Snow and Non-Snow Events Based Winter Traffic Crash Pattern Analysis and Developing Lake Effect Snow Induced Crash Count Prediction Model

Bandhan Dutta Ayon
Western Michigan University, b.dutta.ayon@gmail.com
SNOW AND NON-SNOW EVENTS BASED WINTER TRAFFIC CRASH PATTERN ANALYSIS AND DEVELOPING LAKE EFFECT SNOW INDUCED CRASH COUNT PREDICTION MODEL

by

Bandhan Dutta Ayon

A thesis submitted to the Graduate College in partial fulfillment of the requirements for the degree of Master of Science Geography Western Michigan University June 2017

Thesis Committee:
Benjamin Ofori-Amoah, Ph.D., Chair
Lei Meng, Ph.D.
Laiyin Zhu, Ph.D.
Snow and non-snow events based winter traffic crash pattern analysis and developing Lake Effect snow induced crash count prediction model

Bandhan Dutta Ayon, M.S.

Western Michigan University, 2017

This study aims to (i) find the spatial-temporal traffic crash pattern and (ii) develop a statistical model for predicting LES induced traffic crash counts. Three southwest Michigan counties: Allegan, Kalamazoo and Calhoun are selected to conduct this study. Snow and non-snow event based comparative analysis are conducted using Getis-Ord Gi* statistic to identify high density crash cluster locations in the study area. While several new and oscillating hot spots are detected in Allegan during snow, no hot spot is detected during non-snow events based analysis. In Kalamazoo, traffic crashes do not exhibit much difference in spatial-temporal trend during snow or non-snow weather. Some new hot spots and numerous sporadic and oscillating hot spots are observed in Calhoun County roadways during snow, which are found as persistent hot spot locations from non-snow event based analysis. Negative binomial regression models with temporal random effects are fitted to the data treating daily average traffic crash counts as response variable; and temperature difference between Lake surface and overlaying air, wind speed, and wind direction as explanatory variables. All the variables exhibit statistically significant positive estimates to predict LES induced traffic crash counts for Allegan and Kalamazoo. However, wind direction is found statistically insignificant for Calhoun County model. These results will help further researches to explore LES induced traffic crashes and devise seasonal countermeasures.
ACKNOWLEDGEMENTS

I would like to express my heartiest gratitude to my committee for their expert advice, and encouragement: Dr. Benjamin Ofori-Amoah, my committee chair; Dr. Lei Meng; and Dr. Laiyin Zhu. My completion of this thesis could not have been accomplished without their continued support. However, I am especially indebted to Dr. Benjamin Ofori-Amoah, Chairman of the Department of Geography. As my teacher and mentor, he has provided me extensive personal and professional guidance.

I would especially like to thank Dr. David Lemberg for his extraordinary ideas to carry out research on lake effect snow induced traffic crashes. I cannot express enough thanks to Dr. Gregory Veeck. He has provided me so much time and support that more than I could ever give him credit for. I would also like to acknowledge Dr. Rajib Paul’s valuable assistance for conducting statistical analysis. I would also like to thank Dr. Kathleen Baker for her ideas and support which helped me to use a newly incorporated ArcGIS tool to perform spatial-temporal crash pattern analysis.

My heartfelt thanks to my loving and supportive wife, Prajna, for providing continuous inspiration. I also want to thank my parents. I know their love and guidance are always with me in whatever I pursue. Finally, I offer my sincere appreciation to the Almighty God.

Bandhan Dutta Ayon
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ..................................................................................................................... ii

LIST OF TABLES ................................................................................................................................. vi

LIST OF FIGURES ............................................................................................................................... vii

CHAPTER

I. INTRODUCTION .......................................................................................................................... 1
   Research Objectives ......................................................................................................................... 3
   Study Area ......................................................................................................................................... 4
   Data and Data Collection .................................................................................................................. 5
   Thesis Organization ......................................................................................................................... 6

II. LITERATURE REVIEW .................................................................................................................. 7
   Lake Effect Snow and Traffic Crashes .............................................................................................. 7
   Nature of Traffic Crash Data Analysis .............................................................................................. 8
   Spatial Pattern Analysis of Traffic Crashes ....................................................................................... 8
       Spatial Pattern Analysis Technique ............................................................................................... 9
       Statistical Analysis of Different Factors Associated with Crash Occurrence ............................ 12

III. METHODOLOGY ......................................................................................................................... 16
   Spatial-temporal Pattern Analysis of Winter Traffic Crashes .......................................................... 16
       Creating Space Time Cube ........................................................................................................... 17
       Emerging Hot Spot Analysis ....................................................................................................... 19
       Selection of Neighborhood Distance ........................................................................................... 20
Table of Contents—Continued

CHAPTER

Developing a Statistical Traffic Crash Prediction Model ………………23
  Variable Definition…………………………………………………….23
  Model Description…………………………………………………….25

IV. RESULTS AND DISCUSSION…………………………………………………29
  Spatial-temporal Pattern Analysis of Winter Traffic Crashes…………29
  Allegan County Traffic Crash Pattern………………………………34
  Kalamazoo County Traffic Crash Pattern…………………………38
  Calhoun County Traffic Crash Pattern……………………………40

Developing a Statistical Traffic Crash Prediction Model………………..42
  Model Description……………………………………………………44
  Model for Allegan County Traffic Crash Counts…………………44
  Model for Kalamazoo County Traffic Crash Counts……………45
  Model for Calhoun County Traffic Crash Count…………………46

V. CONCLUSION……………………………………………………………………49
  Major Findings…………………………………………………………50
  Limitations……………………………………………………………..53
  Future Studies………………………………………………………….54

REFERENCES………………………………………………………………..56
APPENDICES

A. Allegan County Traffic Crash Hot Spot Maps
   A1. Spatial-temporal crash pattern during non-snow (north-west part)
   A2. Spatial-temporal crash pattern during non-snow (south-west part)
   A3. Spatial-temporal crash pattern during snow (north-west part)
   A4. Spatial-temporal crash pattern during snow (south-west part)

B. Kalamazoo County Traffic Crash Hot Spot Maps
   B1. Spatial-temporal crash pattern during snow (north-west part)
   B2. Spatial-temporal crash pattern during snow (south-east part)
   B3. Spatial-temporal crash pattern during snow (south-west part)
   B4. Spatial-temporal crash pattern during non-snow (north-west part)
   B5. Spatial-temporal crash pattern during non-snow (south-east part)
   B6. Spatial-temporal crash pattern during non-snow (south-west part)

C. Calhoun County Traffic Crash Hot Spot Maps
   C1. Spatial-temporal crash pattern during non-snow (north-east part)
   C2. Spatial-temporal crash pattern during non-snow (south-east part)
   C3. Spatial-temporal crash pattern during snow (south-west part)
   C4. Spatial-temporal crash pattern during snow (north-east part)
   C5. Spatial-temporal crash pattern during snow (south-east part)
   C6. Spatial-temporal crash pattern during snow (south-west part)
LIST OF TABLES

1. Number of hot spots and cold spots detected from spatial-temporal analysis ........31
2. Daily average traffic crashes in study area ............................................43
3. Model results for Allegan County traffic crash counts ............................45
4. Model results for Calhoun County traffic crash counts ............................46
5. Model results for Kalamazoo County traffic crash counts .......................47
LIST OF FIGURES

1. The location map of study area: Allegan, Kalamazoo, and Calhoun County .................. 4
2. Three-dimensional data structure of space-time bins .................................................. 17
3. Basic illustration of emerging hot spot analysis .......................................................... 19
4. Spatial autocorrelation by distance .............................................................................. 22
5. Areas receiving heavy snow for the given wind direction ............................................. 25
6. Areas receiving heavy snow for the given wind direction ............................................. 25
7. Spatial-temporal traffic crash pattern during snow in Allegan County (north-east part) ................................................................. 34
8. Spatial-temporal crash pattern during non-snow in Allegan County (north-east part) ................................................................. 35
9. Spatial-temporal traffic crash pattern during snow in Allegan County (south-east part) ................................................................. 36
10. Spatial-temporal crash pattern during non-snow in Allegan County (south-east part) ................................................................. 37
11. Spatial-temporal crash pattern during snow in Kalamazoo County (north-west part) ................................................................. 38
12. Spatial-temporal crash pattern during non-snow in Kalamazoo County (south-west part) ................................................................. 39
13. Spatial-temporal traffic crash pattern during snow in Calhoun County (north-west part) ................................................................. 40
14. Spatial-temporal crash pattern during non-snow in Calhoun County (north-west part) ................................................................. 41
15. Observed traffic crashes vs. model fit (Allegan county) ............................................. 45
16. Observed traffic crashes vs. model fit (Kalamazoo county) ........................................ 46
17. Observed traffic crashes vs. model fit (Calhoun county) ............................................. 47
CHAPTER I

INTRODUCTION

Contemporary United States society can be defined as ‘nation on wheels’ because of the residents’ intensive use of motorized vehicle to move people and goods. The observed rise in mobility, since 1945, has increased society’s exposure to potential travel risks (Haight, 1984; Andrey, 2000; WHO, 2004). In this respect, inclement weather is known to be a major factor affecting traffic safety. According to the National Research Council, 1.5 million vehicle crashes occur each year in the U.S., accounting for approximately 800,000 injuries and 7,000 fatalities, are related to adverse weather including rain, sleet, snow, fog, ice, or some combination of these factors. The injuries, loss of life, and property damage from weather related-crashes cost an annual average of $42 billion (Nation Research Council, 2004).

Winter weather in many parts of the U.S. is very severe with heavy snowfall events along with light rain and numerous freezing drizzle and freezing fog events that cause roads to be ice covered (Martinelli, 2004). Blowing snow or fogs also impair visibility for drivers. These inclement winter conditions have many negative impacts on the transportation system including the fact that low friction icy roadway makes traffic operating and maneuvering more difficult; restricted visibility limits drivers’ ability to judge the unexpected circumstances ahead; and the accumulating or blowing snow on the roadway obstructs vehicles as well as covers pavement markings, exacerbating the already worsened situation (Qin et al., 2006). Past studies revealed that, increased crash rate in winter ranges from less than 100% to more than 1,000% during snowfall (Qiu et al., 2008). It conforms with the fact that sixty-nine percent of U.S. residents live in snowy regions (with more than five inches of annual snowfall) and 74% of the nation’s roads are located in snowy areas (Pisano et al.,
But owing to the geographical locations, the nature and extent of the snowfall may vary from states to states or even within the state.

