AN APPLICATION OF MULTI-LEVEL BAYESIAN NEGATIVE BINOMIAL MODELS WITH MIXED EFFECTS ON MOTORCYCLE CRASHES IN OHIO

A Thesis
Presented to
The Graduate Faculty of The University of Akron

In Partial Fulfillment
of the Requirements for the Degree
Master of Science

Thomas V. Flask
May, 2012
AN APPLICATION OF MULTI-LEVEL BAYESIAN NEGATIVE BINOMIAL MODELS WITH RANDOM EFFECTS ON MOTORCYCLE CRASHES IN OHIO

Thomas V. Flask

Thesis

Approved: 

Advisor
Dr. William H. Schneider IV

Department Chair
Dr. Wieslaw K. Binienda

Committee Member
Dr. Ping Yi

Dean of the College
Dr. George K. Harritos

Committee Member
Dr. Anil Patnaik

Dean of the Graduate School
Dr. George R. Newkome

Date

ii
Motorcycle crashes compose an increasing proportion of United States motor vehicle crashes and fatalities. Through Full Bayesian negative binomial models with various types of random effects, this study compiles two distinct types of motorcycle crash models after literature reviews in the fields of motorcycle safety and statistical research in transportation safety. Using Ohio motorcycle crash data, motorcycle crashes are analyzed at the regional and roadway segment levels. At the regional level, single vehicle motorcycle crash models at the Ohio township are found to be improved in goodness-of-fit by both county level and township level spatial random effects with different levels of neighborship. First order township neighbors were found to create the most improvement in the Deviance Information Criterion (DIC). At the segment level, multi-vehicle motorcycle crashes models were found to be improved in goodness-of-fit if neighborship was defined at a sufficiently large radius to include prior knowledge of the surrounding regions. In both types of models, including regional information at the county and township levels helped avoid the pitfalls associated with motorcycle data, which is often unavailable, such as a lack of motorcycle specific Annual Average Daily Traffic (ADT) and Vehicle Miles Traveled (VMT). Each model produced descriptive parameter results that show the whether a predictor has a positive or negative influence on the frequency of motorcycle crashes and its magnitude. Conclusions based on the DIC and parameter results are made in terms of motorcycle riding and safety, as well as recommendations for implementation of the models developed in these studies on other data sets. Finally, the models are compared and contrasted to highlight the advantages and disadvantages of each approach.
ACKNOWLEDGEMENTS

I would like to sincerely thank my family, friends, research team, sponsors, and fiancée Sydney for their support over the duration of this project. My family and friends have shown unwavering support throughout my academic studies. My advisor Dr. William Schneider IV has provided pivotal direction and guidance on this project and many others during my college career. The members of the research team - Curtis Deibel, Brandon Stakleff, Maggie Pekersoy, Will Holik, and Alex Maistros - have each provided valuable insight and suggestions to this study for which I am very grateful. I also would like to thank the members of the thesis committee for this study as well for their valuable comments and suggestions.

Additionally, I would like to thank the United States Department of Transportation (USDOT), National Highway Transportation Safety and Administration (NHTSA) and Ohio Department of Public Safety’s Office of Criminal Justice Services-Traffic Section (OCJS-TS) for their generous support, specifically Mr. Felice Moretti and Mr. Robert Painter of OCJS-TS. I would also like to thank the Ohio Department of Transportation (ODOT) for the data provided for use in this study. I am also grateful for insight given from the riding perspective by Mr. Imre Szauter of the American Motorcyclist Association (AMA).
# TABLE OF CONTENTS

LIST OF TABLES ..................................................................................................................................... iv

LIST OF FIGURES .................................................................................................................................... iv

LIST OF ABBREVIATIONS ..................................................................................................................... v

CHAPTER

I. INTRODUCTION ................................................................................................................................... 1

1.2 Objectives ......................................................................................................................................... 2

1.3 Overview of Thesis ........................................................................................................................... 2

1.3.1 Chapter II: Background Information ......................................................................................... 2

1.3.4 Chapter III: Random Effects Models .......................................................................................... 3

1.3.5 Chapter IV: A Mixed Effects Model at the Ohio Township Level ............................................ 3

1.3.6 Chapter V: A Mixed Effects Model of State-Maintained Roadway Segments in Ohio.......... 3

1.3.7 Chapter VI: Summary and Comparison of Regional and Segment Level Models ............... 4

II. BACKGROUND INFORMATION ....................................................................................................... 5

2.1 Data Sources in this Study ................................................................................................................ 5

2.1.1 OH-1 Reports ............................................................................................................................. 5

2.1.2 Map Data ................................................................................................................................... 7

2.1.3 Derivation of New Fields ........................................................................................................... 8

2.1.4 Quality Assurance and Control ................................................................................................ 10

2.2 Poisson Models and Negative Binomial Models .............................................................................. 11

2.2.1 Characteristics and Assumptions of Basic Models ................................................................. 11

2.2.2 Foundations of Bayesian Analysis ........................................................................................... 13

2.2.3 Poisson Family Models ........................................................................................................... 15

2.3.1 Markov Chain Monte Carlo Simulation .................................................................................. 19
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>NLFID example for State Route 45 in Trumbull County</td>
</tr>
<tr>
<td>2.2</td>
<td>Summary of data used in ArcGIS</td>
</tr>
<tr>
<td>3.1</td>
<td>An explanation of variable definitions for Equations 3.1, 3.2, and 3.3</td>
</tr>
<tr>
<td>4.1</td>
<td>Injury severity of single vehicle Ohio motorcycle crashes based on time and behavior</td>
</tr>
<tr>
<td>4.2</td>
<td>Injury severity of single vehicle Ohio motorcycle crashes based on most harmful event</td>
</tr>
<tr>
<td>4.3</td>
<td>Descriptive statistics of Ohio motorcycle crashes</td>
</tr>
<tr>
<td>4.4</td>
<td>Goodness-of-fit characteristics for mixed effects models with varying spatial random effects specifications</td>
</tr>
<tr>
<td>4.5</td>
<td>Parameter estimates and results from mixed effects models</td>
</tr>
<tr>
<td>5.1</td>
<td>A summary of the descriptive statistics of covariates of potential predictors for multi-vehicle motorcycle crashes in Ohio at the segment level</td>
</tr>
<tr>
<td>5.2</td>
<td>A comparison of model goodness-of-fit for different neighborhood radii through the DIC</td>
</tr>
<tr>
<td>5.3</td>
<td>A summary of the results of the segment level multi-vehicle motorcycle crash model where the neighborhood radius equals 1.75 miles</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>A summary of the data available for motor vehicle crashes in the TRACTapes</td>
<td>6</td>
</tr>
<tr>
<td>3.1</td>
<td>A map of the township and county boundaries of Ohio</td>
<td>25</td>
</tr>
<tr>
<td>4.1</td>
<td>Distribution of injury severity in Ohio single-vehicle motorcycle crashes</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>based on crash type</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>Spatial distribution of Ohio single-vehicle motorcycle crashes</td>
<td>34</td>
</tr>
<tr>
<td>4.3</td>
<td>Difference in expected number of single vehicle motorcycle crashes in</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Ohio townships when township level spatial random effects are specified</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>The change in DIC with neighborhood radius in miles. The horizontal line</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>represents the value of DIC when spatial correlation is not considered. The</td>
<td></td>
</tr>
<tr>
<td></td>
<td>horizontal line signifies a significant difference in DIC.</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>AADT</td>
<td>Annual Average Daily Traffic (in vehicles per day)</td>
<td></td>
</tr>
<tr>
<td>ADT</td>
<td>Average Daily Traffic (in vehicles per day)</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion</td>
<td></td>
</tr>
<tr>
<td>BUGS</td>
<td>Bayesian Analysis Using Gibb’s Sampling</td>
<td></td>
</tr>
<tr>
<td>CAR</td>
<td>Conditional Autoregressive</td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td>Deviance Information Criterion</td>
<td></td>
</tr>
<tr>
<td>FARS</td>
<td>Fatality Analysis and Reporting System</td>
<td></td>
</tr>
<tr>
<td>GIS</td>
<td>Global Information System</td>
<td></td>
</tr>
<tr>
<td>iid</td>
<td>Independent and Identically Distributed</td>
<td></td>
</tr>
<tr>
<td>LM</td>
<td>Lane Miles</td>
<td></td>
</tr>
<tr>
<td>MCMC</td>
<td>Markov chain Monte Carlo</td>
<td></td>
</tr>
<tr>
<td>NHTSA</td>
<td>National Highway Traffic Safety Administration</td>
<td></td>
</tr>
<tr>
<td>NLFID</td>
<td>Network Linear Feature Identifier</td>
<td></td>
</tr>
<tr>
<td>ODOT</td>
<td>Ohio Department of Transportation</td>
<td></td>
</tr>
<tr>
<td>ODPS</td>
<td>Ohio Department of Public Safety</td>
<td></td>
</tr>
<tr>
<td>pdf</td>
<td>Probability Distribution Function</td>
<td></td>
</tr>
<tr>
<td>PUCO</td>
<td>Public Utilities Commission of Ohio</td>
<td></td>
</tr>
<tr>
<td>RD</td>
<td>Road Density</td>
<td></td>
</tr>
<tr>
<td>RM</td>
<td>Road Miles</td>
<td></td>
</tr>
<tr>
<td>SAS</td>
<td>Statistical Analysis Software</td>
<td></td>
</tr>
<tr>
<td>SLM</td>
<td>Straight Line Mileage</td>
<td></td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>State Route</td>
<td></td>
</tr>
<tr>
<td>US</td>
<td>United States</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Motorcycle crashes are known to compose a disproportionately high amount of the overall vehicle fatalities in the United States. Fatalities on motorcycles represented 4,462 deaths of the 30,797 total fatal crashes in the United States in 2009 (FARS, 2012). In 2009, when considering the number of fatalities per 100 million VMT (Vehicle Miles Travelled), the motorcycle fatality rate was over 18 times that of passenger car crashes, which was the second highest crash rate for a specific vehicle class (FARS, 2012).

Nationally, 10.1% of the vehicle crash fatalities were motorcycle crashes in 2009 (FARS, 2012). Despite the limited riding season in Ohio due to the weather, the amount of motorcycle involvement in fatalities in the same year was even higher at 16.3%, or 166 fatalities (FARS, 2012). In more recent years, the number of fatal crashes has decreased slightly due to a proactive approach, with 164 fatal crashes in 2010. At the same time, the number of motorcycle crashes in Ohio increased from 4,165 in 2009 to 4,381 in 2010 (ODPS, 2012).

A total of 13,914 multi-vehicle motorcycle crashes occurred in the state of Ohio between 2006 and the summer of 2011. Of these crashes, 72% resulted in injuries or fatalities to at least one individual (ODPS, 2011). In Ohio, 453 fatalities in 2009 were occupants of cars; in comparison, 166 fatalities occurred on motorcycles in the same year, despite the fact that only 3.3% of registered vehicles in Ohio are motorcycles (FARS, 2012; ODPS, 2011). Although motorcycles represent a relatively small portion of Ohio traffic, they compose a disproportionately high amount of injury and fatal crashes. Therefore, it is important to develop increasingly effective methods of estimating the factors involved in multi-vehicle motorcycle crashes.
1.2 Objectives

This research aims to create a tool that may be used to reduce the frequency of motorcycle crashes in Ohio. The objectives of this study may be summarized in the following points:

- **Review Ohio Crash Data:**
  This includes checks for quality and consistency. The data must be cleaned and reviewed for any mistakes that may have been made (such as motorcyclists “wearing seatbelts”), from the recording of the accident at the scene to the transcription into the digital database where the data are available for use. Additionally, registration data, coordinates of Ohio county and township borders, and crash locations were reviewed.

- **Evaluate the Effectiveness of Mixed Effects Models in Modeling Ohio Motorcycle Crash Data:**
  In this thesis, several types of models are investigated for the crash data. Then, they are compared to find which best represents the known data, and thus which will provide the most reliable prediction of crash locations and hot spots. Checks for goodness-of-fit and statistical reliability will be conducted on each of the models.

- **Provide Recommendations Based on the Findings:**
  With the results of the models known, recommendations and suggestions for changes in policy, practice, or engineering can be made to improve motorcycle safety in Ohio.

1.3 Overview of Thesis

The following subsections briefly describe the content of each chapter of this study. The goals, methods, and outcome of each section are summarized below.

1.3.1 Chapter II: Background Information

This section provides the details of each data source used in this study. The major data sources involved in the creation of this report are Ohio crash data, Ohio map data, and RI-15 Curve Data. The specific use of
each dataset in deriving additional quantities is explained, as well as which quantities were then used to
model Ohio motorcycle crashes. The foundational elements that appear in all the models in this study are
included in this section. These topics that are discussed are regression models, Poisson models, negative
binomial models, Markov chain Monte Carlo models, Gibbs Sampling, and goodness-of-fit measures.

1.3.4 Chapter III: Random Effects Models

Two broad types of random effects terms are discussed in this section: correlated random effects and
uncorrelated random effects. One focus of this study is the appropriateness, strengths, and weaknesses of
various specifications of spatially correlated random effects terms. The definitions and methodologies for
each of the terms included in this study are outlined in this section. Additionally, the general purpose for
uncorrelated random effects terms are discussed.

1.3.5 Chapter IV: A Mixed Effects Model at the Ohio Township Level

This chapter describes a group of models analyzing single vehicle motorcycle crashes in Ohio at the
township level. An analysis of the effectiveness of different definitions and weighting schemes to define
the neighborhood of each region is performed. The models are compared and contrasted in terms of ease of
interpretation and goodness-of-fit. Benefits to modeling spatial data with a regional approach are also
discussed in this chapter.

1.3.6 Chapter V: A Mixed Effects Model of State-Maintained Roadway Segments in Ohio

The models in this chapter show the effects of varying the radius at which neighborship is established
between state-maintained roadway segments on the performance of mixed effects models on multi-vehicle
motorcycle crashes in Ohio. Benefits for modeling crashes at the roadway segment level are discussed, and
the goodness-of-fit measures and interpretability are reviewed for each model. The study finds that spatial
correlation may effectively be used to reduce the error in single vehicle motorcycle crash models at the
township level.
1.3.7 Chapter VI: Summary and Comparison of Regional and Segment Level Models

This chapter summarizes the results of the regional and segment level models in this study. The advantages and disadvantages of each approach are outlined, as well as the impact of model choice on the interpretation for single and multi-vehicle motorcycle crashes. Finally, recommendations and a summary of Bayesian modeling for motorcycle crashes are provided.
CHAPTER II
BACKGROUND INFORMATION

2.1 Data Sources in this Study

Details for each Ohio crash are available through a number of government organizations. Each organization is responsible for the keeping and dispersal of a certain quantity and level of detail of records. In order to create the best possible description of motorcycle crashes that occurred in the state of Ohio between 2006 and 2011, multiple datasets were acquired, checked for quality and feasibility, and combined. The fusion of the datasets created the final dataset used in this study.

