MAPPING THE FUTURE OF MOTOR VEHICLE CRASHES

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

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Brandon A. Stakleff
August, 2015
MAPPING THE FUTURE OF MOTOR VEHICLE CRASHES

Brandon A. Stakleff
Dissertation

Approved:

Advisor
Dr. William Schneider

Civil Engineering Department Chair
Dr. Wieslaw Binienda

Committee Member
Dr. Stephen Duirk

Interim Dean of the College of Engineering
Dr. Rex Ramsier

Committee Member
Dr. Anil Patnaik

Interim Dean of the Graduate School
Dr. Chand Midha

Committee Member
Dr. Scott Sawyer

Date

Committee Member
Dr. Mark Fridline
ABSTRACT

To reduce the occurrence of motor-vehicle crashes, professionals in education, enforcement, and engineering are continually tasked with implementing safety solutions. Identifying locations of high rates of crashes allows safety solutions to more adequately target their intended audience. This research examines advances in identifying hot spots of motor-vehicle crashes. These advancements come from improving: 1) the calculation of spatial autocorrelation and interpolation, 2) the identification of spatio-temporal patterns, and 3) the influence of geographical patterns on the spatial distribution of crashes. Overall, by improving the hot spot analysis, concerned professionals may be better prepared and lower the number of alcohol-related crashes.

The location of hot spots is important in the implementation of enforcement campaigns. A lapse in accuracy may allow a vehicle operator suspected of disobeying traffic laws from being properly disciplined. Improvements in the calculation of spatial autocorrelation and interpolation result from the use of network distances instead of Euclidean based distances. Network based distances increase the accuracy of resulting hot spots.

With the accuracy of hot spots improved, the optimal times to implement safety campaigns in their identified areas become important. Many hot spots purely analyze crashes as if they all occurred at the same time. By investigating crashes in this manner, some key influences may be lost and the efficiency of the implemented campaign may be
reduced. Spatio-temporal hot spot are examined and show that as time progresses, clusters of crashes occur and disappear throughout space. By moving campaign sites as the location of crashes move, the overall efficiency of campaign tactics would benefit.

Hot spots of crashes have continually been scrutinized for their focus on areas of large populations. In an effort to rectify this belief, the normalization of hot spot is examined in relation to population density. It is found that the strict use of population density provides unfavorable results. Instead, the identification of hot spots through either the frequency or societal crash costs varies the resulting hot spot location. Using crash frequency allows for high visibility/mass target campaigns to best be realized. Meanwhile, the use of societal costs best targets high valued crash occurrences.
ACKNOWLEDGEMENTS

I would like to thank my sponsors, committee, research team, family, and friends who have all aided and supported my journey thus far. A special thanks, goes to Dr. Schneider for his support and guidance through my graduate studies. The members of my research team, Alexander Maistros, William Holik, and Mallory Crow, who have provided strategic input and knowledge throughout the course of projects. The support of my family and friends is especially important as I undertook countless hours and long nights of research. I would also like to thank Dr. Patnaik, Dr. Duirk, Dr. Sawyer, and Dr. Fridline for their valuable time, considerations, and suggestions through this research. Additionally, I would like to thank the University Transportation Centers research program for sponsoring my research and the Ohio Department of Public Safety for providing the necessary crash data utilized within this research.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>ix</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER I INTRODUCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Benefits of this Research</td>
</tr>
<tr>
<td>1.2 Organization of the Dissertation</td>
</tr>
<tr>
<td>1.2.1 Chapter II: Background Information</td>
</tr>
<tr>
<td>1.2.2 Chapter III: Comparing the Use of Euclidean and Network Based Distances When Calculating Hot Spots for Law Enforcement Patrol</td>
</tr>
<tr>
<td>1.2.3 Chapter IV: A Spatio-Temporal Hot Spot Examination of Alcohol-Related Single and Multiple Vehicle Crashes</td>
</tr>
<tr>
<td>1.2.4 Chapter V: Examining the Use of Normalization in Mapping of Alcohol-Related Hot Spots</td>
</tr>
<tr>
<td>1.2.7 Chapter VII: Conclusion and Recommendations</td>
</tr>
<tr>
<td>1.1 Benefits of this Research</td>
</tr>
<tr>
<td>1.2 Organization of the Dissertation</td>
</tr>
<tr>
<td>1.2.1 Chapter II: Background Information</td>
</tr>
<tr>
<td>1.2.2 Chapter III: Comparing the Use of Euclidean and Network Based Distances When Calculating Hot Spots for Law Enforcement Patrol</td>
</tr>
<tr>
<td>1.2.3 Chapter IV: A Spatio-Temporal Hot Spot Examination of Alcohol-Related Single and Multiple Vehicle Crashes</td>
</tr>
<tr>
<td>1.2.4 Chapter V: Examining the Use of Normalization in Mapping of Alcohol-Related Hot Spots</td>
</tr>
<tr>
<td>1.2.7 Chapter VII: Conclusion and Recommendations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER II BACKGROUND INFORMATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Point-Based Analysis</td>
</tr>
<tr>
<td>2.2 Segment-Based Analysis</td>
</tr>
<tr>
<td>2.3 Zonal-Based Analysis</td>
</tr>
<tr>
<td>2.4 Spatial Analysis Summary</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER III COMPARING THE USE OF EUCLIDEAN AND NETWORK BASED DISTANCES WHEN CALCULATING HOT SPOTS FOR LAW ENFORCEMENT PATROL</th>
</tr>
</thead>
</table>
CHAPTER V EXAMINING THE USE OF NORMALIZATION IN MAPPING OF ALCOHOL-RELATED HOT SPOTS

5.1 Introduction ..............................................................................................................62
5.2 Data ..........................................................................................................................65
5.3 Methodology ............................................................................................................66
  5.3.1 Population Density ........................................................................................67
  5.3.2 Spatial Autocorrelation .................................................................................68
  5.3.3 Interpolation of Spatial Autocorrelation .......................................................69
5.4 Results ......................................................................................................................70
5.5 Conclusion ................................................................................................................80