In the Great Lakes Region of the United States, the generally worse condition of transportation safety in winter is exacerbated by lake-effect snow (LES). According to National Oceanic and Atmospheric Administration (NOAA), this winter weather event occurs when a cold air mass moves across a large expanse of warmer lake water. The lower layer of the air mass is warmed causing it to pick moisture from the lake. As the moisture rises through the colder upper layer of the air above, it freezes and turns into snow which is then deposited on the leeward or downwind side. What makes LES an additional winter transportation hazard is the sudden onset of the heavy and blinding snowfall, strong winds, falling temperatures, and low visibility associated with it, and driver anxiety and confusion due to the rapidly deteriorating weather conditions (Banacos et al., 2014, DeVoir 2004).

Thus, reports of traffic crashes in the Great Lakes Region attributed to LES during winter are quite frequent (DeVoir 2004; Banacos et al., 2014). For example, in January 2014 a crash involving 40 semis and cars near Michigan City Indiana killing three people was attributed to blizzard caused by lake-effect snow (McClatchy 2014). In the same year, LES was credited with an accident involving 21 vehicles on US 131 in Southwest Michigan. In January 2015, an 18-vehicle chain-reaction crash on I-80 in Clarion County, Pennsylvania that killed two people was attributed to a white-out from lake-effect snow (Schmitz 2015). In January 2015, LES caused 170 vehicle pile-up near Galesburg, Southwest Michigan. In February, the same year 38 vehicles were involved in a crash on Chicago’s Kennedy Expressway due to LES. In January 2016, a pile-up involving 50 vehicles and one death occurred on I-94 near Hartford, Southwest Michigan during LES. However, studies that investigate the linkages between LES and traffic crashes are found in the borderline in scholarly literature on traffic safety or crash analysis.
Thus, while LES is taken as given and is identified as a cause of severe traffic crashes some important questions remain unanswered. For instance, what are the spatial and temporal patterns of winter traffic crashes in this region? Are there any consistent/emerging high density crash locations in the study area? What is the difference between spatial-temporal distribution of snow and non-snow events based incidents? What are the optimal meteorological characteristics behind LES formation that are also associated with traffic crashes? Answers to these questions are very useful for adopting seasonal traffic safety measures and the goal of this research is to explore so.

Research Objectives

Reported weather conditions during traffic incidents may contribute to crash occurrences, but do not necessarily imply causation (National Center for Statistics and Analysis, 2001; Nelson et al., 2002). However, for a traffic safety professional, the first task is to identify and prioritize safety deficient locations to implement efficient safety countermeasures, prior to addressing the root causes of crashes in relation to weather interaction with roadway geometric features and traffic conditions. Recognizing the fact, this study aims to (1) find spatial-temporal pattern of traffic crashes in the winter; and (2) develop a model for predicting traffic crashes in LES that is based on meteorological variables.

The results of the study will enhance our knowledge about LES, inform us about the major winter crash hot spots in the area, and will provide a model that can be used to predict traffic crashes in future LES. This will also help transportation safety professionals to make provisions for safety-deficient spots in the area during winter.
Study Area

Out of 9 counties of Southwest Region of Michigan Department of Transportation, three contiguous counties (Allegan Calhoun, and Kalamazoo) were considered as the study area. The region was chosen because of its (a) proximity to Lake Michigan which makes it more vulnerable to LES (b) the presence of surrounding meteorological data recording stations providing information on Great Lakes and (c) the resemblances between meteorological characteristics of the counties. Moreover, the area serves as a corridor for major highways including US 131 and Interstate 94 and 96.

Figure 1. The location map of study area: Allegan, Kalamazoo, and Calhoun County (Source: Author).
There is therefore a high volume of thru traffic and experiences considerable amount of traffic crashes in winter. Furthermore, there is also a resemblance in winter traffic crash frequency (per 1,000 population) in each county which ranges from 17-18.5 (see Figure 1).

Data and Data Collection

This study utilizes three sets of data – crash data, spatial data involving county boundary and road network of study area, and meteorological data related to LES. Crash data from 2005 to 2014 were collected from Michigan Traffic Crash Fact’s database. The data primarily contained daily information for each and every reported crash such as spatial information (e.g. the location of crash incidents based on latitude and longitude value), street type (e.g. interstate route, U.S. route, Michigan route, County road), relation to roadway (e.g. on the road, on the shoulder), temporal information for instance, the accident year, month, day and time of the day, and corresponding weather related information (e.g. clear, cloudy, rain, snow). GIS shapefiles for county boundary and road network data were obtained from Michigan Geographic Framework (MGF) database, Version 14a, 2014. Meteorological data includes lake surface temperature, overlaying air temperature, wind speed, and wind direction. Daily average data for these variables for corresponding 10 years were collected along with daily mean snowfall data for selected nine counties from the NOAA Great Lakes Environmental Research Laboratory and the Weather Underground respectively. As the goal of this research is to explore the spatial pattern of winter weather related traffic crashes and examine the association of the LES in incidents, data for winter months (November to March) were only considered for overall analysis.
Thesis Organization

The rest of the thesis comprises four chapters. Chapter II identifies researches, models, and case studies supporting the thesis topic and eventually will help to establish a theoretical framework for conducting the research. Chapter III details the research design and methods used in this study. This chapter involves discussion on what tools, measurements and statistical tests were performed for analyzing the data. Chapter IV presents both the spatio-temporal pattern analysis of traffic crashes and results of the statistical analysis performed in this study. The final chapter focuses on research outcomes, implications, and recommendations for future scholarly works.
This chapter reviews the relevant and scientifically sound literature that relate to the subject this thesis. It is divided into four sections. The first section deals with LES and traffic crashes. The second reviews the nature of traffic crash data analysis. The third, spatial pattern analysis of vehicle crashes, and the fourth reviews statistical analysis of different factors associated with crash occurrences.

Lake Effect Snow and Traffic Crashes

Lake effect snow (LES) can generate a local, severe, and narrow band of moderate to heavy snow. Most intense lake effect events can reach 3 to 5 inches per hour (Oakland County, 2013). When LES accompanies with gusty wind, this rapid onset may cause extreme traffic safety hazards. High number of traffic crashes may occur due to occasional or local snowfall (Juga 2010). Juga (2010) examined two sea-effect snowfall (similar to LES) cases occurred on 20 January, 2006 and 8 February, 2007 at a coastal area in Finland. Though the daily average of traffic crashes in the Helsinki metropolitan area was 79 during winter, the number of crashes was 371 and 219 respectively for the same area in two above-mentioned cases.

According to State of Indian crash facts 2011, northern Indiana counties experience higher numbers and rates of winter traffic crashes than southern counties due to LES from Lake Michigan. However, while mass people know anecdotally that LES affects traffic safety, barely any research has examined the relation between LES and traffic crashes. Therefore, methodologies to study LES induced traffic crashes are yet widely explored.
Nature of Traffic Crash Data Analysis

Literature of automobile crash research involves a variety of efforts designated to analyze different types of crashes under different types of conditions and factors. Quddus, (2008) and Siddiqui et al., (2012) attempt to categorize crash safety analysis and safety performance functions (SPFs) into two broad branches. One branch investigates crashes in a ‘disaggregate’ nature as these analyses do not usually consider spatial confinement or aggregation in crash data preparation. This branch analyzes crash occurrences by specific type (such as fatalities, serious injuries, slight injuries, injuries/non-injuries) and/or specific micro-level roadway locations (such as intersections, corridors, and freeways) (Levine et al., 1995; Khattak et al., 2001; Noland et al., 2004; Ghosh et al., 2004; Aguero-Valverde et al., 2006; Mitra, 2009). The other branch of safety studies examines automobile crashes using macro-structural covariates at various levels of area-aggregation (e.g. census blocks, census tracts, and Traffic Analysis Zones/TAZs) (Ng et al., 2002; Li et al., 2007; Siddiqui et al., 2012). Several researchers analyze crash data from different temporal perspectives. Levine et al. (1995); Khattak et al. (2001); Erdogan et al. (2008); and Prasannakumar et al. (2011) all measure the fluctuation of the magnitude and frequencies of crashes, injuries, and fatalities in relation to varied temporal scales, such as hourly, daily, weekly, monthly, seasonally, and yearly.

Spatial Pattern Analysis of Traffic Crashes

Irrespective of the nature of investigation, researchers conducting the majority of traffic accident studies, study the spatial patterns or spatial correlations in conjunction with other data in crash data. Different researchers use varied methodologies to analyze such crash records, but the application of spatial tools and statistical models are the most widely exercised techniques. With the advance of the Geographical Information System (GIS)
techniques, it has become easier to employ spatial analytical technique for crash data processing. Erdogan et al. (2008) note that there have been many studies incorporating GIS applications on traffic safety and accident analyses through intersection analysis, segment analysis, cluster analysis, density analysis, pattern analysis, proximity analysis, spatial query analysis and spatial accident analysis modeling techniques.

Among these analyses, the identification and analysis of crash hot spots or black spots is a standard practice throughout the United States and rest of the world as well (Mandloi et al., 2003; Anderson, 2009; Mitra, 2009). But despite all this work, past research provides no universally accepted definition of a crash ‘hot spot’ for a given set of accident locations. Usually hot spots can be defined as high risk or high frequency collision concentration locations in geographic space (Anderson, 2009; Gundogdu, 2010; Prasannakumar et al., 2011).

