A GIS-BASED BAYESIAN APPROACH FOR ANALYZING SPATIAL-TEMPORAL PATTERNS OF TRAFFIC CRASHES

A Thesis

by

LINHUA LI

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

MASTER OF SCIENCE

August 2006

Major Subject: Civil Engineering

A GIS-BASED BAYESIAN APPROACH FOR ANALYZING SPATIAL-TEMPORAL PATTERNS OF TRAFFIC CRASHES

A Thesis

by

LINHUA LI

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

MASTER OF SCIENCE

Approved by:

Co-Chairs of Committee,	Yunlong Zhang
	Daniel Z. Sui
Committee Members,	Dominique Lord
	Brian Bochner
Head of Department,	David Rosowsky

August 2006

Major Subject: Civil Engineering

ABSTRACT

A GIS-Based Bayesian Approach for Analyzing Spatial-Temporal Patterns of Traffic Crashes. (August 2006)

Linhua Li, B.Eng., Shanghai Maritime University;

M.S., Texas Southern University

Co-Chairs of Advisory Committee: Dr. Yunlong Zhang
Dr. Daniel Z. Sui

This thesis develops a GIS-based Bayesian approach for area-wide traffic crash analysis. Five years of crash data from Houston, Texas, are analyzed using a geographic information system (GIS), and spatial-temporal patterns of relative crash risk are identified based on a hierarchical Bayesian approach. This Bayesian approach is used to filter the uncertainty in the data and identify and rank roadway segments with potentially high relative risks for crashes. The results provide a sound basis to take preventive actions to reduce the risks in these segments.

To capture the real safety indications better, this thesis differentiates the risks in different directions of the roadways, disaggregates different road types, and utilizes GIS to analyze and visualize the spatial relative crash risks in 3-D views according to different temporal scales. Results demonstrate that the approach is effective in spatially smoothing the relative crash risks, eliminating the instability of estimates while maintaining real safety trends. The posterior risk maps show high-risk roadway

segments in 3-D views, which is more reader friendly than the conventional 2-D views. The results are also useful for travelers to choose relatively safer routes.

DEDICATION

Dedicated to

My father, Jialiu Li,

My mother, Jiezhen, Yu, and

My wife, Ying Chen

ACKNOWLEDGMENTS

I would like to thank my Committee Chairs, Dr. Yunlong Zhang and Dr. Daniel Z. Sui for helping me with this thesis. I also appreciate the insightful comments of other committee members, Dr. Dominique Lord and Mr. Brain Bochner. I also appreciate Texas A&M University for giving me the wonderful education and Texas Transportation Institute (TTI) for the research assistantship I received while I was working toward my degree.

I want to thank Dr. Li Zhu for her great help in Bayesian model programming in WinBUGS. I also appreciate Dr. Ned Levine of H-GAC (Houston-Galveston Area Council), Mr. Emmanuel Samson of TxDOT Houston District, Dr. Dennis Perkinson and Mr. Danny Morris from TTI, and many others for providing me crash data, traffic data and valuable comments for this thesis.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

1 INTRODUCTION

1.1 Background

Transportation accidents were the seventh highest leading cause of death in the United States (United States Department of Transportation [USDOT] and Bureau of Transportation Statistics [BTS], 2001). However, motor vehicle crashes, which account for about 95 percent of transportation-related deaths and an even higher percentage of transportation injuries, were the leading cause of death for people between the ages of 3 and 33 (National Highway Traffic Safety Administration [NHTSA], 2005). There were 42,815 fatalities and approximately 2,926,000 injuries in the U.S.A. in 2002, resulting from approximately 6,316,000 police-reported motor vehicle crashes (USDOT and NHTSA, 2004). There is a crash-related death every 12 minutes and crash-related injury every 11 seconds in the U. S.

It was estimated that motor vehicle crashes occurred in 2000 alone cost U.S.A. $230.6 billion (Blincoe et al., 2002). This is equal to about $820 for every person living in USA and 2.3 percent of the U.S. Gross Domestic Product (GDP). The components of the crash costs include productivity losses, property damage, medical costs, rehabilitation costs, travel delay, legal and court costs, emergency services, insurance and administration costs and the costs to employers (Blincoe et al., 2002). Deaths, injuries, and property damage due to these crashes are not only a major cause of personal

---

This thesis follows the style of Accident Analysis and Prevention.

suffering and financial loss to the victims, their families, and friends, but also to society at large.

Despite the rapid progress that has been made during the past 40 years to improve highway safety, motor vehicle crashes have become and remain a major social problem in USA (USDOT and BTS, 2001), and the need to improve traffic safety has become a societal concern at the global level (Evans, 2004).

Decades of interdisciplinary research on traffic crashes has revealed that there are generally three categories of factors affecting traffic safety and efficiency – driver behavior (about 160 million drivers in USA), vehicle types (motorcycles, passenger car, sport utility vehicles, pickup truck to large trucks), environmental conditions (roadway condition, design, number of lanes, capacity, pavement type, traffic flow, speed, density, occupancy, weather, lighting etc.) (Brodsky and Hakkert, 1983; Fridstrøm and Ingebrigtsen, 1991; Golias, 1992; Golob et al., 1990; Haight and Oslen, 1981; Jegede, 1988; Jones et al., 1991; Levine et al., 1995a; Miaou and Lum, 1993; Ng et al., 2002) All these factors interact with each other and influence the occurrences and severity of crashes simultaneously. Among those, errant driver behavior—such as alcohol and drug use, reckless operation of vehicles, failure to properly use occupant protection devices, and fatigue—is a major factor contributing to a high proportion of crashes (USDOT and BTS, 2001).

In USA, the first step for transportation agencies, such as state Departments of Transportation (DOTs) and local government transportation departments to improve transportation safety is to identify and rank hazardous sites on roadways. By focusing on

these identified roadway sites/segments/intersections, funds can be allocated to address critical safety concerns and to develop countermeasures to reduce crash frequency, severity, and risk. Due to limited resources, the question for transportation agencies often is how to accurately determine where the risky roadway sites/segments are and when the risks are the highest.

1.2 Research Objectives

The primary objective of this thesis is to develop a geographic information system (GIS)-based Bayesian approach for area-wide traffic crash analysis. The main focuses of the crash analysis are to estimate relative crash risks and to determine the spatial-temporal patterns of the risks. To better capture the real safety indications, this research:

- differentiates the risks in different directions of the roadways,
- disaggregates different road types,
- integrates hierarchical Bayesian approach to filter the uncertainty in the data with large variance and capture the real safety tendency, and
- incorporates GIS to analyze and visualize the spatial relative crash risks in 3-D views according to different temporal scales.

This thesis uses a Houston case study to demonstrate and validate the proposed method. The results of the analysis can help pinpoint high-risk roadway segments that need attention from transportation agencies. The severe locations could be mapped to alert motorists.

1.3 Thesis Organization

This thesis consists of 5 sections. In Section 1, some background information on transportation safety, especially on automobile crashes, is introduced, and the research objectives are defined. In Section 2, the available literature on spatial and temporal analyses on traffic crashes, GIS applications in crash analysis, crash models, Bayesian approach, and methods of identification of hazardous sites for crashes are reviewed and summarized. In Section 3, the methods used in the research are described, including data collection and preliminary processing, GIS mapping, crash prediction modeling, hierarchical Bayesian modeling, and 3-D mapping technologies are described. Using Houston, Texas as study area, the estimations of parameters of crash models for different road types, results of temporal, spatial, and spatial-temporal patterns of relative crash risks are presented in details in Section 4. The conclusions and proposed recommendations based on the research results are summarized in Section 5.

2 LITERATURE REVIEW

This section contains a review of the literature related to existing studies on traffic crashes in many perspectives. Previous spatial, temporal, spatial-temporal analyses of traffic crashes are summarized in the first section, followed by a detailed review of GIS applications in crash studies. Crash prediction models and Bayesian approach are reviewed in sections 2.3 and 2.4. Section 2.5 introduces the current methods of identifying hazardous roadway sites for crashes. The last section points out the voids in the existing literature that this research is attempting to fill.

2.1 Temporal, Spatial, and Spatial-temporal Patterns Analyses of Crashes

Crashes have been studied from different spatial and temporal perspectives with various methodologies. As for the temporal patterns of motor vehicle crashes, reported studies focused primarily on the fluctuation of the quantity or rate of crashes, injuries, and fatalities according to different temporal scales, such as hourly, daily, monthly, and yearly (El-Sadig et al., 2002; Fridstrøm and Ingebrigtsen, 1991; Levine et al., 1995c; USDOT and BTS, 2001). High levels of temporal aggregation (e.g., yearly) can not easily detect the changes of short-term structural variables, while more disaggregated analysis requires detailed temporal information and has high variability.

Concerning the spatial patterns of motor vehicle crashes, several studies examined crashes that occurred in different environmental settings, such as specific roads (Skabardonis et al., 1998), road types (Brodsky and Hakkert, 1983), intersections

(Golias, 1992; Nicholson, 1985) or corridors (Golob et al., 1990; Okamoto and Koshi, 1989). Additionally, studies on traffic crashes have also been performed at different geographic scales using data aggregated to different administrative units, ranging from census tract/traffic analysis zone (TAZ) (Ng et al., 2002), city (Jones et al., 1991; Levine et al., 1995a), county (Fridstrøm and Ingebrigtsen, 1991; Jegede, 1988), to state and national level (Haight and Oslen, 1981).

Using spatially disaggregated data is another trend in recent crash analyses. Prior to the wide application of GIS, most researchers had used a highly aggregated data set (total number of crashes, total travel distance etc.) or attempted to disaggregate data based on demographic (age, sex, race, etc.), severity (fatality vs. injury vs. property damage only (PDO); long duration vs. short duration, etc.), vehicle types (truck vs. car, etc.) or roadway facility type (urban vs. rural, etc.) rather than spatial aspects (Brodsky and Hakkert, 1983; Fridstrøm and Ingebrigtsen, 1991; Jones et al., 1991; Miaou, 1994). With GIS becoming more and more popular, it was easier or less costly to spatially disaggregate data and conduct analyses at finer resolutions.

Researchers also explored disaggregation of crash data in both time and spatial scales (Black, 1991; Miaou et. al., 2003) to identify the temporal and spatial changes of crash patterns distributions or crash risks. GIS facilitates both spatial disaggregation and temporal disaggregation, and its applications in crash analysis are discussed in the next section.

2.2 GIS Applications in Crash Analysis

"GIS is a collection of computer hardware, software, and geographic data for capturing, managing, analyzing, and displaying all forms of geographically referenced information (Environmental Systems Research Institute (ESRI), 2006)." GIS can provide many more analytical capabilities than many other normal mapping tools. For example, GIS is a database that has geographic information, and it is an intelligent mapping system that can link features with other features. It is also an information transformation tool that can create new geo-database based on existing one by applying analytic functions (ESRI, 2006).

The fundamental difference between GIS and any other information system is "that it has the knowledge of how events and features are geographically located (Goh, 1993)." With spatial (geographic) data stored, GIS enables spatial display, spatial integration, spatial query, spatial analysis and processing, etc., almost all of which had implementations in transportation studies, especially in crash analysis (Miller and Shaw, 2001; Thill, 2000).

2.2.1 Spatial Display

Crashes can be shown at their occurrence locations on digital maps. There are at least three methods to correctly add crashes onto maps. The selection of the method depends on the recorded location data.

- Crashes can be directly added if the exact geographic references of crash locations, like XY coordinates (e.g., latitude and longitude), are available. This applies to the crash locations that were recorded using a GPS receiver.

- Address geocoding can be conducted when the exact address is available. This normally applies to the two format types of crash addresses that were recorded on crash reports. One is the exact address, like street name and number, city, state, and zip code. The other format is an intersection of two streets. Both of these address formats can be recognized and added to the correct locations if the correct address locator manager is selected (Ormsby, 2004). However, this method needs the input of a reference layer, which stores all the geographic information and attributes of each road section.
- Dynamic segmentation (linear referencing) can be performed to locate crashes on routes when the roads on which crashes occurred and their positions relative to the starting points of the roads are known. This method requires a route layer, which includes linear features that store unique identifiers and measurement systems. The crash data should be stored with an attribute indicating the route name and attribute indicating the measurement.

With the development of computer science and GIS techniques, numerous researchers used GIS to display crash locations on digital maps and then perform spatial analysis (including clustering analysis) of crashes (Kam, 2003; Levine et al., 1995a; Petch and Henson, 2000; Steenberghen et al., 2004). The spatial analysis of crashes is discussed in details in section 2.2.4.

2.2.2 Spatial Integration

GIS enhances integration of data from different sources based on geographic locations. In a crash analysis, researchers used GIS to link crash data with traffic data (e.g., traffic volume, speed limit) (Saccomanno et al., 1997), roadway inventory data (e.g., pavement, geometry, road condition, number of lanes) (Siegel and Yang, 1998), environmental data (e.g., land use) (Ng et. al., 2002), demographic data (e.g., employment, population) (Levine et al., 1995b), social-economic and other potential contributing factors and particular locations of interest such as schools (Affum and Taylor, 1995) to better capture the relationship between the crash occurrences and contributing factors. It was also argued that the GIS integration capabilities are fundamental for a project even when GIS is not the focus of the project (Lamm et al., 1995).

2.2.3 Spatial Query

The major advantage of spatial query in GIS is that the results of query from the database can be viewed in a spatial format (Miller et al., 1995), and the results of query from digital maps can be viewed in a tabular format. The linkage between a database format and a spatial format is one of the main characteristics of GIS data.

This capability facilitates contributing factor analysis. One can select crashes based on variables and visually see the spatial patterns of selected crashes and discern geographic relationships. For example, by limiting crashes to those that occurred on Friday and Saturday evenings between 9:00 p.m. and 6:00 a.m. and involved drivers less

than 24 years of age, certain types of spatial cluster could be found (Miller, 1999). One can spatially select crashes to see whether these crashes share common attributes. For example, Affum and Taylor (1995) found that crashes involving children were always within 1 km of schools. The spatial query and tabular query can also be used simultaneously. For example, one can determine whether one crash was caused by an earlier crash by examining a space criterion (within 1600 m of the earlier crash) by spatial query and a time criterion (within 15 minutes of the earlier crash) by tabular query (Raub, 1997).

2.2.4 Spatial Analysis

Spatial analysis of crashes is a highly quantitative, statistical analysis. In crash analysis, the cluster analysis was commonly conducted in order to find out the hot spot of crashes, by either two-dimensional approach or linear approach (Black, 1991; Levine et al., 1995a; Flahaut et al., 2003; Kam, 2003; Petch and Henson, 2000; Steenberghen et al., 2004). One essential problem with the crash cluster analysis is spatial autocorrelation, which had already been discussed by Black (1992), and Black and Thomas (1998). Grid-based crash cluster analysis was suggested and implemented by Choi and Park (1996) and Steenberghen et al. (2004). Following earlier works on network autocorrelation analysis, one recent significant advance is the network-constrained approach to conduct spatial point pattern analysis over a network (Yamada and Thill, 2004). Statistical approaches for cluster analysis are also widely available in a number of software

packages including CrimStat III (Levine, 2004), SAS (SAS Institute Inc., 1999), SPSS (SPSS Inc, 2005), Splus (Insightful Corp., 2001), ArcGIS (ESRI, 2004).

In the linear approach of spatial analysis of crashes, it is common to divide roadways into analysis units (segments) to address the safety. There are different ways to define segments. The Highway Safety Manual (HSM) attempts to make each segment "homogeneous with respect to traffic volume, lane width, shoulder width, shoulder type, driveway density, roadside hazard rating, curvature, grade, presence of passing lanes or short four-lane sections, and presence of center two-way left turn lanes" for two-lane highways (Bellono-McGee Inc. and Midwest Research Institute, 2003). Thus, a new roadway segment starts wherever any of the variables (Annual average daily traffic (AADT), lane width, shoulder width, shoulder type, driveway density, etc.) changes. Furthermore, a new segment begins at intersections, 250 ft before and after the center of each intersection, beginning or end of a horizontal curve, point of vertical intersection for a vertical curve or an angle point at which two different roadway grades meet, etc. (Bellono-McGee Inc. and Midwest Research Institute, 2003). Another way is to define a segment with a specific length, and let the segment dynamically move along the roadway until the segment reaches a specific threshold, like at least such a minimum number of crashes so that the possible hazardous site could be identified (Hovenden et al., 1995; Miller et al., 1995). The third way is sliding window method, which uses a predetermined size of cells/grids sliding on the map. The roadways are cut by the cells into different segments. (Choi and Park, 1996).

2.3 Crash Prediction Models

Researchers over the past two decades have developed a variety of statistical methods to predict crashes in different roadway sections and establish the relationship between crash characteristics (rate, frequency, fatality, injury, duration, severity, etc.) and related variables, such as weather conditions, geometric design of roads, traffic volume, road density, and driver behaviors.

Models that have been widely used since the early days include multivariate linear regression model and log-linear regression model. These two models assume that the random error term in the function is normally and independently distributed with constant variance, so that statistical tests on the model parameters confidence intervals on coefficients and variables can be easily obtained. Unfortunately, there are a lot of practical situations where the assumption of normally distributed error term can not hold. Crash count data, binary responses or other continuous variables with positive and high-skewed distribution can not be modeled with a normally distributed error term (Lord, 2006).

The generalized linear model (GLM) was developed to allow fitting regression models for univariate response data that follow very general distributions, called exponential family, including normal, binomial, Poisson, negative binomial, geometric, gamma distributions etc. (Lord, 2006). GLM usually consists of three components, a random component, a systematic component, and a link function that connects the random and systematic components to produce a linear predictor (Lord and Persaud, 2000). GLM has the advantages over multivariate linear model and log-linear model

because it provides flexibility in choosing a link function and a distribution of the error term (Qin and Lin, 2006). For example, log-linear link function and negative binomial error distribution can be chosen for crash count data analysis. The formal structure of GLM is summarized by Myers et al. (2002). Equations 2.1 and 2.2 show a typical functional form and its log form for statistical models in safety. It is very convenient to use software program to fit the data and estimate coefficients.

$$y_i = \beta_0 * F_i^{\beta_1} \exp(\sum_{j=2}^{k} \beta_j x_j) \tag{2.1}$$

$$\ln y_i = \ln \beta_0 + \beta_1 \ln F_i + \sum_{j=2}^{k} \beta_j x_j \tag{2.2}$$

Where: y_i = outcome or response variable, crash count per unit of time;

F_i = traffic flow;

β_0 = intercept;

β_1 = coefficient for traffic flow;

β_j = unknown coefficients; and

x_j = covariates or explanatory variables.

Miaou and Lum (1993) found that conventional linear regression models were not appropriate for modeling vehicle crash events on roadways. When the mean and variance of the crash frequencies were approximately equal, the Poisson regression was found to be a more appropriate model for examining the relationship between crashes and influential factors (Miaou, 1994). Overdispersion occurs when observed variance of

the data is larger than the predicted variance. When overdispersion was moderate or high, the use of negative binomial regression was found to be more appropriate (Miaou, 1994).

In the recent research, Lord and Persaud (2000) presented an application of a generalized estimating equation (GEE) procedure to develop a crash prediction model that incorporates temporal trends in crash data. The results showed that the GEE model incorporating time trends was superior to the models that did not accommodate trend and/or the temporal correlation in crash data.

The relationship between the number of crashes (by type, severity, etc.) and the amount of exposure on a transportation facility (e.g., traffic flow, travel distance, etc.) was of interest to researchers. Linear relationship ($\beta_1 = 1$) between number of crashes and AADT was supported by Chipman (1982), Gårder (1989), Hauer (1982), and Janssens (1999) and Wolf (1982). This means that the individual probability of being involved in a crash increases linearly as the exposure increases. Persaud and Dzbik (1993) found that $\beta_1 = 1.206 > 1$ for severe crashes on freeway segments that have more than 4 lanes. However, most researchers agree that the relationship between crashes and traffic flow has been shown to actually follow a non-linear relationship ($\beta_1 < 1$), in which, crash counts usually increase at a decreasing speed as traffic flow increases (Tanner, 1953; Hauer and Persaud, 1988). This implies that the individual risk decreases as the flow increases and it is less dangerous for a driver to travel under heavy flow conditions than under light flow conditions (Lord, 2002). The possible reasons include that the variance of the speed becomes lower; drivers are more alert; and more non-reported minor crashes occurred as traffic flow increases.

However, the result that power coefficient of traffic flow is less than 1 can cause problem in the optimization of network safety. Maher et al. (1993) attempted to optimize safety and vehicular delay simultaneously on computerized transportation networks, and found that traffic flow tends to concentrate on a few roads rather than disperse on many roads when the network is solely optimized for safety. This raises doubts on the correctness of $\beta_1 < 1$. Furthermore, common sense would tell us just the opposite of $\beta_1 < 1$. The individual risk of being involved in a crash should increase under heavy flow condition because of limited gaps or spaces, reduced maneuverability and more complicated driving tasks.

Therefore, Lord (2002) argued that perhaps the current model forms in the literature are not appropriate, and the traffic flow may not be a suitable measure of exposure. It is suggested that Gamma function might be more appropriate for the relationship between crash count and traffic flow, and density might be a suitable measure of exposure in crash function (Lord, 2002; Lord et al., 2005).

2.4 Bayes Method

The Bayesian approach has been widely used in statistics and sciences over the past decade. One of the major advantages of the Bayesian approach is that it is capable of predictive forecasting of risks even in the presence of sparse data or rare events (Withers, 2002). The ability to incorporate prior knowledge without the restriction of

classical distributional assumptions makes Bayesian inference a potent forecasting tool in a wide variety of fields (Withers, 2002).