CHAPTER VI CONCLUSION AND RECOMMENDATIONS ...........................................84
6.1 Introduction ..............................................................................................................84
6.2 Calculation of Spatial Autocorrelation and Interpolation ........................................84
6.3 Identification of Spatio-Temporal Patterns ..............................................................85
6.4 Influence of Geographical Patterns on the Spatial Distribution of Crashes ..........86
6.5 Future Recommendations .........................................................................................87
  6.5.1 Linking Safety Campaigns to Hot Spots.......................................................87
  6.5.2 Predicting Future Hot Spot Locations ...........................................................88
6.6 Conclusions ..............................................................................................................88

LITERATURE CITED ......................................................................................................90
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 3.1. Injury Severity for Geocoded Alcohol-Related Crashes in Cuyahoga County</td>
<td>19</td>
</tr>
<tr>
<td>Table 3.2. PAI comparison.</td>
<td>31</td>
</tr>
<tr>
<td>Table 3.3. Percent difference in societal crash costs.</td>
<td>32</td>
</tr>
<tr>
<td>Table 3.4. Percent difference in length of roadway</td>
<td>32</td>
</tr>
<tr>
<td>Table 4.1. Descriptive Statistics of Alcohol-Related Crashes in Cuyahoga County, Ohio.</td>
<td>40</td>
</tr>
<tr>
<td>Table 4.2. Time Groupings for Clusters by Time of Day.</td>
<td>53</td>
</tr>
<tr>
<td>Table 4.3. Time Groupings for Clusters by Day of Week.</td>
<td>58</td>
</tr>
<tr>
<td>Table 5.1. County Geographical Makeup.</td>
<td>65</td>
</tr>
<tr>
<td>Table 5.2. Geographical Coverage of Significant Hot Spots.</td>
<td>79</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 3.1. Comparison of $Gi^*$ z-scores obtained by Euclidean and network-based analysis for Cuyahoga County.</td>
<td>25</td>
</tr>
<tr>
<td>Figure 3.2. Comparison of hot spot areas between Euclidean and network-based analysis.</td>
<td>27</td>
</tr>
<tr>
<td>Figure 3.3. Identification of hazardous roadways.</td>
<td>29</td>
</tr>
<tr>
<td>Figure 4.1. Hot Spots of Alcohol-Related Crashes.</td>
<td>48</td>
</tr>
<tr>
<td>Figure 4.2. Spatio-Temporal Analysis of Alcohol-Related Crashes in Cuyahoga County by Time of Day.</td>
<td>52</td>
</tr>
<tr>
<td>Figure 4.3. Comparison of Single and Multi-Vehicle Hot Spots by Time of Day.</td>
<td>55</td>
</tr>
<tr>
<td>Figure 4.4. Spatio-Temporal Analysis of Alcohol-Related Crashes by Day of Week.</td>
<td>57</td>
</tr>
<tr>
<td>Figure 4.5. Comparison of Single and Multi-Vehicle Hot Spots by Day of Week.</td>
<td>59</td>
</tr>
<tr>
<td>Figure 5.1. County Population Density.</td>
<td>71</td>
</tr>
<tr>
<td>Figure 5.2. Hot Spot Maps of Urban Counties.</td>
<td>73</td>
</tr>
<tr>
<td>Figure 5.3. Hot Spot Maps of Rural Counties.</td>
<td>77</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Motor vehicle crashes claim dozens of lives each day. In 2012 alone, there were 33,561 total motor vehicle fatalities in the United States (NHTSA, 2014). One type of crash that contains high rates of injury and concern to motor vehicle safety officials is alcohol-related crashes. The fatality rate in 2012 for instances when an operator of a motor vehicle has a blood alcohol concentration (BAC) of 0.08% or greater, is 3.29 per 100,000 people (NHTSA, 2015). The use of alcohol, additionally, seems to affect males more than females, as 24 percent of males operating a motor vehicle involved in a fatal crash had a BAC of 0.08% or greater. Meanwhile, the amount of females operating a motor vehicle involved in a fatal crash while having a BAC of 0.08% or greater was much less, at 14 percent. The factors and contributing circumstances vary greatly between crashes, but the bottom line is that a significant number of people die each year in cases where preventable measures could have been taken to avoid the loss of life. The clear issues are defining preventable measures and then implementing them into practice. Saving lives is the goal and responsibility of our collective transportation community. In order to reduce crashes, it is up to law enforcement, educators, engineers, researchers, doctors, lawyers, judges and others to determine the pertinent information that is used to create the grant funding opportunities, educational campaigns, and laws that keep our
roadways safe. The tools that help these developers work together in the flow of information and data are critical.

One such tool is the spatial mapping of motor vehicle crashes. Mapping crash locations allows for a visual identification of high impact locations, trends, and outliers. This visual identification follows the goal of Data-Driven Approaches to Crime and Traffic Safety (DDACTS), set out by the National Highway Traffic Safety Administration (NHTSA), which is to develop a data driven approach to identify geospatial areas with higher crash and crime problem areas (US DOT, 2009). The crashes analyzed from mapping are located either through geocoding the addresses/reference points or latitude/longitude coordinates obtained from crash reports and crime reports. The visual representation of the distribution may draw some initial conclusions; however, multiple crashes that occur near the same location may, at first glance, appear as a single occurrence. Due to the possible misidentification of multiple crashes, a further investigation must be performed before any real solutions may be obtained. In addition, many options are available for identifying trends in spatial data, and the mapping of the crashes must be used to facilitate this form of analysis.