Spatial Pattern Analysis Technique

Kernel density estimation (KDE) is one of the most popular methods for analyzing the first order properties of a point event distribution (Silverman, 1986; Bailey et al., 1995) partially because it is very easy to understand and implement (Xie et al., 2008). Therefore, KDE is widely used in hot spot analysis where crashes are modeled as point events. Incorporation of KDE tools in some leading commercial software packages (e.g. the Spatial Analyst Extension of ESRI’s ArcGIS, CrimeStat) makes the analysis more user friendly and visually accessible. The planar KDE option is mostly used for traffic crashes hot spots analysis and detection. Examples include studies of identification of safety deficient areas in Kannur district, Kerala (Jayan et al., 2010), pedestrian crash zones detection (Nambisan et al. 2007), wildlife-vehicle accident analysis (Krisp et al., 2007), highway accident hot spot analysis (Erdogan et al., 2008) and so on. There is also evidence of using additional technique in conjunction with KDE tool to better comprehend spatial patterns of crashes. Erdogan et al.
examines crash data for Afyonkarahisar City, Turkey comparing GIS based KDE and repeatability analysis (RA) methods. Results from both methods are consistent and indicate the same potential hot spot locations, but the RA method effectively determined more hot spot locations than KDE. Based on accident data for London, UK, Anderson, (2009) uses kernel density estimation to identify hot spots. Hot spot units are further categorized into homogeneous zones using a K-means clustering algorithm based on specific environmental characteristics.

In contrast to the new developments associated with network K-function models, few studies attempt to extend the KDE methods to a network space. Borruso (2005) analyzes patterns of point events distributed through a network with a modified KDE, termed as network density estimation (NDE) in his paper, which considers the kernel as a density function based on network distances (Xie et al., 2008). Borruso (2005) pointed out the possibility of extending the standard 2-D KDE to network spaces for identifying potential ‘linear’ clusters along roadways. However, in his study, the kernel is still area based (using network service area) and the outcome (point density) is still mapped onto a 2-D Euclidian space. The resulting model is still a KDE in a planar space instead of a network space. To eschew the limitations of applying typical 2-D planar KDE methods, Xie et al. (2008) develops a network KDE approach to characterizing the spatial patterns of traffic accidents. This technique is tested with crash data for year 2005 over the road network of Bowling Green, Kentucky metropolitan area.

Though KDE accounts for spatial information but it assumes that all points in the neighborhood are alike regardless of their characteristics. In other words, this point pattern analysis treats each point with equally weight. The spatial autocorrelation (SA) method overcomes this limitation by incorporating not only the locations of point events but also their associated values. This method measures the degree of spatial autocorrelation (cluster or
dispersed distribution) and tests the assumption of independence (or randomness). SA occurs when a set of spatial features and their associated data values exhibit any systematic patterns in space. If the data values associated with nearby features are alike, the SA is positive. Negative autocorrelation applies if neighboring features display dispersed pattern because of unlike feature values, and random patterns exhibit no SA.

Gundogdu (2010) points out that numerical analyses of spatial point distributions are popular among academicians for many years and have been improved over the years. A very early work that resulted from these efforts is the renowned Moran’s I statistic introduced by Moran in 1948. Moran’s I becomes popular mostly because of its ability to specify whether the apparent similarity (or dissimilarity) in feature values (e.g., the number of crashes) at a certain space and its neighbors is greater than expected in a random distribution (Songchitruksa et al. 2010). Moran’s I only confirms the pattern of distribution (whether the features are clustered or dispersed in terms of their associated data values). A positive autocorrelation measured by Moran’s I captures the existence of both high-value clustering (hot spots) and low-value clustering (cold spots), but cannot distinguish between them (Getis et al. 1992; Ord et al. 1995; Mitchell 2005; Songchitruksa et al. 2010).

Moran’s I failure to differentiate high-value dominant and low-value dominant clusters led to consideration of new spatial autocorrelation method known as Getis–Ord (Gi*) spatial statistics, developed and modified by different scholars (Cliff and Ord 1981; Ord and Getis 1995; Anselin, 1995 and Gatrell et al., 1996). The Gi* spatial statistics can recognize and classify the tendency for positive spatial clustering and thus distinguish between the hot spots and cold spots (Ord et al. 1995).

Work by Getis and Ord, 1996 led to large number of more recent studies, such as Prasannakumar et al., 2011; Gundogdu, 2010; Khan et al., 2008. Khan et al. analyze weather related crash data for Wisconsin using GIS aided Getis-Ord Gi*(d) statistic. The results
identify spatial patterns and clusters for crashes at different locations depending upon changes in weather conditions (e.g. snow, rain, and fog). Prasannakumar et al. formulate a spatio-temporal clustering of road accidents of Thiruvananthapuram city, Kerala. ArcGIS-aided Moran's I, and Getis-Ord Gi* statistic is employed for spatial-statistical analysis of crash data for monsoon and non-monsoon seasons in this study.

Statistical Analysis of Different Factors Associated with Crash Occurrence

Some of the previous researches have also dealt with the diverse additional conditions which impact motor vehicle crashes. Investigation of different non-spatial aspects (e.g. traffic volume or speed, deer involvement, driver’s behaviors) as potential factors of crash occurrences are examples of such efforts (Allen et al., 1976; Deery et al. 1999; Chang, 2012). Other researchers explore the correlations between crash characteristics (rate, frequency, fatality, injury, duration, severity, etc.) and related variables, such as roadway geometry, road design, environment, land use (Shankar et al., 2004; Anderson, 2009).

Much of the previous research on traffic accident evaluates the relationship of accident occurrence and above mentioned influential factors by developing crash event models. A variety of statistical methods are employed in different cases. However, multiple linear regression, Poisson regression, and negative binomial regression models are most extensively used in past researches. Some researchers (Miaou et al., 1993; Abdel-Aty et al., 2000; Karlaftis et al., 1998;) reject the suitability of using multiple OLS linear regression models for modeling vehicle crash events on roadways. Some statistical assumptions such as the homoscedasticity and continuity of data make multiple linear regression models inappropriate for crash analysis. Erroneous assumptions can affect associated confidence intervals of results and thus invalidate the hypothesis tests concerning the significance of the parameters (Jovanis et al., 1986; Li et al., 2007).
Poisson regression is useful for exploring the relationship between crashes and contributing factors especially when the mean and variance of the crash frequencies are roughly equal (Miaou, 1994). According to Miaou (1994) both negative binomial (NB) regression and zero-inflated Poisson regression can be effective when over-dispersion is modest or high. When experimental variance of the data is greater than the anticipated variance or mean, the condition is termed over-dispersion (Ng et al., 2002; Li et al., 2007).

Noland et al. (2004) employs native binomial count data models to demonstrate the associations between spatially disaggregate ward level data (land use types, road characteristics) with traffic fatalities in England. Mitra (2009) and Quddus (2008) use both Bayesian statistical model and negative binomial models to detect the relationship between crash injuries and major influential factors. Ng et al. investigate the influences of potential land use causal factors on the occurrence of accidents in Hong Kong using the Empirical Bayes (EB) approach. Bayesian inference has been used in the past in disease mapping and ecological analysis (Aguero-Valverde et al., 2006) and now it is evident that this approach is also quite appropriate for accident analysis over the past decade. According to Withers (2002), the Bayesian approach has the ability to forecast risks precisely even in the presence of sparse data or sporadic events. He believes that the ability to incorporate prior knowledge without the restriction of classical distributional assumptions makes Bayesian inference a potent forecasting tool for a wide variety of research problems (Li et al., 2007). The Bayesian approaches, from empirical Bayes to full Bayes, are frequently implemented in crash analysis studies to estimate crash risk and predict crash frequency (Brüde and Larsson, 1988; Hauer, 1992, 2002; Mountain et al., 1996; Ng et al., 2002; Aguero-Valverde et al., 2006; Quddus, 2008). Another comparison of Full Bayes (FB) hierarchical models and Negative Binomial (NB) estimates is presented using injury and fatal crash data for Pennsylvania by Aguero-Valverde (2006) et al. The authors credited the FB models as more accurate with respect to
the identification of covariates (e.g. demographics, weather conditions, transportation infrastructure) with crash risk.

Extensive studies have also been conducted related to the relationship between adverse weather and crash incidences. (Khattak et al., 2001; Khan et al., 2008; Qiu et al., 2008; Andrey, 2010; Theofilatos et al., 2014;). Qin et al. (2006) claims that weather conditions such as rain, snow, sleet, fog, and ice are accountable for reducing road surface friction, impairing driver visibility, and obstructing roadway and thus engender traffic collisions. Rooney (1968) shows that collisions increased by at least 200% in areas recording 3 to 12 snow days per year. Taking traffic volume into consideration, Zhang et al. (2005) explores that the highest risk occurred at traffic flow rate from 1,200 to 1,500 vehicles per hour per lane (vphpl) under snow conditions.

Beyond simply documenting the fact that snowstorms and other adverse weather conditions have deadly effects on traffic safety, researchers attempt to identify the weather variables that directly or indirectly cause safety problems. Various environmental factors and weather parameters are studied such as pavement temperature, air temperature, atmospheric visibility, wind speed and direction, snow intensity, duration, and coverage (Qin et al., 2006). Khattak et al. (2001) conduct a study of the Interstate highway system in Iowa using detailed crash, weather, traffic exposure, and roadway geometry data. Higher wind speed (gusts) resulted in more injurious crashes, whereas higher snowfall intensity tended to result in less injurious crashes. Although the previous results vary, conclusions for such studies are fairly consistent (Qin et al., 2006). Hanbali (1994), Brown et al. (1997), and Andrey et al. (2001) all emphasize the safety benefits offered by winter maintenance roadway deicing activities. Hanbali (1994) considered the economic impacts of winter road maintenance on roadway users and found a significant decrease in crash rates after deicing maintenance activity when compared with crash rates prior to deicing. Brown et al. (1997) reported a study conducted in
Quebec, Canada, of crash rates and frequencies during winter months (December to March, inclusive) as compared to crash rates and frequencies during summer months. Their research indicated that winter months had higher minor and material damage accident rates but lower severe and fatal crash rates. They acknowledged, however, that these results could change if winter road maintenance activities were modified.

McBride et al. (1977) report an increase in severe injury crash rates in snow-belt states (i.e., Minnesota, Wisconsin, and Michigan) compared with the nonsnow-belt states during winter months. LES can be a determining factor behind the increased crash rates for the Great Lake regions. However, no specific studies have yet been conducted that relate the meteorological phenomena behind LES formation and traffic crashes in the area. Therefore, the aim of the study is to identify spatial patterns of crashes and investigate the impact of LES on crash incidents using the GIS based hot spot analysis and a Bayesian approach to negative binomial regression.
CHAPTER III

METHODOLOGY

To achieve the goal of this research, the entire study will be conducted in two major phases: i) spatial-temporal pattern analysis of winter traffic crashes, ii) developing a statistical crash count prediction model.

