the lower speed locations.

- Number of lanes had a positive and significant effect. This variable might be correlated with the traffic exposure to some degree. Therefore, the positive value of this covariate was attributed to its relationship with the traffic exposure.

- Lane drop was found to be significant and positively related to the crash risk. This might be due to the difficulty of drivers in adapting to changing traffic conditions.

- Population was found to have no impact on the crash risk. Some might expect population would be another indicator of traffic exposure. However, the model uses a fine grid of cells and there is a good chance that the population of the nearest town might not affect the interchange on the roadway.
• The existence of an on ramp variable was found to be insignificant in terms of its effect on the crash risk. Drivers tend to be more cautious while entering the roadway due to the presence of a yield or stop sign. Hence, crash risk was not significantly affected by the on ramp covariate.

• On the other hand, off ramp covariate had a significant and positive impact on crash risk. This might also be another sign that the drivers need to take their time and be more careful in decelerating and merging before exiting the roadway to avoid a higher risk of crashes.

• The spatial dependence of grid cells was found to be a significant variable. Figure 5.10 shows the range of the marginal posterior distribution of the range of the spatial dependence. It can be seen that the range of the highest spatial dependence is around ±0.5 miles and the spatial dependence diminishes around ±1 miles.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D</th>
<th>2.50%</th>
<th>97.50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.4897</td>
<td>1.0859</td>
<td>1.3549</td>
<td>5.6196</td>
</tr>
<tr>
<td>Skid number</td>
<td>0.008</td>
<td>0.0051</td>
<td>-0.0021</td>
<td>0.018</td>
</tr>
<tr>
<td>Surface distress</td>
<td>-0.1615</td>
<td>0.0431</td>
<td>-0.2461</td>
<td>-0.0769</td>
</tr>
<tr>
<td>Rut depth</td>
<td>1.0669</td>
<td>0.4557</td>
<td>0.1731</td>
<td>1.9626</td>
</tr>
<tr>
<td>Speed limit</td>
<td>-0.0479</td>
<td>0.0134</td>
<td>-0.0742</td>
<td>-0.0215</td>
</tr>
<tr>
<td>Number of lanes</td>
<td>0.2871</td>
<td>0.0763</td>
<td>0.1374</td>
<td>0.437</td>
</tr>
<tr>
<td>Lane drop</td>
<td>0.5651</td>
<td>0.2401</td>
<td>0.0942</td>
<td>1.037</td>
</tr>
<tr>
<td>Population</td>
<td>0</td>
<td>0.0001</td>
<td>-0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Onramp</td>
<td>0.1321</td>
<td>0.0804</td>
<td>-0.0258</td>
<td>0.2898</td>
</tr>
<tr>
<td>Offramp</td>
<td>0.2737</td>
<td>0.0837</td>
<td>0.1093</td>
<td>0.4377</td>
</tr>
<tr>
<td>Spatial dependence</td>
<td>0.7655</td>
<td>0.2663</td>
<td>0.3441</td>
<td>1.3735</td>
</tr>
</tbody>
</table>

$D : 2618.56, D(\hat{\theta}) = 2255.05, p_0 = 363.52, DIC = 2982.08$
Figure 5.10 The marginal posterior distribution of spatial dependence

For the Poisson MCMC model, only the covariates at the link level can be included in the model as seen in Table 5.11. Based on the results of the Poisson MCMC model:

- Link length had a positive effect but it is statistically insignificant which highlights that the longer links were not necessarily associated with a higher crash risk.

- The number of lanes had a positive effect but this was found to be statistically insignificant. In the previous model, we claimed that number of lanes might be an indication of exposure. In this model, there is a chance that the traffic exposure is represented more ably by another covariate which possibly makes “number of lanes” variable insignificant.

- Lane drop was also found to have a positive effect, but was also statistically insignificant. This covariate is found to be significant in LGCP model. It is possible that assigning lane drop to a link might average out the effect of a lane drop. On the other hand, for the point level model the effect of this covariate can be more ably captured.
The population covariate was found to be positive and significant. Unlike the LGCP model, the population covered all neighboring census blocks along a link and there was an interchange at the start and at the end of the link. Therefore, the population might be a sign of the demand on the roadway (traffic exposure).

Spatial correlation and spatial heterogeneity terms were both found to be positive and significant. Spatial correlation was larger than spatial heterogeneity which indicates the clustering of crash risk along the roadway considered in this study.