The identification of spatial distributions within motor vehicle crashes allows pertinent safety campaigns to adequately address the relevant motorists on the roadway. Efforts in the campaigns may come in the form of determining which locations are most hazardous to motorists or the contributing factors that are harmful to motorists. The pertinent hazardous locations may then be used as a target area in which to implement a campaign. While using hot spot maps allows for aids in identifying drivers operating a vehicle under the influence of alcohol, the link between crashes and the implemented
safety campaigns need to be strengthened. The relationships linking these two aspects together pertain to which roads are highlighted as a target area, why particular crashes are occurring in highlighted locations, and what type of safety campaigns may be implemented. In order to solve these questions, a deeper understanding of the relationships between alcohol-related crashes and their associated hot spots is examined.

1.1 Benefits of this Research

This research allows for a greater understanding of the relationships between alcohol-related crashes and the locations in which they occur. In the past, the overall location of where crashes are occurring has been developed. This research delves deeper into the spatial distribution of crashes and identifies how the location of these crashes affects safety campaigns implemented in an attempt to reduce the number and severity of crashes. The overall goal of this research is to create geospatial means to analyze motor-vehicle crashes. The geospatial means include the examination of spatial relationships along roadway networks, the spatial analysis of crashes continuously over progressing time, and an analysis of the effects of geographical distributions. These analyses realize important relationships between crashes and their surroundings, which aid in reducing the number and severity of crashes.

The overall objective of this research, to create a geospatial means of analyzing motor-vehicle crashes, is achieved through three different steps. First, the spatial relationship between alcohol crashes and the roadways that they occur on is examined. This identifies a more accurate means of analysis and patterns of roadways that affect spatial analyses. This examination provides a unique perspective in identifying the link between spatial analyses and the legality behind their use in preventing alcohol-related
crashes. The second analysis examines the spatial relationship of crashes through the progression of time. This spatio-temporal analysis identifies movements of hot spots continually throughout time and variances between both single and multi-vehicle alcohol-related crashes. This examination is unique in the ability to continually analyze spatial distributions through a moving window of time. The third analysis examines the association between alcohol-related crashes and the geographical components of population. The relationship of geographies indicates that normalizing for population density does not provide any substantial benefit; however, by investigating varying crash attributes, the focus of crashes in high population areas is reduced. This examination is unique in the identifying the use of various spatial analyses towards targeting crashes and implementing different types of safety campaigns. While this research identifies results for specific areas within the state of Ohio, the ability of the methodologies developed within this research extend beyond those study areas and may be applied to a wide variety of regions.

1.2 Organization of the Dissertation

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

1.2.1 Chapter II: Background Information

Chapter II discusses the current conditions regarding spatial analyses of motor vehicle crashes. The chapter opens with insight into different types of spatial analyses being conducted within research of motor vehicle crashes. The review of previous studies is broken down into three different areas, including: point-based, segment-based, and
zonal-based analyses. This chapter further expands on spatial analyses by presenting the ability to express hot spots through the interpolation of spatial autocorrelation.

1.2.2 Chapter III: Comparing the Use of Euclidean and Network Based Distances When Calculating Hot Spots for Law Enforcement Patrol.

This chapter builds upon the background analyses identified in Chapter II. The influence of distance on the spatial analysis of crashes is investigated towards the application of hot spots in legally implementing alcohol focused safety campaigns. Specifically, relating the use of Euclidean or network based distances to the implementation of hot spots for patrolling of alcohol-related crashes. The investigation of varying distances is applied to the calculation of both spatial autocorrelation and interpolated values.

1.2.3 Chapter IV: A Spatio-Temporal Hot Spot Examination of Alcohol-Related Single and Multiple Vehicle Crashes.

Chapter IV builds upon those findings from the previous chapter by using network based distances to investigate the spatial variation between single and multiple vehicle crashes where an involved driver was intoxicated with alcohol. This spatial variation is examined through a spatio-temporal analysis. Within this analysis clusters of crashes are identified throughout time, across both the time of day and day of the week. The movement of these clusters is examined for the ability to increase the efficiency of safety campaigns.
1.2.4 Chapter V: Examining the Use of Normalization in Mapping of Alcohol-Related Hot Spots.

While Chapter IV identified the presence of movement in clustered crashes as time progresses, the effects of population on the location of clusters has been made a concern. Chapter V explores the effects of population density on the location of clusters and the ability to implement safety campaigns in the location of hot spots. In order to assess these effects, the normalization of hot spots is investigated. The hot spots investigated are determined based on the frequency and societal costs of crashes. The location of the resulting hot spots for both normalized and non-normalized spatial autocorrelation is compared for their use in educational, enforcement, and engineering campaigns.

1.2.7 Chapter VII: Conclusion and Recommendations

This chapter reviews the advancements in spatial analyses pertinent to motor-vehicle crashes examined within this research. The application of these advancements is reviewed. Additionally, a look into the future of crash mapping beyond examined methodologies is discussed. These future recommendations build upon the techniques used within this research.
CHAPTER II
BACKGROUND INFORMATION

The ability to locate where crashes are occurring provides large opportunities to safety officials who aim at reducing the number and severity of crashes. The use and support of spatial modeling within DDACTS allows the further investigation of spatial modeling to aid in the reduction of crashes. The analysis of crashes through DDACTS exploits one option to reduce alcohol-related crashes, by employing safety-related campaigns in high risk locations. The identification of these high risk locations is paramount to the successful implementation of these safety campaigns. Without knowing the ideal location of where these crashes are occurring, safety related efforts may either be imposed upon non-pertinent people or misused in locations where large amounts of crashes are not realized. In order to obtain a better understanding of crashes, their location and attributes are compared to one another. Spatial analyses use Tobler’s first law of geography (Tobler, 1970), that “everything is related to everything else, but near things are more related than distant things” to achieve this understanding. The spatial analysis of crashes allows for the optimal location of implemented safety campaigns to be identified.