In the modeling and data analysis, the Bayes method can be shown in Equation 2.3. Given y as the observed crash frequency, what is the probability of occurrence of θ crashes?

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)} = \frac{p(\theta)p(y|\theta)}{\int p(\theta)p(y|\theta)d\theta} \tag{2.3}$$

Where: $p(\theta|y)$ = posterior probability conditional on y;

$p(\theta)$ = prior distribution (can be informative or non-informative);

$p(y|\theta)$ = likelihood function when it is regarded as a function of θ for a fixed y; and

$p(y)$ = prior projective distribution (also called marginal distribution of y).

2.4.1 Empirical Bayes Method

The empirical Bayes (EB) method is usually employed to simplify the Bayes computation. The prior distribution is estimated from actual data. This goes against the fundamentals of full Bayes method because EB method uses the data twice. For the EB method, different weights are assigned to the prior distribution and standard estimate respectively. The weights are estimated with assumption that the mean for each segment follows a Gamma distribution.

$$\mu_{EBit} = \gamma_{it}\hat{\mu}_{it} + (1-\gamma_{it})y_{it} \tag{2.4}$$

$$\gamma_{it} = \frac{1}{1+\frac{\hat{\mu}_{it}}{\phi}} \tag{2.5}$$

Where μ_{EBit} = empirical Bayes posterior estimate for segment *i*, time *t*;

$\hat{\mu}_{it}$ = expected crash count of segment *i*, time *t*, (maximum likelihood estimate, mean of a, can be estimated from a reference site);

γ_{it} = weight given to expected crash count for segment *i*, time *t*;

y_{it} = observed crash count for segment *i*, time *t*; and

ϕ = inverse dispersion parameter of Poisson-gamma regression (maximum likelihood estimate).

As can be seen from the equations, the posterior estimate is always between the observed crash count and expected crash count. The weight is dependant on the data themselves. For a fixed $\hat{\mu}_{it}$, as ϕ increases from 0 to infinity, the EB estimate would increase the weight on $\hat{\mu}_{it}$ from 0 to 1, and conversely decrease the weight on y_{it} from 1 to 0 (Hauer and Bamfo, 1997; Hauer et al., 1989; Miaou and Lord, 2003). Poisson model can be obtained when ϕ approaches infinity.

The applications of the EB method in crash frequency estimation and before-after studies are abundant (Arnold and Antle, 1978; Hauer, 1992, 2002; Ng et al., 2002; Mountain et al., 1996; Qin and Lin, 2006). It can also eliminate effects due to regression-to-the-mean in crash data (Brüde and Larsson, 1988; Hauer, 2002). It is expected that EB method will continue to be popular because of the relatively simple computation and its superior ability to perform a maximum likelihood estimate.

2.4.2 Full Bayes Method

Full Bayes method is a computationally-intensive process, so it is commonly modeled using computers and Markov Chain Monte Carlo (MCMC) simulation techniques are now frequently used for estimating the posterior distribution. It commonly uses multiple levels of analysis in an iterative way, so it is called hierarchical Bayes model. The hierarchical model allows the modeler to structure some dependence between the parameters under study in a logical manner. The hierarchical procedures, whether from a full Bayes or EB model, usually result in a smoothing of estimates for each unit towards the average outcome rate, and have generally been shown to have better precision and predictive performance (Congdon, 2001). This thesis uses the hierarchical Bayesian approach to model relative crash risks, with detailed modeling process shown in section 3.5.

In the public health area, considerable progress has been made in developing methodologies in disease mapping, risk assessment and prediction, particularly in the application of hierarchical Bayes models (Besag and Newell, 1991; Wakefield and Morris, 2001; Wakefield et al., 2000; Waller, et al., 1997; Zhu and Carlin, 1999). Studies have shown that the risk estimation using hierarchical Bayes models has several advantages over classical methods. Disease is a rare incident, and typically rare for an analysis unit, like census tract or county. Therefore, there is a large variability across analysis units, especially for analysis units with small population size. This makes it difficult to differentiate the chance variability and genuine difference in the estimates.

Hierarchical Bayes methods, with proper spatial-correlated random effects modeled, have been shown to be able to account for the high-variance estimate in low population area and keep the overall spatial trends (Ghosh et al., 1999; Sun et al., 2000).

Traffic crashes are very similar to disease incidents, in its rare occurrence and large variability across the analysis units, so it is appropriate to borrow hierarchical Bayes models from disease mapping and apply to traffic crash analysis, risk assessment and crash mapping. The hierarchical Bayesian approach was already implemented in crash analysis in some studies to estimate crash risk and safety performance (MacNab, 2004; Miaou et al., 2003; Miaou and Song, 2005; Qin et al., 2005). Risk maps for area-based traffic crashes were also explored by MacNab (2004) and Miaou et al. (2003).

2.5 Identification of Hazardous Locations

There are a lot of methods used for the identification of hazardous locations, which is the prerequisite of the engineering study to find out the countermeasures. The hazardous locations are the sites that experience significantly more crashes than other sites with the same characteristics.

2.5.1 Frequency Method

The frequency method (Fitzpatrick et al., 2000; World Road Association (PIARC), 2003) summarizes the number of crashes for each site, and ranks them by descending order. Those sites with more than a predetermined number of crashes are

classified as high-frequency site. This method is commonly used to measure the safety for a spot location (hot spot identification). However, this method does not take into account of vehicle exposure (e.g., traffic volumes), which has a direct relationship with crash frequency. Therefore, the results have bias toward high-volume sites. This method also suffers from the regression-to-the-mean bias in which an unusually high count is likely to decrease subsequently even if no improvement were implemented (FHWA, 2002; Hauer, 1997).

2.5.2 Crash Rate Method

The crash rate method (Fitzpatrick et al., 2000; PIARC, 2003) ranks sites by the ratio between number of crashes and vehicle exposure. Rates are given in crash per million entering vehicles (EV) for spot locations and crashes per million vehicle miles traveled (VMT) for segments. Sites with rates higher than a predetermined rate are classified as high-rate sites. The advantage of this method is that it includes the vehicle exposure as denominator. The importance of using traffic volume as normalization is emphasized by Affum and Taylor (1996).

The sliding window-based ranking approach is also a crash rate method. Crash frequency, VMT and other explanatory covariate values for each cell are aggregated and calculated for each cell. Then the crash rate is calculated for each cell independently and then the cells are ranked by the rates (Choi and Park, 1996; FHWA, 2002).

In order to make rate ranking valid, the linear relationship between number of crashes and EV or VMT, which has been argued in many studies to be questionable, has

to hold (FHWA, 2002). Another problem in this method is that the observed rates have large uncertainties in low vehicle exposure (low EV or short length or low VMT) sites. Because of the high uncertainty, the rates on low exposure segments tend not to be useful since they can be extremely high or extremely low (Miaou and Song, 2005).

2.5.3 Critical Crash Rate Method

The critical crash rate method (Fitzpatrick et al., 2000; PIARC, 2003) identifies those sites where crash rate is greater than the average crash rate for similar sites across the state or similar region. This method also applies a statistical test to determine the significance of each site's crash rate when compared to the mean crash rate of similar sites. This method also has the small area estimation problem as discussed in the crash rate method.

2.5.4 Equivalent Property Damage Only Method

The Equivalent Property Damage Only Method (Fitzpatrick et al., 2000; PIARC, 2003) weights fatal and injury crashes against a base-line of PDO crash. For example, fatal crash and serious injury are given a weight 9.5, and minor injury is given a weight 3.5, and PDO has the weight 1 (Agent, 1973). The Equivalent PDO index value is calculated based on different weights for different severities for each site, and sites are ranked by the Equivalent PDO index.

This method improves on the previous methods in that it includes crash severity into consideration. However, this method does not include exposure and it has the bias toward high-speed sites. Furthermore, the weights are assigned arbitrarily.

2.5.5 Relative Severity Index Method

The relative severity index method (Fitzpatrick et al., 2000; PIARC, 2003) incorporates weighted average cost of crashes at site. In this method, crash frequency at each severity level is multiplied by the average cost for crashes at that severity level. The subtotal for each of these severity-specific costs are summered and the sum is divided by the total crash frequency. This method has the same limitations as the equivalent PDO method.

2.5.6 Combined Criteria Method

In the combined criteria method (Fitzpatrick et al., 2000; PIARC, 2003), more than one previously mentioned method is used (e.g., frequency combined with rate method). Sites are ranked by one method first and sites ranked high are then investigated by another method. Different weights can also be assigned to different methods to select priority locations for investigation. This can avoid the limitations of one single method.

2.5.7 Statistical Model Method

This method (Fitzpatrick et al., 2000; PIARC, 2003) requires the development of statistical models using the reference population, and compares the observed value with the predicted value. Statistical model can account for non-linear relationship between number of crashes and vehicle exposure to produce more accurate rank. However, this method is relatively complex.

2.5.8 Potential for Safety Improvement

The potential for safety improvement method (FHWA, 2002; PIARC, 2003) has also been defined as "identification of sites with promise." This method compares the observed or predicted values at given site with predicted values estimated from a reference population. The difference between the two indicates that the site could potentially reduce its number of crashes to those of the reference population. This method is always used with EB method, which is discussed next.

2.5.9 Empirical Bayes Method

The EB method (FHWA, 2002; PIARC, 2003) uses information from the reference population and the observed information at the site to generate a more accurate estimate of the crashes. This method takes into consideration of long-term mean, which has not been addressed by any of the previous introduced methods. If the reference population is homogeneous, EB-method of moments should be used, in which the mean

crash frequency and variance are calculated directly from reference population. When the reference population is not homogeneous, EB-regression method should be used, in which crash prediction models are developed using statistical regression techniques to serve as reference population (Hauer 1992, 2002, 2004).

The EB method is often used with the Potential for Safety Improvement method. The potential for improvement is calculated as the difference between the EB predicted crash frequency and average crash frequency of reference population (when population is homogeneous) or multivariate model predicted crash frequency (when population is not homogeneous) (Hauer, 2002; PIARC, 2003).

2.5.10 Full Bayes Method

The full Bayes method is a relatively new method, and it ranks sites using posterior probabilities that a site experiences more crashes than expected. This method can include all covariates of the model for the ranking purpose, and it can solve the small area estimation problem and remove the effects of spatial and temporal autocorrelations. Miaou and Song (2005) have used this method to rank intersections and counties. This thesis uses this method to rank roadway segments. This thesis uses a full Bayes method to rank the hazardous roadway segments. Section 3.5 presents the detailed hierarchical Bayesian modeling process.

2.6 Problems in the Current Literature

In the current literature, risk maps for area-based (e.g., county-based) traffic crashes were explored by MacNab (2004) and Miaou et al. (2003). However, the link-based (e.g., roadway segment-based) risk maps and rank risky segments using a hierarchical Bayes model has not been done very often, although full Bayes model has been used for roadway sections (Miaou, et al., 2005). The hierarchical Bayes model can incorporate spatial assumptions and enable the customary Bayesian approach to solve the uncertainty of the relative risks in low exposure areas, and catch similarity of relative risks in nearby or adjacent regions. In link-based crash risk hierarchical Bayes modeling, the spatial autocorrelation structure is different from area-based modeling.

In the analyses of area-wide crash patterns, it is not common to disaggregate crashes in the same location based on different directions (e.g., eastbound/westbound, northbound/southbound). The recent research by Qin et al., (2006) showed that the choice of flow split factor has no significant effect on the other risk factors based on their small samples of data. In other words, they found that the results from the model that considers directional difference is similar to the one that does not take traffic direction as consideration. However in this thesis, relative risks in different directions as well as crashes are differentiated because theoretically opposite directions of roadways may have different risk values due to contrasted crash counts, traffic characteristics (e.g., traffic volumes), roadway conditions (e.g., work zone), and environments (e.g., lighting). Without differentiating directions for automobile crashes, risk values for two directions might have been averaged out, leading to erroneous risk estimation.

Splitting directions will split the data that include spatial autocorrelations apart. Theoretically, the occurrence of a crash will influence the surrounding traffic conditions. For example, the vehicles behind the crashed vehicle might experience unexpected delay due to the crash; the vehicles on the opposite directions might also be influenced because drivers in the other direction tend to look at crashes that occurred in this direction and may lose concentration on their own driving. These are the indications of spatial autocorrelation. The hierarchical Bayesian approach used in this thesis attempts to address spatial autocorrelations between adjacent segments, between opposite directions of roadways and between freeway and frontage road.

The trip-based crash rate analysis was advocated because it was assumed that every trip (route) is unique in terms of crash risk (Kam, 2003). However, each route (e.g., from home to work) includes different road types (e.g., local, collector, arterial, freeway) and different road types should have different roadway, traffic, and environmental conditions, so crash risks on different road types should be different. Therefore, disaggregation of the crash analyses for different road types is needed. Furthermore, it is noticed that even the same road type's physical environmental conditions (e.g., road condition, and lighting) may be different both spatially (by segment) and temporally (by hour/day/week/year). Therefore, it is believed that spatial disaggregation (including road type disaggregation) and temporal disaggregation (hourly, weekly and yearly) are necessary and would produce usable results.

The visualization effect is important for information transfer. In the display of roadways (links) in GIS, researchers could adjust the width and color of the roadways to

produce a 2-D visualization, but a 3-D visualization can provide more information than a 2-D visualization because of the extra dimension. A 3-D map turns points into vertical lines, lines into walls, polygon into blocks and it can be shown from different azimuths and altitudes so it is more reader friendly. This research explores a new 3-D visualization approach for mapping traffic relative crash risks along transportation networks, which can help presentations of research results.

3 STUDY METHODOLOGIES

This thesis attempts to develop a hierarchical Bayesian approach to conduct area-wide link-based relative crash risk analysis, in which different directions, road types are disaggregated along both spatial and temporal scales. Bayesian spatial smoothing among adjacent roadway segments was implemented to filter the uncertainty in the data and better capture the real risk trend. The final stable posterior relative crash risks by different temporal scales are detailed into segment level and displayed with 3-D illustrations. This method could help accurately rank the hazardous sites that need safety improvements.

In this section, the study design, data collection and processing, GIS digital mapping and the modeling process are described. The first section presents the characteristics of the study area and the roadway network. The second section introduces the available data and how they were processed. The next section contains the methods of GIS mapping of the crashes. The fourth section illustrates the development of crash prediction model, and the final section includes step-by-step Hierarchical Bayesian modeling methodologies used in this thesis research.

The overall process in this thesis can be divided into three parts (Figure 3.1): (1) data preparation and preliminary processing; (2) Bayesian spatial smoothing and updating; and (3) GIS mapping and visualization of final results. ArcMap (ESRI, 2004) and ArcScene (ESRI, 2004) developed by ESRI are used for GIS operations and 3-D mapping respectively, and Bayesian modeling is performed using WinBUGS

(Spiegelhalter et al., 2004) and its add-on program GeoBUGS (Thomas et al., 2004) that fits spatial models and produces outputs for maps.

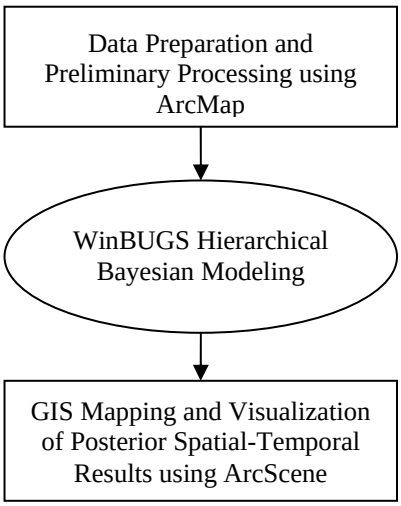


Figure 3.1 Overall Processing Flow Chart

3.1 Study Area

Central portion of Houston, Texas was chosen as the study area for this thesis research. Houston is the fourth biggest city in USA and the biggest in Texas by population from the US 2000 census (demongraphia.com, 2001). However, the problem of traffic safety in Houston is important. There were 45,228 serious crashes within the City of Houston in year 1999, which constituted 14.5% of all serious crashes within Texas, compared to the City's 9.4% share of the Texas population (H-GAC, 2006).

In this research, the state-maintained roadways in the study area are investigated. The study roadways include major freeways, such as I-10, I-45, I-610, US-59, TX-288,

and US-290, as well as their frontage roads, and arterials such as US-90, TX-35, FM-521, and FM-1093 etc.

Each roadway has two directions, represented by two lines in the GIS map (Figure 3.2). The roadways were evaluated separately by direction in order to avoid the averaging effect so that more accurate direction-based results could be obtained.

Before the safety evaluation and crash prediction, each roadway was divided into analysis unit consisting of individual homogeneous roadway segments. In the segmentation process, each urban freeway mainline and each frontage road were divided into segments by entrance ramps and exit ramps. Arterials were divided into segments by major intersecting roads. After the segmentation process, the roadways under evaluation are consisted of 1108 segments with varying length, each of which is homogeneous with respect to traffic volumes, number of lanes, and lane width. It is noted that this proposed method in this research is a macroscopic method, which only evaluates roadway segments, and does not evaluate safety for each individual arterial intersection because of the unavailability of physical characteristics of the intersection and crash data on the intersecting roads.

The segments are grouped into five different types for safety evaluation according to their different functions and characteristics. The types, segment counts, and the statistics of segment lengths are listed in Table 3.1.

Table 3.1 Roadway Segment Types and Statistics

Roadway Segment Types	Number of Segments	Minimum Length (mi)	Maximum Length (mi)	Average Length (mi)
Urban freeway mainlines	471	0.11	1.36	0.45
Freeway system interchange area	112	0.12	1.00	0.36
Frontage roads without diamond interchange	171	0.10	0.89	0.41
Frontage roads with diamond interchange	228	0.13	1.01	0.45
Arterials	126	0.15	2.01	0.69

There is no separate type for freeway ramps. Entrance ramp and exit ramp are the entrance to and exit from the freeway mainline and are considered as a part of the freeway mainline in this study. As a result, crashes that occurred on ramps are considered to have happened on the freeway mainlines. Freeway system interchange area is the area where two freeways interchange, starting from the upstream interchange ramp gore and ending at the downstream ramp gore of the same interchange. Diamond interchanges are common where a freeway crosses a non-major road in this study area. The frontage roads in diamond interchange areas are defined as the frontage road with diamond interchange segment. The other frontage roads are defined as the frontage roads without diamond interchanges. The roadway network in the study area is shown in Figure 3.2 with five types of roadway segments identified.

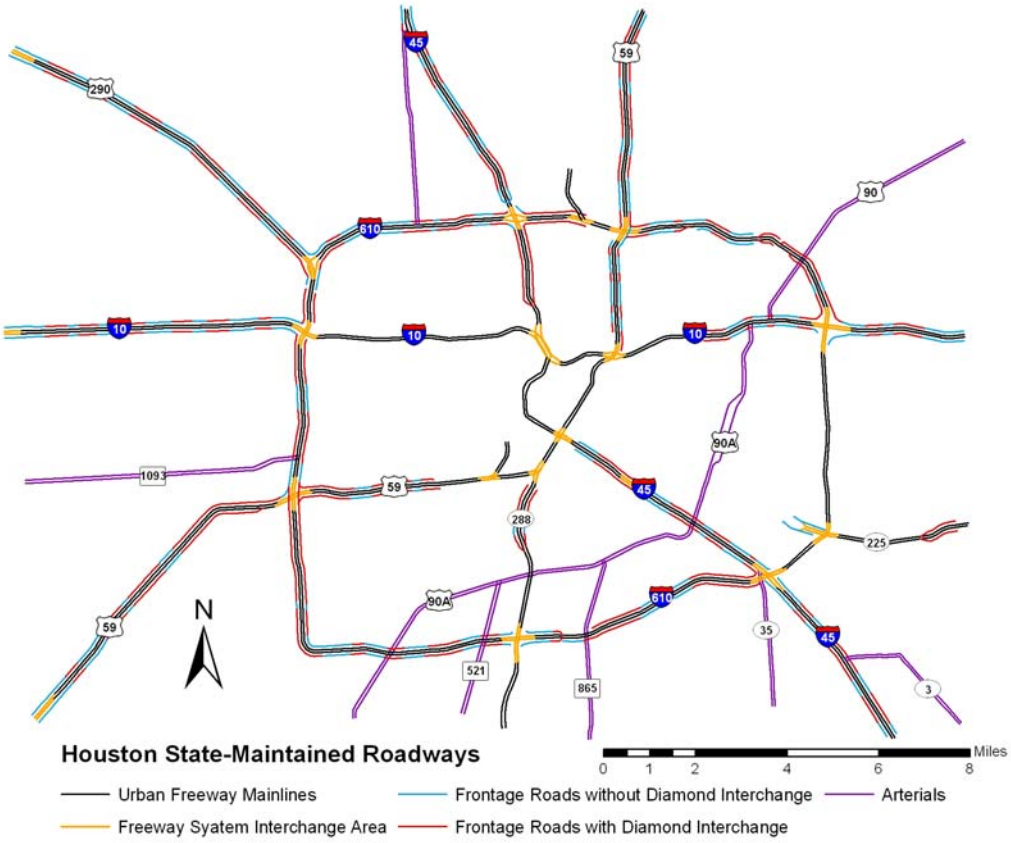


Figure 3.2 Study Network and Segment Types

3.2 Data

The data used in this study include crash data, traffic volume data, volume adjustment factors and GIS roadway data. They were obtained from different data sources and contained much related information for the study area.