2.1.1 OH-1 Reports

The genesis of the description of a crash occurs at the crash site itself. After a crash occurs and is reported, a police officer must come to the scene of the crash and record the details. This information is recorded on a form called the OH-1 Report. Every crash that is reported has a corresponding OH-1 Report, which chronicles many characteristics of the crash, as seen in Figure 2.1 on the following page. After the crash reports are collected, they are compiled and checked for inconstancies and errors. After this process, the crash data becomes available for distribution. It is important to note that although police officers do acquire personal information from the parties involved, such as name, address, and insurance information, all data available publicly and used in this study was entirely anonymous and did not contain any information to link the crash to the parties involved. The data is intended only for statistical and research purposes, for which it was used in this study. One of the most useful fields in the crash database is called the Network Linear Feature Identifier (NLFID). The NLFID is a string fourteen characters that uniquely identifies a roadway, in this case the roadway on which the crash occurred (ODPS, 2011; ODOT, 2011). The NLFID provides a quick, concise summary of a roadway. As an example, consider the following
Figure 2.1: A summary of the data available for motor vehicle crashes in the TRACTapes (ODPS, 2011)
NLFID: STRUSR00045**C. This NLFID represents State Route 45 in Trumbull County. The NLFID can be broken into components as shown:

Table 2.1: NLFID example for State Route 45 in Trumbull County

<table>
<thead>
<tr>
<th>Characters</th>
<th>Segment</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>S</td>
<td>State Maintained</td>
</tr>
<tr>
<td>2-4</td>
<td>TRU</td>
<td>Trumbull County</td>
</tr>
<tr>
<td>5-6</td>
<td>SR</td>
<td>State Route</td>
</tr>
<tr>
<td>7-11</td>
<td>00045</td>
<td>Route 45</td>
</tr>
<tr>
<td>12</td>
<td>*</td>
<td>No Duplicates</td>
</tr>
<tr>
<td>13</td>
<td>*</td>
<td>No Specific Characteristics</td>
</tr>
<tr>
<td>14</td>
<td>C</td>
<td>Cardinal Direction</td>
</tr>
</tbody>
</table>

Therefore, this NLFID represents State Route 45 in Trumbull County.

2.1.2 Map Data

The Ohio crash data contains the latitude and longitude of each recorded crash. The Ohio Department of Transportation (ODOT) has roadway and county data available online, in the form of shape and layer files (ODOT, 2011). These map files complement the latitude and longitude coordinates from the crash data by visually displaying the roads and counties in which the crashes occurred. Although the crash data does include the county and roadway that a crash occurred on, plotting the crashes on a map of Ohio allows basic observations of clusters and trends. The maps also create opportunities for quality analysis and control that will be described in detail later in this chapter. Table 2.2 contains a summary of the information used to create the maps, datasets, and models in this study.

Any of the available categories of information can be used to select subsets of data for more refined analysis. For example, the dataset can be narrowed down into only crashes that occurred in 2009 by use of the YEAR column in the crash data, or it could be narrowed down into only major collector roads through the use of the ODOT map data’s Functional Class criterion.
Table 2.2: Summary of data used in ArcGIS

<table>
<thead>
<tr>
<th>Layer</th>
<th>Notable Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counties</td>
<td>County boundaries, population, elevation, extreme latitude and longitude, land area, abbreviation</td>
</tr>
<tr>
<td>Cities</td>
<td>City boundaries, population, land area</td>
</tr>
<tr>
<td>Interstate Highways</td>
<td>Highway location, NLFID, log points, segment lengths, surface, shoulder, and roadway width, number of lanes, functional class, speed limit, truck and car ADT</td>
</tr>
<tr>
<td>US Routes</td>
<td>Highway location, NLFID, log points, segment lengths, surface, shoulder, and roadway width, number of lanes, functional class, speed limit, truck and car ADT</td>
</tr>
<tr>
<td>State Routes</td>
<td>Highway location, NLFID, log points, segment lengths, surface, shoulder, and roadway width, number of lanes, functional class, speed limit, truck and car ADT</td>
</tr>
<tr>
<td>County Roads</td>
<td>Highway location, NLFID, log points, segment lengths*</td>
</tr>
<tr>
<td>Municipal Roads</td>
<td>Highway location, NLFID, log points, segment lengths*</td>
</tr>
<tr>
<td>Township Roads</td>
<td>Highway location, NLFID, log points, segment lengths*</td>
</tr>
<tr>
<td>Ohio Crash Data</td>
<td>Latitude and longitude of crashes and all information from Table 2.1</td>
</tr>
</tbody>
</table>

*Derived from log points

2.1.3 Derivation of New Fields

In order to create models that best describe the observed crash locations, the data had to be organized and rearranged to create the most useful, versatile dataset possible. At times, the most concise description of a roadway or crash is acquired through a modification or combination of the available datasets. In all cases, the procedures to obtain the new fields were consistent over all samples and did not introduce any new sources of bias or error. The procedures only served to create simpler and more descriptive data fields.

In the Ohio crash data, the position of a crash along an indicated road is defined by the SLM (Straight Line Mileage) Log, which is the straight line distance from the beginning of the roadway to the
crash location (ODPS, 2011). This allows a simple way of marking a specific location along the roadway indicated by the NLFID. In the case of the ODOT map data, each NLFID is divided into segments through the use of log points. Interstate, US, and state routes contain a log point that mark the SLM at the beginning of each segment, another log point that shows the SLM at the end of the segment, and the length of the segment. Therefore, the beginning log point of each segment matches the terminal log point of the previous segment. The county, municipal, and township roads have the same log points as the other routes, but do not have the segment length (ODOT, 2011). However, the segment length for these roads can be derived by the following formula

\[ l_h = SLM_{h+1} - SLM_h \]  

(2.1)

Where \( l_h \) represents the length of segment \( h \), and \( SLM \) denotes the straight line mileage given by the log points. Through this calculation, which was performed via SQL (Structured Query Language) in ArcGIS, all segment lengths were acquired. The sum of the segment lengths is equal to the number of road miles (RM) in a region. In this case, the region \( j \) is the county. This relationship is illustrated as below:

\[ RM_j = \sum_{i=1}^{n} l_{ij} \]  

(2.2)

The number of lanes miles (LM) is often used in lieu of road miles. Since the number of lanes for each segment was given in ArcGIS, the lane miles were calculated as shown:

\[ L_i = l_i N_i \]  

(2.3)

Where \( L_i \) denotes the quantity of lane miles and \( N_i \) represents the number of lanes in both directions in segment \( i \). Again using county level regions, the sum of the lane miles is calculated as shown here:

\[ LM_j = \sum_{i=1}^{n} L_{ij} = \sum_{i=1}^{n} (l_{ij} N_{ij}) \]  

(2.4)

The number of lane miles is typically a more descriptive predictor of crashes than road miles since accounting for the number of lanes gives weight to regions that have many high traffic roads, especially small regions that may have fewer miles of road, but a larger percentage of densely traveled routes.
To adjust further for the variation of size in counties, since more densely populated regions tend to have smaller counties than less populated regions for administrative purposes, the area of each county can be considered. This derived predictor is called road density (RD) and is a ratio of the lane mileage and area of each region, as illustrated here:

\[ RD_j = \frac{LM_j}{A_j} \]  

(2.5)

where \( A_j \) represents the area of each region in square miles. The quantity \( RD_j \) then represents the road density of region \( j \) in lane miles of roadway per square mile. Road density is used for convenience in lieu of using both area and lane miles as separate predictors for each region.

2.1.4 Quality Assurance and Control

To complete this study, the distinct data sources listed previously were combined into one more descriptive database. The end goal of this action was to create a database that displayed as much information as possible for each crash. This information included information on the county, township, and roadway level for each crash. Similarly, the database could be used to visualize information on the county and township levels and the details of each crash that occurred on each roadway. To accomplish this, missing information had to be analyzed and either filled through the use of the other data sources or the observation with missing or incorrect information was removed.

In a number of cases, the same information will appear in multiple data sources. For example, the passenger vehicle and overall ADT was found in the Ohio crash data and Ohio map data. Each time this occurred, it had two distinct impacts on this study. First, it allowed for the comparison of the data between the multiple sources. The numbers could be compared with each other and verified. Any sizeable difference between the data sources would then demand an explanation, as well as basis for the selection of a singular source to consider in the models used in this study. However, in the comparison of the data sets, the only instance in which the data sets appeared to disagree was ADT. After a brief inquiry into the source of the difference, it became clear that the ADT had been measured in different years. The most up-to-date source, the ODOT Map Files, was also the most complete (ODOT, 2011; ODPS, 2011).
Using WinBUGS, missing data can be assigned a distribution in a similar fashion to the parameters of the model. If no comment can be made as to the likelihood of one value over another, a non-informative prior can be assigned. However, a maximum entropy prior can often be assigned by eliminating impossible values. In the models outlined in this study, only state-maintained routes were considered at the segment level, while all routes were considered in the regional cases. All information for the predictors was known in these analyses, and this method for missing data was not necessary.

2.2 Poisson Models and Negative Binomial Models

All the models in this study were based on negative binomial models. Variations were made within the models to find the best fit for the Ohio motorcycle crashes. These variations included different types of spatial random effects, the use of an uncorrelated heterogeneous error term, and datasets that contain different types of crashes. These models will be compared and contrasted using various goodness-of-fit measures in later sections. The aim of the following section is to describe the characteristics and assumptions that are uniform over all models in this study. The details and impacts of random effects terms can be found in Chapter V.

2.2.1 Characteristics and Assumptions of Basic Models

One of the most basic ways to describe or quantify one or more trends in a data set is through the use of a general regression model. The simplest form of the regression model is a linear regression model with a single predictor. It should be noted that traditional and Bayesian inference into a data set will produce an identical linear regression model (Koch, 2007). Such a model simply states the effect of one variable specified as independent on another, which is specified as dependent. For example, one may wish to model the effect of county population on the number of crashes per county. A model to describe the aforementioned effect would appear as follows:

\[ y_i = \beta_0 + \beta_1 x_i + \epsilon_i \]  \hspace{1cm} (2.6)

where \( y_i \) is the number of crashes per county (dependent variable), \( \beta_0 \) is the y-intercept, \( i \) is the index of counties, \( x_i \) is the county population (independent variable), \( \beta_1 \) is the effect of a unit increase in population
on the number of crashes per county, and $\varepsilon_i$ is the unexplained model error (Kutner et al., 2005). Logically, reducing the amount of model error typically results in a more useful model. Thus, a great deal of effort can be put into varying model specification to reduce the error. One of the most common explanations for error in a model is that there are other, unknown factors that also influence the behavior of the dependent variable. To reduce this source of error, additional predictors are added to the model (Kutner et al., 2005).

In the case of crashes per county, perhaps the researcher also wants to study the effect of the number of lane miles and the size of the county, as well as its population. The revised linear regression model would appear as follows:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$$ \hspace{1cm} (2.7)

where the original components of the model have not changed, but $\beta_2 x_{2i}$ represents the effect of the number of lane miles on crashes per county, and $\beta_3 x_{3i}$ represents the effect of the size of the county. This can be extended for any number of predictors (Kutner et al., 2005).

In some cases, the researcher may suspect that the effect of a predictor is not linear, but rather exponential or logarithmic. In that case, the terms can be specified as such and models with linear terms and the adjusted terms can be compared for goodness-of-fit. Sometimes, however, it is not the addition of predictors but the removal of them that leads to a better fit. Despite the wealth of information that may exist in regards to a particular subject, not every category may be an significant predictor. Conversely, some predictors may be correlated to another, in which case, the model parameters may be suspect. An interaction term may be considered to reduce this effect. The interaction term consists of the product of two or more predictors and a new coefficient. For example, an interaction term between $x_{1i}$ and $x_{2i}$ would appear as $\beta_{21} x_{2i} x_{1i}$ (Kutner et al., 2005).

This procedure is completed for both basic and more advanced models. Since random effects terms serve to reduce the amount of unexplained heterogeneity, the model building process is facilitated by compiling a model with the best fitting fixed effects terms prior to the addition of random effects terms. This stands true for both basic models, such as regression models, and the more advanced Bayesian negative binomial models appearing in this study. SPSS V.19 was employed for this purpose. Before final
model was compiled in WinBUGS V.1.4.3, all the potential fixed effects terms were considered and investigated through the negative binomial modeling ability of SPSS.

2.2.2 Foundations of Bayesian Analysis

In the specification and analysis of a model, two types of inference can be applied. The first, traditional statistics views all unknown parameters as inflexible and does not provide an avenue to make use of prior knowledge in their estimation. The other type, Bayesian inference, focuses on infusing prior knowledge into models through the use of probability theory (Koch, 2007). In Bayesian analysis, parameters are used to estimate the posterior distribution of a quantity. When specifying the distribution of a model, the user is simultaneously specifying the posterior distribution (Gelman et al., 2004; Koch, 2007). Thus, a negative binomial model leads to a negative binomial posterior distribution, and likewise a Poisson model leads to a Poisson posterior distribution. When a parameter is unknown, the user can specify a prior distribution for the parameter. Depending on the user’s knowledge of this parameter, different levels of confidence can be shown through the prior distribution (Gelman et al., 2004; Koch, 2007). The type of prior distribution that illustrates the least amount of certainty is called a noninformative prior. A noninformative prior simply states that the value of an unknown parameter is equally likely to be any real number between negative infinity and infinity (Gelman et al., 2004; Koch, 2007).

When only a small amount of information can be contributed to the model about a parameter, the statement is known as a maximum entropy prior. In practice, when little is known about a parameter one may specify a uniform distribution, as shown:

\[ \omega \sim \text{Uniform}(a, b) \]  

(2.8)

where \( \omega \) is the unknown parameter and \( a \) and \( b \) are the upper and lower limits, respectively (Gelman et al., 2004; Koch, 2007). The parameters of a prior distribution, such as the parameters \( a \) and \( b \), are known as hyperparameters. For a maximum entropy, or diffuse, prior, the upper and lower limits would be very far apart, or relaxed (Koch, 2007). For example, if \( \omega \) were the number of miles between randomly occurring events, and all that was known is that the events were all in the continental United States, \( a \) could be
specified as near 0 miles, and \(b\) as 4,000 miles. In simpler terms, this states that the parameter could be the distance between any two points in the continental United States, since no two points in the country are more than 3,500 miles apart. It does however restrict the location of an event from being in China, which would violate the upper bound.

When a significant amount of information is known about a parameter, an informative prior can be specified. An informative prior is a specific, confident statement about the distribution of a parameter. An informative prior may be used if the samples contain sufficient data for the estimation of the hyperparameters of a distribution. For mathematical convenience, when the data is used to estimate an informative prior distribution, a conjugate distribution of the posterior is chosen for the prior, if it is possible. That is, the prior distribution should share the same functional form as the likelihood function if possible (Gelman et al., 2004; Koch, 2007). The likelihood function is defined as the function that describes the distribution of probabilities of an unknown parameter. This differs from the concept of a prior distribution, as the prior distribution refers to the confidence of the hyperparameters and the distribution of the previously known or estimate data, and the likelihood function refers to the probability distribution function (pdf) of the posterior distribution (Congdon, 2003; Gelman et al., 2004; Koch, 2007).

The posterior probability, prior probability, and likelihood can be related by a single function. This relationship, which is shown below in Eq. 2.9, is called Bayes Theorem:

\[
p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \tag{2.9}
\]

In this formula, \(p(\theta|y)\) denotes the posterior probability, \(p(y|\theta)\) the prior probability, and \(p(\theta)\) the likelihood (Koch, 2007). The term \(p(y)\) represents a normalizing constant (Gelman et al., 2004; Koch, 2007). This formula is an illustration of how Bayesian inference provides a medium to allow prior knowledge to influence the final results, the posterior distribution, in a way that traditional statistics does not (Gelman et al., 2004; Kéry, 2010; Koch, 2007). Each parameter may be assigned a distribution which allows the researcher to include prior knowledge in the model.
2.2.3 Poisson Family Models

The Poisson distribution is used in statistics to represent count data. Therefore, when choosing a
distribution to represent the data at hand, the Poisson distribution is only appropriate if the data is discrete
(Congdon, 2003; Gelman et al., 2004). However, crash data is discrete because fractional crashes cannot
occur. As such, the Poisson distribution has often been employed when modeling crash data. The
distribution function for a single Poisson distributed data point $y$ given an expected frequency $\theta$ is as
follows:

$$p(y|\theta) = \frac{\theta^y e^{-\theta}}{y!} \quad (2.10)$$

where $p(y|\theta)$ is the probability of the occurrence of $y$ (Gelman et al., 2004). Over the vector
$\mathbf{y} = (y_1, y_2, ..., y_n)$ where $\mathbf{y}$ is a vector of iid observations, the likelihood function of the Poisson
distribution is:

$$p(y|\theta) = \prod_{i=1}^{n} \frac{1}{y_i} \theta^{y_i} e^{-\theta} \propto \theta^{t(y)} e^{-n\theta} \quad (2.11)$$

where $i$ is the index of observations, $n$ is the number of observations, and $t(y)$ is the sufficient statistic
(Gelman et al., 2004):

$$t(y) = \sum_{i=1}^{n} y_i \quad (2.12)$$

The likelihood function, which is one of the components of Bayes’ Theorem, is used to modify prior beliefs
through knowledge obtained from the posterior density (Congdon, 2003).