Several methods of mapping may be used in spatial analyses. Kim and Levine (1996) identify three different ways to study spatial information, including point, segment, and zonal analyses. While there are three different levels in which to investigate
crash locations, the spatial analyses methods used within each level may often overlap from one to another. One example of this overlap is through the use of Moran’s $I$, Geary’s $C$, and the Getis-Ord $G$ statistic. These methods, which may be used in either point or zonal-based analyses, indicate both global and local levels of clustering. The global indication examines spatial autocorrelation over an entire study area. In other words, an indication is determined for all crashes as a whole. The local indication examines spatial autocorrelation at each specific location. Moran’s $I$ and Geary’s $C$ both investigate features based on their similarity to nearby features. Meanwhile, the $G$ statistic investigates features based on the concentration of high or low feature values.

Boots and Tiefelsdorf (2000) further explain the representation of global Moran’s $I$ as an overall indication of whether similar or dissimilar values are located in close proximity to one another. Whereas, Anselin (1995) further explains the local Moran’s $I$ as an indication of similarity at each specific point, allowing for pockets of crashes to be determined. The global representation of the $G$ statistic is explained by Getis and Ord (1992) as an overall measure of, or lack thereof, concentration of points. Getis and Ord (1992) similarly explain the local $Gi^*$ statistic, where groups of points that have high or low spatial association are identifiable.

2.1 Point-Based Analysis

Point-based map analysis uses the specific locations of crashes, resulting in a series of points on a map. The location of the crashes is often determined based on the longitude and latitude of the crash, an address at which the crash occurred, or an intersection and an estimated distance from the intersection where the crash occurred. Each individual point identified on a map would thus relate to one individual crash.
Crashes occurring in the same location would then be identified by one point overlaid by another. The point-based analyses allows for differentiation between these crashes that occur in the same location.

The latest and most popular measures of spatial distribution for point-based analyses are calculated using Moran’s $I$ and the $Gi^*$ statistic, described in the previous paragraph. Other measures of spatial distribution, such as the nearest neighbor index, are also available for point-based analyses. The nearest neighbor is a global indicator of clustering that indicates if the average distance between neighboring features is more or less than the expected distance separating one another. Applying these analyses to crashes, the overall location of points, the spatial distribution from this overall location, or the spatial distribution from one point to another may then be analyzed. The overall location of crashes and their spatial distribution from this location was examined in Hawaii by Levine et al. (1995). This examination of crashes in Hawaii identified that clustering was present throughout the entire study area using the nearest neighbor index. The spatial distribution was also analyzed from the overall averaged location of points using the standard deviational circle and ellipse. Kang et al. (2012) and Wong (1999) used the latter measures in addition to the mean center to describe the distribution of crashes.

While studies have shown that the use of Moran’s $I$ and $Gi^*$ are useful in the identification of spatial autocorrelation among crashes (Songchitruxka and Zeng, 2010; Truong and Somenahalli, 2011), the analysis of spatial distributions of crashes has expanded to investigate crash densities. The use of kernel density estimation (KDE) allows for locations that have high occurrences of crashes to be realized, as shown
through the research of Backalic (2013), Plug et al. (2011), Pulugurtha et al. (2007), and Schneider et al. (2001). The combination of using KDE, Moran’s $I$, and $Gi^*$ have allowed others, such as Blazquez and Celis (2013), Kuo (2013), Prasannakumar et al. (2011), and Schneider et al. (2004), to compare statistical cluster significance to various density values.

2.2 Segment-Based Analysis

While point-based mapping has shown promise in identifying relationships within crash data sets, other approaches have used segment-based mapping for the identification of factors relating to crashes. The crashes used within segment-based analyses are aggregated to small segments of roadways, and these roadway segments are then analyzed for patterns. Segment-based mapping has been employed by Imprialou et al. (2014) to identify the roadway segments on which crashes have occurred and to determine how the segments may be improved. These types of analysis have allowed roadway segments to be identified through the use of frequency of crashes (Loo and Yao, 2013) and the $K$-function (Yamada and Thill, 2004). Analyses such as KDE, which were identified and used within point-based mapping, have also been used in segment-based mapping, as seen in Erdogan et al. (2008). The analysis of roadway segments has grown to include the association of segments’ neighbors and additional contributing factors into the final analysis using Bayesian statistics, as seen by Aguero-Valverde (2013), Aguero-Valverde and Jovanis (2008), El-Basyouny and Sayed (2009), Li et al. (2007), Mitra (2009), Vandenbulcke et al. (2014), and Yu and Abdel-Aty (2013).
2.3 Zonal-Based Analysis

The final type of mapping, as described by Kim and Levine (1996), is zonal-based mapping. This type of mapping uses a specific defined area, such as counties, traffic analysis zones (TAZ), as well as census block, block group, and tract levels. Zones at each of these levels, which have been created by government entities to group the residing population for various purposes, are treated in a manner similar to a quadrat analysis (Nicholson, 1998), which uses grid-based zonal boundaries to aggregate crashes and test for randomness within the crashes’ dispersal area. The thought is that once these areas are defined, state and local agencies may more efficiently allocate the appropriate resources – including personnel, money, or educational materials – that are required to reduce the number and severity of crashes. The zonal analyses are conducted by aggregating all crashes contained within each zone’s boundary, creating a single frequency value for each zone. Each zone is then analyzed based on the neighboring zones or the distance from the center of that zone to the center of other zones. Many of the analyses used are similar to those used in both the point-based and segment-based mapping. Kim et al. (2010) used quadrat analysis to investigate crashes that were aggregated through a 0.1-m² grid. Similarly, Yiannakoulias et al. (2012) aggregated crashes zonally by census tract to identify the relative risk associated with each zone. An application where the density of crashes was determined within zonal boundaries (Chen et al., 2014) has also been completed, as the use of KDE extends beyond point-based mapping. Spatial autocorrelation for zones was investigated by Erdogan (2009), Khan et al. (2008), and Khan et al. (2009). Many studies, such as Lee et al. (2014), Loukaitou-Sideris et al. (2007), Pirdavani et al. (2012), Scheiner and Holz-Rau (2011), Sukhai and
Jones (2013), Treno et al. (2007), and Wang and Kockelman (2013), have aggregated crashes zonally in order to use the frequency of crashes to investigate the associated factors through the use of regression models. The spatial relationship between one zone and its neighboring zones was also conducted through many studies using Bayesian statistics, as seen in Aguero-Valverde and Jovanis (2006), Karim et al. (2013), Lee et al. (2014), Ng et al. (2002), Pulugurtha et al. (2013), Quddus (2008), Wang et al. (2012), and Xu et al. (2014).