3.2.1 Crash Data

Five years' crash data (1996-2000) from the Texas Department of Transportation (TxDOT) Traffic Operations Division Crashes Data Files (TRF crash files), were

obtained and processed. Only the crashes with fatality (K), incapacitating injury (A) and nonincapacitating injury (B) were used in this thesis research and are referred to as "KAB crashes" in this study. The data include information for crashes on all the state-maintained roadways such as location (by control-section number and milepoint), severities (e.g. fatality, injuries, property damage only), time and date, weather, road condition, etc. Data for crashes on roadways maintained by local government are not included in the database.

3.2.2 Traffic Data

Annual Average Daily Traffic (AADT) information from 1996 through 2000 was retrieved from traffic maps of Harris County, obtained from the Houston District Planning Office of TxDOT. Traffic maps indicate the AADT for each roadway segment (combined both directions) in the study area. However, for freeways, the AADT indicated by traffic maps is the combined volume of both freeway mainlines and frontage roads. In order to separate the volumes for these two road types, a separate freeway mainline volume file was obtained from the Center for Transportation Safety at Texas Transportation Institute (TTI). This file has all the mainline AADT for all state-maintained roadways in the study area during the same study period. The AADT for the frontage roads was calculated as the difference between the combined AADT and mainline AADT. Annual average daily vehicle miles traveled (AADVMT) on any segment was calculated by multiplying the AADT and the length of the roadway segment (Equation 3.1).

$$AADVMT = AADT \times L \tag{3.1}$$

where, $AADT$ = annual average daily traffic (two-way);

$AADVMT$ = annual average daily VMT (two-way); and

L = segment length (miles).

Average Daily VMT (ADVMT) was determined by multiplying daily adjustment factors with AADVMT (Equation 3.2). Hourly VMT (HVMT) was calculated by multiplying hourly adjustment factors with ADVMT (Equation 3.3). These adjustment factors were obtained from Transportation Modeling Program at Texas Transportation Institute (TTI) and shown in Tables 3.2 and 3.3. TTI assumed that Monday through Thursday had the same daily VMT adjustment factor and hourly adjustment factors. It is reasonable because there is mainly commuting traffic on these days and the traffic shows a very similar hourly pattern on each of the four days according to traffic data.

$$ADVMT = F_d \times AADVMT \tag{3.2}$$

$$HVMT = K_h \times ADVMT = K_h \times F_d \times AADVMT \tag{3.3}$$

where, $ADVMT$ = average daily two-way VMT;

F_d = daily VMT adjustment factor, ratio of the daily VMT to the annual average daily VMT;

$HVMT$ = hourly VMT (two-way); and

K_h = hourly VMT adjustment factor, ratio of the two-way hourly VMT to the two-way average daily VMT.

Table 3.2 Daily VMT Adjustment Factors (F_d) (source: TTI)

	Daily VMT Adjustment Factor
Friday	1.15888
Saturday	0.94946
Sunday	0.75293
Monday-Thursday	1.03468

Table 3.3 Hourly VMT Adjustment Factors (K_h) by Day of Week (source: TTI)

	Hourly VMT Adjustment Factors by Day of Week			
Hour of Day	Friday	Saturday	Sunday	Monday-Thursday
12AM-1AM	0.009268	0.019596	0.025397	0.009529
1AM-2AM	0.006234	0.01332	0.018286	0.006143
2AM-3AM	0.005939	0.012402	0.017506	0.005635
3AM-4AM	0.005288	0.00854	0.011188	0.005138
4AM-5AM	0.008397	0.008764	0.008634	0.008693
5AM-6AM	0.02491	0.015995	0.011571	0.027675
6AM-7AM	0.055524	0.026519	0.016586	0.060862
7AM-8AM	0.068679	0.034494	0.020995	0.072897
8AM-9AM	0.056887	0.042011	0.027562	0.060127
9AM-10AM	0.04875	0.049167	0.040078	0.050988
10AM-11AM	0.048288	0.055528	0.051903	0.049274
11AM-12PM	0.051398	0.060432	0.057388	0.051632
12PM-1PM	0.054019	0.063779	0.065177	0.053004
1PM-2PM	0.055247	0.06416	0.070741	0.054463
2PM-3PM	0.058843	0.064205	0.07105	0.057966
3PM-4PM	0.065018	0.064441	0.070007	0.064375
4PM-5PM	0.069372	0.063472	0.068647	0.071503
5PM-6PM	0.070934	0.062825	0.068948	0.077139
6PM-7PM	0.060299	0.059991	0.066978	0.061615
7PM-8PM	0.049222	0.052214	0.057674	0.045984
8PM-9PM	0.039658	0.045855	0.051762	0.036039
9PM-10PM	0.034536	0.042524	0.043862	0.030995
10PM-11PM	0.030332	0.038484	0.034338	0.023266
11PM-12AM	0.022958	0.031278	0.02372	0.015057

Hourly directional volume data at some locations were obtained from the District of Transportation Planning at TxDOT Houston District. The data are from several

permanent traffic counting stations on major freeways such as I-45, I-610, US-59 and US-290, which continuously recorded hourly traffic volumes on both directions. Hourly directional VMT (Equation 3.4) were adjusted based on the directional split factors, which were calculated from the directional volume data collected by count stations. It is assumed that all the traffic, the volumes of which are indicated on each roadway segment by traffic maps, travel through the whole length of each segment. Therefore, the derived traffic direction split factors are the same as VMT directional split factors since the roadways on both directions have the same lengths. It is also noted that the data are from only seven count stations, and the direction split factors of other roadway locations have to been derived based on the existing count station data. Figure 3.3 gives an example of the directional split factors for weekdays from a count station located on US-59 south at Beechnut Street, which is outside I-610 loop. It can be seen that in the morning peak, eastbound traffic is above 60% of the combined volume. In the afternoon the westbound traffic becomes to the majority until the early morning. It is assumed that the US-59 south outside I-610 loop had the same directional distribution as the one obtained from the aforementioned count station.

$$DHVMT = D_i \times HVMT = D_i \times K_h \times ADVMT = D_i \times K_h \times F_d \times AADVMT \quad (3.4)$$

where, $DHVMT$ = hourly VMT in one direction; and

D_i = directional split factor, ratio of hourly volume in one direction to the two-way volume.

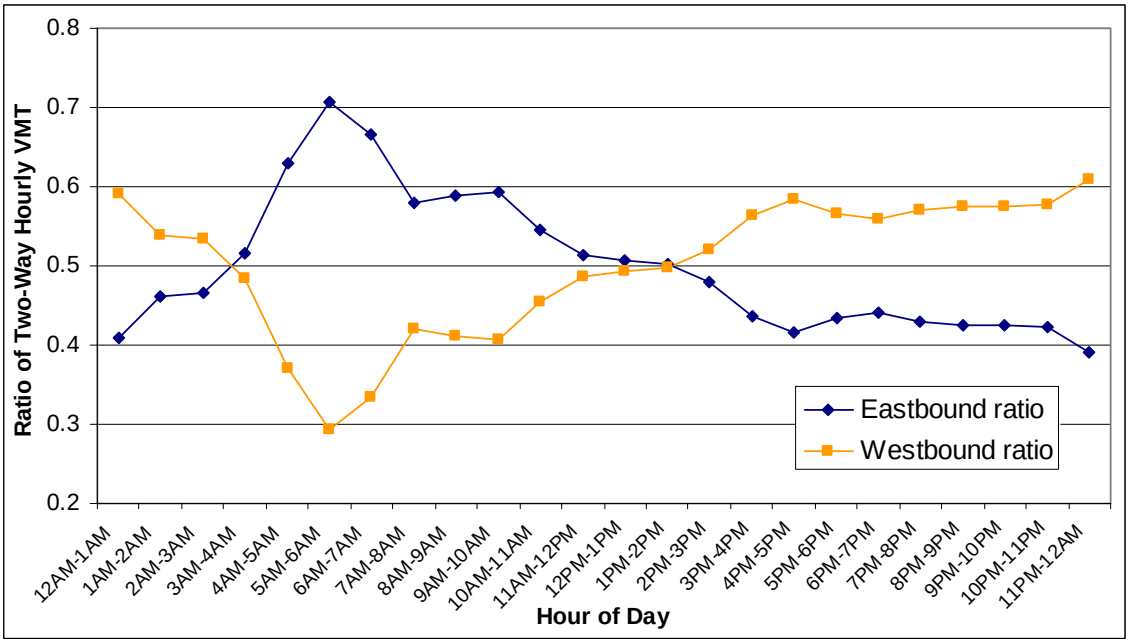


Figure 3.3 Directional Split Factors (D_i) by Hour of Day on Weekdays on US-59 South at Beechnut St.

The roadways inside I-610 loop must be adjusted according to assumptions since there is no count station data inside the loop. Theoretically, the directional difference of traffic inside I-610 loop should not be as big as outside the loop since employment areas are more scattered inside the loop. The directional difference was assumed to end in downtown, the freeway surrounding which was assumed to have had 50/50 traffic directional split all the time. The freeway segments between downtown and I-610 loop were assumed to have had the average directional split of those of downtown and the count station outside the loop, which is shown in Figure 3.4. This traffic adjustment method was applied to all the freeways except I-610 loop since I-610 does not head to downtown. In the study area, four count stations are located on the I-610 loop with each station on each direction. The hourly traffic split on each of the four directions of the

loop was assumed to be the same as the corresponding count station. The traffic on other days of a week was also adjusted by using the same method based on the count station data for corresponding days of a week.

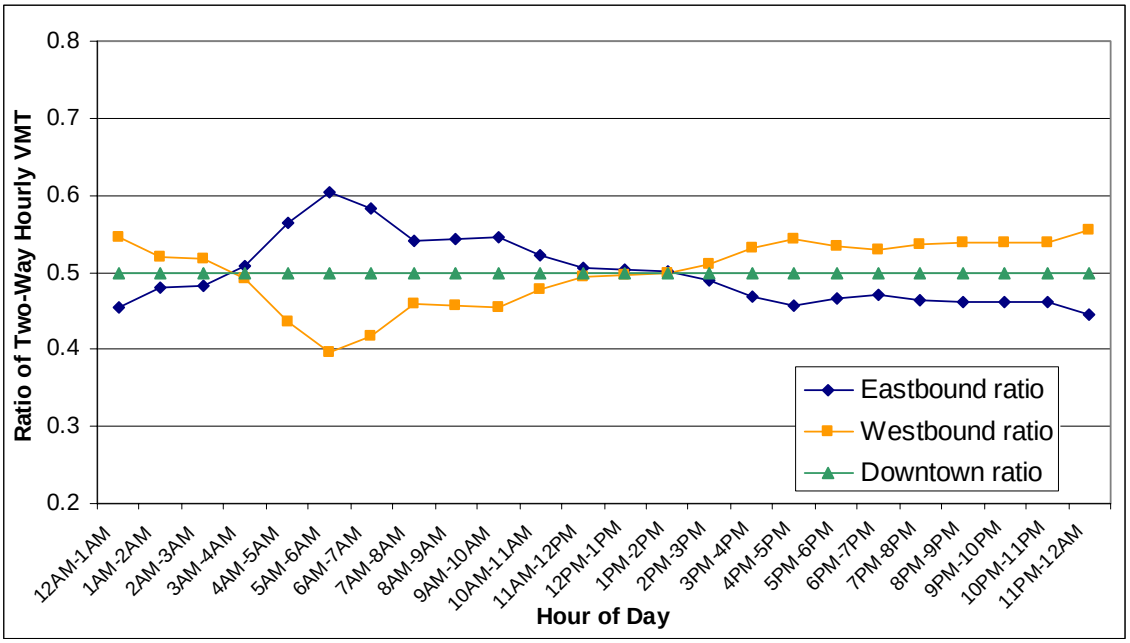


Figure 3.4 Directional Split Factors (D_i) by Hour of Day on Weekdays on US-59 South inside I-610 Loop

It is also assumed that the directional split of each roadway on a daily basis was 50/50 during the study period. This means that the inbound daily volume was the same as outbound daily volume.

3.2.3 GIS Roadway Data

The base state-maintained roadway network was retrieved from TxDOT link shape file. Proper editing (cut, divide, split, merge) was done in accordance with the

roadway segments. In Figure 3.2, every link represents one roadway in one direction. The two lines in the middle are mainlines of freeway. The links that are on both sides of the mainline are frontage roads. Arterials do not have frontage roads beside them. Some freeway segments do not have frontage roads beside them either.

The base control-section map for the study area was obtained from TxDOT map office in Austin. Control-section number is a specific identification number for each roadway section defined by TxDOT. The location of a crash in TRF crash dataset is simply recorded as a control-section number and a milepoint, which describes the position relative to the starting points of the control-section.

3.3 GIS Mapping of Crash Locations

There are at least three methods to locate crashes onto digital maps, as mentioned in the literature review: direct adding based on XY geographic coordinates, address geocoding, and dynamic segmentation. The selection of the method depends on the available location data. In this research, dynamic segmentation was performed to locate crashes on routes because the roads on which crashes occurred (control-section numbers) and their relative positions to the starting points of the roads (milepoints) were known and other location information, like exact address, zip code, latitude or longitude was not available.

The base roadway network shape file was retrieved from TxDOT link data. Proper editing (cut, divide, split, merge) was done again based on the roadway control-sections. Then, routes were created from existing links by giving them corresponding

control-section number, identifying milepoint directions and adding starting and ending measurements. The Houston state-maintained roadway network routes file, which was used as route reference, was successfully created with control-section number as the route identifier and milepoint as the measurement.

The ESRI's ArcMap was used to perform dynamic segmentation. The TRF crash file, which indicates the crash location along the route with control-section number as the route identifier and the milepoint as the measure location field, was used as point event table. Figure 3.5 shows the example of the dynamic segmentation process.

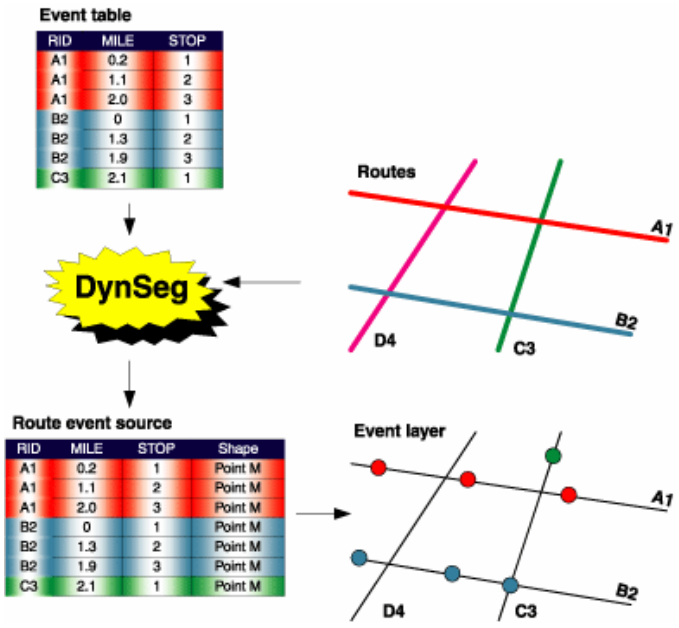


Figure 3.5 Dynamic Segmentation Process in ArcMap
(Source: ESRI's ArcMap Help File)

In ArcMap, the "add route events" function was used to add each crash as a point along the routes on the map layer. The accuracy level of crash location is 0.1 mile, which is believed to provide sufficient accuracy for area-wide macroscopic crash analysis. The

shortest roadway segment length is 0.10 mile, which can guarantee at least one possible crash location for each segment. Having segments longer than 0.10 mile is also consistent with the minimum length commonly used (Resende and Benekohal 1997).

Approximately 98% of crashes were successfully and correctly located on the maps for each of the five years in the study period. For example, 9,305 out of 9,460 (98.36%) crashes were located in the year 2000 map, as shown in Figure 3.6. In the figure, each point represents at least one crash (it is possible that several crashes occurred at the same location). The remaining 1.64% could not be added due to various reasons, such as unknown control-section number, unknown milepoint or error in the milepoint information, etc. It is assumed that less than 2% unallocated crashes do not significantly affect the analysis results.

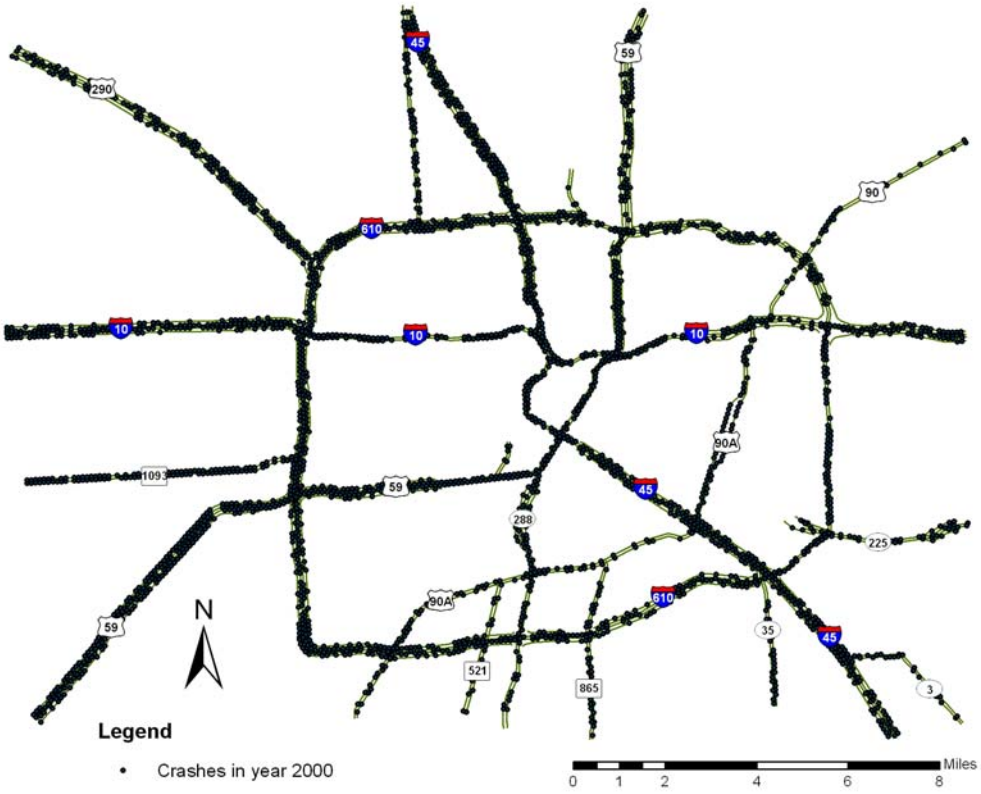


Figure 3.6 Identified Locations of Traffic Crashes, 2000

3.4 Crash Prediction Model

This thesis is to analyze the spatial-temporal patterns of the relative crash risks. Relative crash risk in this thesis is defined as the ratio between observed crash frequency and expected crash frequency (Equation 3.5).

$$Risk_{ijt} = \frac{Y_{ijt}}{E_{ijt}} \tag{3.5}$$

where, $Risk_{ijt}$ = relative crash risk for roadway segment *i*, road type *j*, and at time *t*;

Y_{ijt} = observed crash count for roadway segment *i*, road type *j*, and at time *t*; and

E_{ijt} = expected crash count for roadway segment *i*, road type *j*, and at time *t*.

In the above equation, Y_{ijt} was counted from the GIS crash point event layer by Zonal Spatial Analyst in ESRI's ArcToolbox (ESRI, 2004). Crash prediction models were built to estimate expected crash count E_{ijt} for different segment, different road type, and different time period. When the value of relative risk is greater than 1.0, the segment is riskier than expected. When it is less than 1.0, the segment is safer than expected.

The crash prediction model is a mathematical equation that estimates the average crash frequency of an entity as a function of traffic flow and other road characteristics. Although a crash is a function of various characteristics of the driver, traffic, vehicle, road and the environment, the expected crash count for segment *i* and road type *j*, is a dependant variable based on only segment length and traffic flow in the model (Equation 3.6). Segment length is assumed to have a linear relationship with expected crash count. The relationship between AADT and crash frequency was assumed to be unknown and would be found out according to the data and estimation method.

$$E_{ij} = 5 * \beta_0 * L_i * (DAADT_{ij})^{\beta_1} \tag{3.6}$$

where, E_{ij} = expected crash count for roadway segment *i*, road type *j* per 5 years;

$DAADT_{ij}$ = average of 5 years' directional AADT from year 1996 to 2000, for roadway segment *i*, and road type *j*, assumed to be a half of two-direction AADT;

L_i = segment length (miles); and

β_0, β_1 = coefficients to be estimated.

It is common to use maximum likelihood methods to estimate the coefficients in the crash model. The traditional least squares or weighted least squares regression methods are not used since the crash counts are discrete, non-negative values, and the variance of the number of crashes increases as the traffic flow increases, violating the assumptions of these two models (Lord and Persaud, 2000). Software SAS was used to estimate intercept β_0 and variable parameter β_1 for each road type by using maximum likelihood methods in the thesis.

Hourly crash frequency can also be estimated to calculate the relative crash risks for each hour of day. The expected hourly crash count is estimated by using the following equation.

$$E_{ijh} = DHVMT_{ijh} * \frac{E_{ijd}}{ADVMT_{ijd}/2} \tag{3.7}$$

where, E_{ijh} = expected crash count for roadway segment *i*, road type *j* and hour *h*;

E_{ijd} = expected crash count for roadway segment *i*, road type *j* and day *d*;

$DHVMT_{ijh}$ = directional hourly VMT for roadway segment *i*, road type *j* and hour *h*; and

$ADVMT_{ijd}$ = average daily VMT for segment *i*, road type *j* and day *d*.