In addition to the assumption that the dataset is composed of discrete values only, one also enters
into another critical assumption when employing the Poisson distribution. The second criterion of the
Poisson distribution is that the mean must be equal to the variance. A model that uses the Poisson
distribution and also uses data for which this is not the case is called overdispersed. This means that there is
more variation in the data than a simple model will be able to explain (Congdon, 2003; Lawson, 2009).
Overdispersion, much like spatial correlation, is unlikely to be fully accounted for by only including fixed effects terms (Lawson, 2009).

The Vuong statistic can be employed to test whether a model is overdispersed (Shankar et al., 2003; Vuong, 1989). The Vuong statistic tests the appropriateness of using a Poisson model, zero inflated Poisson model, or a negative binomial model. The Vuong statistic is measured as follows:

\[
\sqrt{n}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_{i}\right]
\]

\[
\frac{\sum_{i=1}^{n} (m_{i} - \bar{m})^2}{\sqrt{n} \sum_{i=1}^{n} m_{i}}
\]

where \(\nu\) is the Vuong statistic, \(n\) is the number of observations, and \(f_{1}(y_{i} | x_{i})\) is the probability of predicting that the number of motorcycle crashes is equal to 1 (zero inflated Poisson model) or 2 (negative binomial model) (Shankar et al., 2003). The Vuong statistic is considered inconclusive between -1.96 and 1.96. Values less than -1.96 indicate that the Poisson model is the correct choice, while values greater than 1.96 indicate that the ZIP model is preferred. Note that ±1.96 is chosen as is the statistic applied for a 95% confidence interval in a t-test (Kutner et al., 2005; Shankar et al., 2003).

A zero inflated Poisson model can be applied in cases where a preponderance of zeros exists in the data. There are two distinct situations that can cause a data point to be recorded as zero. First, the location may have not had an instance of the attribute being considered. Perhaps an intersection or a segment did not have any crashes on it in the given time frame. In a different time frame, the location could possibly have crashes that occur in it. Second, the region or point in question may be entirely free of the attribute being examined. For example, in the case of motorcycle crashes, the segment, point, or region would be considered completely safe. In other words, there is no chance of a crash ever occurring at the given location. Such models have been applied to vehicle crash data in the past with increased measures of fit (Lee and Mannering, 2002; Shankar et al., 2003). However, other studies, especially Lord et al. (2005) and Lord et al. (2007) have demonstrated some conceptual issues with the application of zero inflated Poisson models. For example, the concept of claiming any length of roadway could not possibly have a crash seems
faulty. Even if none of the factors related to the roadway itself would cause an increased likelihood of crashes, vehicle and operator related issues could cause a crash on any segment. Despite better results in measures of fit in a zero inflated Poisson model, a negative binomial model may be a more robust choice to model motorcycle crash data (Lord et al., 2005; Lord et al., 2007).

The negative binomial distribution is used to simulate the number of successes \( r \) of an experiment or action that occur in a number of repetitions \( x \). The probability distribution function of the negative binomial distribution is shown here:

\[
P(x) = \binom{x - 1}{r - 1} p^r (1 - p)^{x-r} \tag{2.15}
\]

where \( P \) is the cumulative probability and \( p \) is the probability of success in each iteration (Congdon, 2003). One of the key differences between negative binomial and Poisson models is that in negative binomial models, the mean is not restricted to being equal to the variance. Instead, the following system of equations describes the mean and variance of a negative binomially distributed quantity:

\[
E(y_i) = \frac{r}{p} \tag{2.16}
\]

\[
var(y_i) = \frac{r}{p^2} (p + 1) \tag{2.17}
\]

where \( E(y_i) \) is the mean, or expected value, of \( y_i \), and \( var(y_i) \) is the variance of \( y_i \) (Gelman et al., 2004). The specification of the mean and variance in this manner allows greater flexibility when creating models. The removal of the \( E(y_i) = var(y_i) \) restriction allows the user to create a more statistically robust model than simply using a standard Poisson model.

Given these relationships, the negative binomial distribution in model form appears as follows:

\[
\ln \lambda_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon_i \tag{2.18}
\]

where \( \lambda_i \) denotes the value of the Poisson parameter in region \( i \), \( \beta_{1,2,...,n} \) represents the effect of a unit increase of the corresponding \( x_{1,2,...,n} \), or the value of the predicting quantity, and \( \epsilon_i \) is the error associated in
the estimation of \( y_i \) (Lawson, 2009). The error term in Eq. 2.18 may be assigned a gamma distribution, with a mean equal to 1 and a variance denoted by \( \sigma^2 \), or as shown here:

\[
\varepsilon_i \sim \text{gamma}(1, \sigma^2)
\]  

(2.19)

The variance of the error term is related to the expected value of \( y_i \) through the following relationship:

\[
\text{var}(y^2) = E(y_i) + \alpha E(y_i)^2
\]  

(2.20)

where \( \alpha \) is the overdispersion parameter, and also the square root of the variance, or standard deviation, of \( \exp(\varepsilon_i) \) (Gelman et al., 2004; Oh et al., 2006). Specifying this relationship in a negative binomial model avoids the pitfalls of the Poisson distribution by quantifying the overdispersion term.

To correct for exposure, the expected number of occurrences term of the Poisson distribution is written as below:

\[
\lambda = \theta_i e_i
\]  

(2.21)

where \( \theta_i \) is the number of occurrences in the specified time interval, which is adjusted by \( e_i \), which is a term that factors the exposure method into the equation (Lawson, 2009). With this adjustment, the Poisson distribution parameter appears as below:

\[
\ln \theta_i = \beta_0 x_i + \beta_0 + \varepsilon_i
\]  

(2.22)

where the right hand side is the familiar form of the general regression model, and the left hand side is the natural logarithm of \( \theta_i \), the expected number of crashes in a specified time interval (Gelman et al., 2004; Lawson, 2009). This combination of exposure correction and the negative binomial model are used together in this study to model motorcycle crash data.

2.3 Bayesian Modeling Methodology

Bayesian modeling has become an increasingly popular modeling option in transportation safety research. The following section describes some specific components of the analysis that make Bayesian modeling an effective option.
2.3.1 Markov Chain Monte Carlo Simulation

One of the reasons for the relatively recent rise in popularity of Bayesian methods is the ability to apply computer simulation methods to complex problems. Previously, Bayesian analysis consumed a great deal of time, since it frequently requires solving very complex equations, often with time consuming iterations (Congdon, 2003). One of the most popular computer methods used to this end is Markov chain Monte Carlo (MCMC) simulation, which is a method used by WinBUGS. Given a prior distribution for each unknown parameter, the user then chooses the number of chains to be used in the estimation of the parameters. Each chain is made up of an initial value for each unknown parameter that is specified in the model by the user. The MCMC simulation of each individual chain is run separately. Starting at the specified initial value, the true distribution of each parameter is estimated and improved with each iteration as the posterior distribution is sampled. The value of the subsequent iteration is based entirely on the value of the previous iteration (Congdon, 2003; Gelman et al., 2004).

However, care must be taken that the iterations are drawing nearer to the true value of each parameter, without knowing what the actual value is. This challenge is frequently overcome through the use of multiple chains. Each chain has a different value specified for each parameter. Typically, one tries to specify at least one value that is both greater than and less than the expected value of the parameter. By running multiple chains independently, the trend in each chain can be compared with the others. One common way of assessing convergence using this comparison is through the Gelman-Rubin statistic (Congdon, 2003). This statistic measures the within-chain variance $W$ and the between-chain variance $B$ as the samples $\psi_{ij}$ are drawn, and then compares them as follows:

$$\bar{\psi}_j = \frac{1}{n} \sum_{i=1}^{n} \psi_{ij} \quad (2.23)$$

$$\bar{\psi}_* = \frac{1}{m} \sum_{j=1}^{m} \psi_j \quad (2.24)$$

$$s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\psi_{ij} - \bar{\psi}_j)^2 \quad (2.25)$$
\[ B = \frac{n}{m-1} \sum_{j=1}^{m} (\tilde{\psi}_j - \bar{\psi})^2 \]  
\[ W = \frac{1}{m} \sum_{j=1}^{m} s_j^2 \]  
\[ GR = \frac{n-1}{n} W + \frac{1}{n} B \]

where \( n \) denote the number of iterations, \( i \) is the index of the iterations, \( m \) is the number of chains, \( j \) is the index of the chains, and \( s_j^2 \) is the variance of each chain (Congdon, 2003; Gelman et al., 2004). From these factors, the Gelman-Rubin statistic is calculated for each parameter as a weighted average of the between-chain and within-chain variance:

\[ GR = \frac{n-1}{n} W + \frac{1}{n} B \]  

where \( GR \) is the Gelman-Rubin statistic (Congdon, 2003; Gelman et al., 2004). The value of this statistic is assessed for each parameter individually. Theoretically, the value of \( GR \) approaches 1 as the parameters converge. A typically accepted value to signify convergence is 1.1 or less (Gelman et al., 2004). The value of the iterations before the model converges should be discarded as burn-in values so that the final results of the model only include values from the converged model to estimate the posterior distribution (Gelman et al., 2004).

Gibbs Sampling is a specific MCMC algorithm. A Gibbs Sampler, such as WinBUGS, functions by dividing each parameter vector into subvectors. At the point at which each subvector has been estimated, one iteration has been completed. Gibbs Sampling uses the results of all other iterations in the estimation of each successive iteration. The conditional distributions of the parameters are continually sampled in this fashion for the specified number of iterations (Congdon, 2003; Gelman et al., 2004).

2.3.2 Goodness-of-Fit

After the model is updated using the MCMC Gibbs Sampling method, the results must be interpreted and compared. The differing magnitudes and orientations of the parameters offer a great deal of insight into the underlying causes of motorcycle crashes. However, without any additional measures of
goodness-of-fit, it is difficult to compare and contrast the models. Additionally, a path must be taken to
determine which model is ultimately the preferred approach for modeling motorcycle crash data.

One frequent quantity recorded in Bayesian modeling is the model deviance, which is calculated
as shown:

\[ D(y, \theta) = -2 \log p(y|\theta) \]  

(2.29)

where \( y \) is the univariate response of the model, given the simulations \( \theta \), and \( \log p(y|\theta) \) is the log-
likelihood (Gelman et al., 2004; Lawson, 2009). The deviance provides a measure of how much
discrepancy exists between the model and the data. As an improved measurement of discrepancy, the
estimated average deviance can be estimated as follows:

\[ \bar{D}_{avg}(y) = \frac{1}{L} \sum_{l=1}^{L} D(y, \theta_l) \]  

(2.30)

where \( \bar{D}_{avg}(y) \) denote the estimated average deviance, \( L \) is the number of simulations, and \( l \) is the index
over the simulations (Gelman et al., 2004; Lawson, 2009). The estimated average deviance is preferred as a
summary of model error since it incorporates all the entire range of possible parameter values. This
prevents the overestimation of model fit that can be displayed by the deviance at a point (Gelman et al.,
2004; Lawson, 2009).

However, a statistician’s sole goal should not be to simply find the model with the lowest
deviance. Care must be taken not to violate the assumptions of distributions or to fit an inappropriate model
to a dataset. For example, as stated earlier in this chapter, applying a zero inflated Poisson model may not
be appropriate for modeling crash data. In addition to creating a model with a better fit, model parsimony
should also be considered. Creating a model with an excessive number of parameters should be avoided,
even if it does offer a slight improvement in measures of fit (Lord et al., 2007). Needlessly complex models
consume more time to run, require an unnecessary amount of data, and most importantly, are not easily
interpreted or applied. The end goal of analyzing a dataset should be two-fold. First, a model should
summarize and explain trends in the dataset. Second, the model should be useful, have a clear application,
and information from it should be easily parlayed to nearly anyone. To this end, unnecessary parameters add clutter and decrease the effectiveness of the model’s message, which is, after all statistical analysis, the only universally understood function of a model (Lord et al., 2007). To quantify the effect of this, the effect number of parameters \( p_D \) can be found as follows:

\[
p_D = \bar{D}_{\text{avg}}(y) - D(y, \theta) \tag{2.31}
\]

The effect number of parameters increases as more parameters are added (Gelman et al., 2004; Lawson, 2009). For a linear model without parameter constraints \( p_D \) is equal to the count of the parameters in the model. Using the deviance and effective number of parameters, three interrelated Bayesian goodness-of-fit measures can be derived. The first is the Deviance Information Criterion (DIC), which is the default measure in WinBUGS. The DIC can be expressed as:

\[
\text{DIC} = D(\theta | y) + 2p_D \quad \text{Eq. 2.32}
\]

In addition to including the summary of the model’s deviance, the DIC also considers how many parameters were used in the specification of the model (Congdon, 2003; Gelman et al., 2004; Lawson, 2009). The model that provides the best fit to the dataset has a lower DIC. However, if too many parameters are applied, the effects of adding the parameters will outweigh a small decline in the overall model deviance (Gelman et al., 2004; Lawson, 2009). In this way, the DIC is not only a measurement of goodness-of-fit, but also model effectiveness.
CHAPTER III
RANDOM EFFECTS MODELS

3.1 Random Effects

A great deal can be learned from a fixed effects model. The intercept and other fixed effects parameters can describe the positive or negative effect of the predictors on the number of crashes in a region and shed light as to which safety improvements should be given priority. However, the addition of random effects terms into a model can highlight more specific trends in a dataset. One of the focuses of this study is to examine and analyze the effects of varying the types of fixed and random effects terms specified on the overall goodness-of-fit and interpretability of motorcycle crash models.

3.2 Applying Multi-level Information through Random Effects Terms

To describe events that occur in a large area, it is often not practical or desirable to sample the events in entirely uniform locations. Often, the events tend to be clustered spatially and temporally, be it for population, season, or another factor that dictates the occurrence. The more diverse the times and locations are, the greater chance that the events did not occur under the exact same conditions. For example, even events only one day apart are not at all unlikely to experience different weather conditions. When random, independent events occur, they can be grouped into spatial and temporal clusters. It is possible that each one of these clusters has characteristics that cause unique trends, or trends dissimilar to other clusters. To account for these various trends, a random effects term can be added to a model (Lawson, 2009). The random effects term can be used to account for spatial, temporal, or uncorrelated random effects in a number of different ways depending on the data used and distributions chosen (Congdon, 2003). Much of this study is devoted to the effect of varying the distribution of this parameter.
Motorcycle crash data can be described at multiple levels. For example, the most precise description would be provided on the basis of an individual crash. At this level, details about the rider, passenger, motorcycle, crash site, and crash location can be examined. Data at this level provides a powerful tool for small-scale analysis of crashes in that several crashes can be compared and contrasted manually. However, upon considering spatial correlation with the goal of predicting the number of crashes at a location, the results are often more easily interpreted when considering a region rather than individual crashes. For this reason, this study analyzes the effect of varying the resolution, or size of each region, on analyzing and interpreting Ohio motorcycle crash data. The levels of resolution as a whole are referred to as the hierarchical structure of the model.