The studies previously mentioned in latter three paragraphs have contributed greatly to the identification of spatial variations in transportation related crashes. These existing methods have proven useful in identifying underlying patterns within a set of data points.

2.4 Spatial Analysis Summary

There are a number of ways these techniques may be applied to the use of safety campaigns. One such method is to target specific points or allow law enforcement to patrol areas based on their ability to pass through significantly clustered points. Point-based analyses are useful because it maintains the integrity of the existing data, allowing each crash location to be spatially related to the contributing results. While this data integrity is important, locations may be missed in the event that a crash did not occur in its exact same place during the study period. Zonal-based analyses may remedy this issue in that all locations would thus have an attributable level of spatial distribution associated with them. While, this allows locations where crashes are likely to occur to be identified, the presence of aggregating crashes based on an arbitrary zone allows a bias from the principal investigator to be realized. This bias may be minimalized or removed during the
use of segment-based analyses; however, due to the aggregation of crashes, the spatial
distribution is not analyzed at the location in which the crash occurred, only a nearby one.
The use of aggregation provides information pertaining to crashes within a specific area,
the difference when locating a crash on one side of the boundary or segment versus
another may create large differences in the indicated outcome. Even though smoothing
techniques may be used to reduce this effect, the elimination of aggregation boundaries
would allow for a smooth transition between all locations, allowing for the spatial aspect
of the crash to be weighted higher than the boundary that it falls within.

In an effort to remove the influence of bias or aggregation, the use of KDE and
interpolation have provided a means to identify locations where safety campaigns may be
implemented. These two methods allow for a level of clustering to be realized throughout
all locations of a study area. The location of safety campaigns is thus identified in an area
where clustering is statistically significant. This may be seen with the use of the $Gi^*$
statistic and an interpolator. The result of the $Gi^*$ statistic is a z-score. That z-score is
then interpolated and distributed throughout the entire study area. Only those locations
that are significantly clustered are identified. Safety campaign implementations may then
take place within the identified area.
CHAPTER III

COMPARING THE USE OF EUCLIDEAN AND NETWORK BASED DISTANCES WHEN CALCULATING HOT SPOTS FOR LAW ENFORCEMENT PATROL

3.1 Introduction

In 2012, 30,800 fatal vehicle crashes occurred throughout the United States, which translates to a rate of 10.69 fatalities per 100,000 people (NHTSA, 2015). Of the 30,800 fatal crashes, a total of 10,322 vehicle operators had a blood alcohol concentration (BAC) of 0.08 or greater (NHTSA, 2015). The effects of alcohol on drivers have been heavily studied. Connor et al. (2004) identified a strong association between those who drink alcohol before driving and crashes with injuries. Peck et al. (2008) investigated the relationship between BAC and drivers under the age of 21, identifying a higher relative crash risk than predicted for the effect of BAC and age. Evans (1990) found that traffic-related fatalities would be reduced by nearly 47% if there were not any alcohol-related crashes.

Educators, engineers, and law enforcement agencies have attempted to reduce the total number of alcohol-related fatalities. Educational efforts may be directed toward a diverse range of drivers, spanning from new or existing drivers to those who have been convicted of operating a vehicle while intoxicated (OVI). The messages presented to each of these different subgroups of drivers may be specifically tailored to the conditions relevant to each operator. The design of roadways may also be altered in an effort to
make roads safer. Additionally, safety campaigns may be implemented through law enforcement in an effort to reduce the number of intoxicated drivers on the roadway. Some of these campaigns are in the form of saturation patrols, corridor enforcement, or checkpoints. The implementations used by educators, enforcement, and engineers may benefit from research studies that disseminate information about hazards to drivers, provide insight into the drivers’ perception of altered roadway, or identifying the location in which to implement safety campaigns.

The identification of locations in which to implement measures such as safety campaigns varies widely. The National Highway Traffic Safety Administration (NHTSA) has proposed and implemented the idea of Data-Driven Approaches to Crime and Traffic Safety (DDACTS) in an effort to reduce the occurrence of crimes, crashes, and traffic violations. This strategy has progressed through an interest in identifying hot spot locations and causative variables associated with incidences in selected areas. The identification of hot spots varies with the type of analysis employed and may include counts of crashes on roadway segments, counts of crashes within a defined grid system, and the use of spatial analysis. The aggregated counts of crashes both on roadway segments and within gridding systems may allow for a simple-to-conduct and easy-to-comprehend examination of alcohol-related crashes. The use of spatial analysis allows for an investigation into the spatial distribution of the crashes and their contributing factors. The distribution and the variability between contributing crash factors is important in addressing the hazardous issues within each specific area. Spatial analysis, through the identification of hot spots, establishes specific areas that may be used for the implementation of enforcement patrols. These hot spots provide a means of identifying
the location in which to implement strategies for reducing the number of crashes and their injury severity. Maistros et al. (2014) described the performance of alcohol-related safety campaigns such as saturation and corridor patrols that were located using hot spots.

There are several types of spatial analyses that may be used to identify hot spots of motor vehicle crashes. Some examples of commonly used analysis methods for identifying the spatial autocorrelation between each crash location rise from the use of kernel density estimation (KDE), Getis-Ord \( Gi \), and Moran’s \( I \). KDE has shown its viability in terms of identifying high risk locations in which crashes occur (Backalic, 2013; Plug et al., 2011; Pulugurtha et al., 2007; Schneider et al., 2001). The use of the \( Gi^* \) statistic and Moran’s \( I \) have also shown exceptional abilities in identifying spatial autocorrelation between crashes and their attributable contributing factors (Songchitruska and Zeng, 2010; Truong and Somenahalli, 2011; Kuo, 2013). One important aspect of using spatial analysis to determine the location of hot spots is for the legal implementation of safety campaigns within the defined areas. The combination of using the \( Gi^* \) statistic and interpolation allows for an unbiased, statistical identification of the location of the hot spot. While this unbiased identification is preferred, there is still some differentiation between the approaches used by some researchers for conducting a spatial analysis.