Note: the directional average daily VMT is assumed to be 50% of the two-way ADVMT. E_{ijd} is estimated from Equation 3.6.

The best way to estimate hourly crash frequency is to build hourly crash models for each of the 5 road types and each hour of the 24 hours of a day. However, hourly

model is not the focus of this thesis and doing 120 hourly models will consume a lot of time. Both the advantages and limitations of this approach will be discussed in Section 4.1.

The relative risk, which is the ratio of observed crash count to expected crash count, can be transformed into the ratio of observed hourly crash rate to average daily crash rate. Equation 3.8 shows the transformation.

$$Risk_{ijh} = \frac{Y_{ijh}}{E_{ijh}} = \left(\frac{Y_{ijh}}{DHVMT_{ijh}} \right) \Big/ \left(\frac{E_{ijd}}{ADVMT_{ijd}/2} \right) \tag{3.8}$$

where, $Risk_{ijh}$ = relative crash risk for freeway segment *i*, road type *j*, and at hour *h*; and

Y_{ijh} = observed crash count for freeway segment *i*, road type *j*, and at time *h*.

3.5 Hierarchical Bayesian Modeling

Bayesian methods infer individual-level parameter estimates by borrowing information from other individuals (Bolstad, 2004; Lee, 2004). Hierarchical Bayesian modeling uses multiple levels of analysis in an iterative way (Carlin and Louis, 1996; Rossi et al., 2006). Unlike conventional statistical inference which derives the average estimates of parameters, hierarchical Bayesian modeling produces parameter estimates for each individual analysis unit. It also identifies and flags "extra variance" (Congdon, 2001; Winkler, 2003). In spatial statistics, if there is high uncertainty in a regression model, the result explains only a small amount of variance. But in a hierarchical

Bayesian model the unexplained "extra variance" is usually identified as either spatially-correlated effects or heterogeneity effects (Best et al., 1999).

Hierarchical Bayesian modeling involves two stages. At the first stage, a likelihood model for the vector of observed crash counts given the vector of relative risks of crashes was specified. At the second stage, a prior model over the space of possible relative risks was specified. Using software packages such as WinBUGS/GeoBUGS or sophisticated computation algorithms could yield a set of posterior means for the relative risks given the observed crash counts. The set of posterior means of the relative risks was then used to create a map to visualize the high- or low-risk segments. Observed risk maps were developed from the likelihood model (the first stage) only, and often feature large outlying relative risks in small areas (where the traffic flow and/or segment length is small). Hence, observed risk maps usually show high uncertainty due to the small sample sizes in the small areas. They also fail to catch similarity of relative risks in nearby or adjacent regions. An appropriately tailored Bayesian approach can incorporate spatial assumptions and help smooth the maps with large variances for those mapping units with small populations by borrowing strength from neighbors. The smooth change of the relative crash risks was assumed here along adjacent roadway segments based on the Tobler's First Law of Geography—"Everything is related to everything else, but near things are more related than distant things (Tobler, 1970)."

At the first hierarchy, the likelihood model (Miaou et al., 2002; Zhu and Carlin, 1999) assumes that the observed crash counts Y_{ijt} for road segment *i,* road type *j,* and in

time period *t* (hour, day, or year) are Poisson distributed with the mean m_{ijt}, which is the product of expected crash frequency for road segment *i,* road type *j,* and in time period *t* and the relative risk. The model can be represented as:

$$Y_{ijt} \mid m_{ijt} \sim Poisson(m_{ijt}),\ i = 1, \ldots, I, j = 1, \ldots J,\ t = 1, \ldots T, \tag{3.9}$$

$$m_{ijt} = E_{ijt} \exp(\mu_{ijt}),\ i = 1, \ldots, I, j = 1, \ldots J,\ t = 1, \ldots T, \tag{3.10}$$

where, Y_{ijt} = observed crash counts for roadway segment *i*, road type *j* and at time *t*;

μ_{ijt} = log relative risk for roadway segment *i*, road type *j* and at time *t*;

E_{ijt} = expected crash count for roadway segment *i*, road type *j* and at time *t*; and

$\exp(\mu_{ijt})$ = $Risk_{ijt}$, relative risk for roadway segment *i*, road type *j* and at time *t*.

The second hierarchy is log-relative risk, which is modeled as

$$\mu_{ijt} = \beta_{jt} + \alpha x + \theta_{ij}^{(t)} + \phi_{ij}^{(t)} \tag{3.11}$$

where, β_{jt} = an overall intercept for road type *j* and at time *t*;

x = the covariate of road type;

α = the corresponding effect of road type *x*;

$\theta_{ij}^{(t)}$ = non-structured random noises for segment *i*, road type *j* and at time *t*; and

$\phi_{ij}^{(t)}$ = spatially-correlated random effects for roadway segment *i*, road type *j* and at time *t*.

The overall intercept β_{jt} captures a main effect for road type *j* at time *t*. Although it is possible to specify a parametric function (e.g., linear or quadratic form) for the time

effect, a qualitative form that allows data to reveal the presence of any temporal trend is preferred.

The next step is to define four components in the second hierarchy. The distribution of non-structured random noise $\theta_{ij}^{(t)}$ is assumed to be normally distributed, while the spatially-correlated random effect $\phi_{ij}^{(t)}$ is assumed to follow a conditional autoregressive (CAR) model. Placing flat (uniform) priors on the main effects β_{jt} and road type effect α are advocated. They are modeled as

$$\theta_{ij}^{(t)} \sim Normal(0, \tau_t) \tag{3.12}$$

$$\phi_{ij}^{(t)} \sim CAR(\lambda_t) \tag{3.13}$$

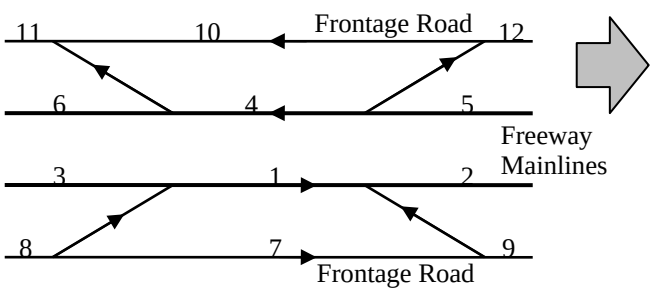
$$\beta_{jt} \sim falt(\,) \tag{3.14}$$

$$\alpha \sim falt(\,) \tag{3.15}$$

According to Besag (1974), the joint distribution of the vector of spatial effects $\boldsymbol{\varphi}_t$ (at time *t*) is proportional to $\exp(-(\lambda_t / 2)\boldsymbol{\varphi}_t^{\mathrm{T}}\mathbf{B}\boldsymbol{\varphi}_t)$, i.e., a multivariate normal density with mean $\mathbf{0}$ and covariance matrix $\mathbf{B}^{-1}$. The elements of matrix $\mathbf{B}$ is determined as $\mathbf{B}_{kk} = a_k$ and $\mathbf{B}_{kl} = -a_k\omega_{kl}$, with a_k denoting the number of neighbors of segment *k*, and ω_{kl} being the elements in the adjacency matrix $\mathbf{W}$. The $(k,l)^{\text{th}}$ element in matrix $\mathbf{W}$ equals to 1 if two segments *k* and *l* are adjacent to each other.

To compute the adjacency matrix, each roadway segment is considered to be adjacent to its upstream segment and downstream segment in the same direction. It is

assumed that the segment is also adjacent to the segment with opposite traveling direction. Figure 3.7 shows an example of adjacent segments on a typical freeway.



Segment	Num. of Adjacent Segments	Adjacent Segments
1	4	2, 3, 4 and 7
4	4	1, 5, 6 and 10
7	3	1, 8 and 9
10	3	4, 11 and 12
…	…	…

Figure 3.7 Adjacent Segments Definition on a Typical Freeway

The similarity among the random effects across time is encouraged by assuming τ_t and λ_t to be Gamma distributed based on the conjugate distribution theory.

$$\tau_t \sim Gamma(a,b) \tag{3.16}$$

$$\lambda_t \sim Gamma(c,d) \tag{3.17}$$

The parameters were set as $a = 0.001$, $b = 0.001$ (i.e., the τ_t have prior mean 1 and standard deviation $\sqrt{1000}$) and $c=0.01$, $d=0.01$ (i.e., the λ_t have prior mean 1, standard deviation 10). These are vague priors designed to allow the data to dominate the allocation of excess spatial variability to heterogeneity and clustering. Note that the constraints $\sum_{ij} \phi_{ijt} = 0$, at $t = 1,...,T$ must be added to identify the time effects β_{jt}, due to the location invariance of the CAR prior. The MCMC implementation in WinBUGS ran two parallel sampling chains for 20,000 iterations each, and discarded the first 5,000

iterations as pre-convergence burn-in based on the evaluation of convergence statistics, such as MCMC error.

3.6 Three-dimensional Mapping

This thesis attempts to display relative crash risk maps in a new way, the 3-D views. A 3-D map has not only all the power of a 2-D map, but also the expression from the third dimension, so it is believed that a 3-D map can show more information and readability than a 2-D map. For example, links can be displayed not only with different color and width, which is common in a 2-D view, but also with its third dimension—height in a 3-D view. In the relative risk maps in this thesis, every segment will be displayed with corresponding height, which represents its relative risk value. ESRI's software ArcScene is used to produce 3-D maps.

The overall methodological flow diagram in this thesis can be shown in details as Figure 3.8. In the preliminary data processing and analysis part, GIS is needed in segmentation process, locating of the crashes, counting the crashes, and integrating volume and road type data. In the hierarchical Bayesian modeling part, GIS is needed to generate adjacency matrix for the model. In the last part, GIS is definitely needed to draw the maps. Therefore, the approach developed in this thesis is a GIS-based approach.

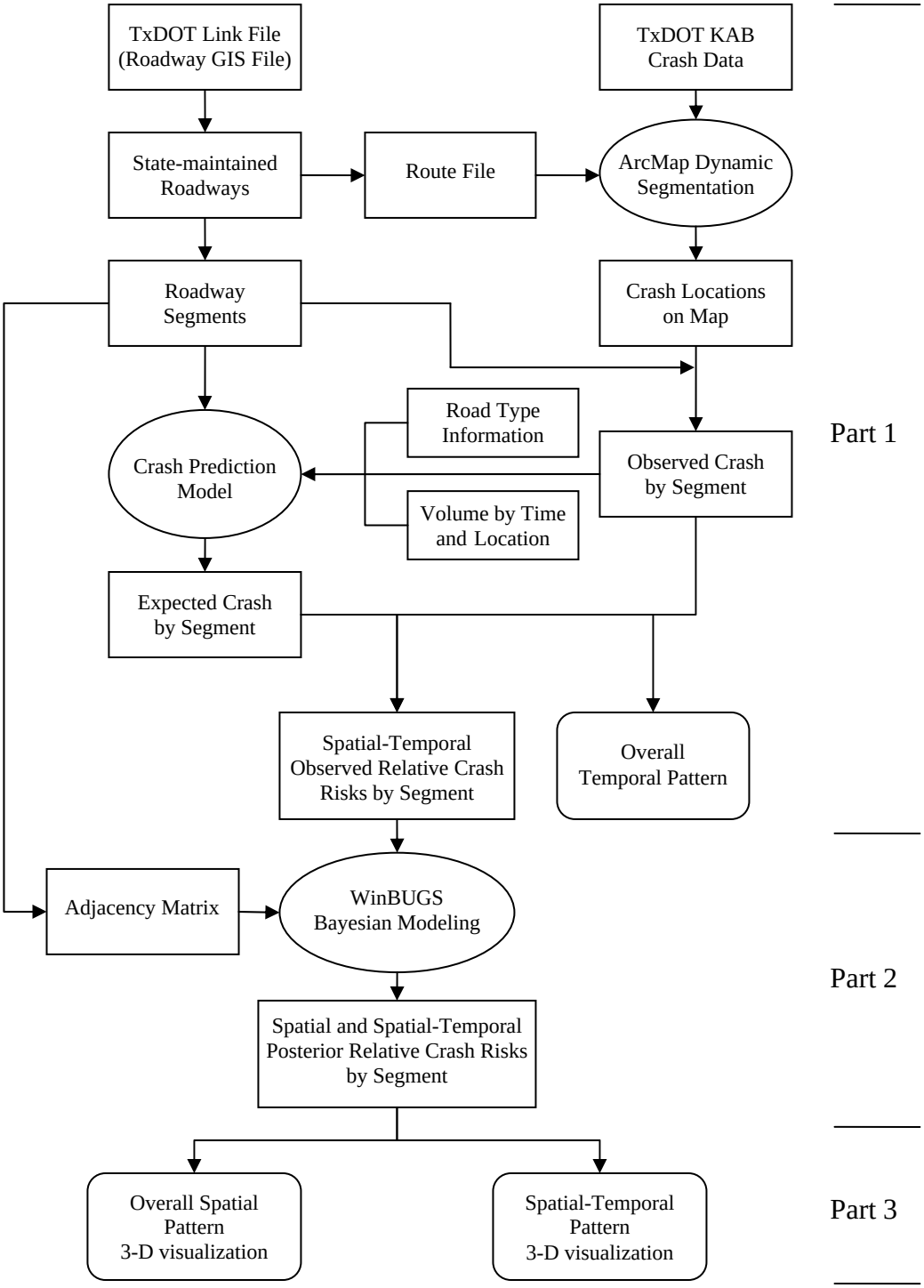


Figure 3.8 Flow Diagram of Relative Crash Risk Assessment

4 RESULTS AND DISCUSSIONS

This section begins with the results of the crash prediction models for five road types, and their implications. Overall temporal results of crash risks, calculated from the raw crash and traffic data are then presented. This is followed by the statistical spatial results, which describe the spatial distributions of posterior relative crash risks, calculated from WinBUGS, for the 5-year period. Spatial-temporal results are presented in the final section, targeting spatial distribution of posterior relative crash risks for every day of a week and every hour of a day.

4.1 Crash Model Results

The crash model was used to estimate the relationship between traffic flow and crash frequency. A simple model was built with the functional form of $E_{ij} = 5 * \beta_0 * L_i * (AADT_{ij})^{\beta_1}$ (Equation 3.6) in section 3.4. Segment length is assumed to have a linear relationship with crash frequency. Coefficient β_1 captures the flow's influence on the crash count and β_0 represents the overall intercept for a particular road type. This model was applied to each of the five road types respectively, i.e., urban freeway mainlines (471 segments), freeway system interchange area (112 segments), frontage roads without diamond interchange (171 segments), frontage roads with diamond interchange (228 segments), and arterials (126 segments). The maximum likelihood method was used to estimate the coefficients in the crash models. The results

provided by SAS are shown in Table 4.1. Figure 4.1 illustrates the relationship between directional AADT and crash frequency for these five road types.

Table 4.1 Estimated Coefficients for Different Road Types

	β_0	β_1	Dispersion
Urban Freeway Mainlines	3.59×10^{-6} (5.40×10^{-7}, 2.40×10^{-5})	1.3777 (1.2102, 1.5452)	0.2375 (0.2045, 0.2705)
Freeway System Interchange Area	0.00059 (2.76×10^{-5}, 0.01279)	0.9585 (0.6864, 1.2307)	0.4468 (0.3303, 0.5633)
Frontage Roads without Interchange	0.15423 (0.03508, 0.67801)	0.4447 (0.2804, 0.6089)	1.0055 (0.7882, 1.2228)
Frontage Roads with Diamond Interchange	2.21732 (0.79303, 6.19969)	0.2339 (0.1183, 0.3496)	0.7041 (0.5751, 0.8331)
Arterials	0.00103 (0.00015, 0.00720)	0.9870 (0.7818, 1.1923)	0.3780 (0.2738, 0.4821)

Note: 95 % confidence intervals are in parenthesis.

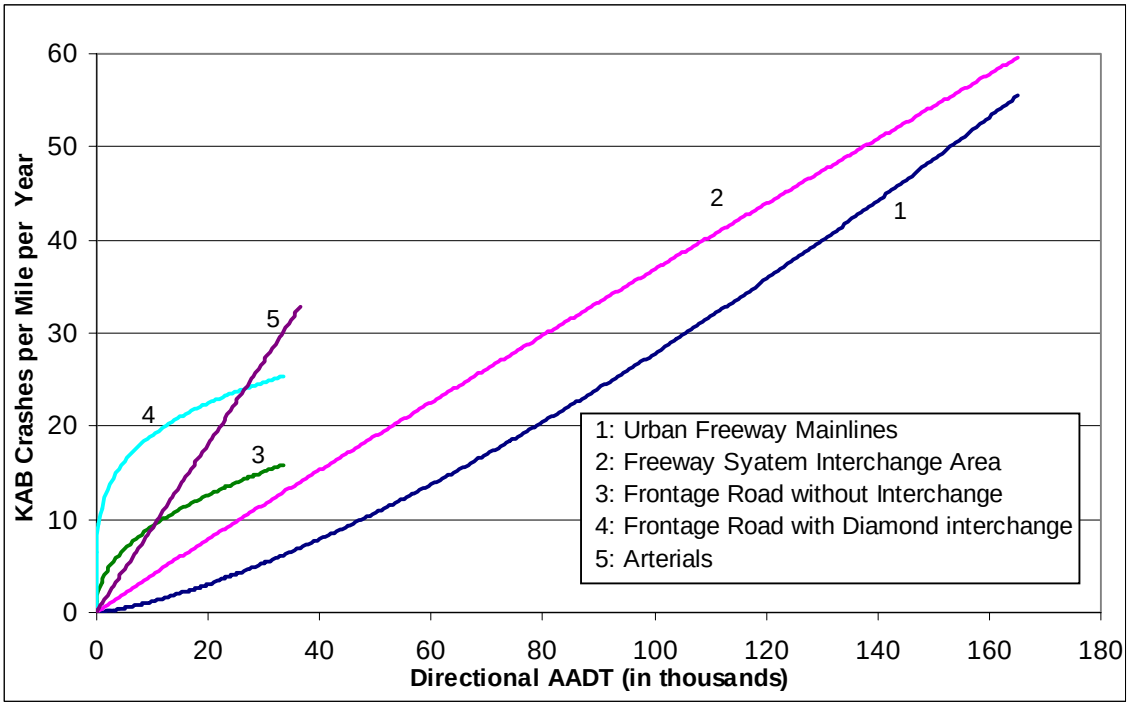


Figure 4.1 Results of KAB Crash Prediction Models by Different Road Types

It can be seen from Table 4.1 and Figure 4.1, the crash count on freeway system interchange area and arterials have linear relationships with traffic flow, because their 95% respective confidence interval for β_1 cover 1.0. This means that with the increase of traffic flow, the crash count will increase with a constant rate. For urban freeway mainlines, β_1 is bigger than 1, implying that crash frequency increases with a higher rate than the increase of AADT, which matches the research result by Persaud and Dzbik (1993). However, the crash rate for urban freeway mainlines is always lower than any other road types before it reaches maximum AADT, so urban freeway mainlines are always the safest road type among the five types.

Frontage roads shows different relationships, because their β_1 are both less than 1, implying the crash counts increase with decreasing rates when their AADTs go up. The possible reasons can be that the speeds of vehicles are high at low volumes while low at high volumes. Frontage roads with diamond interchange have the highest crash rate before its fitting line intersects with arterials at 27,000 directional AADT. Frontage roads without diamond interchange experience less crashes per unit of exposure than frontage roads with diamond interchange. In comparison with arterials, frontage roads without diamond interchange have a higher crash rate at low AADT (<10,000) and lower crash rate at moderate and high AADT (>10,000) than arterials.

Overall, a freeway (including urban freeway mainlines and freeway system interchange areas) is safer than a frontage road and arterial for all ranges of AADT, and freeway mainlines are always safer than freeway system interchange areas. This is expected since freeways have better road condition, wider lane width, and fewer traffic

conflicting points than frontage roads and arterials. The safety of the other three road types depends on the AADT. Frontage roads are relatively less safe at low AADT and safer at high AADT than arterials. Section 4.3 will show the calculation results of crash rates for these five road types for this study area.

There are a few ways to calculate the hourly expected crash frequency. Using one general hourly model will produce bias when more safety performance functions should have been used separately for daytime, nighttime, morning peak and afternoon peak etc. (Mensah and Hauer, 1998). To avoid this “function averaging” problem, the ideal way is to build hourly models for each hour of day and each road type separately. However, doing that will cost a lot of time and cause another problem, which is the ignorance of temporal correlation between the adjacent hours. These issues are all important and worth more research in the future. Since it is not the focus of this thesis research, the hourly crash frequency was estimated by Equation 3.7 in order to simplify this process.

In this hourly crash prediction model, daily average crash rate for each segment is used to multiply hourly VMT to get expected hourly crash frequency. The underlying assumption for this method is the linear relationship between flow and crash frequency, which is supported by Martin (2002).

These daily expected crash counts and hourly expected counts are denominators in the calculation of relative crash risks. The relative risks are presented in later spatial analysis section and spatial-temporal analysis section.

4.2 Temporal Analysis

The overall temporal analysis is based on the observed data, so it does not necessarily need GIS or Bayesian model, but it is useful for understanding the problem, and the result is a necessary precursor for the later spatial-temporal analysis, which targets the spatial patterns in different time periods. According to the results, the monthly numbers of KAB crashes do not display any significant differences in this study area, so the temporal analysis only focuses on temporal change of relative crash risks by day of week and hour of day.