The first and smallest region considered in this study is the roadway level. The roadway level is made up of segments of Ohio roads. By combining the data sources used in this study and using the NLFID’s and log points described in Chapter 2, the number of motorcycle crashes can be described through the use of roadway level predictors. Although the predictors will be outlined thoroughly in Chapter 5, a brief summary of the unique roadway level predictors is provided here. Analysis at the roadway level requires crashes to be assigned to a particular roadway segment. Since multiple crashes occur within many segments, it is not practical to examine the characteristics of individual riders, passengers, and vehicles. However, data unique to each roadway segment is available. Among other characteristics, this includes the lane width, shoulder width, and segment length.

The next level of the hierarchical structure is the township level. Ohio is divided into 1,459 townships (PUCO, 2011). The township boundaries are an ideal choice to use as regions due to the fairly uniform size and shape throughout most of the state, as seen in Figure 3.1. At this second level of analysis, individual roadway characteristics cannot be considered, since many roads and the crashes that occurred on them are summarized over the region. However, a combination of the first and second hierarchical levels allows the consideration of individual roadway characteristics as well as township level characteristics, such as township population or area. The third and largest hierarchical level considered in this study is the
Figure 3.1: A map of the township and county boundaries of Ohio

Ohio is divided into 88 counties (ODOT, 2011). Although not as uniform in size or shape as the townships, the counties are still a useful administrative boundary to apply to motorcycle crash data. Several townships are included in each county. Therefore, only county level statistics such as county population or number of lane miles per county can be considered at this level. As shown later in this study, a mix of all three levels provides information from the roadway, township, and county levels. The introduction of the second and third level random effects terms allows the identification and explanation of regional trends through data.

The system described above can be illustrated as follows, where variables are defined as in Table 3.1 (Congdon, 2003):

\[
\ln y_{hij} = \alpha_1 x_{1hij} + \alpha_2 x_{2hij} + \cdots + \alpha_n x_{nhij} + T_{ij} + C_j
\]

(3.1)

\[T_{jk} = \beta_1 \psi_{1ij} + \beta_2 \psi_{2ij} + \cdots + \beta_{mij} \psi_{mij}\]

(3.2)
### Table 3.1: An explanation of variable definitions for Equations 3.1, 3.2, and 3.3

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_{hij} )</td>
<td>Expected number of crashes on roadway ( ijk )</td>
</tr>
<tr>
<td>( h )</td>
<td>Roadway index</td>
</tr>
<tr>
<td>( i )</td>
<td>Township index</td>
</tr>
<tr>
<td>( j )</td>
<td>County index</td>
</tr>
<tr>
<td>( \alpha_{0,1,2...n} )</td>
<td>Parameter values for roadway level predictors</td>
</tr>
<tr>
<td>( \chi_{(1,2...n)hi} )</td>
<td>Values of roadway level predictors</td>
</tr>
<tr>
<td>( T_{ij} )</td>
<td>Township random effects term</td>
</tr>
<tr>
<td>( \beta_{1,2...m} )</td>
<td>Parameter values for township level predictors</td>
</tr>
<tr>
<td>( \psi_{(1,2...m)ij} )</td>
<td>Values of township level predictors</td>
</tr>
<tr>
<td>( \varsigma_j )</td>
<td>County random effects term</td>
</tr>
<tr>
<td>( \delta_{1,2...l} )</td>
<td>Parameter values for county level predictors</td>
</tr>
<tr>
<td>( \omega_{(1,2...l)j} )</td>
<td>Values of county level predictors</td>
</tr>
</tbody>
</table>

\[
C_k = \delta_1 \omega_{1j} + \delta_2 \omega_{2j} + \cdots + \delta_l \omega_{lj} \quad (3.3)
\]

3.3 Conditional Autoregressive Correlation

Adding higher level parameters is not the sole method of introducing hierarchical structure into a model. A negative binomial model, as outlined in Chapter 2, with a spatial random effects term appears as follows:

\[
\ln y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + u_i \quad (3.4)
\]

where \( u_i \) represents the spatial random effects term (Lawson, 2009). This term attempts to explain the model error due to spatial correlation in a number of different ways depending on the assigned distribution. One popular distribution for spatial random effects is normal conditional autoregressive (CAR) correlation,
also known as Intrinsic CAR (Congdon, 2003; Lawson, 2009). Consider each region as having a relationship with each other region. In the simplest case, regions that share boundaries are considered neighbors, and those that do not share boundaries are not considered neighbors. This is the basis of the CAR distribution. With the neighbors defined, a matrix is created of length \( n \), where \( n \) is the number of regions, and the neighbors of each region fill the row of the corresponding element (Congdon, 2003; Lawson, 2009). This matrix is termed the adjacency matrix. With the neighbors defined, the CAR distribution is specified such that:

\[
u_i \sim \text{Normal}\left(\mu_i, \sigma^2\right) \tag{3.4}\]

where \( \sigma^2 \) denotes the variance of the term \( u_i \) and \( \mu_i \) its variance (Congdon, 2003; Lawson, 2009). Then consider a square, symmetric matrix \( C \) of dimension \( n \) and composed of elements \( c_{ij} \). In the case of equal weights for all neighbors, each \( c_{ij} \) is assigned a value of one in the case of neighborship. In all other cases, including that of \( i = j \), each \( c_{ij} \) receives a value of zero (Congdon, 2003). Then, the mean of the error for each region can be written as:

\[
\mu_i = \rho \sum_j (c_{ij} e_j), \quad i \neq j \tag{3.5}
\]

where \( \rho \) is defined in the range of the minimum and maximum eigenvalues of \( C \) and \( e_j \) is the value of the corresponding neighbor in the state of comparison against \( e_i \) (Congdon, 2003).

The preceding description outlined the most common application of the CAR random effects term. Variations to the hyperparameters of the CAR distribution may be used to adjust the term for certain situations. One instance of this is the case of second order neighbors. First order neighbors are regions which are contiguous, or share a common boundary. Similarly, second order neighbors are regions which share boundaries with a region that shares a common boundary with a region in question. This can also be extended to third order neighbors and other higher order neighbors (Aguero-Valverde and Jovanis, 2010; Wang et al., 2009). The weight matrix \( C \) can be adjusted for this scenario as well. So long as the matrix remains symmetric, different weights can be assigned for each neighbor (Aguero-Valverde and Jovanis, 2010). This is typically done to reduce the influence of higher order neighbors.
Alternatively, the neighborhood of a region or segment may also be defined by the distance between that unit and surrounding units. A radius may be specified in which all spatial units that fall inside the radius, which creates a circle surrounding the unit of interest, are defined as neighbors (Aguero-Valverde and Jovanis, 2010). Any units that are located outside this boundary are not defined as neighbors. This approach, which was applied in the segment level models in Chapter 5, may be implemented by preparing the data in ArcGIS and continuing with a Full Bayesian analysis in WinBUGS.

3.4 Uncorrelated Random Effects

Spatially correlated random effects terms are often added to models of datasets when one suspects that the data may be spatially correlated. The spatially correlated random effects term serves to minimize the amount of error in the estimation of the posterior distribution by explaining the error through the location of each region. However, some additional error is still present after this process. For this reason, when a spatial random effects term is included in a model, it is often accompanied by an uncorrelated random effects term. The purpose of using two different random effects terms is to explain not only the spatially correlated heterogeneity, but also the uncorrelated heterogeneity (Lawson, 2009; Congdon, 2003). With these random effects terms, a negative binomial model appears as follows:

\[
\ln y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + u_i + \nu_i
\]

(3.6)

where \( \nu_i \) is the uncorrelated random effects term (Lawson, 2009). This term is typically assigned a noninformative normal prior distribution with zero mean:

\[
\nu_i \sim Normal(0, \tau_{\nu})
\]

(3.7)

where \( \tau_{\nu} \) is the unknown variance of \( \nu_i \). The variance of the uncorrelated random effects term is often assigned a diffuse gamma prior (Lawson, 2009). One of the focuses of this study is the evaluation of the inclusion of an uncorrelated random effects term in the model.
4.1 Introductory Literature Review

Many approaches may be taken to analyze data effectively to reduce the severity and frequency of motorcycle crashes. Some studies are undertaken specifically to identify the factors of injury severity using discrete outcome models (Pai and Saleh, 2008; Quddus et al., 2002; Savolainen and Mannering, 2007; de Lapperent, 2006). Discrete outcome models estimate the probability of various crash outcomes given crash characteristics. Consider Table 4.1, Table 4.2 and Figure 4.1, which show injury severity trends of Ohio single vehicle motorcycle crashes based on various circumstances. Unlike negative binomial models, discrete outcome models may include behavioral characteristics, where negative binomial models are unable to utilize this information, since it is unknown how many vehicles successfully traversed a route despite poor behavior. For example, alcohol and drug use has been shown to have a heavy impact on the frequency and severity of all types of vehicle crashes (Begg, Langley, and Stephenson, 2003; Creaser et al., 2009; Branas and Knudson, 2001; Huang and Lai, 2011), and helmet use is shown to reduce injury severity, despite the movement by state administrations away from universal helmet laws for riders (Coben et al.,

Table 4.1: Injury severity of single vehicle Ohio motorcycle crashes based on time and behavior

<table>
<thead>
<tr>
<th>Injury Severity</th>
<th>Single Vehicle Motorcycle Crash</th>
<th>Alcohol Related</th>
<th>Speed Related</th>
<th>Night Related</th>
</tr>
</thead>
<tbody>
<tr>
<td>Property Damage Only</td>
<td>2599</td>
<td>95</td>
<td>16.5%</td>
<td>209</td>
</tr>
<tr>
<td>Possible Injury</td>
<td>580</td>
<td>35</td>
<td>6.1%</td>
<td>82</td>
</tr>
<tr>
<td>Non-incapacitating</td>
<td>2013</td>
<td>180</td>
<td>31.3%</td>
<td>461</td>
</tr>
<tr>
<td>Injury</td>
<td>Incapacitating Injury</td>
<td>196</td>
<td>34.0%</td>
<td>356</td>
</tr>
<tr>
<td></td>
<td>Fatality</td>
<td>70</td>
<td>12.2%</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>6543</td>
<td>-</td>
<td>1166</td>
</tr>
<tr>
<td>Most Harmful Event</td>
<td>Event</td>
<td>Percent Occurrence</td>
<td>PDO</td>
<td>Possible Injury</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-------</td>
<td>--------------------</td>
<td>-----</td>
<td>----------------</td>
</tr>
<tr>
<td>Motor Vehicle in Transport</td>
<td>3134</td>
<td>44.7%</td>
<td>1870</td>
<td>254</td>
</tr>
<tr>
<td>Overturn/Rollover</td>
<td>1518</td>
<td>21.6%</td>
<td>147</td>
<td>102</td>
</tr>
<tr>
<td>Other Non-Collision</td>
<td>355</td>
<td>5.1%</td>
<td>161</td>
<td>33</td>
</tr>
<tr>
<td>Ditch</td>
<td>269</td>
<td>3.8%</td>
<td>40</td>
<td>24</td>
</tr>
<tr>
<td>Animal - Deer</td>
<td>209</td>
<td>3.0%</td>
<td>59</td>
<td>22</td>
</tr>
<tr>
<td>Ran Off Road Right</td>
<td>155</td>
<td>2.2%</td>
<td>21</td>
<td>31</td>
</tr>
<tr>
<td>Ditch</td>
<td>269</td>
<td>3.8%</td>
<td>40</td>
<td>24</td>
</tr>
<tr>
<td>Other Fixed Object (Wall, Building,</td>
<td>133</td>
<td>1.9%</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>Tunnel, Etc)</td>
<td>127</td>
<td>1.8%</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Curb</td>
<td>101</td>
<td>1.4%</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Tree</td>
<td>69</td>
<td>1.0%</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Embankment</td>
<td>68</td>
<td>1.0%</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Ran Off Road Left</td>
<td>54</td>
<td>0.8%</td>
<td>11</td>
<td>7</td>
</tr>
</tbody>
</table>

2007; Houston and Richardson, 2008; Mayrose, 2008; McCartt et al., 2011). It is notable that Ohio does not require adult riders to wear helmets after they have have carried a motorcycle endorsement, or motorcycle riding license, for over one year (ORC 4507, 2012).

Since motorcycles compose a lesser portion of vehicle miles travelled (VMT), data that are commonly employed to describe overall vehicle crashes and traffic are not as consistently available in a motorcycle specific setting. For instance, while overall average daily traffic (ADT) is often shown to correlate strongly with vehicle crashes, motorcycle-specific ADT is not consistently available, especially at a reasonable level of analysis across an entire state. As a result of this limitation, general traffic ADT is frequently considered in the prediction of motorcycle crashes. For example, Haque et al. (2010) conducted a study which included overall ADT in an analysis of motorcycle crashes at signalized intersections, while Paulozzi (2005) conducted a study which used registration and motorcycle VMT.

The use of hierarchical, also known as multilevel, Bayesian analysis modeling is widespread in safety research. For example, Song et al. (2006) considered several Bayesian multivariate spatial models
Figure 4.1: Distribution of injury severity in Ohio single-vehicle motorcycle crashes based on crash type using Texas crash data to estimate crash rates. Wang and Quddus (2009) applied a hierarchical model with a spatial random effects term to assess the impact of traffic congestion on crash frequency on the M25 Freeway in London. A hierarchical Bayesian model with site specific random effects was used to estimate crash frequency in Utah by Schultz et al. (2011). Abdalla (2005) employed a hierarchical Bayesian model to analyze the effectiveness and use of safety belts in United Arab Emirates.

Hierarchical models at a regional level are effective ways to optimize the implication of data that are often available at an administrative level into an organized summary of the factors and the magnitude of their effects on vehicle crashes (Quddus, 2008; Eksler and Lassarre, 2008). For crashes involving motorcycles, Haque et al. (2010) developed hierarchical models to explain the extra variation in motorcycle crashes at signalized intersections and demonstrated that crashes occurring at the same intersection tend to be more similar than otherwise.

Spatial random effects terms are often deployed alongside uncorrelated random effects terms to prevent the inference of undue spatial correlation (Mitra, 2009; Quddus, 2008). Guo et al. (2010) used
conditional autoregressive spatial effects to model corridor-level spatial correlations in Florida. Aguero-Valverde and Jovanis (2006) used Bayesian hierarchical methods with spatial and temporal effects to model county level crash frequency in Pennsylvania. In a separate study, Aguero-Valverde and Jovanis (2010) investigated the effectiveness of various spatial random effects methods in multi-level data, applying the spatial effects to the first level of analysis by specifying the spatial correlation at the roadway segment level. Despite the amount of research done in both spatial analysis and motorcycle safety, little to no research specifically addresses spatial analysis of motorcycle crashes. However, this approach is ideal for motorcycle data because this approach allows the researcher to include more descriptive predictors of motorcycle crashes.

4.2 Research Objectives

The objective of this study is to develop a model to predict single-vehicle motorcycle crashes in Ohio at a regional level. Although some types of motorcycle specific data are largely unavailable, such as motorcycle-specific ADT or VMT, the predictors in this model were selected to capture aspects of motorcycle activity and the demographics of each region. The hierarchical negative binomial model with mixed effects that is developed in this study suits the availability of data by including two regional levels of predictors and spatial correlation, which reduces model error caused in part by unobserved factors. This approach reduces the impact of the lack of motorcycle-specific data, such as motorcycle-specific measures of exposure. Finally, recommendations are made for application of the model by practitioners and administrators planning campaigns directed at reducing the frequency of single-vehicle motorcycle crashes in Ohio.