Spatial-temporal Pattern Analysis of Winter Traffic Crashes

According to Science Hijinks, a joint NOAA and NASA educational website, LES usually occurs between November and February. However, there is a lot of evidence of LES in the month of March including this year’s (12 – 14 March, 2017) heavy LES in Illinois (National Weather Service). Therefore, in this study winter was defined as consisting of five months (November – March). Collected crash data from 2005 to 2014 were classified as snow and non-snow events based incident. Weather conditions reported at time of crash occurrence was considered for this classification. Daily average meteorological data collected from Weather Underground were used to identify whether the incident occurred on a snowy day or not.

A newly incorporated GIS tool: Space-Time Pattern Mining Tool was used to perform the spatial-temporal pattern analysis of the traffic crashes. This tool summarized a set of points (crash locations) into a netCDF (network Common Data Form) data structure by aggregating them into space-time bins. Within each bin, all points were counted. For all bin locations, the trend for counts over time were evaluated by Getis-Ord Gi* statistic. Moreover, this tool integrated the temporal component in the analysis that the other traditional tools (KDE, Moran’s I, Getis-Ord Gi*) do not take into account. This spatial-temporal pattern
analysis was conducted using two ArcGIS tools: Creating Space Time Cube and Emerging Hot Spot Analysis.

Creating Space Time Cube

Incident points of ten years’ crash locations were used as Input Features. 22,031 incident points were used as input for snow events based crash pattern analysis. On the other hand, 61,490 non-snow event based crash locations were analyzed for identifying their inherent spatial-temporal distribution. This tool aggregated these points into space-time bins which can be conceptualized as a three-dimensional (3-D) cube. In this 3-D data structure, X and Y dimensions represented space and the t dimension represented time (please see figure 3). Every bin adheres to a static location in space (x,y) and in time (t). Therefore, bins encompassing the identical (x, y) area and the identical duration corresponded to same location ID and the same time-step ID respectively. If any location contains zero point counts for all time steps, that location would not be included in the analysis.

Figure 2. Three-dimensional data structure of space-time bins (Source: ESRI).
To precisely measure distances, all the data points were first projected from WGS 1984 Geographic Coordinate System to NAD 1983 Projected Coordinated System. These projected data were then used as input features. Output from this tool was a netCDF representing the input points as well as messages summarizing cube characteristics written to the Results window. The netCDF file created was used as input to the Emerging Hot Spot Analysis tool.

The time step interval is the incremental change in time for which the spatial-temporal trend analysis is being conducted. It helps to aggregate incident points across time. Time-step intervals are always fixed durations, and the tool requires at least ten time steps. Here the time step interval was 1 year as the study aims to capture the annual variation of winter traffic crashes in space over the decade. Date of each day associated with individual point feature was considered as the Time Field parameter.

To aggregate data points within each bin following the roadway, 250 meters by 250 meters bins were created. The length of the bin corresponds to the length of the road segments. Therefore, the distance Interval was decided carefully. Distance interval shouldn’t be too large or too small. Larger bin may cause losing the underlying patterns in crash point data. Moreover, it may include the superfluous areas outside the zone of analysis (i.e. road segment). On the other hand, smaller distance interval may lead to too many empty bins (cube filled with zero crash count). This selection of the length of road segments conforms with the study on Real-Time Crash Prediction at Tokyo Metropolitan Expressway, Japan where the crash data were aggregated for each 250 meters road segment (Hossain & Muromachi 2010). However, there is no standard for segment length for crash aggregation and it may vary from 100 m to several kilometers (Kweon et al. 2011).
Emerging Hot Spot Analysis

netCDF files created by the Create Space Time Cube tool from earlier phase was then used to identify the pattern of point distributions. Emerging Hot Spot Analysis tool helps to classify cluster of incident densities into eight specific hot or cold spot trends: new, consecutive, intensifying, persistent, diminishing, sporadic, oscillating and historical hot and cold spots. These patterns were detected using a space-time implementation of the Getis-Ord Gi* statistic. Getis-Ord Gi* considers the value for each bin within the context of the values for neighboring bins. In this study, a bin was considered a neighbor if its centroid falls within the neighborhood distance of 8597 meters and its time interval was within the neighborhood time step which was 1 year. Selection of neighborhood distance was vital and thus discussed in detail in the next section. Basic illustration of this process is presented in figure 3.

Figure 3. Basic illustration of emerging hot spot analysis (Source: ESRI).

Getis-Ord Gi* identifies whether bins with higher crash counts or bins with lower counts tend to cluster in the road segment. This tool works by looking at each space time cube within the context of adjacent cubes. If a bin's crash count value is high, and the counts for all of its neighboring bins is also high, the cluster of crashes will generate a hot spot. The statistical equation for calculating Getis-Ord Gi* can be written as,
\[ G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{\sum_{j=1}^{n} w_j^2 - (\sum_{j=1}^{n} w_{i,j})^2}{n-1}}} \]  

(1)

where, \( x_j \) is the attribute value for bin \( j \); \( \omega_{i,j} \) is a spatial weight vector between bins \( i \) and \( j \); \( n \) is equal to the number of bins and

\[ \bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \]  

(2)

\[ S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2} \]  

(3)

The result from this analysis is a z-score, p-value, and binning category for every bin in the space-time cube. The \( Gi^* \) statistics is actually a Z score. In this study, Getis-Ord \( Gi^* \) statistic computes the ‘Z’ value for each 250 meters road segment. If the crash counts at a certain bin differs from the Z values calculated from the median, and if the difference is too large to be the result of random chance, a statistically significant Z score is derived (Prasannakumar et al. 2011). p-value indicates whether the cluster is statistically significant or not. For statistically significant positive Z scores, larger Z score denotes more intense clustering of high values (hot spot). For statistically significant negative Z scores, the smaller the Z score, the more intense the clustering of lower number of crashes within the bin (cold spot). The time series of these z-scores at 250 meters road segments was then analyzed using the Mann-Kendall statistic to identify the temporal pattern.

Selection of Neighborhood Distance

Neighborhood distance is the spatial extent of the analysis. This value regulates which crashes are analyzed together in order to assess local space-time clustering. While a bin with a large point count may be interesting, unless its space-time neighbors also have large point counts, it probably won’t be a statistically significant hot spot. Since crashes within a bin are
analyzed in terms of its neighboring bins defined by a distance threshold, it is necessary to find an appropriate neighborhood distance where spatial autocorrelation is maximized. ArcGIS’s Incremental Spatial Autocorrelation (ISA) tool was used to find the neighborhood distance. Prior to run ISA tool, following tasks were exercised using the tools discussed below.

Integrate with the Collect Events tools were used to aggregate the traffic crash data in 250 meters. The Integrate tool snap incident points together within the specified distance. Collect Events create a new feature class containing a point at each unique location with an associated count attribute specifying the number of crash occurrences. The resultant ICOUNT field is then used as input field to calculate distance band from neighbor count. The ICOUNT field must contain assorted values because this statistic requires some variability in the features being analyzed. If all input values were similar, this tool fails to execute. Calculate distance band from neighbor count is ArcGIS’s spatial-statistical tool which returns three numbers: the minimum, the maximum, and the average distance to a specified number of neighbors. The maximum distance is the furthest distance and minimum distance is the closest distance between a feature and its n<sup>th</sup> neighbor for that feature. The reported average value is the average distance between all the features and their n<sup>th</sup> neighbor. The maximum and average values are then exploited to find distance band using ISA tool.

ISA evaluates spatial autocorrelation for a series of distances. Moreover, it creates a line graph of those distances and their corresponding z-scores. The magnitude of spatial clustering is indicated by z-scores. Statistically significant peak z-scores indicate distances where clustering of traffic crashes is most pronounced. To ensure that the all of the aggregated crash point has at least one neighbor, the maximum value is used as the Beginning Distance. Average value is used as the Distance Increment as increasing the distance needs to ensure the growth in number of neighbors. The ISA tool runs the Spatial Autocorrelation
(Global Moran’s I) tool for a series of increasing distances, measuring the intensity of spatial clustering for each distance.

To generate an output graph, an Output Table location should be specified. The X-axis of the graph represents the distance and Z-score results are reported on Y-axis. At some particular distance, however, the Z-score generally peaks. Peaks reflect distances where the spatial processes promoting clustering were most pronounced. One strategy for identifying an appropriate scale of analysis is to select the distance associated with the peak and often this is the first peak. Figure 4 shows the interface of the ISA tool with all the necessary input field. In this study, the peak of the z-scores were found at distance 8597 meters.

Figure 4. Spatial autocorrelation by distance (Source: Author’s calculation).
Variable Definition

Three negative binomial regression models with temporal random effects were fitted to the data treating daily average traffic crash counts as response variable; and temperature difference between Lake surface and overlaying air, wind speed, and wind direction as explanatory variables. However, to evaluate the impact of LES on traffic crashes, understanding the variables contributing to the formation of LES would be very crucial. Therefore, the following section briefly describes the variables that were used in model calibration.

*Temperature Difference.* This variable represents the temperature difference between lake surface and overlaying air at Lake Michigan. The daily average data for lake surface and overlaying air temperature were collected from NOAA Great Lakes Environmental Research Laboratory. The air moving across the lake must be significantly cooler than the surface air to form LES. Specifically, the air temperature at an altitude where the air pressure is 850 millibars (85 kPa) should be 13 °C (23 °F) lower than the lake surface temperature for significant lake-effect snow to occur. The differences between these temperature datasets were calculated and categorized into two categories: temperature difference less than 13° C and greater than 13° C. It is worth noting that, the air pressure level was not considered in this analysis.

*Wind Speed.* While relatively light winds cause snowfall closer to the shore, strong winds tend to blow the snow further inland and produce a snow maximum. The winds blow the developing snow leeward from the lakeshore. Wind speed plays a key role in generating snow bands which determine the horizontal spreading of lake-effect snow and how far inland it will reach. According to Department of Atmospheric and Environmental Science, State University of New York, the optimal wind speed near the surface for LES should be 10 kts or
more, with most lake effect snows occurring with speeds of about 10-20 kts. Therefore, collected daily average wind speed data were categorized in binary classes: less than 5 m/s (approximately 10 kts) and more than 5 m/s.