4.2.1 Day of Week Analysis

Average KAB crash counts in the study area varies by day of week. Figure 4.2 illustrates that in terms of crash count during 1996 to 2000, Sundays were the lowest, whereas Fridays were the highest and Saturdays were the second highest days. Mondays through Thursdays had very close values, which were between those of Sundays and Saturdays.

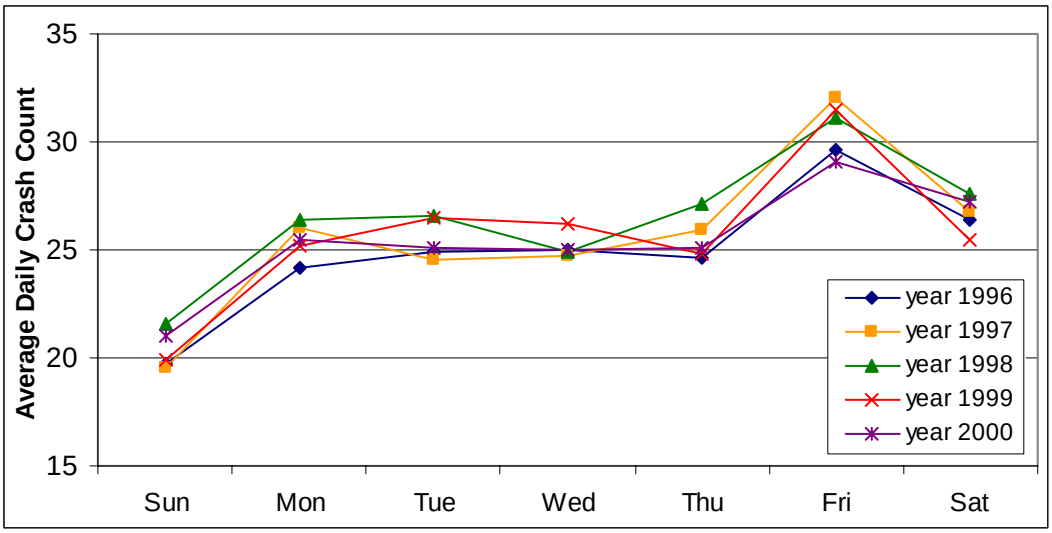


Figure 4.2 Average Daily KAB Crash Count in the Study Area, 1996-2000

VMT showed a different pattern from the crash count, shown in Figure 4.3. It is assumed that different years have same VMT daily patterns, which are introduced in Table 3.1, obtained from TTI. These average daily VMT were calculated by multiplying the daily adjustment factors with AADVMT, which varies by years. In terms of VMT, Sundays were the lowest and Fridays were the highest, whereas Mondays through Thursdays had similar values, which were the second highest. Since weekday traffic is likely work-related, drivers had more consistent driving habits. Therefore, the VMT is relatively stable from Monday through Thursday. On Friday not only may more people drive to work, but they may also like to travel for leisure activities after work. VMTs on Saturdays were lower than weekdays and Fridays, but higher than Sundays probably because there were some work-related traffic, shopping, and party traffic on Saturdays.

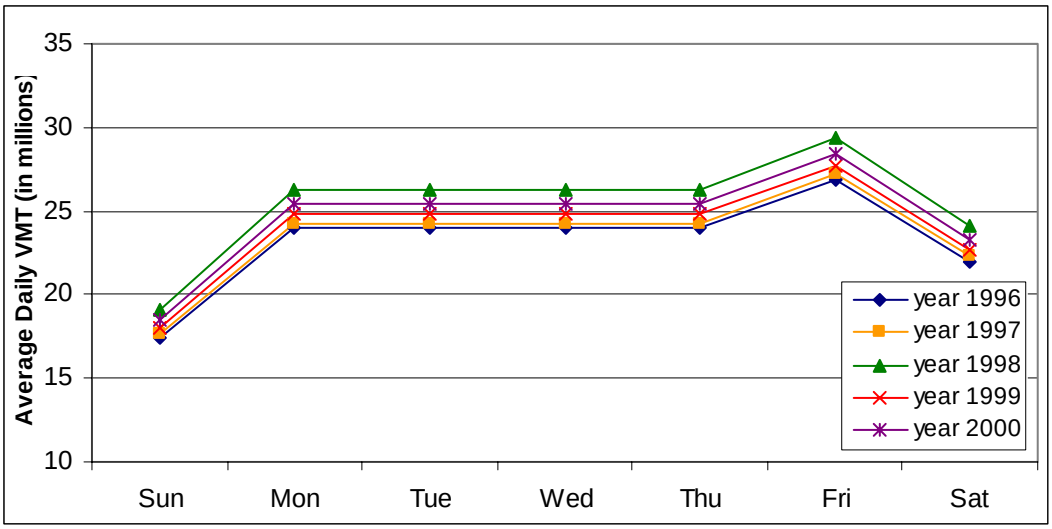


Figure 4.3 Average Annual Daily VMT by Day of Week, 1996-2000

The plot of relative crash risk by day of week (Figure 4.4) shows that although different years had different risks, the risks in each year followed a similar "U" shape pattern. Saturdays had the highest overall relative crash risk in a week since Saturday has the moderate crash counts and relatively small expected crash frequency based on its relatively low traffic volumes. The relative risk for Sundays was the second highest. In each year, Mondays through Thursdays had similar risk values, which were below 1, implying that they were safer than the average. The relative risk on Fridays was in the middle, safer than Saturday and Sunday, less safe than weekdays. The reasons for this "U" shape might be that drivers were more familiar with roads and had more consistent driving habits during weekdays, while on weekends there were more leisure trips, higher speed, and high percentages of intoxicated drivers.

Relative crash risks also fluctuated by year. The relative risks on Friday and Saturday in 1997 were higher than those in other years. Year 2000 had a relatively low

risk during the working days. However, generally, the weekly patterns of relative risks through these years were very similar.

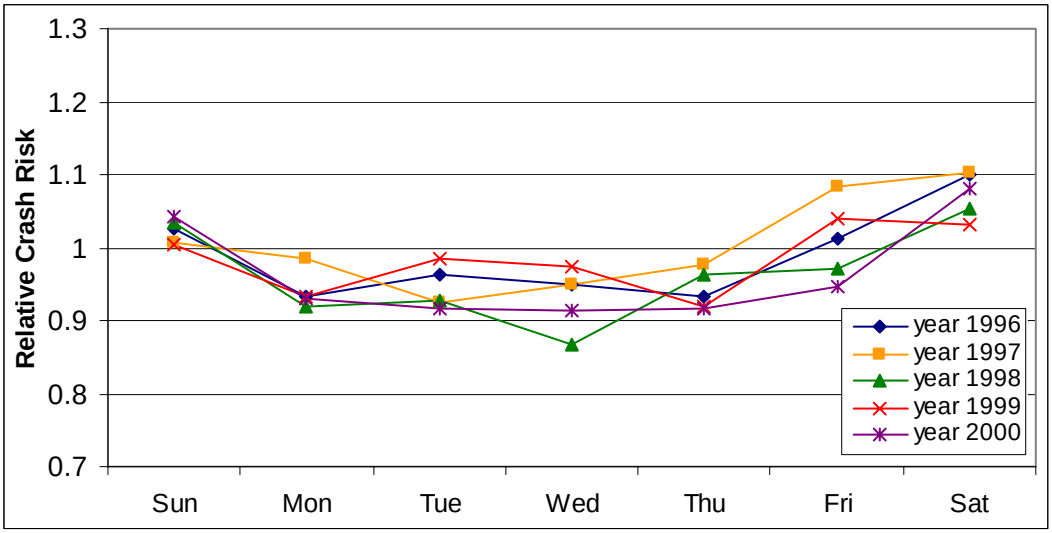


Figure 4.4 Observed Annual Relative Crash Risks by Day of Week, 1996-2000

4.2.2 Hour of Day Analysis

The hour of day analysis aims at assessing the patterns of relative crash risks by each hour of a day. This analysis was performed in four categories: weekdays (Monday through Thursday), Friday, Saturday, and Sunday, since each category shows a similar hourly pattern both on crash count and VMT.

The hourly patterns of crash count were different by day of week, as shown in Figure 4.5. There are two crash count peaks (7:00 a.m.—9:00 a.m. and 3:00 p.m.—7:00 p.m.) on Weekdays and Friday. However, the two peaks in weekend were at different time, which were 1:00 a.m.—3:00 a.m. and 2:00 p.m.—5:00 p.m. There is no morning peak (7:00 a.m.—9:00 a.m.) of crash counts in weekend and no early morning peak

(1:00 a.m.—3:00 a.m.) on weekdays or Friday. The time of afternoon peak in weekend was earlier than that on weekdays and Friday, and the peak crash count was less. From Figure 4.5, crash counts during weekdays were higher than that on Friday in the morning peak (6:00 a.m.—10:00 a.m.), but lower than that on Friday during the rest of a day. Crash counts on Saturday was higher than that on Sunday all the time except the early morning peak (1:00 a.m.—5:00 a.m.).

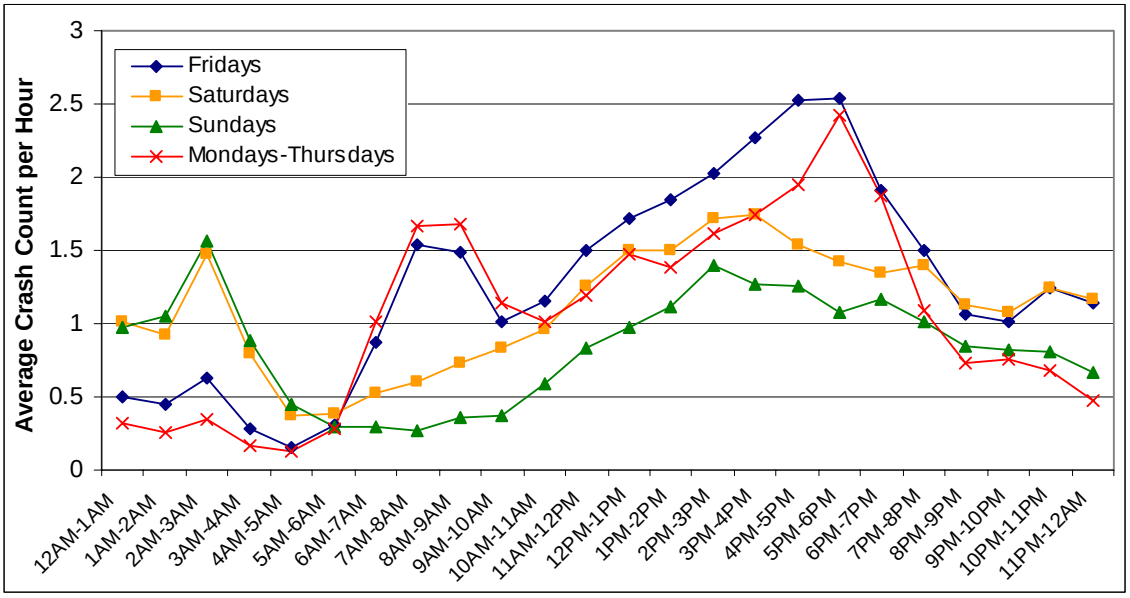


Figure 4.5 Average Hourly KAB Crash Count by Hour of Day, 1996-2000

VMT had a close hourly pattern to crash count for those four categories, as shown in Figure 4.6. VMT during weekdays and Friday had two peaks (7:00 a.m.—9:00 a.m. and 3:00 p.m.—7:00 p.m.), which were at the same time as crash count peaks. However, the hourly VMT during weekends had only one flat peak (1:00 p.m.—7:00 p.m.), which was in the afternoon. There was no early morning peak in VMT for

weekend. VMT on weekdays was always lower than that on Friday and VMT on Saturday was bigger than that on Sunday except early morning (1:00 a.m.—4:00 a.m.).

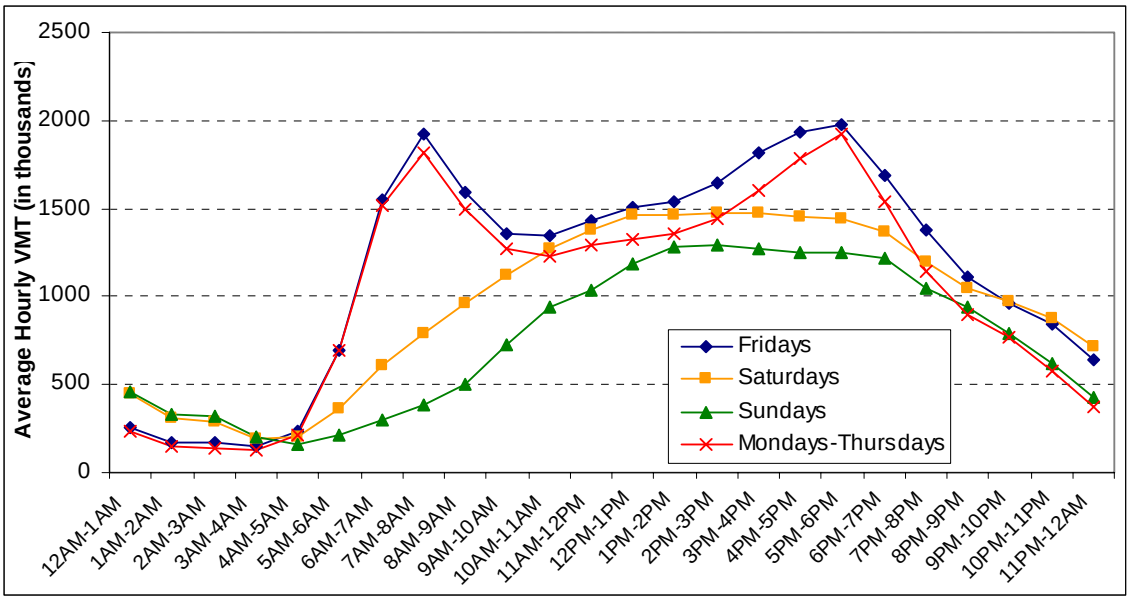


Figure 4.6 Average Hourly VMT by Hour of Day, 1996-2000

The plot of the relative crash risk by hour of day (Figure 4.7) shows that different categories have very similar trends although they have different distributions for crash count and VMT. Generally, the risk at nighttime (10 p.m.-5:00 a.m.) was higher than daytime. The risk was very high in the early morning, especially between 1:00 a.m. and 4:00 a.m. The highest risk of around 4.0 existed between 2:00 and 3:00 a.m. on Sundays and Saturdays. The risks in this same period on weekdays and Fridays were also much higher than 1.0. Therefore, the early morning was the least safe time in a day, with weekend being less safe than weekdays and Friday. It was also noticed that the relative risks were higher than 1.0 from 11:00 p.m. to 4:00 a.m. for all categories. The

contributing factors probably include the dark environment without good lighting conditions, driver fatigue, and drunk drivers. The plot also shows that during weekdays and Friday, morning (risk<1) is safer than the afternoon (risk>1) although morning peak and afternoon peak had almost the same VMT. This is perhaps because drivers are more attentive after a good night sleep. The lowest risk for Mondays through Fridays existed at 5:00-6:00 a.m., but the lowest-risk period lagged several hours on Sundays and Saturdays. It can also been seen that Friday always had higher relative risk than weekdays except 6:00 a.m. to 10:00 a.m. and Saturday always had higher risk than Sunday except from 4:00 a.m. to 7:00 a.m.

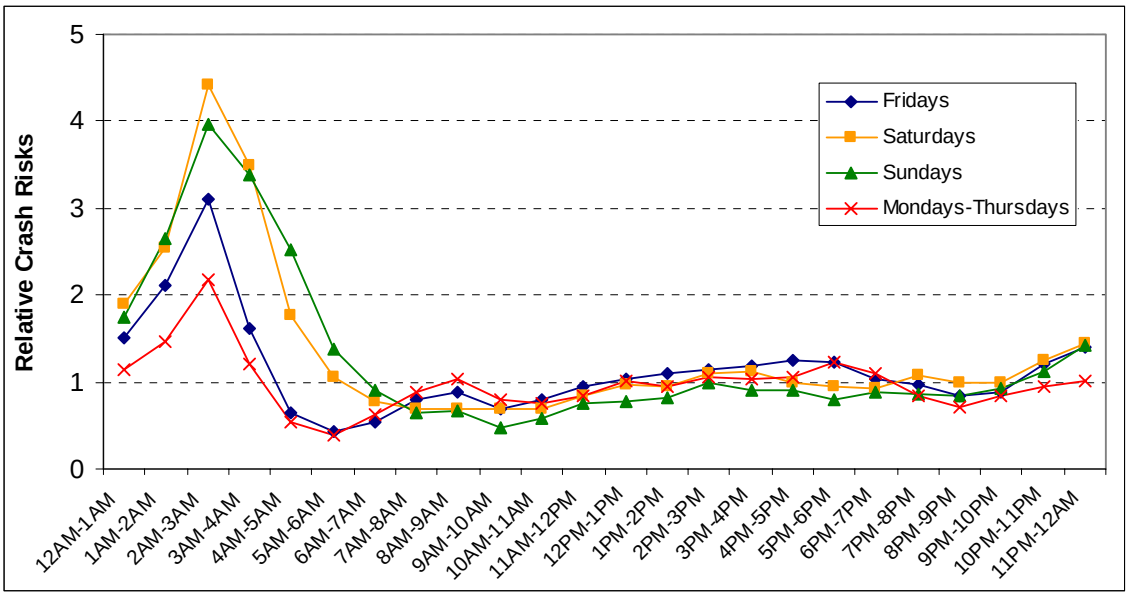


Figure 4.7 Observed Average KAB Relative Crash Risks by Hour of Day, 1996-2000

4.3 Spatial Analysis

Spatial analysis is used to determine the spatial distribution of relative crash risks for the study area from year 1996 to year 2000. GIS and the Bayesian approach are two necessary tools to identify risky segments. In this section, the crash rates for different road types are calculated and compared. Then, overall spatial patterns are presented with both observed risk map, which used the raw data directly and a posterior relative risk map, which used the posterior means of the estimations from WinBUGS. The comparison of these two maps is used to show the strength and advantage of Hierarchical Bayesian approach.

4.3.1 Crash Rate Analysis by Road Types

This subsection evaluates the safety for all the five road types in the study area. Crash rates were calculated by dividing the observed crash count by VMT.

$$Crash\ \ Rate_{jt} = \frac{Y_{jt}}{VMT_{jt}} \tag{4.1}$$

where, $Crash\ \ Rate_{jt}$ = crash rate for road type j, and at time t (number of crashes per million VMT);

Y_{jt} = observed crash count for road type j, and at time t; and

VMT_{jt} = VMT for road type j, and at time t, in millions.

Figure 4.8 shows the overall KAB crash rates for different road types as well as day of week in the study area. Urban freeway mainlines are the safest road type with

overall crash rate 0.71 crashes per million VMT, followed by freeway system interchange area, whose crash rate was 1.02 crashes per million VMT. This could be due to the fact that freeways have better design standards, wide lanes, and with limited access. The reason of lower crash rate for urban freeway mainline than system interchange area might be less merging, diverging, or weaving maneuvers and less speed variance. Frontage roads and arterials had much higher crash rates than those of freeways, especially where diamond interchange exists. Frontage roads with diamond interchange (crash rate equaled to 4.59 crashes per million VMT), which was the least safe road type under evaluation. This might be because of more traffic conflicting points, turning movements, frequent lane changing, and pedestrians. The crash rate for frontage roads without diamond interchange was 2.19 crashes per million VMT, about 1/2 of the rate of the ones with diamond interchanges. This is obvious because speed changes and conflicting movements at diamond interchanges always accompany more crashes. Arterials had crash rate of 2.49 crashes per million VMT, a slight higher than frontage road without interchange but much lower than frontage road with diamond interchange, because arterial crashes combined the intersection-related crashes and non-intersection-related crashes. In terms of crash rate by day of week, it is seen from Figure 4.8 that generally, Saturday and Sunday had higher risks than weekdays and Friday, and Friday was riskier than weekdays.