4.3 Data

The data in this study consist of single-vehicle motorcycle crashes in Ohio and the characteristics of the region in which the crashes occurred. A summary of the data sources and their respective contributions to the study is available in Table 4.3. Two different administrative regions in Ohio are considered: counties and townships. Ohio is divided into 88 counties that are somewhat similar in size. Each county is further divided into several townships, of which there are 1,459 throughout the state.
Table 4.3: Descriptive statistics of Ohio motorcycle crashes

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall motorcycle crashes per county</td>
<td>180</td>
<td>234</td>
<td>12</td>
<td>1293</td>
</tr>
<tr>
<td>Overall motorcycle crashes per township</td>
<td>12</td>
<td>23.833</td>
<td>0</td>
<td>332</td>
</tr>
<tr>
<td>Overall motorcycle fatalities per county</td>
<td>7</td>
<td>8.658</td>
<td>0</td>
<td>49</td>
</tr>
<tr>
<td>Overall motorcycle fatalities per township</td>
<td>0.44</td>
<td>1</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Single vehicle motorcycle crashes per township</td>
<td>5</td>
<td>10</td>
<td>0</td>
<td>149</td>
</tr>
<tr>
<td>Single vehicle motorcycle fatalities per township</td>
<td>0.21</td>
<td>0.59</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>County population (2010)</td>
<td>140,353</td>
<td>230,136</td>
<td>13,435</td>
<td>1,280,122</td>
</tr>
<tr>
<td>Motorcycle registration per county (2010)</td>
<td>4,371</td>
<td>4,839</td>
<td>417</td>
<td>27,100</td>
</tr>
<tr>
<td>Motorcycle endorsements per county (2010)</td>
<td>8,253</td>
<td>9,179</td>
<td>982</td>
<td>47,003</td>
</tr>
<tr>
<td>Percent over age 65 per county</td>
<td>15%</td>
<td>2%</td>
<td>9%</td>
<td>19%</td>
</tr>
<tr>
<td>Mean travel time to work in minutes per county (2005-2009)</td>
<td>24</td>
<td>4</td>
<td>17</td>
<td>34</td>
</tr>
<tr>
<td>Percent below poverty level per county (2009)</td>
<td>15%</td>
<td>4%</td>
<td>5%</td>
<td>35%</td>
</tr>
<tr>
<td>Township Area (sq. mi)</td>
<td>(73 km²)</td>
<td>(28 km²)</td>
<td>(0.03 km²)</td>
<td>(225 km²)</td>
</tr>
<tr>
<td>Township lane miles</td>
<td>128</td>
<td>80</td>
<td>0.01</td>
<td>813</td>
</tr>
<tr>
<td>Maximum Degree of Curve in Township</td>
<td>0</td>
<td>158</td>
<td>16.6</td>
<td>14.2</td>
</tr>
<tr>
<td>Maximum Grade in Township</td>
<td>0</td>
<td>20</td>
<td>7.7</td>
<td>5.4</td>
</tr>
<tr>
<td>Number of Horizontal Curves in Township</td>
<td>27</td>
<td>1061</td>
<td>214.3</td>
<td>220.3</td>
</tr>
<tr>
<td>Number of Vertical Curves in Township</td>
<td>1</td>
<td>789</td>
<td>231</td>
<td>187.6</td>
</tr>
<tr>
<td>County roadway density (lane mile per sq. mi)</td>
<td>3.6</td>
<td>1.47</td>
<td>2.2</td>
<td>9.2</td>
</tr>
</tbody>
</table>

* Data from the Ohio Department of Public Safety (ODPS, 2011)
* Data from the Ohio Department of Transportation (ODOT, 2011)
* Data from the Public Utilities Commission of Ohio (PUCO, 2011)
* Data from the US Census Bureau (US Census Bureau, 2011)

(ODOT, 2011; ODPS, 2011). The townships are ideal for spatial analysis due to their uniformity in size, and the grid-like structure of the townships also helps ensure that potential bias will not be introduced due to boundaries. The analysis of the motorcycle crashes at two levels helps to prevent modifiable areal unit
problems (MAUP) that can arise when point data are aggregated into spatial regions (McGrew and Monroe, 2000).

The Ohio Department of Public Safety (ODPS) Crash Database is a compilation of OH-1 police reports that are filed for each reported crash in the state of Ohio. Of the 11,113 reported single-vehicle motorcycle crash records from 2008 through summer 2011, a subset of 8,355 records contain geographic
coordinates; records in this subset are assigned to a township and county through a spatial join in ArcGIS between the crash locations and township boundaries (ODPS, 2011). A map of single-vehicle motorcycle crash distributions is available in Figure 4.2. Spatial relationships are also employed for obtaining geometric properties of the predictors such as the total area, number of lane miles, and road densities, which are derived from the administrative area and roadway data files. The quantity of lane miles per township was derived as follows:

\[ LM_i = \sum_{h=1}^{N_h} (l_h/N_h) \]  

(4.1)

where \( l_h \) represents the length of segment \( h \), \( N \) is the number of lanes in both directions, and \( LM_i \) is the number of lane miles in each township \( i \). Similarly, the predictor road density is applied at the township level, and is calculated as shown:

\[ RD_j = \frac{\sum_{i=1}^{LM_{ij}} LM_{ij}}{A_j} \]  

(4.2)

where \( A_j \) is the area of the county \( j \), \( LM_{ij} \) is the number of lane miles in township \( i \), which is to be summed with all townships in county \( j \), and \( RD_j \) is the quantity road density for each county. This extra step is applied for the county level so that the area can be factored into the predictor; it is beneficial to include this step in the model due to the tendency of some smaller counties to be more urbanized and thus have a more highly developed roadway system. With the area and the lane miles considered, road density is an appropriate measure of how developed the roadway system is in each county (Aguero-Valverde and Jovanis, 2006). As the townships are more uniformly shaped, and an indicator variable for urban areas is present, the lane miles are applied to the models separately from the township area. The urban indicator for each township is assigned by using the city incorporated area boundaries that are provided by the Ohio Department of Transportation (ODOT; 2011). A total of 940 townships contain at least one city and thus are designated as urban.

Roadway alignment characteristics are also considered at the township level. The variables that describe horizontal and vertical curves capture the alignment in terms of maximum degree of curvature
(horizontal curves) or absolute value of grade (vertical curves) and the number of horizontal and vertical curves per township. The alignment data are produced from the ODOT RI-15 database (ODOT, 2011). Measures of motorcycle traffic and demographics are not typically available to a researcher in a consistent fashion. To partially account for the shortage of such motorcycle data, several county level predictors may be used. The county population is considered in the model to assess the impact of activity generated from that region, as done in studies by Erdogan (2009) and Siddiqui et al. (2011). The number of registered vehicles in each county and the number of motorcycle endorsements on driver licenses, which was specified as an offset, were used as regional measures of riding activity. These measures are similar to measures of vehicle travel in current research. For example, a study by Teoh and Campbell (2010) demonstrated that the license status was a significant predictor of motorcycle crashes, and registration was an important normalizing factor. Quddus (2008) calculated a statistic estimating the number of cars traveling through a region in a study of vehicle crashes in London that factored in the number of registered vehicles in neighboring areas and the distance to those areas. These two measures were also employed in a study of motor vehicle crashes in Pennsylvania by Aguero-Valverde and Jovanis (2006) as potential predictors of crashes. Accordingly, the percentages of residents over 65 years old and below the poverty level are considered in this study to estimate the impact of the demographics of the resident population on the number of single-vehicle motorcycle crashes.

Only statistically significant variables (p-value ≤ 0.05) are included in the final models. Factors included in the model are township lane miles, the number of horizontal curves, the number of vertical curves, the maximum grade in the township, county road density, county population, motorcycle registration per county, the percentage of residents over age 65, and the urban indicator variable for townships. In addition, the number of motorcycle endorsements, or motorcycle riding licenses, was included as a model offset to account for exposure. These variables represent factors that describe the motorcycle crash patterns in Ohio. Although, from a bottom-up approach, some of the standard quantities for overall crash models are not available specifically for motorcycles, these variables, combined with the exposure given by the number of endorsements, provide a number of useful insights into factors behind single-vehicle motorcycle crashes.
4.4 Methodology

This study focuses on estimating the single-vehicle motorcycle crash frequency in townships and counties in Ohio through the use of a full Bayesian hierarchical random effects model. Negative binomial models are often used to predict the frequency of crashes as a robust alternative to a Poisson model for count data (Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Mitra, 2009). The probability density of the negative binomial distribution is defined in the following function:

\[
p(n_i|\mu_i, k) = \frac{\Gamma(n_i + k^{-1})}{\Gamma(k^{-1})\Gamma(n_i + 1)} \left( \frac{k\mu_i}{1 + k\mu_i} \right)^{n_i} \left( \frac{1}{1 + k\mu_i} \right)^{1/k}
\]

where \( k \) is the overdispersion parameter \( k > 0 \), \( n_i \) is the number of samples, and \( \mu_i \) is the mean value (Noland and Quddus, 2004).

In Bayesian inference, unknown quantities, including the vectors \( \beta_k \) and \( \gamma_k \) of coefficients, are viewed as random variables that compose the posterior distribution of the model. In traditional statistics, the data are fitted to the model, and the parameters are fixed (Congdon, 2003). A Bayesian approach allows the researcher to apply knowledge about the behavior of the phenomena being modeled into the model itself. This is illustrated through Bayes theorem as seen in Guo et al. (2010) and Congdon (2010):

\[
p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}
\]

where \( p(\theta|y) \) denotes the posterior density, \( p(y|\theta) \) represents the model likelihood, \( p(y) \) represents the unconditional density of the data, and \( p(\theta) \) represents the known background information.

In this study, data are available at both the township and the county level. A hierarchical model provides an avenue for including both levels of information, as well as including influence from neighboring regions through spatial correlation. This type of model may provide a better estimate without the geographic limitations of providing all information at the same administrative level, since that limitation sometimes restricts the availability of predictor information. The hierarchical negative binomial model used in this study may be expressed by the following system of equations:
\[
\ln \lambda_i = \sum_{k=1}^{\beta_k} (x_{ijk} - \bar{x}_k) + \alpha_j + u_i + v_i \quad (4.4)
\]
\[
\alpha_j \sim \text{Normal}(\mu_{\alpha,j}, \sigma_{\alpha}) \quad (4.5)
\]
\[
\mu_{\alpha,j} = \sum_{k=1}^{\gamma_k} (z_{jk} - \bar{z}_k) \quad (4.6)
\]

where in the first level, \( i \) is the first level index (in this case, the township level index), \( \beta_k \) is the vector of estimated parameters, \( \bar{x}_k \) is the vector of predictor means that is subtracted from the matrix of known predictors \( x_{ijk} \), \( \lambda_i \) is the expected rate of occurrence, \( u_i \) is the township level spatial random effects term, \( v_i \) is the uncorrelated random effects term, and \( \gamma_i \) is the observed rate of occurrence. In the second level, \( j \) is the second level index (in this case, the county level index), \( \alpha_j \) is a term that links the two equations together, \( \gamma_k \) represents the vector of second level parameters, \( z_{jk} \) is the matrix of second level predictors, and \( \bar{z}_k \) is the vector of those predictor means (Yanmaz-Tuzel and Ozbay, 2010; Congdon, 2003). The hierarchical model allows each township to be modeled individually, while still considering the characteristics of each county; this encourages the flexibility necessary to include significant predictors in motorcycle crash models.

In many cases, regions that are close to each other geographically are more similar than regions that are farther away. The amount of this spatial correlation may be quantified on a regional level through the introduction of a spatial random effects term (Miaou and Song, 2005). A spatial random effects term may account for some or all of the error through the effect of neighboring regions (Quddus, 2008), which can have two levels: first order neighbors (where neighbors share borders) and second order neighbors (where neighbors are not adjacent to one another). The spatial random effects term \( u_i \) is given a conditional autoregressive distribution. As specified for \( u_i \), the conditional autoregressive prior distribution is defined as follows:

\[
P(u|\tau) \propto \frac{1}{m} \exp \left( -\frac{1}{2\tau} \sum_i \sum_{j \in \delta_i} (u_i - u_j)^2 \right), \quad i \neq j \quad (4.6)
\]

where \( r \) is the prior knowledge of the spatial region, \( m \) is the index of the number of neighbors in the adjacency matrix, \( \delta_i \) is the neighborhood of the region \( i \), and \( j \) represents the neighbor of region \( i \) (Lawson,
At both levels of neighborship, first and second order spatial correlation is specified for comparison. In addition, models with second order neighbors are run with two weighting systems: one in which first and second order neighbors are weighted equally, and the other in which second order neighbors are given half the influence of first order neighbors, as suggested by Aguero-Valverde and Jovanis (2010) and Wang et al. (2009) in previous research. Such a weighting scheme acknowledges that first order neighbors are more closely related to each region. Regions that are on the state border received influence from the neighboring counties and townships in Ohio but did not receive zero values for absent neighboring regions.

Uncorrelated random effects are often specified alongside spatial random effects (Mitra, 2009; Wang, 2009). By placing this additional random effects term in the model, the uncorrelated random effects term prevents the spatial random effects term from inferring spatial correlation that is not present. The result is a gain in the confidence of the estimate of spatial random effects and fixed effects parameters estimates with only a small increase to the complexity of the model (Aguero-Valverde and Jovanis, 2008). The uncorrelated random effects term is assigned a normal distribution, as shown below:

\[ v_i \sim \text{Normal}(0, \tau_v) \]  

(7)

where \( \tau_v \) is the precision, or inverse variance, of the uncorrelated random effects term, \( v_i \). The precision of \( v_i \) is assigned a diffuse gamma prior distribution (Wang et al., 2009). Note that for ease of comparison, the standard deviation of \( u_i \) and \( v_i \) is displayed rather than \( \tau_v \) or the inverse variance.

Bayesian analysis in this study is carried out through the use of WinBUGS. The results are not considered reliable until the sampler has achieved convergence, which is verified by running multiple Monte Carlo Markov Chains (MCMC). In this study, all models are hierarchically centered for efficient convergence with three chains, as evidenced by the Gelman-Rubin statistic (Lawson, 2009). Three chains were specified and received 20,000 iterations each, for a total of 60,000 iterations. Seven models in total are presented using Bayesian analysis in WinBUGS. Table 4.4 provides a summary of model fit, along with the abbreviated names used throughout this study.

The Deviance Information Criterion (DIC) is a basis of comparison between models. The DIC is based on the model deviance, which is derived from the log likelihood (Congdon, 2010). To summarize this
Table 4.4: Goodness-of-fit Characteristics for Mixed Effects Models with Varying Spatial Random Effects Specifications

<table>
<thead>
<tr>
<th>Model</th>
<th>Order of Neighbors</th>
<th>Dbar</th>
<th>Dhat</th>
<th>pD</th>
<th>DIC</th>
<th>DIC difference</th>
<th>η</th>
</tr>
</thead>
<tbody>
<tr>
<td>County Neighbors (2008-2011)</td>
<td>0</td>
<td>5132</td>
<td>4479</td>
<td>653</td>
<td>5785</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5134</td>
<td>4482</td>
<td>652</td>
<td>5786</td>
<td>-0.4</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5133</td>
<td>4479</td>
<td>655</td>
<td>5788</td>
<td>-2.8</td>
<td>0.49</td>
</tr>
<tr>
<td>2W</td>
<td>0</td>
<td>5132</td>
<td>4479</td>
<td>653</td>
<td>5785</td>
<td>0.1</td>
<td>0.51</td>
</tr>
<tr>
<td>Township Neighbors (2008-2011)</td>
<td>0</td>
<td>5132</td>
<td>4479</td>
<td>653</td>
<td>5785</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5121</td>
<td>4478</td>
<td>643</td>
<td>5763</td>
<td>21.8</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5131</td>
<td>4478</td>
<td>653</td>
<td>5784</td>
<td>0.9</td>
<td>0.86</td>
</tr>
<tr>
<td>2W</td>
<td>2</td>
<td>5118</td>
<td>4460</td>
<td>657</td>
<td>5775</td>
<td>9.9</td>
<td>0.85</td>
</tr>
</tbody>
</table>

\[ DIC = \bar{D} + p_D = D(\bar{\theta}) + 2p_D \]

\[ \eta = \frac{\sigma_u}{\sigma_u + \sigma_v} \]

Note: A significant change in DIC may be considered to be 10 or more.