**Wind Direction.** As goal of this research is to develop a statistical model of traffic crashes relating meteorological variables which are favorable to LES formation, optimal wind direction range was considered as baseline category and rest of the daily average wind direction data were grouped together as unfavorable category. For Allegan county, daily average wind direction data were categorized into two groups: one ranging from 235 to 360 degrees (baseline) and another from 0 to 234 degrees. For developing crash prediction model for Kalamazoo county, wind direction ranging from 280 to 325 degrees were grouped together in one class (baseline) and rest of them were under another class. Similarly, wind flow ranging from 290 to 340 degrees were considered as baseline category for Calhoun county’s model calibration. Wind direction ranges were chosen and classified in binary categories based on information discussed below.

Wind direction was considered as the most important factor which determines where the heaviest snow will fall over land. Wind direction refers to the direction where it originates or coming from. It’s usually measured in degrees as on a compass where 360 degrees is north, 90 degrees is east, 180 degrees is south, and 270 degrees is west. Since Lake Michigan is elongated north-south and southwest lower Michigan is on the eastern edge of it, LES mostly affects this region when the wind has a westerly component to it. How does the prevailing wind direction determine the location of the heaviest LES in these three selected counties are discussed below.

Grand Rapids, Michigan weather forecast office of National Weather Service (NWS) states that at near northerly flow, where the wind direction is from 325 to 350 degrees, the heaviest LES will occur across Allegan. Allegan and Kalamazoo counties are in moderated
threat of experiencing heavy LES when the flow is more northwest, from 305 to 325 degrees and west-northwest, from 280 to 305 degrees. West-Southwest flow, from 235 to 260 degrees, typically produces heavier LES across the Allegan county.

![Map showing wind directions and threat levels]

Figure 5. Areas receiving heavy snow for the given wind direction. Source: NWS, Grand Rapids

According to Gaylord, Michigan weather forecast office of NWS, Calhoun county typically experiences LES with wind direction ranging from 290 to 340 degrees.

![Map showing wind directions and threat levels]

Figure 6. Areas receiving heavy snow for the given wind direction. Source: NWS, Gaylord

Model Description

Negative binomial regression models with temporal random effects helped to characterize the number of crashes on meteorological variables that are responsible for LES formation. This also unfolded the associated temporal trend over 1512 days spanning the decade. Three models are fitted to the data, separately for three different counties, using Bayesian computation technique, such as, Markov Chain Monte Carlo (MCMC) algorithm. Specifically, the R-package ‘tscount’ was used to conduct this statistical analysis.
This study models the number of traffic crashes that occurred on the roadways of three selected counties over a decade. As daily average crash data aggregated at county level are discrete and non-negative integer values, the Poisson regression technique was initially attempted. However, the mean and variance of the response variables (crash counts) were different, suggesting a significant over dispersion in the data. Therefore, the Poisson distribution was rejected and a Negative Binomial (NB) model was used. The NB modeling approach is an extension of the Poisson regression method and allows the mean to differ from the variance. The NB model, derived from the Poisson model can be defined by the following equation:

\[ P(n_i) = \frac{\lambda_i^{n_i} \exp(-\lambda_i)}{n_i!} \]  

(4)

where \( P(n_i) \) is the probability of \( n \) accidents occurring on a roadway \( i \) over a ten-year time period, and \( \lambda_i \) is the expected accident frequency (i.e., \( E(n_i) \)). When applying the Poisson model, the expected accident frequency is assumed to be a function of explanatory variables and formulation of this distribution is:

\[ \lambda_i = \exp(\beta X_i) \]  

(5)

where \( X_i \) is a vector of explanatory variables that includes the meteorological variables associated with crash occurrence that impact crash frequency; and \( \beta \) is a vector of estimable coefficients. With this form of \( \lambda_i \), the coefficient vector \( \beta \) can be estimated by the maximum likelihood method. The corresponding likelihood function is:

\[ L(\beta) = \prod_i \frac{\exp[-\exp(\beta X_i)] \exp(\beta X_i)^{n_i}}{n_i!} \]  

(6)

To overcome the over-dispersion problem, negative binomial regression can be applied by relaxing the assumption that the mean of accident frequencies equals the variance.
To do this, an error term is added to the expected accident frequency ($\lambda_i$) and NB model is derived by rewriting eq (2) as,

$$\lambda_i = \exp(\beta X_i + \varepsilon_i)$$  \hspace{1cm} (7)

where $\exp(\varepsilon_i)$ is a gamma-distributed error term with mean one and variance $\alpha$. This gives a conditional probability distribution as follows

$$P(n_i|\varepsilon) = \frac{\exp[-\lambda_i \exp(\varepsilon_i)](\lambda_i \exp(\varepsilon_i))^{n_i}}{n_i!}$$  \hspace{1cm} (8)

Integrating $\varepsilon$ out of this expression produces the unconditional distribution of $n_i$. The formulation of this distribution (the negative binomial) is

$$P(n_i) = \frac{\Gamma(\theta + n_i)}{\Gamma(\theta) \cdot n_i!} \cdot u_i^\theta (1 - u_i)^{n_i}$$  \hspace{1cm} (9)

where $u_i = \theta/(\theta + \lambda_i)$ and $\theta = 1/\alpha$, and $\Gamma$ is a value of gamma distribution. The corresponding likelihood function is

$$L(\lambda_i) = \prod_{i=1}^{N} \frac{\Gamma(\theta + n_i)}{\Gamma(\theta) n_i!} \left[ \frac{\theta}{\theta + \lambda_i} \right]^\theta \left[ \frac{\lambda_i}{\theta + \lambda_i} \right]^{n_i}$$  \hspace{1cm} (10)

This function is maximized to obtain coefficient estimates for $\beta$ and $\alpha$. This model structure tolerates the variance of the process to differ from the mean such that,

$$\text{var}[n_i] = E[n_i][1 + \alpha E[n_i]]$$  \hspace{1cm} (11)

where $\alpha$ is the variance of the gamma-distributed error term. This additional parameter $\alpha$ is used as a measure of dispersion. The selection of the model largely depends on the statistical significance of the estimated coefficient $\alpha$. If $\alpha$ is significantly different from zero, the NB model is the right pick, otherwise the NB model simply reduces to a Poisson model with $\text{var}[n_i] = E[n_i]$. 
To recapitulate, ArcGIS’s space-time pattern mining tools were used to identify the spatial-temporal pattern of winter traffic crashes in the study area. Snow and non-snow events based crashes were analyzed separately to reveal the differences in the temporal trends and spatial pattern. Detailed findings and crash hot spot maps were presented in the next chapter. Negative binomial regression models with temporal random effects were fitted to predict daily traffic crash counts based on three explanatory variables namely, temperature difference between lake surface and overlaying air, wind speed, and wind direction. The R-package ‘tscount’ was used to develop the traffic crash count prediction model. Three different models were developed to fit the data separately for three selected counties. Results of these statistical models are documented in chapter IV.
CHAPTER IV

RESULTS AND DISCUSSION

Spatial-temporal Pattern Analysis of Winter Traffic Crashes

Snow and non-snow events based traffic crash hot spots were resulted from emerging hot spot analysis. While the hot spot can be defined as the high-density crash occurrence location or the safety deficient roadway segments, cold spot can be considered as area less prone to crash occurrences. The resulted spatial pattern of traffic crashes is portrayed county wise. Each county is divided in four quarters to display the emerging hot spot or cold spot. For each quarter, snow and non-snow events based crash hot spot maps were presented and discussed together. The Emerging Hot Spot Analysis tool classifies the pattern of traffic crashes into eight different trends of hot spot or cold spot within aggregation areas. According to ESRI’s ArcGIS, these trends can be defined as follows:

New Hot Spot (New Cold Spot). A location that is a statistically significant hot spot (cold spot) for the final time step (i.e. year 2014) and has never been a statistically significant hot spot (cold spot) before (i.e. from 2005 to 2013).

Consecutive Hot Spot (Consecutive Cold Spot). It is basically the New Hot Spot (New Cold Spot) location where less than ninety percent of all bins are statistically significant hot (cold) spots.

Intensifying Hot Spot (Intensifying Cold Spot). A location that has been a statistically significant hot spot (cold spot) for ninety percent of the time-step intervals (i.e. for 9 years), including the final time step (i.e. year 2014). In addition, the intensity of clustering of high (low) counts in each time step (i.e. year) is experiencing statistically significant increase.
**Persistent Hot Spot (Persistent Cold Spot).** A location that has been a statistically significant hot spot (cold spot) for ninety percent of the time-step intervals (i.e. for 9 years) without any obvious increasing or decreasing in the intensity of clustering over time.

**Diminishing Hot Spot (Diminishing Cold Spot).** A location that has been a statistically significant hot spot (cold spot) for ninety percent of the time-step intervals (i.e. for 9 years), including the final time step (i.e. year 2014). In addition, the intensity of clustering of high (low) counts in each time step (i.e. year) is experiencing statistically significant decrease.

**Sporadic Hot Spot (Sporadic Cold Spot).** An on-again then off-again high density crash location. Less than ninety percent of the time-step intervals have been statistically significant hot (cold) spots and none of the time-step intervals have been statistically significant cold (hot) spots.

**Oscillating Hot Spot (Oscillating Cold Spot).** The location representing statistically significant hot spot (cold spot) for the final time-step interval (i.e. in year 2014) which has been a statistically significant cold spot (hot spot) during a prior time step. Less than ninety percent of the time-step intervals have been statistically significant hot spots (cold spots).

**Historical Hot Spot (Historical Cold Spot).** The most recent time-step (i.e. year 2014) is not hot (cold), but at least ninety percent of the time-step intervals have been statistically significant hot spots (cold spots).

**No Pattern** will be detected if any location does not fall into any of the hot or cold spot patterns defined above.

Snow and non-snow events based spatial pattern analyses were conducted to explore the differences in hot spot locations. This study reported a lot of variation in the pattern of traffic crashes as well as their number. Number of hot spots and cold spots detected from spatial-temporal analysis were reported in Table 1. While there was no new hot spot detected
for non-snow events based analysis, 547 new hot spot locations were identified as a result of snow events based traffic crashes.