Table 5.11 The results of link level MCMC model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D.</th>
<th>MC Error</th>
<th>2.50%</th>
<th>97.50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.225</td>
<td>0.05588</td>
<td>3.19E-04</td>
<td>4.112</td>
<td>4.334</td>
</tr>
<tr>
<td>Link length</td>
<td>0.1106</td>
<td>0.06939</td>
<td>0.002209</td>
<td>-0.0254</td>
<td>0.2494</td>
</tr>
<tr>
<td>Number of lanes</td>
<td>0.2044</td>
<td>0.1486</td>
<td>0.004937</td>
<td>-0.09799</td>
<td>0.4936</td>
</tr>
<tr>
<td>Lane drop</td>
<td>0.233</td>
<td>0.1113</td>
<td>0.0063</td>
<td>-0.1961</td>
<td>0.6433</td>
</tr>
<tr>
<td>Population</td>
<td>0.08772</td>
<td>0.03398</td>
<td>0.001125</td>
<td>0.02004</td>
<td>0.1538</td>
</tr>
<tr>
<td>SD of Spatial Correlation</td>
<td>0.3137</td>
<td>0.09144</td>
<td>0.004171</td>
<td>0.1044</td>
<td>0.4795</td>
</tr>
<tr>
<td>SD of Spatial Heterogeneity</td>
<td>0.2954</td>
<td>0.07618</td>
<td>0.00383</td>
<td>0.02174</td>
<td>0.4148</td>
</tr>
</tbody>
</table>

\[
\hat{D} = 255.696, \hat{D}(\theta) = 223.007, p_D = 32.688, DIC = 288.384
\]

The two models estimated in this study had different numbers of parameters and covariates. Hence, it is not possible to compare the model fit based on the DIC values used in the analysis of model fit for Bayesian hierarchical models. Instead, their ability to accurately map the crash risk was investigated. For this purpose, the crash risk maps shown in Figure 5.11 were generated. While the map on the left side of Figure 5.11
represents the posterior mean of crash risk estimated from the LGCP model, the model on the right side shows the posterior mean of the crash risk based on a Poisson MCMC model. To derive the same scale in both cases, the crash risk in link level model was normalized by its length*10, as the point level grids cells are 0.1 mile long. Figure 5.11 shows five categories of crash risks based on natural breaks, which is a data clustering method used to determine the best arrangement of classes by minimizing each class mean based on its standard deviation, in the LGCP model. From Figure 5.11, the high crash risk locations are presented in both cases. However, the Poisson MCMC model aggregates the crash risk over each link which makes it impossible to pinpoint locations with the highest risk along an individual link. On the other hand, since the LGCP model uses a fine grid for estimation, the locations with higher crash risk can be determined with more accuracy. To demonstrate this difference more ably, a section of the roadway between mileposts 0 and 13.5 was zoomed as shown in Figure 5.12.
Figure 5.11 Point level (left) and link level (right) crash risks in 2011
Figure 5.12 Crash risks from LGCP model (top) and Poisson MCMC model (bottom) between mileposts 0 and 13.5
CHAPTER 6. CONCLUSIONS AND FUTURE RESEARCH

In this chapter, the summary of the methodology is presented.

In Chapter 2, current methodologies used in incident duration prediction and estimation are summarized and spatial models for accident mapping reviewed. In Chapter 3, first, different stages of incident management are explained. Then, the need for improving performance of incident management activities is discussed. Chapter 4 outlines an incident duration model based on Bayesian networks. Then, it investigates the adaptive learning methods for improving the performance of Bayesian networks. Chapter 5 describes a series of spatial models for mapping using spatial data with different resolutions.

6.1 Conclusions on Bayesian Networks and Adaptive Learning for Predicting Incident Duration

In this chapter, we describe how we used three different “structure learning algorithms” to estimate a BN for predicting incident durations. To construct the models and illustrate the individual performance of these algorithms, a unique incident dataset was created by combining records from different incident / accident dataset in New Jersey. Although all of the algorithms provided promising results, the BN generated K2 algorithm was selected as the best representation of the data according to its BIC score. The connection strengths between variables in the BN were also calculated in the study.
It is clear that this property will help to better understand the effect of variables on each other that are not visually provided in the BNs.