<table>
<thead>
<tr>
<th>Short Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0º</td>
<td>No spatial random effects</td>
</tr>
<tr>
<td>C1</td>
<td>Only counties that share borders are neighbors</td>
</tr>
<tr>
<td>C2</td>
<td>Neighbors of neighboring counties are also neighbors</td>
</tr>
<tr>
<td>C2W</td>
<td>Neighbors of neighboring counties are less influential neighbors</td>
</tr>
<tr>
<td>0º</td>
<td>No spatial random effects</td>
</tr>
<tr>
<td>T1</td>
<td>Only townships that share borders are neighbors</td>
</tr>
<tr>
<td>T2</td>
<td>Neighbors of neighboring townships are also neighbors</td>
</tr>
<tr>
<td>T2W</td>
<td>Neighbors of neighboring townships are less influential neighbors</td>
</tr>
</tbody>
</table>

The posterior deviance statistic for the entire model space, consider the posterior expected deviance, which is the average of the deviance, or \( \bar{D} = E_{y|x}(D) \). To further describe the characteristics of each model, the DIC is also a function of the effective number of parameters, or the effective model dimension, which represents model complexity (Congdon, 2010). Finally, the DIC is obtained through the combination of these terms:

\[ DIC = \bar{D} + p_D = D(\bar{\theta}) + 2p_D \]  (4.10)
Table 4.5: Parameter Estimates and Results from Mixed Effects Models

**No Spatial Random Effects (Model 0)**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>2.5% CI</th>
<th>97.5% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lane Miles ($\beta_1$)</td>
<td>0.008</td>
<td>0.000</td>
<td>0.008</td>
<td>0.009</td>
</tr>
<tr>
<td>No. Horizontal Curves ($\beta_2$)</td>
<td>0.003</td>
<td>0.001</td>
<td>0.001</td>
<td>0.004</td>
</tr>
<tr>
<td>No. Vertical Curves ($\beta_3$)</td>
<td>-0.002</td>
<td>0.001</td>
<td>-0.003</td>
<td>0.000</td>
</tr>
<tr>
<td>Maximum Grade ($\beta_4$)</td>
<td>0.026</td>
<td>0.007</td>
<td>0.012</td>
<td>0.039</td>
</tr>
<tr>
<td>Urban ($\beta_5$)</td>
<td>0.504</td>
<td>0.061</td>
<td>0.383</td>
<td>0.621</td>
</tr>
<tr>
<td>Constant ($\gamma_1$)</td>
<td>-2.360</td>
<td>0.090</td>
<td>-2.537</td>
<td>-2.179</td>
</tr>
<tr>
<td>Population ($\gamma_2^b$)</td>
<td>-0.002</td>
<td>0.033</td>
<td>-0.067</td>
<td>0.064</td>
</tr>
<tr>
<td>Road Density ($\gamma_3$)</td>
<td>-0.240</td>
<td>0.050</td>
<td>-0.337</td>
<td>-0.143</td>
</tr>
<tr>
<td>Registration ($\gamma_4^{bc}$)</td>
<td>-0.001</td>
<td>0.009</td>
<td>-0.018</td>
<td>0.016</td>
</tr>
<tr>
<td>Percent Over 65 ($\gamma_5^{ke}$)</td>
<td>0.359</td>
<td>2.083</td>
<td>-3.732</td>
<td>4.490</td>
</tr>
<tr>
<td>SC ($\sigma_u$)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>UH($\sigma_v$)</td>
<td>0.334</td>
<td>0.103</td>
<td>0.182</td>
<td>0.544</td>
</tr>
</tbody>
</table>

**First Order Neighbors (Model T1)**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>2.5% CI</th>
<th>97.5% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lane Miles ($\beta_1$)</td>
<td>0.008</td>
<td>0.000</td>
<td>0.008</td>
<td>0.009</td>
</tr>
<tr>
<td>No. Horizontal Curves ($\beta_2$)</td>
<td>0.003</td>
<td>0.001</td>
<td>0.001</td>
<td>0.004</td>
</tr>
<tr>
<td>No. Vertical Curves ($\beta_3$)</td>
<td>-0.002</td>
<td>0.001</td>
<td>-0.004</td>
<td>0.000</td>
</tr>
<tr>
<td>Maximum Grade ($\beta_4$)</td>
<td>0.027</td>
<td>0.007</td>
<td>0.012</td>
<td>0.042</td>
</tr>
<tr>
<td>Urban ($\beta_5$)</td>
<td>0.525</td>
<td>0.059</td>
<td>0.402</td>
<td>0.641</td>
</tr>
<tr>
<td>Constant ($\gamma_1$)</td>
<td>-2.481</td>
<td>0.096</td>
<td>-2.659</td>
<td>-2.293</td>
</tr>
<tr>
<td>Population ($\gamma_2^b$)</td>
<td>-0.027</td>
<td>0.035</td>
<td>-0.097</td>
<td>0.043</td>
</tr>
<tr>
<td>Road Density ($\gamma_3$)</td>
<td>-0.266</td>
<td>0.053</td>
<td>-0.371</td>
<td>-0.162</td>
</tr>
<tr>
<td>Registration ($\gamma_4^{bc}$)</td>
<td>0.004</td>
<td>0.009</td>
<td>-0.015</td>
<td>0.022</td>
</tr>
<tr>
<td>Percent Over 65 ($\gamma_5^{ke}$)</td>
<td>-0.079</td>
<td>2.327</td>
<td>-4.653</td>
<td>4.528</td>
</tr>
<tr>
<td>SC ($\sigma_u$)</td>
<td>0.889</td>
<td>0.132</td>
<td>0.600</td>
<td>1.124</td>
</tr>
<tr>
<td>UH($\sigma_v$)</td>
<td>0.270</td>
<td>0.073</td>
<td>0.156</td>
<td>0.438</td>
</tr>
</tbody>
</table>

Smaller DIC values identify models with lower posterior expected deviance as well as models with a lower effective number of parameters (Congdon, 2010; Yanmaz-Tuzel and Ozbay, 2010). A significant change in the DIC is typically considered to be 7 to 10 points or more (Spiegelhalter, 2002).

To measure the importance of spatial correlation in the data, the spatial correlation coefficient is calculated. This statistic shows the proportion of model error explained by spatial random effects and uncorrelated random effects through the following relationship:
where $\sigma_u$ is the standard deviation of the spatial random effects term $u$, and $\sigma_v$ represents the standard deviation of the uncorrelated random effects term $v$ (Eksler and Lassarre, 2008; Aguero-Valverde and Jovanis, 2010). As this statistic approaches unity, the spatial random effects term becomes increasingly influential.

4.5 Results and Discussion

The seven models are specified with three different kinds of spatial random effect; a summary of Model 1 and Model T1, which is the most improved, is provided in Table 4.5. First order, unweighted second order, and weighted second order neighbors are considered at the township level and at the county level; in addition, a model with no spatial random effects (Model 0) is included as a basis for comparison. The DIC values are calculated for each model; the DICs, along with the components involved in the calculation for each, are presented in Table 4.4. A comparison of DICs demonstrates that most models with spatial random effects show improvement. The exceptions to this trend are the first (C1) and unweighted second order (C2W) models. This finding suggests that spatial correlation is a significant factor in single-vehicle motorcycle crashes, especially given the inherent lack of motorcycle-specific measures of traffic. Of the seven models in this study, two show significant improvement over the model without spatial random effects (Model 0).

The first order neighbor model (Model T1) performed the best, showing significant improvement over any other model in this study. Also of note is the improved performance of the second order neighbor models at the township level when a weighting structure was included. Both of these finding suggest that spatial correlation is a more significant factor in regions that directly neighbor the unit of interest. From the results, it is clear that simply specifying spatial random effects does not guarantee a significant change in goodness-of-fit. In county neighbor models with township level predictors, spatial random effects were not able to explain as much error as township spatial correlation. This is consistent with one of the key rationales behind considering spatial random effects: that spatial random effects help protect against unknown or unobserved predictors that have been unintentionally omitted from the model, such as
Figure 4.3: Relative Risk map of Ohio single-vehicle motorcycle crashes based on Model T1

motorcycle-specific VMT. The finding that townships are spatially similar to those that directly share borders may reflect the travel patterns of motorcyclists. As the distance from the township where the crash occurs increases, more townships become involved in the analysis and are included as neighbors. However, given the relatively small size of the townships, a motorcyclist could ride through each one in a matter of minutes. It would not be at all unrealistic for a rider to cross an entire county, ending up three or more townships from the township in which the trip began. Nonetheless, first order neighbors are effective in the
analysis; this is likely a consequence of the fact that in order to reach any destination more than a few miles away from the township of the trip generation, the rider must pass through one of the adjacent townships. However, at greater distances from the township of the trip generation, the regions likely become too diffuse, and too many neighbors become considered, resulting in less significant spatial correlation.

Figure 4.3 shows a map of the Relative Risk (RR) results of Model T1. The RR map shows the regions in which the crash risk is higher than would be expected, given the exposure in the area. To adjust for the amount of rider activity without motorcycle specific ADT or VMT, the number of endorsed, or licensed, riders per county is entered into the model as an offset term. Many different quantities are used in research in an induced exposure framework to adjust the model for how exposed each sample is (Stamatiadis and Deacon, 1997; Haque et al., 2012). Although ADT or VMT are commonly used in this capacity, such figures are not available for motorcycles in Ohio. Therefore, the number of endorsed riders follows as a natural choice of motorcycle activity in each area (Wasielewski and Evans, 1985).

The impact of the number of lane miles in a township remains fairly constant between models. The impact is positive in all cases; as the number of lane miles in a township increase, the frequency of motorcycle crashes increases. This is consistent with prior research of regional level models and with logical expectations. Areas with more roads tend to generate and draw more traffic, including motorcycles; as more motorcycles are traveling on the roadway, the potential for crashes in general is greater. For example, Aguero-Valverde and Jovanis (2006) found that as infrastructure mileage increased, the number of motor vehicle crashes increased as well.

Three parameters that describe roadway alignment are included in each model. The results show that townships with more horizontal curves tend to experience higher crash rates. A fair portion of motorcycle riding is done for enjoyment, and regions with challenging, curvy roads are often a favorite destination for riders. For example, consider one of the curviest routes in the United States, an 11-mile section of US Route 129 known as the Tail of the Dragon. The parameter for the number of vertical curves in a township shows that hillier segments, in terms of how many but not how steep, tend to be associated with fewer single vehicle crashes. The parameter for the maximum grade of each township complements
this finding by showing that more crashes tend to occur in townships with steeper grades. The combination of the two vertical curve parameters yields a statement that steeper vertical curves, rather than more numerous ones, may instigate more crashes. These findings are consistent with previous studies, such as curve frequency in Shively et al. (2010) or grade in Ahmed et al. (2011).

The final township level predictor, the urban designation, is implemented in the model as an indicator variable. If a township was designated as an urban township, the frequency of crashes in the township increased. Because there are more destinations and residents in urban areas, and thus more motorcyclists and vehicles in general will be on the road in these regions, the potential for single-vehicle motorcycle crashes is increased in these areas. Like characteristics are considered with similar results in other studies, such as land use in Ivan et al. (2000) or the ramp density on freeways in Malyshkina et al. (2009).

The parameter describing county population shows a negative impact on the number of single-vehicle motorcycle crashes in a region. Of particular note is the magnitude of the parameter, which is larger in the township level neighbor models than county level neighbor models. The negative correlation between population and single-vehicle motorcycle crashes is intuitive. This finding also highlights the distinction between areas with high population and urban buildup. While urban areas are found to have more single-vehicle crashes, counties with high population, which may be more spread out than in the urban townships, in general tend to experience lower single-vehicle motorcycle crash rates. A negative correlation between road density and crashes implies that with more roads to choose from in an area dense in roadways, traffic is more diffused than it may be in an area with a less developed roadway system, even if the sizes of the areas are the same.

The parameter for motorcycle registration per county has differing signs based on the specified neighborhood. The model with the best fit (Model T1) shows that counties with more registered motorcycles tend to experience higher single-vehicle motorcycle crash rates. The differing signs are an example of the benefits of considering spatial correlation, as some source of model error appears to interfere with the parameter estimates in models without township level spatial correlation.
The model findings suggest that fewer single vehicle motorcycle crashes tend to occur in counties with a higher percentage of residents over the age of 65. This finding is similar to previous research, such as Carr (1969), which found that single-vehicle crash risk decreased with both age and experience. The differing signs of the parameters across models show that spatial correlation helps reduce model error.

4.7 Conclusions and Recommendations

The size of the areal unit should be chosen to match the preferred level of analysis. At the county level, using predictors of finer scale than the spatial random effects term appears to dilute the benefits of considering spatial correlation in this study. However, significant improvements in the estimates and interpretation of the parameters were gained through township level spatial random effects. Applying spatial random effects and multi-level analysis reduces the error created by missing predictors, which is especially of interest for data sets describing alternative modes of transportation such as motorcycles, since in general less data are available. This type of model is applicable to other cases of motorcycle crashes which face the same data limitations in terms of availability, or to other crash types or phenomena that experience similar difficulties due to lack of data. Regional random effects still may, and frequently are, able to reduce error in models in situations where the ADT and VMT are known, such as overall crash models.

By viewing the expected number of crashes per region, the researcher may identify the areas with the highest density of single-vehicle motorcycle crashes, an approach that may be especially beneficial from the perspectives of enforcement and riding safety advertising programs. As these regions are identified, an administrator may examine the individual regions and clusters to find the underlying causes. The fixed effects in this study, when considered by themselves, may shed some light onto the causes for the regions of higher crash density, but including spatial correlation may enhance model performance. For most purposes, the first order township model is the best model to use; the substantial reduction in the DIC and ease of interpretation for this model would be of the greatest use to practitioners and administrators planning campaigns directed at motorcycle safety and enforcement. In addition, the generality of the regions are an appropriate size to allow flexibility for sign placement and increased patrols and would not constrict these efforts into a specific segment of roadway.
CHAPTER V
A SEGMENT LEVEL ANALYSIS OF MULTI-VEHICLE MOTORCYCLE CRASHES IN OHIO USING BAYESIAN MULTI-LEVEL MIXED EFFECTS MODEL

5.1 Introductory Literature Review

A motorcycle crash, as any other vehicle crash, is a complex event with many influential factors and characteristics. Since the mechanisms that lead to each crash may be dramatically different, it is natural to assume that the factors that are behind the two distinct crash types are different as well (Haque et al., 2012; Geedipally and Lord, 2010; Jonsson et al., 2007; Ivan, 2004; Savolainen and Mannering, 2007; Yau, 2004). Therefore it is reasonable to separate single and multi-vehicle crashes.

Two common approaches to model motorcycle crashes are discrete outcome and negative binomial models. Some common findings from discrete outcome models show that injury severity is significantly affected by factors such as helmet use, speeding, alcohol use, and operator age (Chang and Yeh, 2006; Savalainen and Mannering, 2007; Haque et al., 2009). While discrete outcome models perform well in estimating the impact of behavioral and crash characteristics on the type of crash that occurs, negative binomial models are more commonly used to predict the number of crashes based on information such as geometric, demographic, or infrastructural characteristics (Chin and Quddus, 2003; Haque et al., 2010; Harnen et al., 2003; Houston, 2007; Schneider et al., 2010).