Table 1. Number of hot spots and cold spots detected from spatial-temporal analysis

<table>
<thead>
<tr>
<th>Traffic Crash Pattern</th>
<th>Non-snow based</th>
<th>Snow based</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Hot Spot</td>
<td>0</td>
<td>547</td>
</tr>
<tr>
<td>Consecutive Hot Spot</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Intensifying Hot Spot</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Persistent Hot Spot</td>
<td>3391</td>
<td>1364</td>
</tr>
<tr>
<td>Diminishing Hot Spot</td>
<td>374</td>
<td>0</td>
</tr>
<tr>
<td>Sporadic Hot Spot</td>
<td>0</td>
<td>664</td>
</tr>
<tr>
<td>Oscillating Hot Spot</td>
<td>146</td>
<td>1271</td>
</tr>
<tr>
<td>Historical Hot Spot</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>New Cold Spot</td>
<td>36</td>
<td>2</td>
</tr>
<tr>
<td>Consecutive Cold Spot</td>
<td>2942</td>
<td>504</td>
</tr>
<tr>
<td>Intensifying Cold Spot</td>
<td>3220</td>
<td>0</td>
</tr>
<tr>
<td>Persistent Cold Spot</td>
<td>1298</td>
<td>135</td>
</tr>
<tr>
<td>Diminishing Cold Spot</td>
<td>164</td>
<td>0</td>
</tr>
<tr>
<td>Sporadic Cold Spot</td>
<td>1443</td>
<td>936</td>
</tr>
<tr>
<td>Oscillating Cold Spot</td>
<td>164</td>
<td>87</td>
</tr>
<tr>
<td>Historical Cold Spot</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No Trend Detected</td>
<td>773</td>
<td>1933</td>
</tr>
</tbody>
</table>

Snow events based new hot spot locations were mainly detected in east part of Allegan county along US 131 in between ramp 49A to 52, 59 to ramp 61 and ramp 68 to 72. New hot spots were also found along 142nd Ave from Division S to 22nd St, along 18th St from 136th Ave to 108th Ave and along state highway M-89 from city of Plainwell to city of Otsego; and along 106th Ave in between 6th St to N 16th St. There was no new hot spot detected for either Kalamazoo or Calhoun county. These 547 road segments (250-meter) were statistically significant high cluster crash location only in year 2014. As these locations had never been statistically significant earlier, what caused them to be more crash prone in 2014 should be evaluated. These same locations were found statistically significant persistent cold spot when non-snow events based crashes were analyzed. It suggests that snowy weather
had significant impacts on crashes that occurred in these particular road segments. Seasonal safety measures are therefore urgent in these hot spot locations. Whether there was any deficiency in snowplowing or deicing should be also examined thoroughly.

All the snow-events based persistent hot spots were in Kalamazoo County. On the other hand, non-snow events based persistent hot spot were detected in Kalamazoo and Calhoun Counties. There were 3391 and 1364 persistent crash hot spot locations were detected for non-snow and snow events based analysis respectively. These locations are 250-meter road segments which had been statistically significant crash hot spot for 9 years over the decade (2005 – 2014) without any obvious increasing or decreasing in the intensity of clustering over time. Therefore, persistent hot spot locations should be scrutinized with more emphasis. On-site survey may also reveal the road condition or geometry related problem (if any).

There were 374 demising hot spot identified in the study area during non-snow period. These 250-meter road segments had been statistically significant hot spot for nine years over the decade including year 2014 but the intensity of clustering of high crash counts in each year showed statistically significant decrease. These locations should be inspected to understand the reasons behind this decrease. Knowledge from these hot spots could be a basis for safety improvement measures for persistent hot spots.

A total of 664 sporadic hot spot locations were detected from snow events based spatial-temporal analysis. These locations should also be examined carefully. Though these on-again then off-again high density crash locations had been statistically significant less than ninety percent of the time-step intervals, none of the time-step intervals had been statistically significant cold spot.

The study also found 146 and 1,271 oscillating hot spots from non-snow and snow events based analysis respectively. These locations had been statistically significant hot spots
for less than ninety percent of the time-step intervals. Though they had been statistically significant hot spot in year 2014, during prior time step they had also been statistically significant cold spot as well. These locations are less predictable. It is also evident that there were more oscillating hot spots detected from snow events based analysis than non-now events based. As LES is a rapid-onset and difficult to predict where it falls in the land, it might have significant impacts during crashes on these locations.

There were different patterns of cold spots detected too. Persistent cold spot and intensifying cold spot locations can be examined carefully. Studying factors that make these locations less prone to traffic crashes could help to diagnose and solve the problem associated with hot spot locations.

In the following section, crash hot spot locations are represented with county-wise separate maps along with brief discussion.
Allegan County Traffic Crash Pattern

The results of spatial-temporal crash pattern analysis are presented below. Snow and non-snow events based crash occurrences were analyzed and resulted hot spot locations are displayed in the map. To make the hot spot location discernible, the county is divided into quarters for mapping crash density distribution. The quarters are named as north-east (top right), south-east (bottom right), south-east (bottom left) and north-west quarter (top left portion of the county).

Figure 7. Spatial-temporal traffic crash pattern during snow in Allegan County (north-east part)

Figure 7 suggests that there are mostly consecutive cold spots and new hot spots in north east Allegan resulting from crash occurrences during snow. There are also some persistent cold spots and oscillating hot spots too. New hot spots were detected along US 131 in between ramp 59 to ramp 61 and ramp 68 to 72. New hot spots were also found along
142\textsuperscript{nd} Ave from Division S to 22\textsuperscript{nd} St and along 18\textsuperscript{th} St from 136\textsuperscript{th} Ave to 108\textsuperscript{th} Ave. Oscillating hot spots were found along US 131 in between ramp 61 and ramp 68 and along 135\textsuperscript{th} Ave from 10\textsuperscript{th} St to 18\textsuperscript{th} St.

On the other hand, spatial-temporal crash patterns from figure 8 shows that there was no hot spot in this part of the county during non-snow. Several persistent cold spots were detected for the road segments (e.g. US 131 segments in between ramp 68 to 72; road segments along 142\textsuperscript{nd} Ave from Division S to 22\textsuperscript{nd} St and along 18\textsuperscript{th} St from 136\textsuperscript{th} Ave to 108\textsuperscript{th} Ave) that were found as new hot spot locations from snow-events based analysis. This means that snowy weather has adverse impact on these particular road segments in year 2014 that result in statistically significant high density traffic crash clusters.

Figure 8. Spatial-temporal crash pattern during non-snow in Allegan County (north-east part)
Similarly, for south-east part of Allegan county, several new hot spots were detected along the road segments of US 131 in between ramp 49A to 52; road segments along state highway M-89 from city of Plainwell to city of Otsego; and along 106th Ave in between 6th St to N 16th St due to high number of traffic crashes during snow. These locations were found as consecutive and sporadic cold spots from non-snow events based analysis (see figure 9 and figure 10 below).

Figure 9. Spatial-temporal traffic crash pattern during snow in Allegan County (south-east part)
Figure 10. Spatial-temporal crash pattern during non-snow in Allegan County (south-east part)

Like north-east and south-east part, for other part of Allegan County, no hot spot was detected during non-snow events based analysis (see maps A1, A2 in Appendix A). There were some sporadic hot spot and very few new hot spot also resulted for traffic crashes in these areas during snowy weather condition (see maps A3, A4 in Appendix A). Heaviest bands of LES barely occur along the immediate shoreline, but tend to fall several miles inland (Bob, n.d.). This could be the reason behind very less hot spots detection in west part of Allegan county.
Figure 11 and 12 reveals that, in Kalamazoo county, traffic crashes don’t show much difference in spatial-temporal trend during snow or non-snow weather. Snowfall intensity or severity, number of population, traffic volume etc. can be studied thoroughly to understand this similarity in spatial pattern.

Figure 11. Spatial-temporal crash pattern during snow in Kalamazoo County (north-west part)

However, there were more hot spots from non-snow events based crash pattern analysis. Road segments along US highway-131 (in between business loop 34 A to 38 B), interstate highway-94 (in between ramp 72 to ramp 80), state highway M-43 (in between Nazareth Rd to N 4th St) and most of the other local roadway segments (including Stadium Drive, Okland Drive, S Westnedge Ave, Milham Ave etc.) passing through City of
Kalamazoo and City of Portage were detected as persistent hot spots during snow as well as in non-snow.

Figure 12. Spatial-temporal crash pattern during non-snow in Kalamazoo County (north-west part)

Hot spot map of other parts of the county (see Map A5, A6, A7, A8, A9 and A10 in Appendix A) suggest that, roads passing through the townships and villages of Kalamazoo county are less prone to high density traffic crash clusters.
Calhoun County Traffic Crash Pattern

Several new hot spots were detected along interstate-69 (in between ramp 38 and ramp-42) during snow events and near Marshall township along Michigan Ave. Numerous sporadic and oscillating hot spots were observed in City of Battle Creek and its nearby places along Interstate highway I-94, State highway M-66, other major roads such as Verona Rd, Michigan Ave, Dickman Rd, Helmer Rd S, Capital Ave SW, Riverside Drive, Main St (see figure 13).