This differentiation in the approach to the analysis may be seen in the calculation of the distances separating each crash, which is essential to the calculations included in the spatial analyses. The results vary when using a Euclidean versus network-based distance in the calculation of the hot spot. In the use of the Euclidean distance, a straight-line calculation from one crash location to another is observed. This relationship is often
also called “as the crow flies.” The network-based distance, on the other hand, follows along the path of existing roadways. In this approach, the calculation follows between two crash locations and must follow a pattern that a vehicle may travel. The only exception to this path of travel is that the path may not include parking lots or private roads, which a driver of a vehicle is not likely to use.

Euclidean distances have been used in the calculation of spatial autocorrelation when routing law enforcement patrol operations (Kuo et al., 2013). Euclidean analyses are often used within the development of patrol operations for a number of reasons, including increased flexibility to patrol routes, or software/computer capabilities. One argument against the use of Euclidean analyses is the idea that an analysis that includes a field or parking lot may be misrepresentative. However, to those patrolling the roadways, the use of Euclidean analyses may allow law enforcement officers to broaden their search efforts to patrol locations that may otherwise not be indicated within a hot spot. An analyst trying to identify hot spots for law enforcement patrol may encounter limitations when analyzing crash patterns along a network is not often supported. Thus, the analyst would have to write his or her own program to accomplish this task.

Even though Euclidean based calculations are still currently used in spatial analyses, the use of a roadway network to constrain spatial analyses is on the rise. Some researchers, such as Young and Park (2014), use this type of analysis in an effort to identify heightened areas of crash occurrence. Even though the use of a network distance theoretically seems more beneficial to use, continued research and applications in practice still revert to the use of Euclidean distances. While the use of Euclidean distances does provide the abovementioned benefits, the variations in use within applied
implementations are of the most concern. The important aspect to consider is the way each method affects the identification of roadways that law enforcement, aiming to reduce alcohol-related crashes, may legally patrol.

When using hot spot maps in safety campaigns, the main requirement is for the locations of the enforcement patrols to withstand scrutiny in court hearings when a driver suspected of OVI is under investigation. In cases such as this, the drivers may claim they were illegally targeted. In an effort to maintain the legality of a particular traffic stop, the map identifying the location of the traffic stop must be accurate. Spatial analyses conducted using both Euclidean and network-based distances require accurate identification of the roadways in which law enforcement may patrol. The differing methods may produce results with large ramifications concerning the legality of a traffic stop involving a driver who is suspected of OVI. This research investigates the variation between each of the two types of analysis and compares the resulting roadways identified as hot spots.

3.2 Data

This investigation focuses on alcohol-related crashes occurring in Cuyahoga County, Ohio, from January 1, 2012, through April 9, 2015. The crash data used in this study were obtained from the crash report database maintained by the Ohio Department of Public Safety (ODPS). A total of 3,469 crashes were reported within the studied time period and geographic area in which the reporting officer identified the crash to be alcohol-related. Of these, a total of 3,365 crashes are able to be geocoded by using the longitude and latitude of the reported location of the crash.
The ODPS database contains all reported vehicle crashes in the state of Ohio and includes the injury severity levels of occupants of the vehicles involved in the crash. The range of injury for the highest injury severity realized for all parties involved in a geocoded crash in Cuyahoga County in which a driver was suspected of OVI may be seen in Table 3.1.

Table 3.1. Injury Severity for Geocoded Alcohol-Related Crashes in Cuyahoga County

<table>
<thead>
<tr>
<th>Injury Severity</th>
<th>Number of Crashes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Property Damage Only</td>
<td>1900</td>
</tr>
<tr>
<td>Injury</td>
<td>1400</td>
</tr>
<tr>
<td>Fatal</td>
<td>65</td>
</tr>
<tr>
<td>Total</td>
<td>3,365</td>
</tr>
</tbody>
</table>

Note: Dataset includes alcohol-related crashes that occurred from January 1, 2012, through April 9, 2015. The injury severity relates to the highest severity realized for all parties involved in a crash.

3.3 Methodology

The general methodology of calculating the hot spots for both the Euclidean and network-based distances is essentially the same. This process includes 1) a weighting of the crash severities, 2) the development of spatial weights matrices, 3) the calculation of spatial autocorrelation, and 4) an interpolation of the autocorrelation. The difference between the two analysis approaches resides in the development of the spatial weights matrices, where distances separating one crash from another are calculated using different methods. With the differences in the spatial weights matrices, the resulting hot spot locations obtained for both analysis approaches may then be compared.
3.3.1 Crash Severity Weighting

This study weighs each of the crashes based on the highest injury severity of all members involved in a crash where a driver is suspected of OVI, similar to the process used by Truong and Somenahalli (2011), in which an increasing value was associated to higher injury severities. The weighting system used in this research places a greater importance on higher severity crashes. These weights are based on the societal crash costs, as determined by the American Association of State Highway Transportation Officials (AASHTO) in the Highway Safety Manual (AASHTO, 2010). AASHTO divides the societal crashes into three general severity categories: fatality (K), injury (A/B/C), and property damage only (O). The associated costs in 2001 dollars are $4,008,900 for a fatal crash, $82,600 for a crash with injuries, and $7,400 for a crash with property damage only.