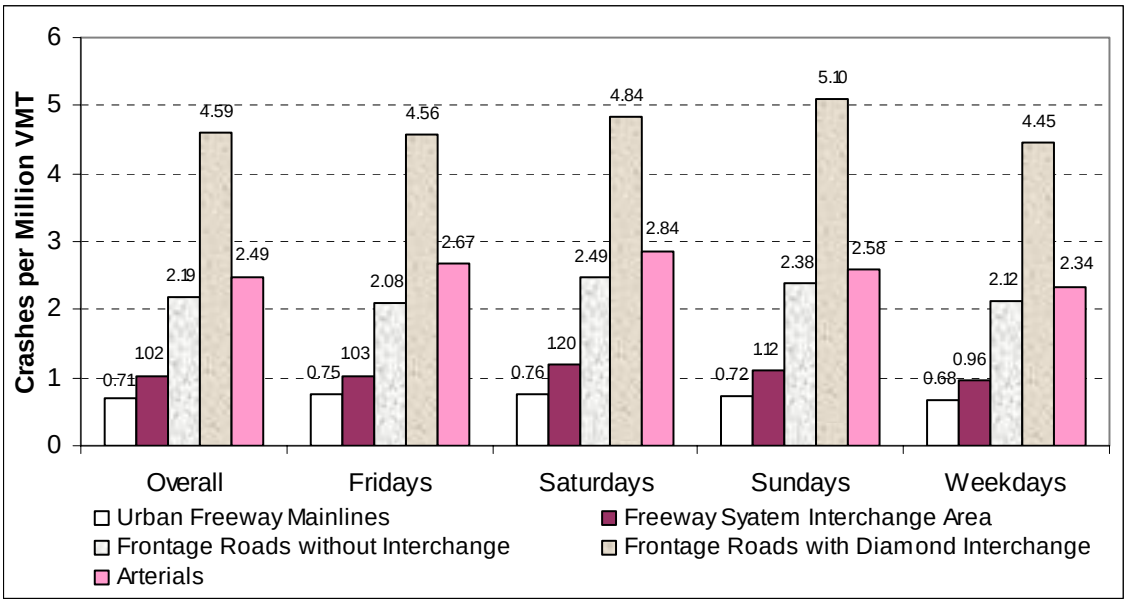


Figure 4.8 KAB Crash Rates by Road Types and Day of Week, 1996-2000

4.3.2 Overall Spatial Pattern

Figure 4.9 and 4.10 show the distribution of overall observed and posterior relative crash risks in Houston respectively during the years 1996 through 2000. The posterior relative risk map (Figure 4.10) shows some characteristic Bayesian smoothing of the observed relative risks (Figure 4.9). After careful comparison, no segment was assigned a risk of exactly zero, and the high risks on the segments with low traffic volumes or short lengths were reduced so that the risk surface became smoother. For example, the minimum and maximum risk values were 0 and 8.2806 respectively in the observed risk map, but they became 0.092 and 7.5768 in the posterior risk map; the variance of the observed risks was 0.674, while the variance of posterior relative risks was 0.573, 15% less. This is an expected result of the Bayes method as it reduces variances. However, the observed high risks in the frontage road of US-59 south

remained to be high, as the method properly recognized the much higher sample sizes in this area. Actually, it seems that the overall observed risk map is very similar to the overall posterior risk map. This is because the 5-year dataset has a lot of data points and has small variance, which the Bayesian approach is supposed to filter. One of the advantages of Bayesian approach is that it can automatically detect the uncertainty in the dataset and then filter it. When dealing with one-year, monthly, daily or hourly crash data with high variance level, the smoothing effects will be more obvious because the approach will filter much uncertainty that exists in the dataset. In the next section, the smoothing effects of this hierarchical Bayesian approach on a daily relative crash map and hourly relative crash risk map are shown and discussed.

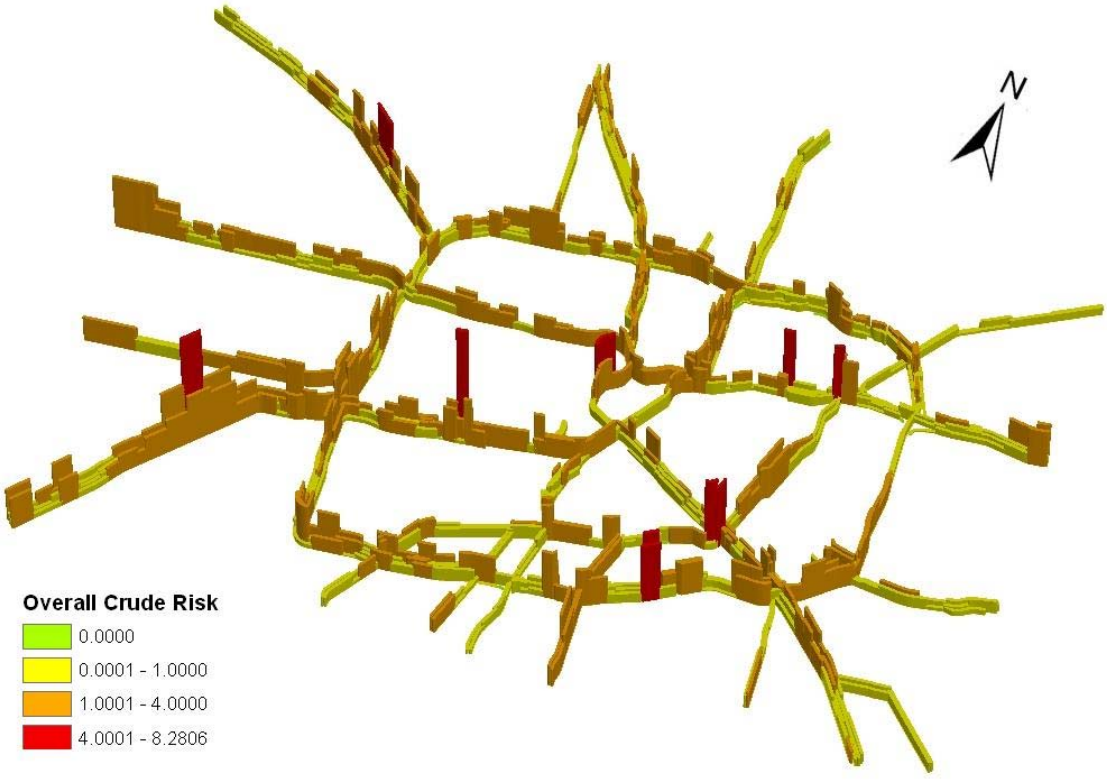


Figure 4.9 Overall Observed Relative Crash Risk, 1996-2000

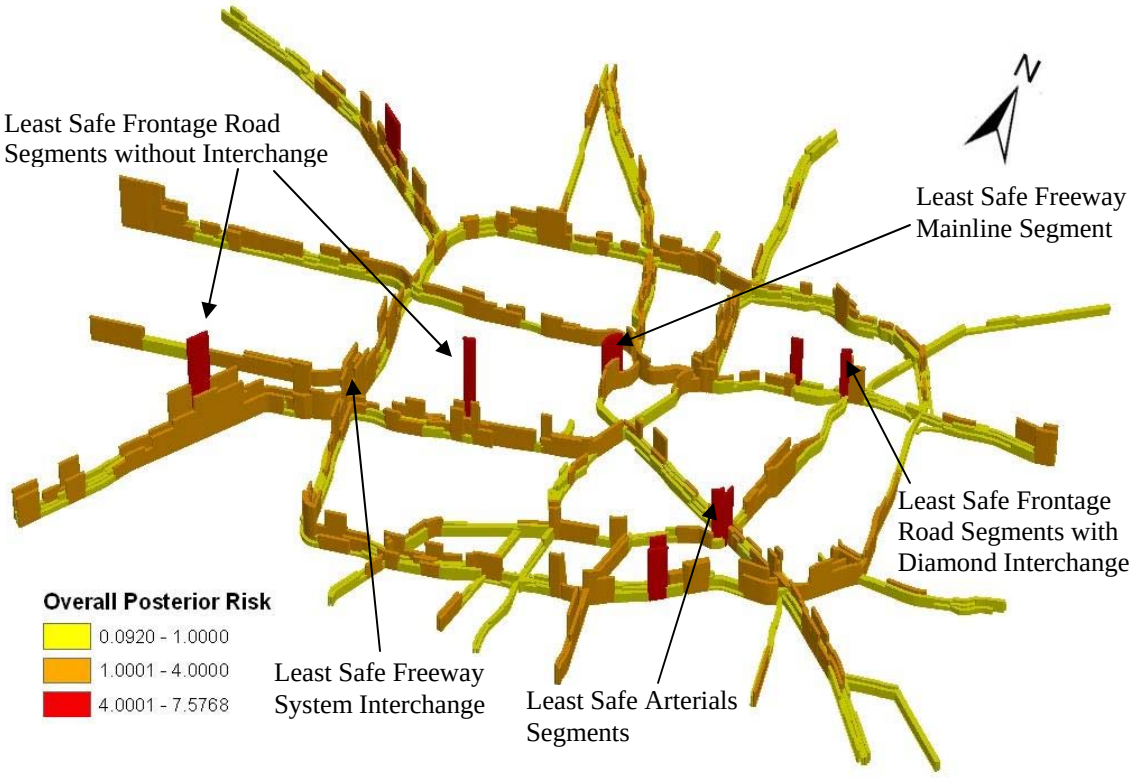


Figure 4.10 Overall Posterior Relative Crash Risk, 1996-2000 (3-D)

The Posterior relative risk map (Figure 4.10) shows there were several risky roadway segments in this area after uncertainty was filtered. The least safe urban freeway mainline segment is located on I-45 downtown. The least safe freeway system interchange area is the interchange of US-59 and I-610 west. For frontage roads without diamond interchange, US-59 westbound has one less safe segment inside I-610 loop and one outside the loop, both of which had relative risks over 7. For frontage roads with diamond interchange, the frontage road of I-10 eastbound at the diamond interchange with US-90 was least safe. The least safe arterial segment was a segment of US-90, at the interchange area with I-45, where a big curve is located. In addition, two frontage road segments on I-610 south, and one frontage road segment on I-10 east also displayed

abnormally high relative risk values. For those roadway segments with high posterior relative risks, there must be something unusual that caused high risks, so the next step for transportation agencies is to do diagnostic analysis to identify contributing factors.

The posterior relative risk map also shows that different directions of some roadways had different risk levels. For example, the least safe urban freeway mainline segment –southbound segments of I-45 in downtown has a posterior relative risk value of 4.03, much bigger than that of its opposite direction segment, which had a posterior relative risk value 2.30. I-10 west inside I-610 loop had an eastbound segment with relative risk less than 1, but westbound segment had relative risk over 1. These suggest the need of direction differentiation in analyzing relative crash risks. Otherwise, the relative risks on two directions will be averaged out so that no direction difference will be observed, leading to mistakes in decision making.

Figure 4.11 shows a 2-D posterior risk map for this 5-year period. After comparison between Figure 4.10 and Figure 4.11, it is obvious that a 3-D map carries more information and is more reader friendly. A 3-D map uses not only link's color and width to express information, but also link's height in the third dimension, which represents the value of the relative risk. In a 2-D map, different color and size shows different categories, but readers can not know the difference of the risk values within one category. However, in a 3-D map, it is obvious that the least safe frontage segment had a higher risk value than the least safe freeway segment from the different heights of the segments. Therefore, 3-D maps help the presentation of relative risks and are advocated.

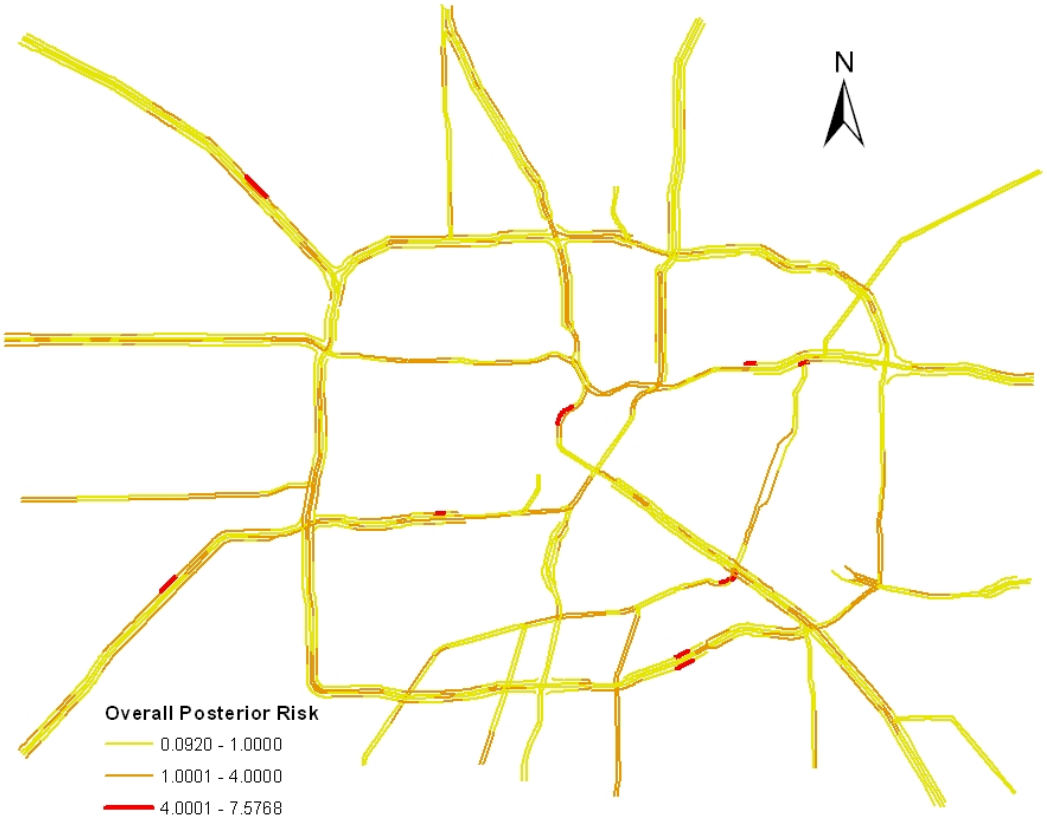


Figure 4.11 Overall Posterior Relative Crash Risk, 1996-2000 (2-D)

4.4 Spatial-Temporal Analysis

The objective of the spatial-temporal analysis is to show the variation of spatial distribution of relative crash risks by day of week and by hour of day. The spatial analysis by day of week was conducted on different days such as Sunday, weekdays (Monday to Thursday), Friday and Saturday for different years (1996 to 2000). The spatial analysis by hour of day was done on each of the 24 hours for different day of a week (Sunday, weekdays, Friday and Saturday). 3-D maps were drawn to illustrate the

spatial distribution of posterior relative crash risks, which were the smoothed results from Hierarchical Bayesian approach.

4.4.1 Spatial Analysis by Day of Week

Figure 4.12 and 4.13 show the distribution of observed and posterior relative crash risks respectively for Fridays in year 1996. Because of the limited data volume, the observed risk map displayed a high variance level, which led to some no-risk segments (with relative risk value 0) and a lot of abrupt changes of relative risks between adjacent segments. Hierarchical Bayesian approach filtered data uncertainty so that real safety tendency is revealed in Figure 4.13. It is very clear that no segment had a risk of exactly 0, high observed risks on those segments with low traffic volume or short length were reduced and the surface of the map became smoother with less abrupt risk changes. After the uncertainty was filtered, the remaining high-risk segments were the segments that were truly less safe and proper safety improvements were needed on those segments. For example, the segment with the highest observed relative risk was the US-90 arterial segment at the interchange area with I-45. Since it is 0.25 miles long (shorter than average) and had AADT 4400 (very low), the Bayes approach detected this small area, and smoothed its observed risks 11.9391 into posterior risk value 4.25 by incorporating its adjacent segments (including upstream, downstream and opposite direction segments), two of which had zero observed risks.

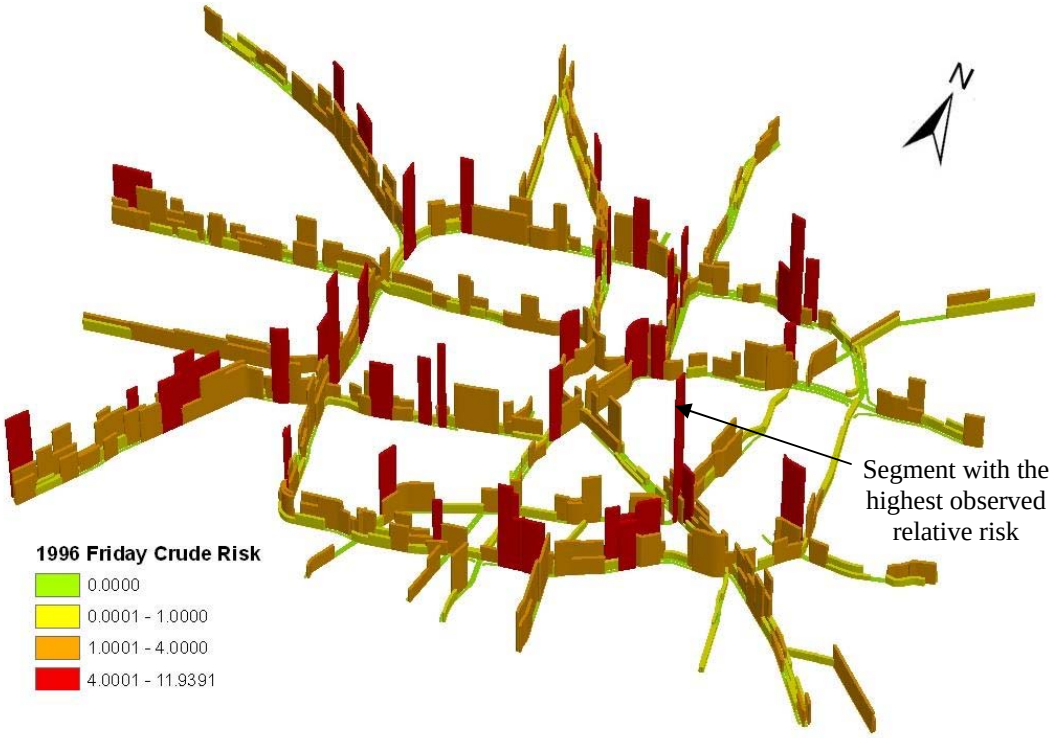


Figure 4.12 Observed Relative Crash Risk for Fridays in Year 1996

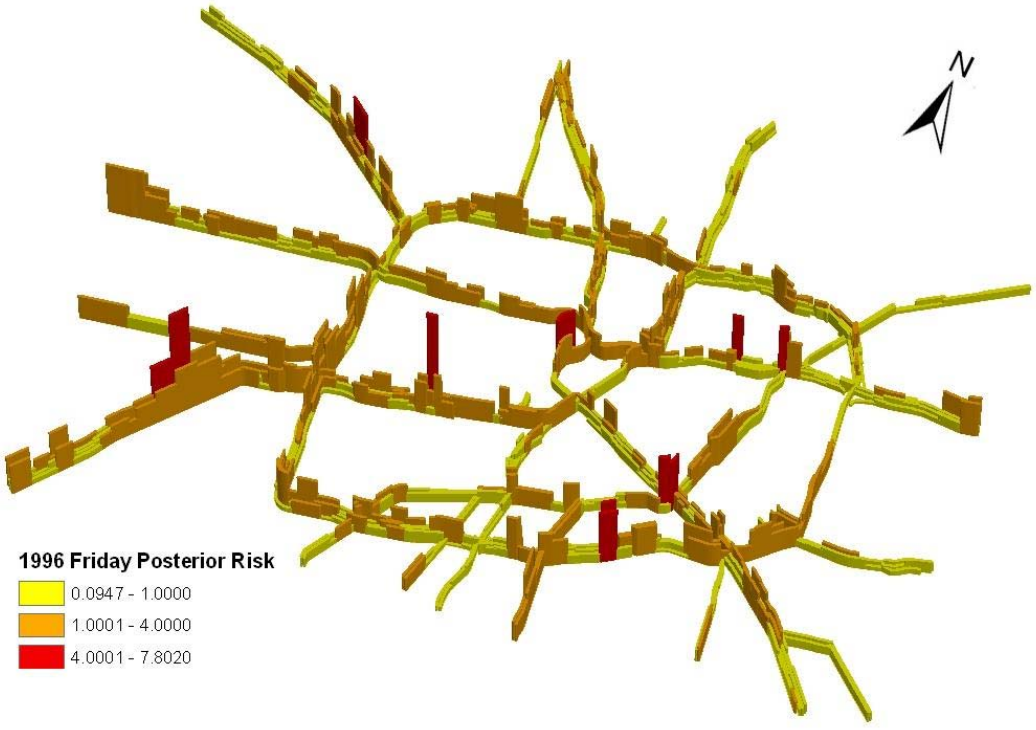


Figure 4.13 Posterior Relative Crash Risk for Fridays of Year 1996

Figure 4.14, 4.15 and 4.16 illustrate the distribution of posterior relative crash risks for Saturdays in year 1997, Sundays in year 1998 and weekdays in year 2000. Actually the same day in a week in different years showed very similar spatial risk patterns. However, different days of a week showed a little different pattern. The relative risks on Saturday were higher than those on other days in a week, and more high-risk segments were observed. For example, I-10 west at Sam Houston Tollway, had a frontage road segment with risk over 4.0 and US-59 south westbound had three more frontage segments with relative risks over 4.0 than the overall posterior risk map. Sunday had less high-risk segments than Saturday, but more than other days of a week, as shown in Figure 4.15. Figure 4.13 indicates that Friday had similar spatial pattern with Sunday, but less risky segments. High-risk segments on weekdays were not as concentrated as those on other days and less high-risk segments were observed on weekdays, as shown in Figure 4.16. Figures 13-16 also verified the validity of overall temporal relative risk pattern by day of week in Section 4.2.1.