![Figure 13. Spatial-temporal traffic crash pattern during snow in Calhoun County (north-west part)](image)

However, these road segments were found as persistent hot spot locations when traffic crashes were analyzed during non-snow events based analysis (see figure 14).
Figure 14. Spatial-temporal crash pattern during non-snow in Calhoun County (north-west part)

For other part of Calhoun County, no hot spot was detected during non-snow events based analysis (see maps A11, A12, A13, A14, A15 and A16 in Appendix A).
Developing a Statistical Traffic Crash Prediction Model

Prior to model traffic crashes for selected three counties, the daily average crash counts were analyzed. Table 2 exhibits the daily average traffic crashes over the years for winter months (November to March). Traffic crashes for each day of these five months were averaged and compared against the days when the three meteorological variables were favorable to form LES. When the cold air mass with speed greater than 5 m/s move across 13° C or more warmer lake surface and wind direction is favorable to carry the LES to particular land areas, daily average traffic crash counts were usually observed to be increased for the all three counties. In this study, favorable wind direction refers to wind direction ranges from 235° to 360°, 280° to 325° and 290° to 340° for Allegan, Kalamazoo, and Calhoun county respectively. Daily average traffic crash counts were increased by 114.66%, 121.51% and 53.34% for these counties respectively when all the three above-mentioned LES forming conditions were met. For few of the years, none of the days from winter months met all the conditions concurrently and hence these years were disregarded in the analysis. Form table 2, it’s clear that all the three meteorological variables favorable to LES formation impacts traffic safety and these were proportionally related to traffic crashes. However, the following analysis reveals whether this was occurring by random chance or there was a statistical evidence to upheld this relationship.
Table 2. Daily average traffic crashes in study area

<table>
<thead>
<tr>
<th>County</th>
<th>Condition*</th>
<th>Daily Average Traffic Crashes (aggregated at county level) in following years</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ii)</td>
<td>18 10 19 21.17 22.33 - 41 - 42.33 28 25.23</td>
<td></td>
</tr>
<tr>
<td>Kalamazoo</td>
<td>i)</td>
<td>30.87 25.1 28.93 32.18 27.56 24.03 24.3 24.01 25.54 28.87 27.14</td>
<td>121.51</td>
</tr>
<tr>
<td></td>
<td>ii)</td>
<td>18 22 71 61.8 95.5 - - - 68 84.5 60.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ii)</td>
<td>23 10 36 26.4 23 - 25 - 33.67 31 26.01</td>
<td></td>
</tr>
</tbody>
</table>

*Condition i): All the days from winter months are considered

ii): Days are considered when the three discussed meteorological variables were favorable to form LES
Model Description

The R-package ‘tscount’ was used to develop negative binomial regression model with temporal random effects. Daily average traffic crash counts aggregated at county level was considered as response variable. Meteorological factors namely temperature difference between lake surface and overlaying air, wind speed, and wind direction were three explanatory variables. ‘tscount’ provides likelihood-based estimation methods for analyzing and modelling of count time series following generalized linear models. The observations used in this model were nonnegative integers. This package employs the generalized linear model (GLM) methodology for modeling the traffic crash counts conditionally on the past information, using a negative binomial distribution and log link function.

Three negative binomial regression model with temporal random effects were fitted to the data separately for three selected counties. These county wise models were discussed below.

*Model for Allegan County Traffic Crash Counts*

Result from the model for Allegan county traffic crash data were shown in Table 3. The model retains all the three significant explanatory variables at the 95% significance level. These variables were temperature difference more than 13° C between lake surface and overlaying air, wind speed greater than 5m/s, and wind direction ranging from 235 to 360 degrees. All these three variables had statistically significantly positive estimates. It means that when the temperature difference between lake surface and overlaying air was more than 13° C or wind speed was greater than 5m/s or wind direction of blowing air over Lake Michigan was in the range of 235° to 360°, more traffic crashes were occurred in Allegan County.
Table 3. Model results for Allegan County traffic crash counts

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.7819</td>
<td>0.1348</td>
<td>0.5176</td>
<td>1.046</td>
</tr>
<tr>
<td>beta_1</td>
<td>0.2517</td>
<td>0.0323</td>
<td>0.1885</td>
<td>0.315</td>
</tr>
<tr>
<td>alpha_1</td>
<td>0.3670</td>
<td>0.0708</td>
<td>0.2282</td>
<td>0.506</td>
</tr>
<tr>
<td>Temp. Difference (&gt;13° C)</td>
<td>0.2243</td>
<td>0.0847</td>
<td>0.0582</td>
<td>0.390</td>
</tr>
<tr>
<td>Wind Speed (&gt;5 m/s)</td>
<td>0.1691</td>
<td>0.0376</td>
<td>0.0954</td>
<td>0.243</td>
</tr>
<tr>
<td>Wind Direction (235°-360°)</td>
<td>0.0983</td>
<td>0.0357</td>
<td>0.0284</td>
<td>0.168</td>
</tr>
</tbody>
</table>

Figure 15 shows the plot of observed traffic crashes and predicted traffic crashes. Vertical Y-axis represents the number of traffic crashes and horizontal X-axis represents the number of observations (1512 days over the decade).

Figure 15. Observed traffic crashes vs. model fit (Allegan county)

Model for Kalamazoo County Traffic Crash Data

Model estimates for Kalamazoo county traffic crash model reveals the similar results (see table 4). The model retains all the three significant explanatory variables at the 95% significance level. Following table summarizes the model results.
Table 4. Model results for Kalamazoo County traffic crash counts

<table>
<thead>
<tr>
<th>Coefficients:</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Confidence Interval</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.4630</td>
<td>0.2131</td>
<td>1.0452</td>
<td>1.881</td>
<td></td>
</tr>
<tr>
<td>beta_1</td>
<td>0.2502</td>
<td>0.0334</td>
<td>0.1847</td>
<td>0.316</td>
<td></td>
</tr>
<tr>
<td>alpha_1</td>
<td>0.2784</td>
<td>0.0779</td>
<td>0.1258</td>
<td>0.431</td>
<td></td>
</tr>
<tr>
<td>Temp. Difference (&gt;13°C C)</td>
<td>0.4149</td>
<td>0.0825</td>
<td>0.2532</td>
<td>0.577</td>
<td></td>
</tr>
<tr>
<td>Wind Speed (&gt;5 m/s)</td>
<td>0.0956</td>
<td>0.0321</td>
<td>0.0328</td>
<td>0.158</td>
<td></td>
</tr>
<tr>
<td>Wind Direction (280°-325°)</td>
<td>0.1372</td>
<td>0.0429</td>
<td>0.0930</td>
<td>0.221</td>
<td></td>
</tr>
</tbody>
</table>

When northwest or west-northwest cold air mass with a speed more than 5 m/s flow across 13°C or more warmer lake water, LES forms and may affect Kalamazoo county. Model suggests that more traffic crashes occurred when these three conditions were true. Therefore, it can be concluded that, LES induced traffic crash counts can be predicted successfully using the meteorological variables that were responsible for LES formation. Figure 16 shows the plot of observed traffic crashes and predicted traffic crashes for Kalamazoo county.

![Figure 16. Observed traffic crashes vs. model fit (Kalamazoo county)](image)

Model for Calhoun County Traffic Crash Counts

Model estimates for Calhoun county traffic crash count prediction model is presented in Table 5. The result suggests that temperature difference and wind speed were statistically
significant variables to predict traffic crash counts in Calhoun county. But, the wind direction ranging from 290 to 340 degrees was found statistically insignificant from zero at 95% level even though it’s exhibiting positive relation with traffic crash counts. So, wind direction (290° to 340°) was not a significant predictor or explanatory variable to predict traffic crash counts in Calhoun. Figure 17 suggests that Calhoun was at moderate risk when LES forms with wind direction 290° to 320°.

Table 5. Model results for Calhoun County traffic crash counts

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimate</th>
<th>Std Error</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.4070</td>
<td>0.0958</td>
<td>0.2193</td>
<td>0.5947</td>
</tr>
<tr>
<td>beta_1</td>
<td>0.1784</td>
<td>0.0265</td>
<td>0.1264</td>
<td>0.2303</td>
</tr>
<tr>
<td>alpha_1</td>
<td>0.6612</td>
<td>0.0512</td>
<td>0.5609</td>
<td>0.7616</td>
</tr>
<tr>
<td>Temp. Difference (&gt;13° C)</td>
<td>0.1128</td>
<td>0.0519</td>
<td>0.0111</td>
<td>0.2144</td>
</tr>
<tr>
<td>Wind Speed (&gt;5 m/s)</td>
<td>0.0657</td>
<td>0.0225</td>
<td>0.0216</td>
<td>0.1097</td>
</tr>
<tr>
<td>Wind Direction (290°-340°)</td>
<td>0.0315</td>
<td>0.0298</td>
<td>-0.0269</td>
<td>0.0899</td>
</tr>
</tbody>
</table>

Even though, one of the three explanatory variables were found statistically not significant, the model fits fairly (see figure 17).

Figure 17. Observed Traffic Crashes vs. Model Fit (Calhoun county)

Wind direction is the most important factor which determines where the heaviest snow will fall. However, the distance from the lake shore also influences the intensity of the
snowfall. According to Great Lakes Integrated Sciences and Assessments (GLISA), cities close to Lake Michigan (e.g. Traverse City, Muskegon, Grand Rapids, Kalamazoo) experience heavy LES. These snowfall events can reach far across but typically diminish significantly in severity. Calhoun county is the farthest from Lake Michigan in comparison to Allegan or Kalamazoo county. This could be a reason why LES impacts less traffic crashes in Calhoun county. This is also conforming with the Table 2 which shows that the increase of traffic crashes in Calhoun during probable LES events were lesser when comparing with two other counties. Traffic crash hot spots map (figure 13) also reveals that there were no persistent hot spots when traffic crashes were analyzed during snow events in north-west part of Calhoun. But, there were numerous persistent hot spot locations for the same area when the analysis was performed for non-snow events based traffic crashes. So, it can be concluded that snowfall has less impact in traffic crash counts as well as in spatial-temporal pattern for Calhoun county than two other counties analyzed.

This thesis led to detect traffic crash hot spots in the study area and uncovered the relationship between LES and traffic crashes, however, it has some limitations. These are discussed in the next chapter along with some recommendations for similar future researches.
CHAPTER V

CONCLUSION

LES induced traffic crashes are frequent in each winter in southwest Michigan. However, there are no specific studies in the academic literature that examine the latent relationship between this meteorological phenomenon and crashes. The spatial-temporal pattern of winter traffic crashes is also unidentified for this region. As a result, some important relevant facts are overlooked. For instance, what are the high-density crash cluster locations in this region during winter season? What is the long-term trend of crashes? Is there any difference between snow and non-snow involved traffic crash patterns? Which safety deficient locations or road segments should be fixed at priority? Where are the seasonal countermeasures required? What are the optimal meteorological characteristics behind LES formation that are also associated with traffic crashes? This thesis attempted to answer these questions by (i) exploring the spatial-temporal pattern of snow and non-snow events based traffic crashes and (ii) developing a statistical model which not only established the relationship between LES and crashes but also predicted LES induced traffic crash counts. Three contiguous southwest Michigan counties: Allegan, Kalamazoo and Calhoun were selected to conduct this study. Traffic crash data and meteorological data related to LES from year 2005 to 2014 were analyzed.