3.3.2 Spatial Weighting

Differences between the Euclidean and network-based analysis approaches first become apparent in the development of the spatial weights matrix. The matrix for each type of analysis is developed using a binary system dependent upon a threshold distance. This threshold distance is the distance where all crashes have at least one neighbor. All crashes that occur within the threshold distance receive a value of 1, while all crashes that occur beyond the threshold distance receive a value of 0. As a result of the variation in the distance calculation used for each approach, the resulting spatial weights matrices may differ.
3.3.3 Spatial Autocorrelation

The method of calculating the spatial autocorrelation does not change based on the type of analysis being conducted when using either the Euclidean or network-based distances. However, due to the differing spatial weights matrices, the resulting values of the spatial autocorrelation may vary from one analysis approach to another. The measure of spatial autocorrelation used for this study is the Getis-Ord $G_i^*$ statistic. This statistic has previously been shown to identify the areas where crash risk is of concern (Khan et al., 2008; Sonchitruska and Zeng, 2010; Truong and Somenahalli, 2011; Prasannakumar et al., 2011; Kuo et al., 2013). The $G_i^*$ statistic is calculated using the following equation:

$$G_i^* (d) = \frac{\sum_j w_{ij}(d) x_j - W_i^* \bar{x}}{s \left[ \frac{W_i^*}{(n-1)} \right]^{1/2}}$$

(3.1)

where:

$$W_i^* = \sum_j w_{ij}(d)$$

(3.2)

$$s^2 = \frac{\sum_j x_j^2}{n} - \bar{x}^2$$

(3.3)

where $w_{ij}(d)$ is the spatial weights matrix, $x_j$ is the cost associated with the injury severity, $\bar{x}$ is the average of all studied societal costs, and $n$ is the total number of crashes (Prasannakumar et al., 2011).

The result of the $G_i^*$ statistic is a z-score describing the dispersion of crashes based on the weighted injury severity and the distance separating each crash from one another. The null hypothesis for this statistic is that the spatial distribution of crashes and their severities are randomly distributed. The locations that are positive and statistically significant are regarded as clusters of high severity crashes, “hot spots”. Meanwhile, the
locations that are negative and statistically significant are regarded as clusters of low severity crashes, “cold spots”.

3.3.4 Interpolation of Spatial Autocorrelation

Once the spatial autocorrelation of the crashes and severities is known at each crash location, a means to patrol each significantly clustered location may be developed. This could be accomplished by either having a law enforcement officer drive a specific road or path through each significant cluster or by identifying an area in which the officer may travel. By allowing the officer to only focus on patrolling points, the legality of stops made at locations that were not spatially investigated may come into question. On the other hand, when an area is defined within a hot spot for an officer to patrol, the legality of stops is statistically backed. In order to provide a statistically backed area (instead of a list of specific points), the value of the spatial autocorrelation must be interpolated throughout the entire study area. Inverse distance weighting (IDW) interpolation is used to identify the z-score along all sections of roadway. Mehdi et al. (2011) describes IDW as an interpolation method that predicts unknown values based on their distance from known values. IDW is calculated through the following equation:

\[
 z_0 = \frac{\sum_{i=1}^{s} z_i \frac{1}{d_i^k}}{\sum_{i=1}^{s} \frac{1}{d_i^k}} 
\]  

(3.4)

where, \( z_0 \) is the estimated value at point 0, \( z_i \) is the measured value at point \( i \), \( s \) is the number of points used to estimate the unknown value, \( d_i \) is the distance between points \( i \) and 0, and \( k \) is the power identifying the influence of distance (Ansari and Kale, 2014). The interpolation of \( Gi^* \) values is calculated using both the Euclidean and network-based distances. This allows for the effect of distance relationships to also be investigated.
3.3.5 Comparison

A comparison between the two analysis approaches is conducted through an examination of the societal crash cost of crashes located on high risk roads and the length of roadways identified as high risk. The first comparison is completed using the prediction accuracy index (PAI), initially presented by Chainey et al. (2008). This index allows for an examination of the accuracy of hot spots (Tompson and Townsley, 2010), which presents a ratio of the crashes occurring within a hot spot to the size of the hot spot. Thakali et al. (2015) updated the PAI by modifying the denominator of the equation to account for the length of roadway for the identified hot spots. A further modification to the numerator of the equation is conducted through this research, in which the aggregated societal crash cost of crashes is analyzed instead of the aggregated number of crashes. The equation used in this research to calculate the PAI may be seen in the following equation:

\[ PAI = \frac{c \times 100}{l \times 100} \]  

where, \( c \) is the societal crash cost of crashes in hot spots, \( C \) is the total societal crash cost of all crashes within the study area, \( l \) is the length of roadways identified as being located in the hot spot, and \( L \) is the total length of roadways within the entire study area. Thakali et al. (2015) indicates that the mapped hot spot that contains a larger PAI is more beneficial. This benefit comes from having a hot spot with a higher crash potential identified in a smaller area of concern. This would provide an increase in efficiency as the patrolling law enforcement officer(s) would attend to more a concentrated location, while not traveling on unnecessary roads.
Once a comparison of the PAI is completed, an investigation into which factors contributed the greatest influence to the resulting PAI values may be conducted. This investigation is conducted through the percent difference of both the societal crash cost of crashes located within hot spots and the length of roadways identified as hot spots. The percent difference for the societal crash cost would compare the total societal crash cost of crashes that occur within the hot spot as determined though each type of analysis, both Euclidean and network-based. Similarly, the percent difference for the length of roadway would compare the total length of roads included in the hot spot for each analysis approach.

3.4 Results

The calculation of spatial weights matrices for the Euclidean and network-based analyses is crucial for facilitating a comparison between the two approaches. The threshold distance was calculated so that each crash has at least one neighbor, found to be 7,414.7 feet and 16,364.8 feet for Euclidean and network-based analysis, respectively. The difference in length resides in the fact that the network-based distance is restricted to following along the path of the roadways, while the Euclidean distance is permitted to follow a straight-line path from one crash location to another. This may lead to large variations in the distance between two points, as one distance may travel through a city block and another may be at least twice as long, traveling around the block. The difference in distance measurements may expound even further within rural areas, as the distance required to travel around a subdivision may be much longer than through a back yard. Since spatial analyses examine the distribution of crash locations, any large variations in distance vastly change the results.
Using the developed spatial weights matrix for each analysis approach, the spatial autocorrelation of the crashes and their injury severities was able to be determined through the calculation of the $Gi^*$ statistic. The significance of clustering for Cuyahoga County, determined by the value of the z-score at each crash location for both the Euclidean and network-based distances, is shown in Figure 3.1.