BNs can also deal effectively with missing data and this feature makes them a suitable tool for real-time incident management applications where quick and accurate real-time decisions are needed to be made with incomplete data. To demonstrate the effectiveness of this new incident duration model, its prediction accuracy with different levels of missing data was also computed. It was found that although its performance decays as the amount of missing information increases, the rate of decay is almost constant which can be an indication of the stability of the model. The prediction accuracy of the BN model was also compared to a CART model estimated using the same data set. It was observed that even though the CART model predicts incident durations with a comparable accuracy, nearly 21%, the BN clearly outperforms this method with 53% accuracy rate. Although the BN performed better than traditional models from a quantitative perspective, there are other important advantages of using BNs which can be summarized as follows:

- BNs are very useful for investigating and understanding the relationships between the variables in the data as presented visually. The traditional methods lack this important property.
- The results obtained from BNs are probabilistic, hence, not only a point estimate but also possible range of outcomes can be observed which provides additional information to decision-makers.
- Categorical variables can easily be incorporated into BNs as they are nonparametric in nature so no probability distribution assumptions about the
variables are required. This property clearly gives an edge to BNs over linear regression and other similar statistical methods where the parametric estimation approaches for which assumptions about the distributions of the parameters are required.

- Another limitation of traditional statistical modeling approaches is that they are generally developed for a single decision variable. However, in the case of BNs every variable can be considered as a decision variable and its distribution can be computed using the observed values of other variables.

To improve the future performance of the BN model, an adaptive learning scheme was introduced. The prediction performance of the chosen model with and without adaptive learning was examined using 2011 incident data from a major highway in New Jersey. It was found that the model performance in predicting future incident durations improved significantly if the proposed adaptive learning scheme was employed because the model can adjust its parameters based on the new data to consider time-dependent changes. Based on the prediction results of the model for 2011 data with and without learning data, the model performance can be improved up to 22% if it is fed with the monthly data and up to 10.1% if it is fed with the quarterly data. This shows that the developed BN has the capability to automatically adapt itself by learning the emerging patterns of new incident durations.

Finally, a complete adaptive learning framework, developed by Castillo and Gama (2006), is introduced for predicting incident duration. It is shown that that performance of the BN with integrated adaptive learning improves over time. While the degree of the improvement in performance is initially steep, the degree of improvements
decreases in time. This shows that the model stabilizes as increasing number of training samples become available over time. It was also observed that the performance of the static models, which were estimated once with the initial dataset, decayed over time.

6.2 Conclusions on Spatial Models for Analyzing and Mapping the Locations with Higher Crash Risks for Incident Management

This chapter, first, presents a county level hierarchical Bayesian model framework for New Jersey. Unlike regular NB models, use of hierarchical Bayesian model enabled us to include spatial correlation and global heterogeneity simultaneously. To develop a truly spatial model, only spatial covariates related to roadway conditions were considered in the model development stage. However, due to limitations of the data, spatial covariates were generated from a set of surrogate variables. It is recognized that the surrogate variables may not represent the effects of the considered spatial covariates which is a limitation of this study. The spatial variation of crashes was also analyzed by roadway type. However, due to data limitations, DVMT in a county was used instead of the DVMT by roadway type. In addition to common practice of modeling based on raw crash counts, we employed the use of crash counts with severity weights in the hierarchical crash models. The severity weights proposed in FHWA’s Five Percent Report were used to represent the impact of the different severity crashes. Finally, the crash rate maps were developed from modeling results for raw crash counts, weighted crash counts, and crashes by roadway type. These maps can be valuable for visualizing the spatial meaning of models that are not always easy to understand by just observing
estimated model equations. Furthermore, from a practitioner’s perspective, these maps enable transportation planners to understand and interpret the model results without having to possess advanced mathematical knowledge.

The results of the study indicate that the most influential covariate for the crashes is the curve crashes, followed by roadway mileage and by roadway defects. “Wet roadway” is found to be the least influential factor. By applying severity weights to the crashes in the hierarchical models for the first time, it was found that it is possible to represent the crash risk better through this new model. The developed risk maps based on the raw crash rates and the weighted crash rates from the model identified locations which are prone to more severe crashes. It is believed that the results of this study will help transportation professionals to identify and rank the locations at an aggregate level, which needs further monitoring. The crash rate maps can be used to pin-point local needs and allocate funds by the state or federal governments. The maps of crash rate by roadway type can also be used to understand the effectiveness of local governments in terms of highway safety.

Next, a municipality level hierarchical Bayesian model for the downed trees during hurricanes in New Jersey was presented. The use of ZIP models enabled us to reduce the effect of over-dispersion in the data. Moreover, hierarchical Bayesian modeling framework enabled us to consider the effects of spatial correlation and global heterogeneity simultaneously. To develop a truly spatial model, only spatial covariates related to roadways and weather conditions were considered in the model specification stage. A map of the estimated mean number of downed trees was developed from the
modeling results for the posterior mean of downed trees and the posterior mean of downed trees normalized by roadway density.