A newly incorporated GIS tool: Space-Time Pattern Mining Tool was used to perform the spatial-temporal pattern analysis of the traffic crashes. While crash hot spot locations were detected using Getis-Ord Gi* statistic, Mann-Kendall statistic is used to identify temporal crash patterns. To achieve the second objective of this study, negative binomial regression models with temporal random effects were fitted to the data treating daily average traffic crash counts as response variable; and temperature difference between Lake surface and overlaying air, wind speed, and wind direction as explanatory variables.
Major Findings

While there are several new and oscillating hot spots were detected in Allegan County during snow, no hot spot was detected during non-snow events based analysis. Non-snow events based spatial-temporal crash patterns analysis of Allegan county revealed several persistent cold spots were detected of road segments (e.g. US 131 segments in between ramp 68 to 72; road segments along 142nd Ave from Division S to 22nd St and along 18th St from 136th Ave to 108th Ave). These were conversely found as new hot spots from snow-events based analysis. Snow events based new hot spots occurred along the road segments of US 131 in between ramp 49A to 52; road segments along state highway M-89 from city of Plainwell to city of Otsego; and along 106th Ave in between 6th St to N 16th St. which were found as consecutive and sporadic cold spots from non-snow events analysis. This apparent difference in snow and non-snow events based analysis suggested that snowy weather had more impact on winter traffic crashes in Allegan. Hence, seasonal road safety measures are required in the identified hot spot locations.

There were more hot spots from non-snow events based crash pattern analysis for Kalamazoo County than snow events based. However, the patterns did not exhibit much difference spatially. Persistent hot spot locations along freeway or highway occurred along road segments in between business loop 34 A and 38 B of US-131, in between ramp 72 and ramp 80 of I-94, in between Nazareth Rd and N 4th St of M-43. Some of the local roadway segments passing through City of Kalamazoo and City of Portage were also had as persistent hot spots during snow as well as non-snow events. There were 3,391 non-snow and 1,364 snow persistent crash hot spot locations for Kalamazoo. These 250-meter road segments should be scrutinized in priority basis as these had been statistically significant crash cluster location for 9 years over the ten years (2005 – 2014) time-period. Detailed study and survey may help to identify any existing safety deficiency.
Some new hot spots and numerous sporadic and oscillating hot spots were observed in Calhoun County roadways during snow, which were also persistent hot spot locations during non-snow events. These new hot spots occurred along Intersate-69 (in between ramp 38 and ramp-42) and near Marshall township along Michigan Ave. Sporadic and oscillating hot spots are observed in City of Battle Creek and its nearby places along Interstate 94, M-66, other major roads such as Verona Rd, Michigan Ave, Dickman Rd, Helmer Rd S, Capital Ave SW, Riverside Drive, Main St. These results indicated that snowy weather had inconsistent impact on traffic crashes in Calhoun County. Snowfall intensity or variability might be studied to explain these inconsistencies. However, all the hot spots occurred on the west part of the County. As discussed earlier, severity LES usually diminishes with the increasing travel distance to reach to Calhoun. This could be the reason what makes the east part of the County less crash prone.

To achieve the second objective of this study, negative binomial regression models with temporal random effects were fitted to the data treating daily average traffic crash counts as response variable; and temperature difference between Lake surface and overlaying air, wind speed, and wind direction as explanatory variables.

The statistical models revealed that all the three explanatory variables exhibit statistically significant positive estimates to predict LES induced traffic crash counts for Allegan and Kalamazoo. It means that when the temperature difference between lake surface and overlaying air is more than 13° C or wind speed is greater than 5m/s or wind direction of blowing air over Lake Michigan is in favorable range, more traffic crashes will occur in Allegan and Kalamazoo counties. Wind direction is considered as the most important factor which determines where the heaviest snow will fall over land and favorable wind directions to affect Allegan and Kalamazoo are 235 to 360 degrees and 280 to 325 degrees respectively. Temperature difference and wind speed were also found proportionally related to Calhoun
County traffic crash counts at 95% confidence interval. However, the wind direction (290 – 340 degrees) was found statistically insignificant even though exhibits positive relation with traffic crash counts.

The intensity of lake effect snowfall diminishes significantly after travelling certain distances depending on the wind fetch and speed. Calhoun is the farthest county from Lake Michigan in comparison to two other counties discussed. This could be the reason why LES impacts less traffic crashes in Calhoun county. Moreover, traffic crashes were aggregated at county level to develop statistical model but the impact of explanatory variables are not uniform over different parts of the county. This is also conforming from the Calhoun County crash hot spot maps which revealed that eastern part of the county is less crash prone. These are suggestive facts which may explain why wind direction were found insignificant. Conducting the same study after dividing the whole county into several smaller scale TAZs may confirm the notions.

However, this research successfully establishes the relationship between different meteorological variables responsible to form LES and traffic crashes. The three explanatory variables used in model development are important factors which contribute forecast LES. It means that, if LES can be forecasted, LES induced traffic crashes can also be predicted. LES is a rapid onset and very complex phenomena and its impact on road safety cannot be pinpointed. However, the developed crash count model can be used to undertake some road safety precautions. This model can be replicated for smaller TAZs and this may help use to decide where the emergency management is urgent during LES.
Limitations

There are several limitations of this study. First, traffic crash occurrence locations were analyzed at county level. Conducting the spatial pattern analysis by dividing the whole county in several traffic analysis zones (TAZs) or grids may capture variation in crash clusters more efficiently. Second, all the traffic crashes were analyzed together irrespective of the difference in their cause, types, or nature. This study did not consider road geometry or crash types (e.g. rear-end collision, head on, sideswipe, rollover, side-impact collision, single vehicle or multi vehicle involved crashes) which may result in different crash cluster locations based on the nature of the crashes. Comparative study based on crash severity (e.g. fatal versus non-fatal) or crash location (e.g. crashes in intersection versus non-intersection) could also be very interesting. Third, temporal analysis was conducted to capture the year wise trend over the decade hence it ignored the monthly or diurnal trends.

Developing a statistical traffic crash count prediction model also had several limitations. For example, to develop the statistical model, daily traffic crashes aggregated at county level were used as response variable. Diving the county into several analysis zones may help to predict the traffic crash counts at local level. This may also help to consider the spatial random effect or developing the model considering spatial correlation. Moreover, the collected explanatory variables were also daily average data. Daily average data fail to capture the hourly variation and assume the impact of the variables on traffic crashes were uniform over different time of the day. Finally, this study incorporated all the days from winter months (November to March) in the analysis. If the data were collected exclusively from snowy days and taken into consideration, more robust model could be developed.

Although this study has several limitations, nevertheless it paves the way for conducting future research by addressing these gaps. This study led to detect high density traffic crash clusters in the study area over the decade. This long term spatial pattern could be...
very useful for seasonal safety campaign. Moreover, a statistical relationship between traffic crashes and LES is successfully established in this study. Expectantly this will help further researches to explore LES induced traffic crashes and devise effective countermeasures. Some of the future research directions are discussed below.

Future Studies

This study has multidirectional prospect for future studies which may reveal more interesting and useful results. Spatial-temporal pattern analysis of winter traffic crashes can be advanced by dividing the counties in several traffic analysis zones so that the analysis may capture the local variation in crash clusters. Identification of winter crash hot spots based on crash severity, types and road geometry could be a subject of future study. Crash hot spots which are identified in this research could be studied in detail through examining the crash location, road geometry and design. This may lead to specific remedial measures.

Emerging Hot Spot Analysis tool produces crash cluster surface of spatial point events over a 2-D geographic space. However, the planar cluster detection process may not appropriate for characterizing traffic crashes, as these point events usually occur in the roadway network, which is a 1-D linear space (Xie et al. 2008; Tang et al. 2016). Network based hot spot analysis would be more suitable for this kind of analysis.

For hot spot analysis, 250 meter by 250 meter bins were created to aggregate traffic crashes. Crash hot spots were detected on this square. A 250 meter segment along the roadway may be suitable for crash cluster analysis. However, this analysis assumed all the roads are 250 meter wide which are actually not. This does not affect the hot spot detection but representation. However, careful consideration of bin size is recommended for future study. Future traffic crash prediction model should incorporate the spatial variables in model calibration. Traffic crash data aggregated at small spatial scale or at specific road segments might be a good practice in that case. Aggregating traffic crashes into smaller TAZs may also
help to develop the prediction model considering the spatial random effect. Incorporation of real-time meteorological data in model developing or validating would be quite revealing too. This may lead to develop more robust and effective prediction model.

Spatial-temporal pattern analysis of traffic crashes can be conducted only for those days when the meteorological variables were favorable to form LES. This may result in LES induced traffic crash hot spots for this area. Findings from the statistical model developed in this study could be used for sorting out the days with optimal meteorological conditions.
REFERENCES


56


Juga, I. (2010). Sea-effect snowfall—a special hazard for road traffic in the coastal areas of Finland. In 15th International Road Weather Conference (pp. 5-7).


Oakland County (2013, November). Winter Hazards FAQs [PDF].


Appendix A
Allegan County Traffic Crash Hot Spot Maps
Figure A1. Spatial-temporal traffic crash pattern during non-snow (north-west part)

Figure A2. Spatial-temporal traffic crash pattern during non-snow (south-west part)
Figure A3. Spatial-temporal traffic crash pattern during snow (north-west part)

Figure A4. Spatial-temporal traffic crash pattern during snow (south-west part)
Appendix B

Kalamazoo County Traffic Crash Hot Spot Maps
Figure B1. Spatial-temporal crash pattern during snow (north-west part)

Figure B2. Spatial-temporal crash pattern during snow (south-east part)
Figure B3. Spatial-temporal crash pattern during snow (south-west part)

Figure B4. Spatial-temporal crash pattern during non-snow (north-west part)
Figure B5. Spatial-temporal crash pattern during non-snow (south-east part)

Figure B6. Spatial-temporal crash pattern during non-snow (south-west part)
Appendix C

Calhoun County Traffic Crash Hot Spot Maps
Figure C1. Spatial-temporal crash pattern during non-snow (north-east part)

Figure C2. Spatial-temporal crash pattern during non-snow (south-east part)
Figure C3. Spatial-temporal crash pattern during snow (south-west part)

Figure C4. Spatial-temporal crash pattern during snow (north-east part)
Figure C5. Spatial-temporal crash pattern during snow (south-east part)

Figure C6. Spatial-temporal crash pattern during snow (south-west part)