More recently, negative binomial models have been improved by using random effects terms, which offer the prospect of including data and relationships that may be difficult to apply in a standard model configuration. Some examples include structured random effects that estimate the impact of crashes being in the same intersection (Kim et al., 2007), corridor (Guo et al., 2010), region (Yannis et al., 2007), or year (Majumbar et al., 2004). In each case, the random effects are included to improve both model fit and interpretation of the findings, showing which intersections or regions are more prone to crash...
occurrence. Not only can random effects be used with assigned groups of similar crashes (such as county or
time) as shown previously, but they may also be used by comparing the crash frequency of nearby
segments or regions (Eksler and Lassarre, 2008; Mitra, 2009; Quddus, 2008; Wang et al., 2009). The
researcher may choose from a variety of methods to define which segments or regions are considered the
neighbor of another, such as contiguity or a fixed distance. In this case, conditional autoregressive (CAR)
random effects are shown to reduce the model error by adding the prior knowledge of neighboring regions
and segments, leading to better parameter estimates. Additionally, CAR random effects are also frequently
paired with an uncorrelated random effects term, which quantifies the model error that is not related to the
nearby regions or segments, but rather unknown or unmeasured influences (Aguero-Valverde and Jovanis,
2008; Eksler and Lassarre, 2008; Guo et al., 2010; Mitra, 2009).

Ultimately, random effects may be used to reduce model error that is caused by data that is
unavailable or unrecorded. Random effects models are compatible with this limitation of motorcycle crash
data because random effects may reduce the model error due to data that is commonly missing, such as
motorcycle specific ADT or VMT (Kyrychenko and McCartt, 2006; Pai et al., 2009; Sass and Leigh, 1991).
This benefit, combined with the opportunity to include additional descriptive data and gain an interpretable
result of the parameter effect, makes Bayesian hierarchical models a strong candidate to model motorcycle
-crash data.

5.2 Data

Three datasets (ODOT, 2011; ODPS, 2011; US Census Bureau, 2011) were used in this study. The
first dataset is provided by the Ohio Department of Transportation (ODOT) and is composed of 32,289
interstate, US route, and state route segments. This dataset, shown in Table 5.1, includes the following:
pavement type, lane width, shoulder width, number of lanes, median presence, horizontal and vertical curve
related statistics, the overall vehicle ADT, and the length of the segment. Secondary information may be
extracted from the initial dataset to calculate the number of horizontal curves per segment, horizontal
curves per mile, maximum degree of curve, and percent of the segment that is a horizontal curve. Similarly,
Table 5.1: A summary of the descriptive statistics of covariates of potential predictors for multi-vehicle motorcycle crashes in Ohio at the segment level

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-vehicle Motorcycle Crashes (^a)</td>
<td>0.1</td>
<td>0.4</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>ADT (veh/day) (^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 to 1,000</td>
<td>8.3%</td>
<td>2,664</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,001 to 5,000</td>
<td>37.1%</td>
<td>11,982</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5,001 to 10,000</td>
<td>21.3%</td>
<td>6,892</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10,001 to 30,000</td>
<td>23.0%</td>
<td>7,433</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥30,000</td>
<td>10.3%</td>
<td>3,318</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length (miles) (^b)</td>
<td>0.6</td>
<td>0.9</td>
<td>0.01</td>
<td>8.2</td>
</tr>
<tr>
<td>Number of Lanes (^b)</td>
<td>2.8</td>
<td>1.3</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Shoulder Width (ft.) (^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;4 ft.</td>
<td>47.9%</td>
<td>15,454</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 to 8 ft.</td>
<td>23.7%</td>
<td>7,639</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥8 ft.</td>
<td>28.5%</td>
<td>9,196</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Divided Highway (ft.) (^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undivided</td>
<td>77.8%</td>
<td>25,129</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 to 30 ft. (divided)</td>
<td>7.6%</td>
<td>2,450</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥30 ft. (divided)</td>
<td>14.6%</td>
<td>4,710</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane Width (ft.) (^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 to 11 ft.</td>
<td>36.7%</td>
<td>11,863</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥11 ft.</td>
<td>63.3%</td>
<td>20,426</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed Limit (mph) (^a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;35 mph</td>
<td>5.0%</td>
<td>1,618</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35 to 45 mph</td>
<td>25.3%</td>
<td>8,156</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥45 mph</td>
<td>69.7%</td>
<td>22,515</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Horizontal Curves (^b)</td>
<td>1.6</td>
<td>9.4</td>
<td>0</td>
<td>327</td>
</tr>
<tr>
<td>Maximum Degree of Curve (^b)</td>
<td>2.5</td>
<td>6.7</td>
<td>0</td>
<td>158</td>
</tr>
<tr>
<td>Horizontal Curves/mile (^b)</td>
<td>1.0</td>
<td>3.8</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Percent Horizontal Curve (^b)</td>
<td>5.5%</td>
<td>16.7%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>Number of Vertical Curves (^b)</td>
<td>1.6</td>
<td>8.3</td>
<td>0</td>
<td>294</td>
</tr>
<tr>
<td>Maximum Grade (+/-) (^b)</td>
<td>1.5</td>
<td>3.3</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Vertical Curve/mile (^b)</td>
<td>1.1</td>
<td>3.7</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Percent Vertical Curve (^b)</td>
<td>9.9%</td>
<td>24.8%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>Lane Miles (ln*mi.) (^b,c)</td>
<td>127.6</td>
<td>79.8</td>
<td>.01</td>
<td>813.2</td>
</tr>
<tr>
<td>Area (mi(^2)) (^c)</td>
<td>28.3</td>
<td>11.2</td>
<td>.01</td>
<td>86.9</td>
</tr>
<tr>
<td>Road Density(ln*mi./mi(^2)) (^b)</td>
<td>3.6</td>
<td>1.5</td>
<td>2.2</td>
<td>9.2</td>
</tr>
<tr>
<td>% Over 65(^d)</td>
<td>14.6%</td>
<td>2.1%</td>
<td>8.8%</td>
<td>19.1%</td>
</tr>
<tr>
<td>Travel Time to Work (min)(^d)</td>
<td>23.6</td>
<td>3.8</td>
<td>17.4</td>
<td>34.1</td>
</tr>
<tr>
<td>Motorcycle Registration (^a)</td>
<td>4,371</td>
<td>4,839</td>
<td>417</td>
<td>27,100</td>
</tr>
<tr>
<td>Motorcycle Endorsements (^a)</td>
<td>8,253</td>
<td>9,2</td>
<td>982</td>
<td>47,003</td>
</tr>
<tr>
<td>Population (^d)</td>
<td>140,353</td>
<td>230,136</td>
<td>13,435</td>
<td>1,280,122</td>
</tr>
</tbody>
</table>

\(^a\)Data from the Ohio Department of Public Safety (ODPS, 2011)

\(^b\)Data from the Ohio Department of Transportation (ODOT, 2011)

\(^c\)Data from the Public Utilities Commission of Ohio (PUCO, 2011)

\(^d\)Data from the US Census Bureau (US Census Bureau, 2011)
the number of vertical curves, vertical curves per mile, maximum grade, and the percent of the segment that is a vertical curve is extracted from the vertical curve data. In addition to the roadway segments, township information from the ODOT dataset, including the number of lane miles, area of the township, and the urban status of the township, is used to capture information about the 1,459 townships in Ohio. Of the 1,459 townships, 940 are considered urban townships in this study. A township is designated as urban if an incorporated city, based on the city boundaries provided by ODOT, is located inside the region (ODOT, 2011). All the variables listed above are considered as fixed effects parameters with the exception of the ADT and segment length, which are entered into the model as an offset as shown:

\[
x_{\text{offset},h} = \frac{(ADT_h \times L_h)}{10^6}
\]

where \(ADT_i\) is the ADT of segment \(i\), \(L_i\) is the segment length, and \(x_{\text{offset},i}\) is the value of the offset for the segment. The offset accounts for the exposure of each segment to multi-vehicle motorcycle crashes. Although the data in this study are missing specific measures of motorcycle travel, these two means of exposure along with each random effects term reduce the error due to the lack of data.

US Census data (US Census Bureau, 2011) was used to include demographic information that described the different regions of Ohio in a manner similar to Aguero-Valverde and Jovanis (2006). Knowledge about the household demographics – such as the percent of residents over age 65, percentage of residents under the poverty level, and the mean travel time to work – was included in the model. In addition to demographic information, the county population, number of motorcycle endorsements (motorcycle licenses), and number of registered motorcycles were used as measures of motorcycle and motor vehicle traffic and were compiled at a regional level.

The number of multi-vehicle motorcycle crashes that occurred between 2006 and the summer of 2011 on each segment is determined by combining Ohio crash data as reported by the Ohio Department of Public Safety (ODPS) with the ODOT geographic locations of each roadway segment (ODPS, 2011; ODOT, 2011). A total of 3,804 non-intersection related multi-vehicle motorcycle crashes are found to have occurred on state-maintained roadways from 2006 through the summer of 2011; this total includes 68 fatal crashes and 1,163 injury crashes, and the crashes involve as many as eight vehicles in a single incident.
Geographic coordinates, which are used to identify the segment on which each crash occurred, are available for 3,379 crashes (ODPS, 2011). Segments with no geographic coordinates and those having unrealistic characteristics (such as excessive lane widths) are removed, and the remaining 3,119 multi-vehicle motorcycle crashes are considered in this study.

5.3 Methodology

Negative binomial modeling with Bayesian inference is commonly used to predict crashes. Within this practice, Bayesian inference differs from traditional statistics in that the parameters are estimated using prior knowledge, as shown in Bayes Theorem:

\[ p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \quad (5.2) \]

where \( p(\theta|y) \) represents the posterior density, \( p(y|\theta) \) denotes the model likelihood, \( p(\theta) \) is the known background information, and \( p(y) \) represents the unconditional density of the data. Guo et al. (2010) and Congdon (2010) present a detailed description of Bayes’ theorem, and the application of Bayesian negative binomial models may be found in examples such as Haque et al. (2010), Mitra (2009), Noland and Quddus (2004), or Quddus (2008).

Within the last few years, researchers such as Aguero-Valverde and Jovanis (2010) and Wang et al. (2009) have been introducing random effects into models so as to include information that may be either unavailable or may be difficult to express in the form of fixed effects. This form of modeling is particularly advantageous in studies such as this, since motorcycle-specific ADT and VMT are difficult to measure. In order to measure the impact of the random effects, two types of negative binomial models are considered in this study: an uncorrelated heterogeneity model (UH model) and spatially correlated random effects models (SC model). In both cases, full Bayesian negative binomial models are specified with normally distributed random effects terms at the county and township levels. These models are selected to improve the estimate of the multi-vehicle motorcycle crash parameters by acknowledging that some of the variation in crash frequency may be attributed to regional characteristics. To maximize the potential model improvement, multiple SC models will be specified, each controlling the amount of information about nearby segments.
by changing the radius at which segments are considered neighbors: the neighborhood radius. This will allow comparisons to be made between SC models and between the UH model and each of the SC models.

5.3.1 Uncorrelated Heterogeneity Model:

The fixed effects terms of the uncorrelated heterogeneity model are measured on two spatial levels: segment and township. The township level fixed effects parameters contribute regional information into the model and are used to calculate their respective random effects terms. This structure provides a mechanism to include a variety of information on two different spatial levels:

$$\ln \lambda_h = \sum_{k=1}^{K} \alpha_k (x_h - \bar{x}_k) + b_t + c_j + v_h$$ (5.3)

$$b_t \sim Normal(\bar{b}_t, \sigma_b), \quad \bar{b}_t = \sum_{l=1}^{L} \beta_l (z_l - \bar{z}_l)$$ (5.4)

$$c_j \sim Normal(\bar{c}_j, \sigma_c), \quad \bar{c}_j = \sum_{j=1}^{J} \gamma_j (w_j - \bar{w}_j)$$ (5.5)

where:
- $h$: segment level index
- $i$: township level index
- $j$: county level index
- $k$: predictor index
- $\lambda_h$: expected number of crashes per segment
- $\bar{b}_t$: mean of township level random effects
- $\bar{c}_j$: mean of county level random effects
- $v_h$: uncorrelated random effects
- $\alpha_k$: segment level parameters
- $\beta_k$: township level parameters
- $\gamma_k$: county level parameters
- $x_h$: segment level predictors
- $z_l$: township level predictors
- $w_j$: county level predictors
- $\sigma_b$: standard deviation of township means
- $\sigma_c$: standard deviation of county means
- $\bar{x}_k$, $\bar{z}_l$, $\bar{w}_j$: means of respective levels

The values of the random effects terms for each township and county also give insight into the magnitude and sign of the regional influence. These random effects terms help to interpret models developed using large statewide datasets. For more information on multi-level models, consider Yanmaz-Tuzel and Ozbay (2010) and Congdon (2003).

The uncorrelated random effects term also describes error that is caused by uncorrelated heterogeneity. The uncorrelated random effects term in this study is defined as follows:

$$v_h \sim Normal(0, \tau_v)$$ (5.6)
where \( \nu_{h} \) represents the normally distributed uncorrelated random effects term and \( \tau_{\nu} \) is the precision, or inverse variance, of the uncorrelated random effects, which is given an uninformative gamma prior distribution (Wang et al., 2009). The UH term prevents the other random effects terms from inferring undue correlation, and therefore improves the parameter estimates. In turn, the Bayesian credible interval (BCI) for each parameter is narrower after removing some of the model error.

5.3.2 Spatially Correlated Random Effects Model:

A second method of reducing the error of a model is to introduce a spatial random effects term. By drawing inference from the structure of the model error, spatial random effects allow information from neighboring segments to reduce the model error. Additionally, transportation safety researchers often consider spatial correlation in order to reduce model error due to omitted or unavailable crash predictors (Quddus, 2008). There is a potential drawback of using this modeling technique: the spatial random effects term might assume that spatial correlation is present in situations where it may not be. To account for this, an uncorrelated random effects term as shown in Equation 5.5 is added to help prevent undue inference of spatial correlation by explaining model error caused by uncorrelated heterogeneity between segments or regions (Mitra, 2009; Wang et al., 2009).

The spatial correlation in this study is structured as a Gaussian conditional autoregressive (CAR) prior distribution, such that:

\[
P(u|r) \propto \frac{1}{r m^{2}} \exp \left( -\frac{1}{2r} \sum_{i} \sum_{j \in \delta_{i}} (u_{i} - u_{j})^{2} \right), \quad i \neq j
\]  

(5.6)

where \( m \) is the index of the number of neighbors in the adjacency matrix, \( \delta_{i} \) is the neighborhood of the segment \( i \), \( r \) is the prior knowledge of the segment, and \( j \) denotes the neighbor of region \( i \) (Lawson, 2009).

This yields a normally distributed spatially correlated error term as proposed by Besag (1974):

\[
u_{h} \sim \text{Normal} \left( \frac{\sum_{j} w_{hj} u_{hj}}{\sum_{j} w_{hj} \sum_{j} w_{hj}}, \frac{\tau_{u}}{\sum_{j} w_{hj}} \right), \quad h \neq j
\]  

(5.7)
where $u_h$ is the spatial random effects term for segment $h$, $u_j$ is the neighbor of region $h$ given index $j$, $w_{ij}$ is the weight of the neighbor (in this case 1 for neighbors, 0 otherwise), and $\tau_u$, the precision of $u_h$, given an uninformative gamma prior distribution, as applied in a study of London crash data by Quddus (2008).

In this study, neighborship is defined at the segment level and is based on the distance between two segments in any direction. To assess the effect of changing the neighborhood radius, 11 radii are chosen between 0 and 7 miles.

5.3.3 Model Evaluation:

The deviance information criterion (DIC) is often used to assess the goodness-of-fit of hierarchical Bayesian models (Aguero-Valverde and Jovanis, 2008; Guo et al., 2010; Mitra, 2009). The DIC is derived from deviance-based model statistics:

$$DIC = \bar{D} + p_D = D(\bar{\theta}) + 2p_D$$  \hspace{1cm} (5.9)

where $\theta$ is the parameter set, $\bar{D} = E_{\theta|y}(D)$, or the average of the deviance, $p_D$ is the effective number of parameters for a hierarchical model, and $D(\bar{\theta}) = D\left[ E_{\theta|y}(\theta) \right]$, which is the deviance at the posterior means of the parameters (Aguero-Valverde and Jovanis, 2010; Congdon, 2010; Spiegelhalter, 2002). Since the DIC increases as model deviance grows, a model that shows a significant reduction in DIC, typically considered 7 to 10 points or more, performs better than the model with a greater DIC (Spiegelhalter, 2002). Ultimately, the lower the DIC, the more efficient the model will be.