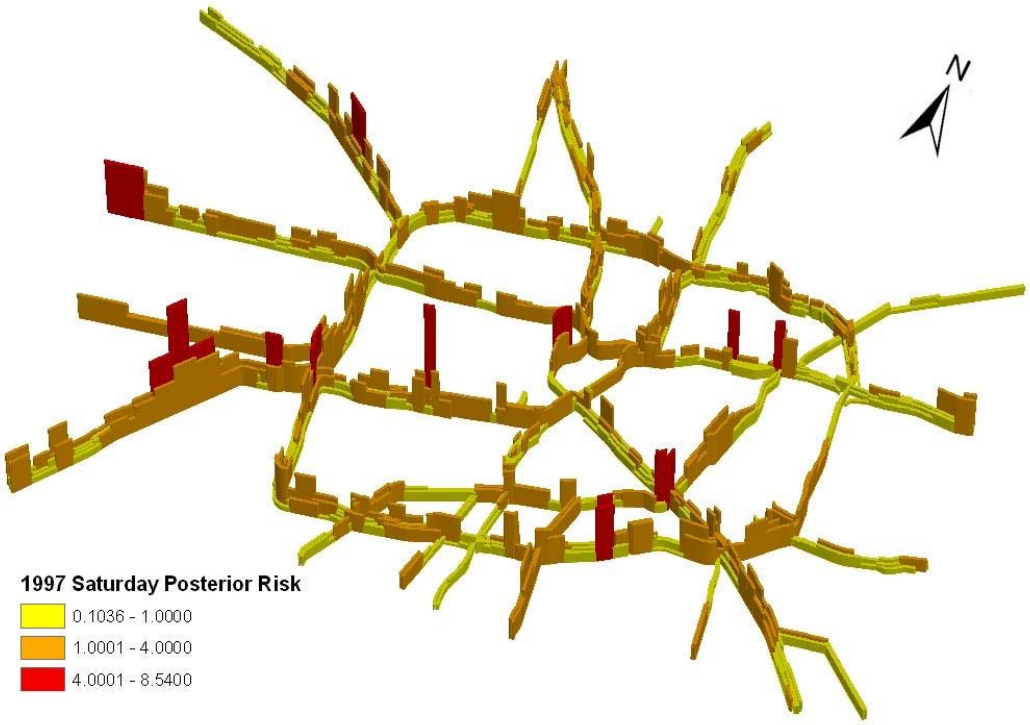


Figure 4.14 Posterior Relative Crash Risk for Saturdays of Year 1997

Figure 4.15 Posterior Relative Crash Risk for Sundays of Year 1998

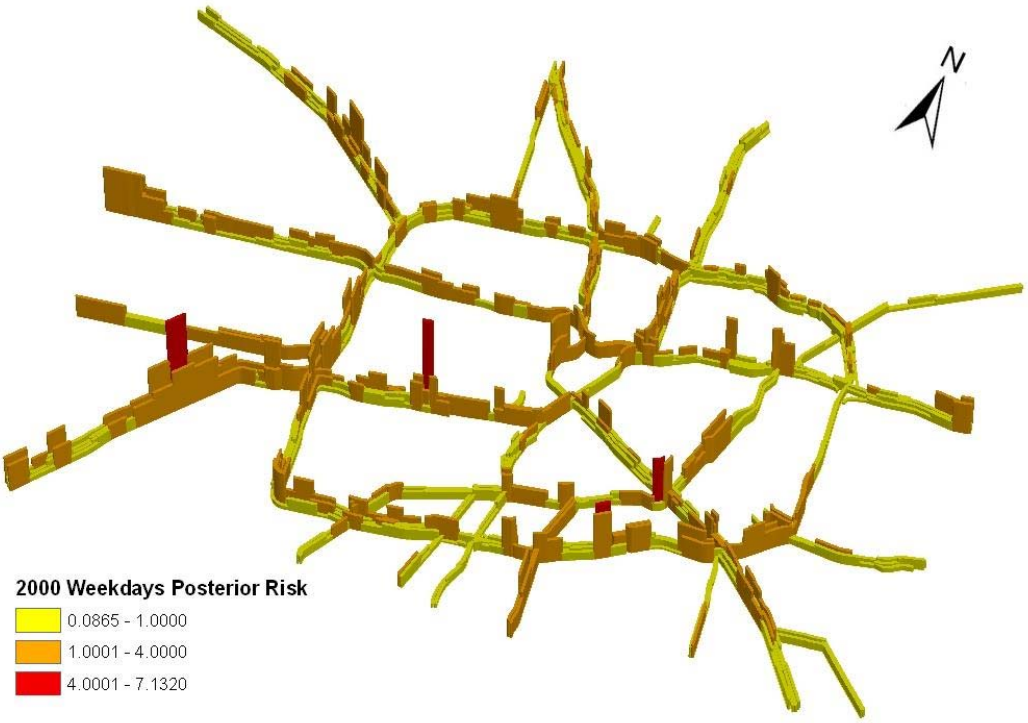


Figure 4.16 Posterior Relative Crash Risk for Weekdays of Year 2000

Whatever the difference of daily pattern was, the less safe segments on one day of a week were probably not safe on other days. These close relative risk distributions suggest the stability of posterior results of Bayesian spatial smoothing due to the lack of the variance and the explicitness of high-risk segments no matter which day or year was analyzed.

4.4.2 Spatial Analysis by Hour of Day

The spatial analysis by hour of day is a higher-level disaggregated analysis, which needs very detailed temporal information about crash and hourly traffic flow information for every roadway segment. In order to calculate expected hourly crash

frequency, the hourly traffic volume were calculated from AADT by using corresponding daily traffic adjustment factors, hourly adjustment factors, and directional traffic split rates as mentioned in sub-section 3.2.2. The crash frequencies were counted by different day of week and hour of day. Because of the high-level disaggregation, more variance is expected in the observed risk map.

Figure 4.18 shows the posterior relative risk distribution between 7:00 a.m. and 8:00 a.m. on weekdays, which is the smoothed result of observed relative risk distribution, shown in Figure 4.17. The observed relative risks for several roadway segments were zero or extremely high because of the zero observed crashes or very low expected crash counts on those segments. The Bayesian approach showed its ability again in filtering the uncertainty in the data with large variance and smoothing the spatial relative risks for all segments. The segment with the highest observed relative risk was the I-10 west eastbound frontage segment without interchange. This segment is 0.2 miles long (shorter than average) with hourly volume 185 vehicles (relatively low) at 7:00 a.m.—8:00 a.m. on weekdays. Bayes approach detected this small area, and smoothed its observed risks 23.2861 to posterior risk value 3.73 by incorporating its adjacent segments (including upstream and downstream frontage segments and adjacent freeway segment).

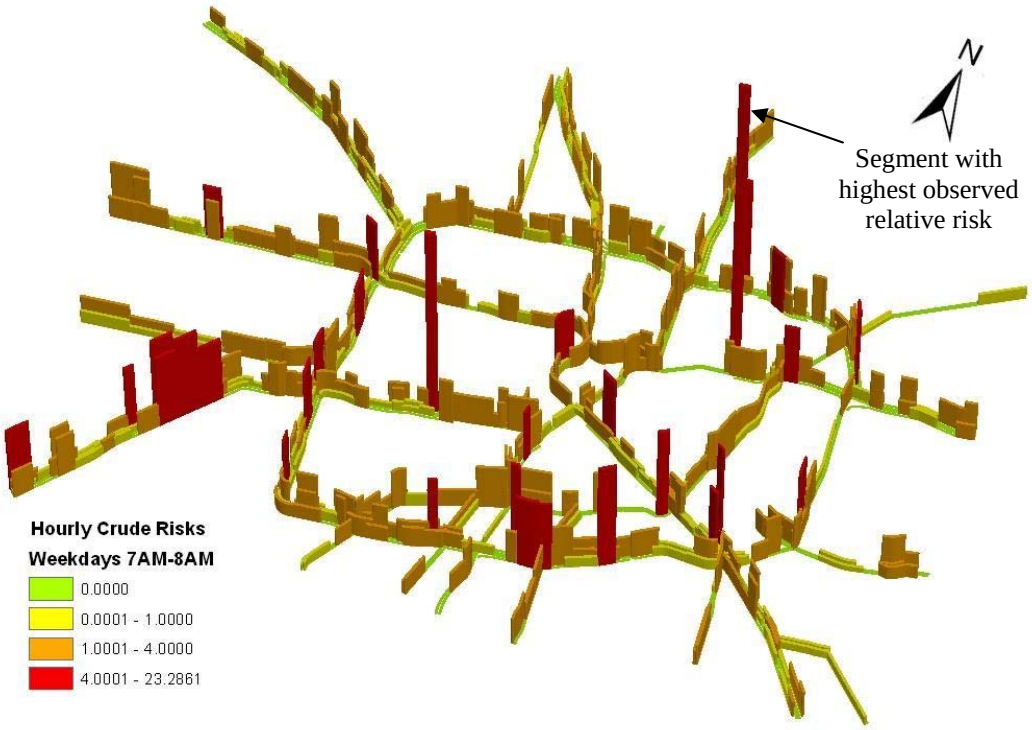


Figure 4.17 Observed Relative Crash Risk, 7:00-8:00 AM, Weekdays

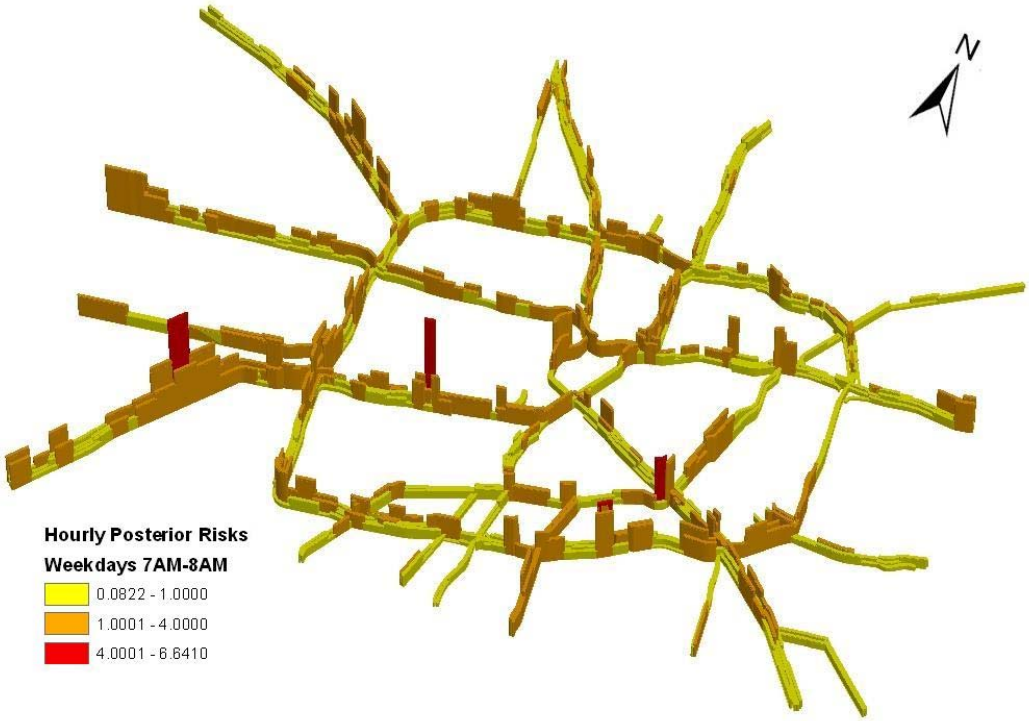


Figure 4.18 Posterior Relative Crash Risk, 7:00-8:00 AM, Weekdays

The above figures show the spatial risk pattern for the morning peak time of weekdays. It can be seen that the general posterior pattern was close to the overall spatial pattern, but there were less high-risk segments than the overall posterior risk pattern. Figure 4.19 shows the spatial pattern of posterior relative crash risks of the afternoon peak hour (between 5:00 p.m. and 6:00 p.m.) on weekdays.

Figure 4.19 Posterior Relative Crash Risk, 5:00-6:00 PM, Weekdays

Figure 4.19 shows that relative risks during the afternoon peak were higher and distributed broader than those at the morning peak. This implies that people had less patience and were more tired when driving after work than before work if the traffic flow in the morning was similar to that in the afternoon. The distribution of relative risks on Fridays is similar to that on weekdays.

The relationship between relative risks in different directions was stable. The majority of the traffic moved from inbound direction to outbound direction from morning to afternoon, but the high-risk segments did not move to the opposite direction from Figure 4.18 and 4.19. This phenomena suggests the stability of the posterior estimations of Bayesian approach, also implies that there were factors other than traffic volume had influences the relative risk values.

It is shown in Figure 4.7 that the lowest risk within a week appeared at 5:00-6:00 AM on weekdays. Figure 4.20 shows the spatial distribution of posterior relative crash risks for that hour period. It can be seen that all the segments were safer than those at morning peak and afternoon peak. The highest risk value during that hour period is 2.999, much less than that in other hour period.

Figure 4.20 Posterior Relative Crash Risk, 5:00-6:00 AM, Weekdays

From Figure 4.7, the least safe hour within a week was at 2:00-3:00 AM on Saturday, whose posterior relative risk distribution is shown in Figure 4.21. It can be seen that the high risks occurred everywhere, from downtown to outside of the loop, and from freeways to arterials, and the relative risk value of each segments was much higher than that at other times. The same hour period on Sundays had the same pattern of relative risks as Saturdays.

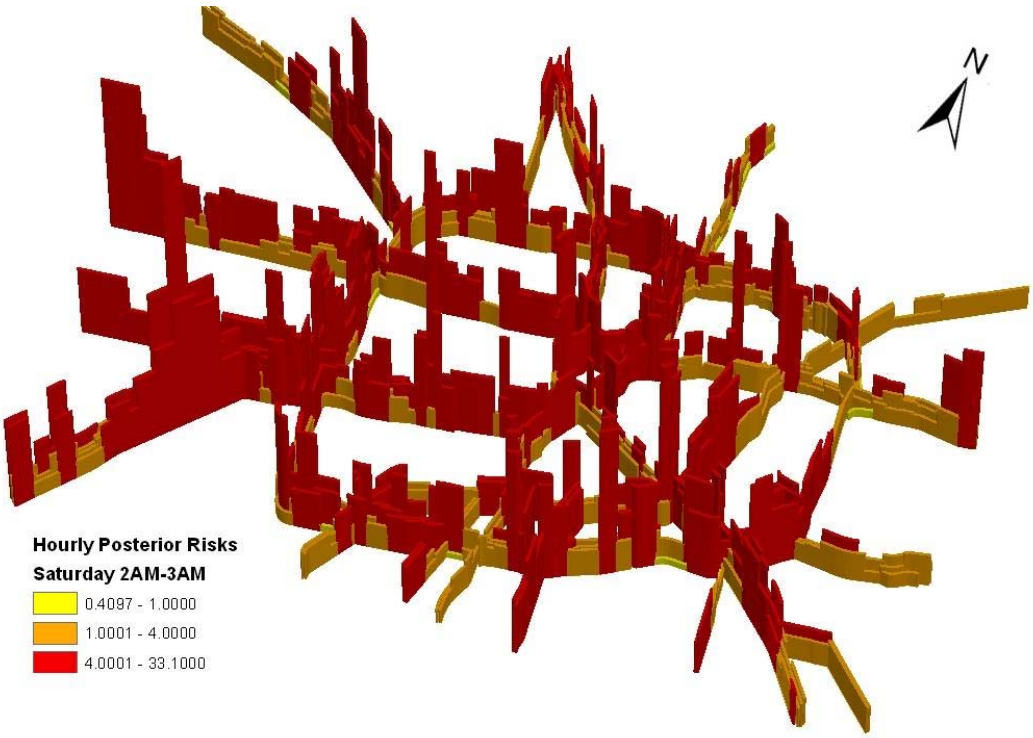


Figure 4.21 Posterior Relative Crash Risk, 2:00-3:00 AM, Saturdays

In the afternoon of weekend, relative crash risks had a flat peak from 2:00 p.m. to 5:00 p.m. The distribution of relative risks at 3:00 p.m. to 4:00 p.m. on Saturdays is shown in Figure 4.22. The typical high-risk segments were the same as weekdays, but

the number of high-risk segments (risk>4) were not as many as there were in the weekday afternoon. The spatial pattern during the same period on Sunday was similar.

Figure 4.22 Posterior Relative Crash Risk, 3:00-4:00 PM, Saturdays

4.5 Discussions

The hierarchical Bayesian approach is a very powerful approach to capture the real safety tendency behind the crash data with large variances. In this thesis research, Bayesian modeling involves two stages: a) a likelihood model as assumptions for the distributional characteristics of the observed events (e.g. traffic crashes) and b) a prior model which is based on prior beliefs over the parameter spaces. The observed risk map is based on the observed count data alone and often feature large outlying relative risks in segments with very low traffic volumes or very short segment length, so that visually

the map shows a high uncertainty. The posterior estimates of the relative risks incorporate both variability in the observed risks and spatial autocorrelation assumptions, and enable the typical Bayesian “borrow of strength” from adjacent segments. The relative risk map based on the posterior estimates preserves the “truly” high-risk segments while smoothing out the random variance. For example, Figures 4.9, 4.12, and 4.17 are observed risk maps for 5 years, one day in a year, and one hour in a day, respectively, with different variance level. However, after Bayesian estimation, the uncertainty was detected and filtered. The posterior risk maps (Figures 4.10, 4.13, and 4.18) are quite similar, with the almost the same high-risk segments. This means that no matter how much the variance is in the raw data, the Bayesian approach will detect it and capture the real safety tendency. Furthermore, from all the posterior risk maps, the risky segments that had been identified from the overall risk maps were always relatively riskier than other segments, although the different day of a week and different hour of a day had different risk patterns. This also suggests the stability of the posterior results.

From all the posterior relative risk maps shown above, some roadways at the same location but in different directions have very different risk values although many of them have similar values. Not differentiating directions in risk assessment might average the two directional risks and make the problem unnoticeable. Therefore, direction differentiation is necessary in relative crash risk evaluations. For those segments with different risk values in different directions, only improving a the direction with a risk value of over 1 might be a wiser idea than improving both direction if the other direction

has a risk value less than 1. Therefore, risky segments can be identified more accurately and limited funds can be saved by differentiating directions.

It should be noticed that same risk values on different road segments do not mean the same impact when a crash occurred. For example, in the morning peak, one lane blockage due to the crash on an inbound lane will cause more congestion, delay, and idle emissions than one on an outbound lane if the capacity of the remaining inbound lanes could not accommodate the inbound traffic. Furthermore, it would take more time for emergency personnel to clear the crashed vehicles from a congested roadway than from an uncongested roadway.

5 CONCLUSIONS

5.1 Summary

This research seeks to develop a GIS-based Bayesian approach to perform area-wide link-based relative crash risk analysis. The approach was used to capture the spatial and temporal patterns of relative risks of traffic crashes and identify the real high-risk segments. To better capture the real safety indications, this thesis differentiates traffic directions, disaggregates different road types, integrates hierarchical Bayesian approach to filter the uncertainties based on the spatial assumptions and presents the results in 3-D visualizations. The results of spatial-temporal crash patterns could be used to determine risky roadway segments that need attention by transportation agencies as well as motorists.

Compared to the conventional statistical inference which derives the point and interval estimates for parameters, hierarchical Bayesian modeling produces the full distribution on parameters taking account into heterogeneity effects, spatial autocorrelation, and covariate effects. Hierarchical Bayesian approach is effective to spatially estimate the observed risks with large variance and high uncertainty, and capture the real safety tendency underneath the crash data. Small mapping units like road segments sometimes are necessary for discerning possible risk factors. But the use of small units causes unstable risk estimates due to small sample sizes and hence produces risk maps with high variability. Smoothing is therefore helpful in producing maps displaying true spatial patterns. The hierarchical Bayesian modeling that was applied in

this thesis demonstrated that the approach performs as expected since the relative risk maps based on the posterior estimates preserve the “truly” high-risk segments for areas with high traffic volumes while smoothing out variability in low traffic segments or short segments due to random variance.

In this study, different values of parameters for crash models are estimated for different road types, and the crash frequency for each road type has a different relationship with its directional AADT. The number of crashes on urban freeway mainlines will increase with a higher speed than AADT, but crash rate for urban freeway mainlines is always lower than any other road types given the same traffic volume. The freeway system interchange area is the second safest road type, and its number of crashes has a linear relationship with AADT. Crash count on frontage roads will increase at a lower rate than their AADT, but the frontage roads without diamond interchange are always safer than the ones with diamond interchange. The number of crashes on arterials has a linear relationship with AADT and its average crash rate is lower than both frontage road types when AADT is small (<10,000), and higher than both frontage road types when AADT is big (>27,000).

Temporal results show the variation of relative crash risks for the study area by day of week and hour of day. In a week, Saturday is the least safe day, followed by Sunday and Friday, and weekdays (Monday-Thursday) are relatively safer. In hour of day, nighttime is less safe than daytime, especially the early morning (1:00 a.m.—4:00 a.m.), which is the least safe time for all days. During weekdays and Friday, there are two flat risk peaks in the daytime, morning peak (7:00 a.m.—9:00 a.m.) and afternoon

peak (3:00 p.m.—7:00 p.m.). During daytime of the weekend, there is only a flat peak in the afternoon from 1:00 p.m. to 5:00 p.m.

The spatial analysis and spatial-temporal analysis of relative crash risks point out the risky segments for different day of week and hour of day. Hierarchical Bayesian approach filtered the uncertainty in the data with large variance and captured the similarity between the risk values of adjacent segments, so the results of spatial analysis are more convincing and accurate. The identification of risky segments is the necessary prerequisite for the next step, contributing factors identification.

Three-dimensional visualization is more effective and carrying more information than two-dimensional maps. Therefore, 3-D maps are advocated in presenting relative crash risks in crash analysis.

5.2 Limitations

The hierarchical Bayesian approach has its limitations. It is complex and performs well only when the variance is moderate or high in the data. When a lot of data are available (e.g. 5 years KAB crashes for Houston), the result of hazardous site identification by using hierarchical Bayesian approach (Figure 4.10) is very similar to the one by using common method (Figure 4.9) since the variance and uncertainty are small in the big dataset And both methods identify the same high-risk segments. However, when look at the relative crash risks by day of week or hour of day, more variance is observed and Bayesian advantages of filtering uncertainty can be shown (Figures 12, 13, 17, and 18). Therefore, the hierarchical Bayesian approach is most

suitable for the data that have large variances. For example, the crashes with speeding or drunk driving, which normally have large variances, can be analyzed by the Bayesian approach so that high-risk locations can be identified to enforcement measures can be allocated at appropriate locations and time periods.