The results of the study indicate that the most influential covariate for the downed trees is the roadway density, followed by maximum wind speed and roadway mileage. Average wind speed and precipitation were found to be the least influential factors. The risk map developed based on posterior mean of downed trees has low explanatory power. On the other hand, the normalized mean of number of downed trees based on roadway density identified a cluster of municipalities which are more prone to the occurrence of downed trees following the landing of the hurricane. It is believed that the results of this study will help transportation professionals on identifying and ranking the most vulnerable locations at the municipality level that requires closer attention. These maps can be used to identify local needs and effectively allocate resources by the responsible agencies.

Finally, a point level LGCP model for interstate highway I-287 in New Jersey was developed. This model specification defined at a more disaggregate level enabled us to include covariates at different resolutions in the model. Moreover, the roadway was represented as a continuous entity without predefined links in this model. Such a model captures the continuous nature of the roadways better than any existing model that pre-discretizes the highway into links between interchanges or some other major monuments. Although this model is used in other fields, to our knowledge, this is the first attempt to employ it in the traffic safety literature.

Many spatial covariates were incorporated into the model and their effects analyzed. It was found that pavement characteristics, which are at point level such as,
surface distress and rut depth significantly affect crash risk. Moreover, it was observed that speed limit, number of lanes, lane drop and off ramp covariates were found to be significant and they might be associated with the drivers’ difficulty to adjust to changing traffic and geometric conditions. The population covariate was found insignificant and this result is interpreted to reflect the fact that the point level model might not be correlated with the population of surrounding municipalities since the model is developed for a limited access interstate roadway that does not directly interact with local municipalities / towns at each point. For comparison purposes, a Poisson MCMC model was also estimated at the link level using link level covariates. Poisson MCMC model was unable to capture the significant effects captured in the LGCP model. However, in this model, population was found to be significant. Since this model is at the link level and the interchanges are located at the start and end of link, it is natural for the population at these entry and exit locations to affect the traffic exposure and the crash risk. Finally, the crash risk maps were developed based on the modeling results. These maps showed that it is easier to visually detect the locations with higher crash risk using the LGCP model.

6.3 Practical Implementations

6.3.1 Incident Duration Estimation and Its Use for Real-world Operations

The time between the occurrence of the incident and its clearance from the roadway is called incident duration. There are many factors affecting overall of incident duration during different stages of incident management operations. Using statistical
methods such as regression, decision trees or Bayesian methods and relevant incident variables, it might be possible to make a prediction about incident duration. An accurate estimate of incident duration, especially during the initial stages of incident, might be useful to effectively manage the incident.

In Chapter 4, an incident prediction model, based on Bayesian networks, was developed. Using this model it was possible to estimate the incident duration even if the information on the incident was limited. The model gives the estimates as the probability distribution of possible incident durations. Therefore, instead of a point estimate of incident duration, it is possible to get an overall idea of the incident duration. Moreover, the estimates of the model can be improved, if additional information is provided.

It is possible to use this model during real-time incident management operations to estimate incident duration. Later, this estimate can also be used for delay estimation. Based on incident duration and delay estimates, the traffic managers at the Traffic Operations Center can make more accurate and reliable decisions about directing resources to the incident scene, traffic diversion, disseminating incident information to the motorists and updating advisory information displayed by variable message signs. When and if additional information on the incident is received, these estimates can be updated along with the decisions.

6.3.2 Spatial Distribution of Incidents and Allocation of Resources

Although incidents are random events in space and time, their distribution along a roadway may not be uniform. For some road characteristics which are changing along a roadway such as number of lanes, AADT can affect the likelihood of occurrence of an
incident. As a result, incidents might be clustered at some locations. It is important to identify those locations and the reasons which, have led to clustering of the incidents at those locations.

In this dissertation, a statistical model to represent the spatial distribution of the frequency and severity of incidents along the roadway was examined. Several spatial characteristics of the roadway were taken into account and different models were developed for different types of roadways such as state, authority, county, etc. The results of the models were used to generate predictive heat maps which showed the distribution of incidents and cluster location with their impact areas.