Using IBM SPSS Statistics V. 19, an initial model is run to assess the statistical significance (p-value less than 0.05) of each predictor. WinBUGS V. 1.4.3 is used to run subsequent analyses of each of the 11 models with neighborhood radii between 0 and 7 miles. Changing the neighborhood radius caused significant changes in the DIC, as shown in Table 5.2. The rationale behind considering multiple models is to explore the impact of changing the amount of spatial knowledge included in the spatial random effects term’s neighborhood. In each of the models, convergence is verified through the Gelman-Rubin statistic (Lawson, 2009). The analysis of each of the models included 20,000 iterations after convergence. In each of the 11 models, the covariates are the same and are statistically significant.
Table 5.2: A comparison of model goodness-of-fit for different neighborhood radii through the DIC

<table>
<thead>
<tr>
<th>Model Number</th>
<th>Neighborhood Radius</th>
<th>Dbar</th>
<th>Dhat</th>
<th>pD</th>
<th>DIC$^a$</th>
<th>DIC difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1$^b$</td>
<td>0</td>
<td>17320</td>
<td>16831</td>
<td>488</td>
<td>17808</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>16292</td>
<td>14671</td>
<td>1621</td>
<td>17913</td>
<td>105</td>
</tr>
<tr>
<td>3</td>
<td>0.75</td>
<td>15899</td>
<td>13997</td>
<td>1902</td>
<td>17802</td>
<td>-7</td>
</tr>
<tr>
<td>4</td>
<td>1.25</td>
<td>17088</td>
<td>16382</td>
<td>706</td>
<td>17795</td>
<td>-13</td>
</tr>
<tr>
<td>5</td>
<td>1.5</td>
<td>16243</td>
<td>14744</td>
<td>1499</td>
<td>17741</td>
<td>-67</td>
</tr>
<tr>
<td>6</td>
<td>1.75</td>
<td>16422</td>
<td>15103</td>
<td>1319</td>
<td>17741</td>
<td>-68</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>17134</td>
<td>16477</td>
<td>657</td>
<td>17792</td>
<td>-17</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>16944</td>
<td>16058</td>
<td>886</td>
<td>17830</td>
<td>22</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>17129</td>
<td>16468</td>
<td>661</td>
<td>17791</td>
<td>-17</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>17069</td>
<td>16364</td>
<td>705</td>
<td>17774</td>
<td>-34</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>17059</td>
<td>16331</td>
<td>729</td>
<td>17788</td>
<td>-20</td>
</tr>
</tbody>
</table>

$^a$DIC = $\bar{D} + pD = D(\hat{\theta}) + 2pD$

$^b$Base model

Note: A significant change in DIC may be considered to be 7 to 10 points or more (Spiegelhalter, 2002). The bold values are statistically different from the base model.

5.4 Results and Discussion

Each model showed the same covariate trends in terms of sign and magnitude. The results for the 1.75 mile radius, or Model 6 shown in Table 5.3, are selected on the basis of efficiency and a smaller DIC value. The neighborhood radius has a significant impact on the goodness-of-fit of each model, as measured through the DIC, which is shown graphically in Figure 5.1. Almost all models with radii over 0.75 miles are significantly improved from the UH model. For smaller radii, such as those less than 0.5 miles, the DIC is not improved, but rather the model error is increased, showing that small neighborhood radii may constrict the knowledge regarding spatial correlation between the segments and may not deliver an entirely representative picture about the way that the segments are related spatially. The DIC is less reduced for 2 mile radii and larger. Simultaneously, the cost in terms of computational power and data preparation increases at an exponential rate as its radius increases. Therefore, since the 1.75 mile radius achieves the largest reduction in DIC with the greatest efficiency, Model 6 is selected.

While motorcycle ADT is not available, total ADT is commonly considered as an offset, such as in Wang and Abdel-Aty (2008), Aguero-Valverde and Jovanis (2006), or Eksler and Lassarre (2008).
Table 5.3: A summary of the results of the segment level multi-vehicle motorcycle crash model where the neighborhood radius equals 3 miles

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>S.D.</th>
<th>95% BCI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>Geometry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-6.5</td>
<td>0.18</td>
<td>-6.8</td>
</tr>
<tr>
<td>ln(ADT)</td>
<td>0.45</td>
<td>0.02</td>
<td>0.42</td>
</tr>
<tr>
<td>ln(Length) (mi)</td>
<td>0.81</td>
<td>0.02</td>
<td>0.77</td>
</tr>
<tr>
<td>Lanes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Lanes</td>
<td>0.04</td>
<td>0.02</td>
<td>-0.006</td>
</tr>
<tr>
<td>Shoulder Width</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;4 ft.</td>
<td>0.55</td>
<td>0.06</td>
<td>0.44</td>
</tr>
<tr>
<td>4 to 8 ft.</td>
<td>0.46</td>
<td>0.07</td>
<td>0.33</td>
</tr>
<tr>
<td>≥8 ft.</td>
<td>Base</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Divided Highway</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undivided</td>
<td>Base</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 to 30 ft.</td>
<td>0.07</td>
<td>0.08</td>
<td>-0.09</td>
</tr>
<tr>
<td>≥30 ft.</td>
<td>-0.01</td>
<td>0.06</td>
<td>-0.22</td>
</tr>
<tr>
<td>Lane Width</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 to 11 ft.</td>
<td>0.12</td>
<td>0.06</td>
<td>0.006</td>
</tr>
<tr>
<td>≥11 ft.</td>
<td>Base</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horizontal Curves</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max Degree</td>
<td>0.02</td>
<td>0.004</td>
<td>0.01</td>
</tr>
<tr>
<td>Horizontal Curve per mile</td>
<td>0.02</td>
<td>0.008</td>
<td>0.005</td>
</tr>
<tr>
<td>Vertical Curves</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max Grade (+/-)</td>
<td>0.05</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Percent Vertical Curve</td>
<td>-0.30</td>
<td>0.13</td>
<td>-0.54</td>
</tr>
<tr>
<td>Township Level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane Miles (ln*mi.)</td>
<td>0.002</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>Area (mi²)</td>
<td>-0.01</td>
<td>0.003</td>
<td>-0.02</td>
</tr>
<tr>
<td>Urban</td>
<td>-0.007</td>
<td>0.06</td>
<td>-0.12</td>
</tr>
<tr>
<td>Random Effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.D. of CAR</td>
<td>0.02</td>
<td>0.002</td>
<td>0.01</td>
</tr>
<tr>
<td>S.D. Uncorrelated</td>
<td>0.05</td>
<td>0.02</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Segment length was also considered as an offset as suggested in Lord et al. (2005) or Ma et al. (2008). The parameter of the offset is fixed to one in order to acknowledge that the variables account for exposure to crashes and are not causative factors.

Concrete paved segments are found to have lower crash rates than segments paved with asphalts, similar to the findings of Geedipally et al. (2012). Segments with six or more lanes are found to experience lower crash rates, such as in Wang et al. (2009). This result may be explained by the increased possibility of changing lanes to avoid a crash, as well as higher design standards. Other findings show that
Figure 5.1: The change in DIC with neighborhood radius in miles. The horizontal line represents the value of DIC when spatial correlation is not considered. The horizontal line signifies a significant difference in DIC.

Note: A significant change in DIC, which is indicated by the dotted line, may be considered to be 7 to 10 or more (Spiegelhalter, 2002).

narrower shoulders may offer fewer chances to move away from a dangerous situation. Divided highways tend to have fewer multi-vehicle motorcycle crashes, and reduce the frequency of multi-vehicle motorcycle crashes, again suggesting that more available space in this area allows the motorcyclist and the driver the ability to escape a crash. This finding is similar to the results of Ahmed et al. (2011), which found wider medians to reduce the frequency of crashes. Smaller lane widths are associated with an increase in the frequency of multi-vehicle crashes in each model. Similar to shoulder widths, narrower lanes suggest that there may not be sufficient space for a motor vehicle or motorcycle to find an escape route. The results for the lane width and shoulder width are consistent with current motor vehicle research, such as Aguero-Valverde and Jovanis (2010).

In regard to horizontal and vertical curves, the model results are found to be consistent with Ma et al. (2008), showing an increasing correlation between the maximum degree of curvature and the frequency of motorcycle crashes. Likewise, the results show that the maximum grade, which is the highest absolute
value of the grade, is associated with an increase in the frequency of motorcycle crashes, which is consistent with Wang (2009). The percentage of the segment that is considered a vertical curve is found to be a significant predictor of multi-vehicle motorcycle crashes. There is an inverse relationship between the topography and the number of multi-vehicle crashes; one possible explanation for this finding may be related to the reduced sight distances, as drivers and riders may consider these segments to be inherently dangerous and will, in turn, travel more conservatively.

At the township level, the frequency of multi-vehicle motorcycle crashes is found to increase with the number of lane miles, suggesting that multi-vehicle motorcycle crashes tend to occur more frequently in areas with a larger amount of developed infrastructure. Urban townships are found to have fewer multi-vehicle motorcycle crashes. This may be caused by higher design standards generally in place in urban road segments.

At the county level, three covariates were found to be significant: number of motorcycle endorsements (motorcycle licenses), county population, and mean travel time to work for residents. The findings suggest that counties with more endorsed riders tend to have lower motorcycle crash rates. The parameter for county population, which is positive, complements this value. The findings show that multi-vehicle motorcycle crashes tend to occur more often in regions that are more densely populated. Finally, the mean travel time to work for residents of a county was found to be a significant, positive-valued predictor of multi-vehicle motorcycle crashes.

5.5 Conclusion

In this study, consideration of spatial random effects improved the estimates of multi-vehicle motorcycle crashes at the segment level. In general, the inclusion of random effects reduced the standard deviation of the parameters and led to a narrower BCI. The analysis of motorcycle crashes at the segment level is found to be useful for estimating the impacts of many geometric factors of roadways on the number of crashes. Hierarchical Bayesian modeling provides more confident estimates of the impact of geometric properties, especially with spatial correlation.
In this report, two separate model styles were developed to describe motorcycle crashes in the state of Ohio. The two different styles, as outlined in Chapter 4 and Chapter 5, yield two distinct levels of interpretation. Each of these interpretations has advantages and disadvantages. Although a single vehicle data set was applied in Chapter 4 and a multi-vehicle data set was applied in Chapter 5, each method could be applied to either data set. The methods in this study could also be applied to overall motor vehicle crashes or another specific type of crash.

Comparison of the two methods provides a few challenges. While the DIC provided an excellent mode of comparison between models of the same type in each study, it is not an appropriate measure of comparison between the two different types. A cursory glance shows that each DIC at the segment level is several times as large as any DIC from the regional level models. Between models of the same type, this finding would show that the regional level models were significantly better tools to model motorcycle crashes. However, this would be a misinterpretation of the DIC statistic and the components of its calculation. The number of samples and the effective number of parameters are significant factors in the magnitude of the DIC. At the regional level, 1,459 distinct regions were used to describe the distribution of motorcycle crashes across the state of Ohio. At the segment level, 32,289 roadway segments were considered across the entire state of Ohio. Therefore, the amount of error per unit of analysis varies significantly across studies. Many more segments were considered than regions, and each region contributes to the amount of error, of which DIC is a measure. There is no true method of identifying one method as the model of choice.
The true answer to the question of model choice is the desired interpretation of the final results. Regional level models provide an excellent overview of the frequency of crashes in each unit of areal analysis. As an administrator over a large area that is broken into regions, such as the townships and counties considered in this study, this may be the most beneficial method of analysis. The regional models are also computationally simpler than the segment level models. This may be an advantage if the goal of applying statistical models is to gain an understanding of the relationship between each region. The segment level model would give far more information than necessary for this type of interpretation. Additionally, modeling at the segment level may eliminate regional parameters that are of interest to the researcher in lieu of segment level parameters that describe the segments better. However, this would not be advantageous if the segment level characteristics are not of primary concern. For example, county level population was a significant predictor in segment level models, but was at the regional level. This is likely because the effects of a larger population are better described by other segment level predictors, such as ADT or the number of lanes on each segment. In some cases, the parameter information at the segment level may not be useful information, and thus would reduce the effectiveness of the interpretation of the results.

The segment level models do provide a wealth of detailed information about the study area when desired. At this level, the crashes are sorted into many more units of analysis. This may identify smaller areas of increased motorcycle crash risk that may have been passed over by a regional level model. All of this additional information may provide a clearer understanding of the distribution of crashes across the state. The segment level parameters also may have valuable interpretations that are not included in the regional level models. For example, this approach could be shown to validate a proposed improvement to a segment, such as widening the shoulders, or increasing the median width. Segment level models are also the best model choice when considering a smaller study area. Although both study areas included the entire state in this report, if only a few counties or townships were considered, the township level and county level regional information may not provide results that are very meaningful. However, a fair amount of roadway segments would likely be present in each region, which would allow for more conclusive results about the spatial distribution of crashes. Alternatively, smaller regions could be specified, such as block
groups. However, information at this level is not always available consistently, especially for areas with smaller populations.

The most thorough method of application for the models in this study can be described in three parts. First, the data set should be analyzed visually. For spatial data, this may include both descriptive statistics and cartographic methods. This initial step is important because it allows the researcher to identify trends in the data set, and potentially discover incorrect or missing data. This provides a baseline for comparison with the model results to assess whether the results reflect the real-world situation. This step also may trigger ideas for additional predictors that may enhance the model results. Next, the regional methods outlined in Chapter 4 could be applied. As stated previously, these methods may provide an excellent overview of a study area and the demographic, infrastructural, or other characteristics and their effect on the frequency of crashes in a region. Applying this method yields the benefits described in Chapter 4. Finally, a researcher could complete the analysis of the data set by applying the segment level procedures outlined in Chapter 5. Applying both procedures to a data set allows a researcher to consider trends that occur on the regional and segment levels. Considering spatial correlation throughout both of these procedures helps ensure that the parameters included in the model are not affected by missing or omitted predictors. Although both methods are similar in terms of spatial and uncorrelated random effects and the negative binomial methodology, both Bayesian models are useful for describing the spatial distribution and frequency of vehicle crashes.
LITERATURE CITED


Houston, D. J. & Richardson, L. E., Motorcyclist fatality rates and mandatory helmet-use laws. Accident Analysis and Prevention, 200-208, 2008.

Huang, H., Chin, H.C., and Haque, M., Severity of Driver Injury and vehicle damage in traffic crashes at intersections: A Bayesian hierarchical analysis, Accident Analysis & Prevention, Volume 40, Issue 1, pp.45-54, 2008.


Jonsson, T., Ivan, J. N., & Zhang, C., Crash prediction models for intersections on rural multilane highways: Differences by collision type, Transportation Research Record 2019: Journal of the Transportation Research Board, pp. 91-98, 2007.


http://www.dps.state.oh.us/CrashRequests/Common/TTDoc.pdf

Ohio Department of Public Safety (ODPS). (2012). Crash TRACTAPE. OH.

Ohio Department of Transportation (ODOT), ESRI/Statewide Digital Data,
http://www.dot.state.oh.us/Divisions/Planning/TechServ/prod_services/Pages/ESRI DwnLds.aspx as of July 2011.


Public Utilities Commission of Ohio (PUCO), Statewide Digital Data,
http://www.puc.state.oh.us/gis/pucodata.cfm as of July 2011.


Schneider IV, W.H., Savolainen, P.T., & Moore, D.N., Effects of horizontal curvature on single-vehicle motorcycle crashes along rural two-lane highways, Transportation Research Record: Journal of the Transportation Research Board, No. 2194, pp. 91-98, 2010.