The direction differentiation cannot address some crashes accurately, although it can avoid averaging effect. The direction differentiation performs well when the traffic or crashes are unevenly distributed on different directions. However, this requires that every crash should be assigned to one of the two directions. It is difficult to assign cross-median crashes (a vehicle crosses the median area of a divided roadway, enters the opposing traffic lanes, and then collides with one or more vehicles on the opposing lanes) to one of the two directions. Therefore, the direction differentiation can not address cross-median crashes accurately.

This research does not address the temporal correlation by day of week and hour of day. The results of temporal analysis by day of week and hour of day will be more stable and reliable if temporal correlation is included in the analysis.

The contributing factors have not been identified. This research found that some segments had a relative risk value different from the opposite direction, but it did not answer what caused the directional difference. Furthermore, risk maps allow determining where and when each roadway segment has a high relative crash risk, but this research did not answer why those segments are riskier or what can be done to improve safety for these segments.

5.3 Future Work

There are many directions in which this study can be extended. Here are some promising extensions:

- The contributing factors need to be determined individually for a certain high-risk roadway segments due to the fact that each roadway segment has its own traffic, roadway, and environmental conditions. The contributing factors need to be identified for these specific locations so that proper preventive measures can be implemented.
- Different functional forms should be tested for the crash prediction models. Incorporating geometric conditions, road conditions, and number of lanes etc. may generate more accurate models for predicting crashes.
- Different prior values can also be tried for hierarchical Bayesian model coded in WinBUGS. It is expected better prior values can help convergence of the two sampling chains with MCMC simulation.
- The crash data with large variability, such as drunk driving crashes and speed-related crashes, can be used for hazardous site identification by the hierarchical Bayesian approach. It is expected that this approach performs well in filtering the uncertainty and capture the hazardous sites for these types of crashes so that proper enforcement measures can be carried out.

REFERENCES

Affum, J.K., Taylor, M.A.P., 1995. Integrated GIS database for road safety management. Technology Tools for Transportation Professionals–Moving into the 21st Century: Resource Papers for the 1995 International Conference. Institute of Transportation Engineers, Washington, D.C., pp.189-193.

Affum, J.K., Taylor, M.A.P. 1996. GIS-based system for the systematic determination of road accident problems and causal factors. Roads 96 Conference Proceeding–Part 5, ARRB Transport Research Ltd., Victoria, Australia.

Agent, K.R., 1973. Evaluation of the High-Accident Location Spot-Improvement Program in Kentucky. Frankfort: Kentucky Department of Highways Research Report no. 357.

Antle, C.E., Arnold, S., 1978. An empirical Bayes solution for the problem considered by Williford and Murdock. Accident Analysis and Prevention 10 (3), 185-188.

Bellono-McGee Inc., Midwest Research Institute, 2003. Highway Safety Manual, Prototype Chapter—Two-Lane Highways, NCHRP project 17-18 (4).

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems (with discussion), Journal of the Royal Statistical Society, Series B, 36, 192-236.

Besag, J., Newell, J., 1991. The detection of clusters in rare diseases. Journal of the Royal Statistical Society, Series A, 154, 143–155.

Best, N.G., Waller, L.A., Thomas, A., Conlon, E.M., Arnold, R.A., 1999. Bayesian models for spatially correlated diseases and exposure data. In: Bernardo et al. (Eds), Bayesian Statistics 6, Oxford: Oxford University Press.

Black, W.R., 1991. Highway accidents: a spatial and temporal analysis. Transportation Research Record, 1318, 75-82.

Black, W.R., 1992. Network autocorrelation in transport network and flow systems. Geographical Analysis, 24 (3), 207- 222.

Black, W.R., Thomas, I., 1998. Accidents in Belgium's motorways: a network autocorrelation analysis. Journal of Transport Geography, 6 (1), 23-31.

Blincoe, L., Seay, A., Zaloshnja, E., Miller, T., Romano, E., Luchter, S., Spicer, R., 2002. The economic impact of motor vehicle crashes, 2000, Report DOT HS 809 446. National Highway Traffic Safety Administration, Washington, D.C.

Bolstad, W.M., 2004. Introduction to Bayesian Statistics. New York: John Wiley & Sons.

Brodsky, H., Hakkert, A.S., 1983. Highway crash rates and rural travel densities. Accident Analysis and Prevention, 15 (1), 73-84.

Brüde, U., Larsson, J., 1988. The use of prediction models for eliminating effects due to regression-to-the-mean in road accident data. Accident Analysis and Prevention, 20 (4), 299-310.

Carlin, B.P, Louis, T.A., 1996. Bayes and Empirical Bayes Methods for Data Analysis. New York: Chapman & Hall/CRC.

Chipman, M.L., 1982. The role of exposure, experience and demerit point levels in the risk of collision. Accident Analysis and Prevention, 14 (5), 475-483.

Choi, K., Park, I., 1996. Traffic accident analysis with GIS and statistical package. Ninth Symposium on Geographic Information Systems for Transportation. American Association of State Highway and Transportation Officials, Washington, D.C.

Congdon, P., 2001. Bayesian Statistical Modeling. John Wiley & Sons, Chichester, U.K.

Demongraphia website, 2001. 2000 Census: US Municipalities over 50,000: Ranked by 2000 Population. Wendell Cox Consultancy. <http://www.demographia.com/db-2000city50kr.htm> (accessed 05/29/06)

El-Sadig, M., Norman, J.N., Lloyd, O.L., Romilly, P., Bener, A., 2002. Road traffic accidents in the United Arab Emirates: trends of morbidity and mortality during 1977–1998. Accident Analysis and Prevention, 34 (4), 465-476.

Environmental Systems Research Institute (ESRI), 2004. Getting to Know ArcGIS Desktop: The Basics of ArcView, ArcEditor, and ArcInfo Updated for ArcGIS 9, ESRI Press. Redlands, California.

Environmental Systems Research Institute (ESRI), 2006. What is GIS? <http://www.gis.com/whatisgis/index.html> (accessed 05/29/06)

Evans, L., 2004. Traffic Safety. Bloomfield Hills, Michigan: Science Serving Society.

Federal Highway Administration (FHWA), 2002. SafetyAnalyst: Software Tools for Safety Management of Specific Highway Sites. Task K, White Paper for Module 1—Network Screening.
<http://www.safetyanalyst.org/whitepapers/module1.pdf> (Accessed 06/20/06)

Fitzpatrick, K., Harwood, D.W., Potts, I.B., 2000. Accident Mitigation Guide for Congested Rural Two-Lane Highways. NCHRP Report 440. National Academy Press, Washington, D.C.

Flahaut, B., Mouchart, M., San Martin, E., Thomas, I., 2003. The local spatial autocorrelation and the kernel method for identifying black zones: a comparative approach. Accident Analysis and Prevention, 35 (6), 991-1004.

Fridstrøm, L., Ingebrigtsen, S., 1991. An aggregate accident model based on pooled, regional time-series data. Accident Analysis and Prevention, 23 (5), 363-378.

Gårder, P., 1989. Pedestrian safety at signalized intersections: a study carried out with the help of a traffic conflict technique. Accident Analysis and Prevention, 21 (5), 435-444.

Ghosh, M., Natarajan, K., Waller, L. A., Kim, D., 1999. Hierarchical Bayes for the analysis of spatial data: an approach to disease mapping. Journal of Statistical Planning and Inference, 75, 305-318.

Goh, P.C., 1993. Traffic accident analysis using geoprocessing techniques. Road and Transport Research, 2 (2), 76-85.

Golias, J.C., 1992. Establishing relationships between accidents and flows at urban priority road junctions. Accident Analysis and Prevention, 24 (6), 689-694.

Golob, T.F., Recker, W.W., Levine, D.W., 1990. Safety of freeway median high occupancy vehicle lanes: a comparison of aggregate and disaggregate analyses. Accident Analysis and Prevention, 22 (1), 19-34.

Haight, F.A., Olsen, R.A., 1981. Pedestrian safety in the United States: some recent trends. Accident Analysis and Prevention, 13 (1), 43-55.

Hauer, E., 1982. Traffic conflicts and exposure. Accident Analysis and Prevention, 14 (5), 359-369.

Hauer, E., 1992. Empirical Bayes approach to the estimation of "unsafety": the multivariate regression method. Accident Analysis and Prevention, 24 (5), 457–477.

Hauer, E., 1997. Observational Before-After Studies in Road Safety. Oxford: Pergamon Press.

Hauer, E., 2002. Estimating safety by the empirical Bayes method. Transportation Research Record 1784, 126-131.

Hauer, E., 2004. Statistical safety modeling, Paper presented at the 2004 Annual Meeting of the Transportation Research Board, Washington, D.C.

Hauer, E., Bamfo, J., 1997. Two tools for finding what function links the dependent variable to the explanatory variables. In Proceedings of the ICTCT 1997 Conference, Lund, Sweden.

Hauer, E., Ng, J.C.N., Lovell, J., 1989. Estimation of safety at signalized intersections. Transportation Research Record 1185, pp. 48-61.

Hauer, E., Persaud, B.N., 1988. How to estimate the safety of rail-highway grade crossings and the safety effects of warning devices. Transportation Research Record, 1114, 131-140.

Houston-Galveston Area Council (H-GAC), 2006. Fact sheet on safety in the City of Houston. H-GAC Traffic Safety Planning Program web page, H-GAC. <http://www.h-gac.com/NR/rdonlyres/eusx7va6rnjs6acmcomtvhjloffsbpniycyp7seqac7vrly35lpi7v3z747rvx5c42345cbzt7wwmsfxqjentptx5ea/Safety+in+the+City+of+Houston_1999.pdf > (accessed 05/05/06)

Hovenden, E., Sligoris, J., Walker, C. 1995. A roads spatial information system: the victorian experience. Road and Transport Research, 4 (1), 110-121.

Insightful Corporation, 2001. S-PLUS 6 for Windows User's Guide, Insightful Corporation, Seattle, Washington.

Janssens, R., 1999. Long-term trend of modeling accident risk using time series data. World Transport Research: Selected Proceedings of the 8th World Conference on Transport Research, Antwerp, Belgium. 2, 71-84.

Jegede, F.J., 1988. Spatio-temporal analysis of road traffic accidents in Oyo State, Nigeria. Accident Analysis and Prevention, 20 (3), 227-243.

Jones, B., Janssen, L., Mannering, F., 1991. Analysis of the frequency and duration of freeway accidents in Seattle. Accident Analysis and Prevention, 23 (4), 239-255.

Kam, B.H., 2003. A disaggregate approach to crash rate analysis. Accident Analysis and Prevention, 35 (5), 693-709.

Lamm, R., Guenther, A.K., Choueiri, E.M. 1995. Safety module for highway geometric design. Transportation Research Record, 1512, 7-15.

Lee, P.M, 2004. Bayesian Statistics: An Introduction, Third Edition. London: Arnold.

Levine, N., 2004. CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (Version 3.0). Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, D.C.

Levine, N., Kim, K., Nitz, L., 1995a. Spatial analysis of Honolulu motor vehicle crashes: I. spatial patterns. Accident Analysis and Prevention, 27 (5), 663-674.

Levine, N., Kim, K., Nitz, L., 1995b. Spatial analysis of Honolulu motor vehicle crashes: II. zonal generators. Accident Analysis and Prevention, 27 (5), 675-685.

Levine, N., Kim, K., Nitz, L., 1995c. Daily fluctuations in Honolulu motor vehicle accidents. Accident Analysis and Prevention, 27 (6), 785-796.

Lord, D., 2002. Issues related to the application of accident prediction models for the computation of accident risk on transportation networks. Transportation Research Record, 1784, 17-26.

Lord, D., 2006. Regression analysis of count data and development of statistical models. CVEN 626 Class notes. Zachry Department of Civil Engineering, Texas A&M University. <http://ceprofs.tamu.edu/dlord/CVEN_626_Course_%20Material.htm> (accessed: 03/05/06)

Lord, D., Persaud, B.N., 2000. Accident prediction models with and without trend: application of the Generalized Estimating Equation (GEE) procedure. Transportation Research Record, 1717, 102-108.

Lord, D., Manar, A., Vizioli, A., 2005. Modeling crash-flow-density and crash-flow-V/C ratio for rural and urban freeway segments. Accident Analysis and Prevention. 37 (1), 185-199.

MacNab, Y., 2004. Bayesian spatial and ecological models for small-area accident and injury analysis. Accident Analysis and Prevention, 36 (6), 1019-1028.

Maher, M.J., Hughes, P.C., Smith, M.J., Ghali, M.O., 1993. Accident- and travel time-minimizing routing patterns in congested networks. Traffic Engineering and Control, 34 (9), 414-419.

Martin, J.-L., 2002. Relationship between crash rate and hourly traffic flow on interurban motorways. Accident Analysis and Prevention, 34 (5), 619-629.

Mensah, A., Hauer, E., 1998. Two problems of averaging arising in the estimation of the relationship between accidents and traffic flow. Transportation Research Record, 1635, 37-44.

Miaou, S.-P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binominal regressions. Accident Analysis and Prevention, 26 (4), 471-482.

Miaou, S.-P., Bligh, R.P., Lord, D., 2005. Developing median barrier installation guidelines: a benefit/cost analysis using Texas data. Transportation Research Record, 1904, 3-19.

Miaou, S.-P., Lord, D., 2003. Modeling traffic-flow relationships at signalized intersections: dispersion parameter, functional form and Bayes vs Empirical Bayes. Transportation Research Record 1840, 31-40.

Miaou, S.-P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accident Analysis and Prevention, 25 (6), 689-709.

Miaou, S.-P., Song, J., 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence. Accident Analysis and Prevention, 37 (4), 699-720.

Miaou, S-P., Song, J.J., Mallick, B.K., 2002. Roadway traffic-crash mapping: a space-time modeling approach. Journal of Transportation and Statistics, 6 (1), 33-57.

Miller, H.J., Shaw, S-L., 2001. Geographic Information Systems for Transportation: Principles and applications. New York: Oxford University Press.

Miller, J.S., 1999. What Value May Geographic Information Systems Add to the Art of Identifying Crash Countermeasures? Virginia Transportation Research Council, VTRC 99-R13, Charlottesville, Virginia.

Miller, S., Johnson, T., Smith, S., Raymond, L., 1995. Design and development of a crash referencing and analysis system. Proceedings of the 1995 Geographic Information Systems for Transportation (GIS-T) Symposium. American Association of State Highway and Transportation Officials, Washington, D.C.

Mountain, L., Fawaz, B., Jarrett, D., 1996. Accident prediction models for roads with minor junctions. Accident Analysis and Prevention, 28 (6), 695-707.

Myers, R.H., Montgomery, D.C., Vinning, G.G., 2002. Generalized Linear Models: With Application in Engineering and Sciences. New York: Wiley.

National Highway Traffic Safety Administration (NHTSA), 2005. Motor vehicle traffic crashes as a leading cause of death in the United States, 2002. Traffic Safety Facts Research Note. NHTSA Washington D.C. <http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/RNotes/2005/809831.pdf> (accessed 09/27/05)

Ng, K., Hung, W., Wong, W., 2002. An algorithm for assessing the risk of traffic accident. Journal of Safety Research, 33 (3), 387-410.

Nicholson, A.J., 1985. The variability of accident counts. Accident Analysis and Prevention, 17 (1), 47-56.

Okamoto, H., Koshi, M., 1989. A method to cope with the random errors of observed accident rates in regression analysis. Accident Analysis and Prevention, 21 (4), 317-332.

Qin, X., Ivan, J. N., Ravishanker, N., Liu, J., 2005. Hierarchical Bayesian estimation of safety performance functions for two-lane highways using Markov Chain Monte Carlo modeling. Journal of Transportation Engineering, 131 (5), 345-351.

Qin, X., Ivan, J. N., Ravishanker, N., Liu, J., Tepas, D., 2006. Bayesian estimation of hourly exposure functions by crash type and time of day. Accident Analysis and Prevention, (forthcoming).

Qin, X., Lin, W-H, 2006. Safety performance measure in the urban regional transportation plan, Transportation Research Board 85th Annual Meeting CD. Washington, D.C.

Petch, R.O., Henson, R.R., 2000. Child road safety in the urban environment. Journal of Transport Geography, 8 (3), 197-211.

Persaud., B., Dzbik, L., 1993. Accident prediction models for freeways. Transportation Research Record, 1401, 55-60.

Raub, R.A. 1997. Occurrence of secondary crashes on urban arterial roadways. Transportation Research Record, 1581, 53-58.

Resende, P.T.V., Benekohal, R.F., 1997. Effects of roadway section length on accident modeling. In Traffic Congestion and Traffic Safety in the 21st Century: Challenges, Innovations, and Opportunities. American Society for Civil Engineers. Chicago, Illinois, pp.403-409.

Rossi, P.E., Allenby, G.M., McCulloch, R., 2006. Bayesian Statistics and Marketing. New Jersey: Wiley.

Saccomanno, F.F., Chong, K.C., and Nassar, S.A. 1997. Geographic information system platform for road accident risk modeling. Transportation Research Record, 1581, 18-26.

SAS Institute Inc., 1999. SAS OnlineDoc, Version 8. Cary, North Carolina. <http://v8doc.sas.com/sashtml/> (Accessed 06/19/06)

Siegel, J., Yang, S., 1998. TMS: Indianapolis's data traffic cop. Geographic Information Systems. 8 (5), 38-40.

Skabardonis, A., Chira-Chavala, T., Rydzewski, D., 1998. The I-880 field experiment: effectiveness of incident detection using cellular phones. California PATH Research Report, UCB-ITS-PRR-98-1.

Spiegelhalter, D.J., Thomas, A., Best, N., Lunn, D., 2004. WinBUGS user manual Version 1.4.1, Medical Research Council Biostatistics Unit, Cambridge, UK.

SPSS Inc., 2005. SPSS 14.0 Brief Guide, 1st Edition. Prentice Hall, Chicago, Illinois.

Steenberghen, T., Dufays, T., Thomas, I., Flahaut, B., 2004. Intra-urban location and clustering of road accidents using GIS: a Belgian example. International Journal of Geographical Information Science, 18 (2), 169-181.

Sun, D., Tsutakawa, R.K., Kim, H., He, Z., 2000. Spatio-temporal interaction with disease mapping. Statistics in Medicine, 19, 2015-2035.

Tanner, J.C., 1953. Accidents at rural three-way junctions. Journal of the Institution of Highway Engineers, 2 (11), 56-67.

Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography, 46, 234–240.

Thill, J-C. (Eds), 2000. Geographic Information Systems in Transportation Research. New York: Pergamon.

Thomas, A., Best, N., Lunn, D., Arnold, R., Spiegelhalter, D.J., 2004. GeoBUGS user manual Version 1.2, Medical Research Council Biostatistics Unit, Cambridge, U.K.

United States Department of Transportation (USDOT), Bureau of Transportation Statistics (BTS), 2001. Transportation Statistics Annual Report 2000, BTS01-02.

Washington, D.C. <http://www.bts.gov/publications/tsar/2000/index.html> (accessed 10/10/04)

United States Department of Transportation (USDOT), National Highway Traffic Safety Administration (NHTSA), 2004. Traffic Safety Facts 2002: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System. Washington D.C. <http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/TSFAnn/TSF2002Final.pdf> (accessed 03/12/05)

Wakefield, J.C., Best, N.G., Waller, L., 2000. Bayesian approaches to disease mapping. In: Elliott, P., Wakefield, J.C., Best, N.G., Briggs, D.G., (Eds), Spatial Epidemiology: Methods and Applications. Oxford: Oxford University Press, pp.104-127.

Wakefield, J.C., Morris, S.E., 2001. The Bayesian modeling of disease risk in relation to a point source. Journal of the American Statistical Association, 96 (453), 77-91.

Waller, L., Carlin, B., Hong, X., Gelfand, A., 1997. Hierarchical spatio-temporal mapping of disease rates. Journal of the American Statistical Association, 92, 607–617.

Winkler, R.L., 2003. An Introduction to Bayesian Inference and Decision, Second Edition. Boston, Massachusetts: Probabilistic Publishing.

Withers, S.D., 2002. Quantitative methods: Bayesian inference, Bayesian thinking. Progress in Human Geography, 26, 553–566.

World Road Association (PIARC), 2003. Road Safety Manual. Route to Market, Quebec, Canada.

Wolf, A.C. 1982. The concept of exposure to the risk of a road traffic accident and an overview of exposure data collection methods. Accident Analysis and Prevention, 14 (5), 337-340.

Yamada, I., Thill, J-C., 2004. Comparison of planar and network K-functions in traffic accident analysis. Journal of Transport Geography, 12, 149-158.

Zhu, L., Carlin, B.P., English, P., Scalf, R., 2000. Hierarchical modeling of spatio-temporally misaligned data: relating traffic density to pediatric asthma hospitalizations. Environmetrics. 11 (1), 43-61.

VITA

Name: Linhua Li

Address: Wilbur Smith Associates, 9800 Richmond Ave. Suite 400, Houston, TX 77042

Email Address: lilinhua@gmail.com

Education: B. Eng., International Containerized Transportation Management, Shanghai Maritime University, China, 2001
M.S., Transportation Planning and Management, Texas Southern University, 2004
M.S., Civil Engineering, Texas A&M University, 2006