We believe that this method can help transportation professionals about decisions on allocating their valuable resources. Here, the resources refer to incident response teams such as tow trucks; service patrols and ambulances, etc. Currently, transportation agencies do not have robust models that can help their resource allocation decisions. However, if it can be predicted that the incidents are more likely to happen at a certain location, more resources could be assigned ahead of time to that location. Furthermore, this model can be used to predict what types of incidents are more likely to occur at a given location. Hence, types of resources likely to be needed at that location could also be identified beforehand. For example, if the frequency of disablement incidents is high at some location, more tow trucks can be positioned at that location. The spatial distribution of incidents can also be used for more effectively determining the resource allocation during special events such as football games, concerts, etc. During these events, the roadways which are used to access the event location are expected to be heavily congested especially before and after the event. Incidents can cause extra delays
on these roadways. It is thus important to respond in a timely manner and clear the incidents during the special events. However, already congested roadways may make it difficult for incident response teams to reach an incident scene especially if they are located far from a specific incident location. Our model can this be used by transportation professionals to pre-determine strategic locations to place the resources and their type.

The model proposed for studying incidents during hurricanes can also be applied to future hurricanes by only using known roadway characteristics and weather forecast data. Since this model was developed using hierarchical Bayesian modeling framework, it is also possible to make best/worst case scenario analysis based on the lower/upper bound values of the predicted parameters. Then based on this scenario analysis, it is possible to allocate resources / response teams to regions that are most likely to be affected by these types of incidents. This model specification is also applicable to the other regions since the data used in the estimation is generally available for other regions as well. While roadway miles and density data can be obtained from state DOTs, the weather related data for the past hurricanes can be obtained from NOAA.

6.3.3 Use of the Proposed Methodology for the Performance Evaluation of Incident Management Operations

The methodology developed in this study aimed to improve the performance of incident management operations. Therefore, it can also be used for evaluating the performance of incident management strategies employed by a transportation agency.

When new strategies or technologies for different stages of incident management are implemented, they are expected to improve the performance of the management
operations. Therefore, it is necessary to evaluate their performance. This can be done by using average incident duration or average delay due to an incident. However, such an approach only provides a coarse picture of the real conditions since the results are averaged out. For a detailed analysis, it may be more useful to analyze incidents at a more disaggregate level because some incidents can demonstrate the underlying weaknesses of the strategy when they are analyzed separately.

The incident duration estimation/prediction methodology employed in this study can be utilized to evaluate the performance of a new strategy on the incident management operations. This requires an approach similar to before-and-after analysis used when different countermeasures are implemented to improve roadway safety. For example, Yanmaz-Tuzel and Ozbay (2010) studied the application of Full Bayes (FB) approach for before-and-after analysis of road safety countermeasures. In this study, Bayesian approaches were used for crash reduction factor (CRF) estimation. CRF can be replaced by reduction in incident duration for purposes of helping to estimate the performance of a new strategy on incident management operations. The incident duration model can be estimated using the data before the implementation of the strategy. Then, for each incident which occurred after the implementation of the strategy, the model can be used to estimate incident duration. The estimated incident duration from the model can be compared with the actual incident duration after the implementation of the strategy. If the strategy is successful, it can be expected that the actual incident duration will be lower than the duration estimated from the model.

Another important issue in incident management is the determination of locations where the likelihood of an incident is higher than the other segments of the roadway.
Generally, the frequentist approach is used to identify these locations. However, this approach cannot fully account for the contributing factors for clustering of the incidents at a specific location, because the spatial correlation is not recognized in this approach.

In this methodology, hotspots and their impact areas are not solely estimated based on the incident frequency. Spatial characteristics changing along the roadway such as roadway geometry, number of lanes, AADT, existence of a ramp are also incorporated into the model. These characteristics can be also used for the evaluation of the performance of operations along the roadway.

6.4 Future Research

The future research directions can be summarized as follows:

- Although a novel and robust incident duration model was developed in this study, the performance of the proposed model should be investigated using new accident data. This will enable us to better understand location specific factors that might affect the prediction accuracy of BNs.

- For the severity weights, different approaches such as economic impact approach in which property damage only equivalents of accidents are also considered (Oh et al., 2010) should be investigated. Although, an example is given for identifying the most relevant variables for incident duration in a dataset, this method can be extended to crash databases that have a very large number of variables. The merits of these methods in crash studies should be further investigated.
For future studies, it is recommended that the Cox model described in this study is extended to analyze point level crashes in a spatio-temporal setting. This will allow for the observation of temporally varying effects of crash risk. Future efforts should also focus on incorporating the severity of crashes in the crash risk model. Different crash severities can be treated as marks and their inclusion might provide better insights about the variation of crash risk among crashes of different severities.
REFERENCES