![Cluster Significance](image)

Euclidean calculation  Network-based calculation

Note: Hot spots represent locations where high injury severity crashes are close in distance to other high severity crashes. Cold spots represent locations where low severity crashes are close in distance to other low severity crashes.

Figure 3.1. Comparison of $Gi^*$ z-scores obtained by Euclidean and network-based analysis for Cuyahoga County.

The cluster significance shown in Figure 3.1 provides a basis for law enforcement agencies to use in focusing their patrol activities. While these points identify locations where incidents are known to have occurred and their related risks, it may be difficult to legally back the traffic stops a law enforcement officer may make while traveling to and from each identified location. Another option would be to allow law enforcement agencies to patrol an area designated by specific boundaries in which a high risk for crashes occurs. In an effort to achieve suitable boundaries, interpolating the z-score of each known cluster would aid in defining an operable area, which identifies where a
similar crash is likely to occur. Even though a crash has not occurred at every location within the study area, it is assumed that locations may share similar characteristics when they are in close proximity to one another.

Once the interpolation of the z-scores is completed, a comparison of the two analyses may be made. While analyses may include distance measurements obtained via two approaches in the calculation of the $Gi^*$, there are also two interpolation methods that may be conducted based on the distances used to determine the IDW. Consequently, three different analysis combinations are investigated: 1) Euclidean $Gi^*$ calculations and Euclidean interpolation (represented as EE), 2) Network-based $Gi^*$ calculations and Euclidean interpolation (represented as NE), and 3) Network-based $Gi^*$ calculations and Network-based interpolation (represented as NN). The results for the network-based interpolation used in NN are obtained through the use of SANET (ver. 4.1). The resulting significantly clustered areas may be seen in Figure 3.2.
Note: Network interpolation completed with the use of SANET ver. 4.1.

Figure 3.2. Comparison of hot spot areas between Euclidean and network-based analysis.

The hot spots resulting from the various combinations of Euclidean and network-based analyses appear to be similar, as may be seen in Figure 3.2. However, it is important to determine the exact boundaries of the hot spots and whether each spot includes an additional 1, 10, or more roadways. The variation between the boundaries identified using the two approaches may present enough of a legal rationale for a case against a suspected driver OVI to be dropped due to an inappropriate stop.
When comparing the three analysis combinations, it is important to identify which roadways are deemed to be high risk in both the Euclidean and network-based roadways. These high risk roadways are ones which, when interpolated, contain a crash severity with a cluster significant z-score greater than or equal to 1.96, which relates to a 95% level of statistical significance. The roadways in Cuyahoga County that were identified to be of high risk based on significant clusters of high severity crashes, from the Euclidean, network, and both Euclidean and network-based analyses may be seen in Figure 3.3.
Note: The abbreviations EE, NE, and NN represents the type of distance used within the calculation of the $Gi^*$ and interpolation. EE indicates Euclidean based $Gi^*$ and Euclidean based interpolation. NE indicates network based $Gi^*$ and Euclidean based interpolation. NN indicates network based $Gi^*$ and network based interpolation. Network interpolation completed with the use of SANET ver. 4.1.

Figure 3.3. Identification of hazardous roadways.

From Figure 3.3, it may be seen that these identified roadways are very similar from one analysis combination to another. However, small differences appear when looking at the overlap in the resulting locations. It has been determined that when comparing EE to NE, 64.3% of roadways identified by NE are also identified by EE; in contrast, only 43.8% of
roadways identified by EE are also identified by NE. This indicates that only about half of the roadways are similar between EE and NE. By only having approximately half of the significant roadways overlapping, there would be a major discrepancy in the location of an implemented safety campaign. This discrepancy plays a large role in the legality of such safety campaigns, as incorrectly targeting a driver suspected of OVI may be a cause for case dismissal. This trend may also be seen when comparing EE to NN; 63.1% of roadways identified by NN are also identified by EE, while only 42.2% of roadways identified by EE are also identified by NN. However, the overlap between the different types of analyses increases drastically when comparing NE to NN. A total of 91.2% of roadways identified by NN are also identified by NE, while 89.7% of roadways identified by NE are also identified by NN. The results from the third combination indicate that the roadways identified by NE and NN are very similar and cover nearly all of the same roadways. This relationship may be seen in Figure 3.3, where the hot spot areas identified by NE or NN cover many of the same locations as that of EE. However, when examining the comparison in the reverse order, the area of concern in EE includes a larger area that extends beyond that of NE or NN. In other words, the network-based calculation of the $Gi^*$ identifies similar areas as the ones obtained for the Euclidean $Gi^*$ analysis; meanwhile, the Euclidean $Gi^*$ analysis may be unnecessarily large and include roadways that may be inappropriately patrolled.

The relationship between the crashes in the dataset and the identified high-risk crash locations (hot spots) was also examined to facilitate a comparison between the three analysis combinations. This examination, through an investigation of the PAI, provides a parameter that permits the comparison of the two analyses for evaluating crashes and
allows the resulting high-risk locations to be identified. The total societal crash cost for all geocoded crashes within the study period is $390,278,500. The total length of roadways in the study area is 5,419.6 miles. These two values, the total cost and roadway length, are compared to the societal crash costs and roadway lengths included in the hot spots to obtain the PAI value for each analysis combination. The societal crash costs, roadway lengths, and PAI values for each analysis combination are presented in Table 3.2.

Table 3.2. PAI comparison.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Societal Crash Cost of Crashes in High Risk Area</th>
<th>Length of Roadway Identified as High Risk</th>
<th>PAI Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EE</td>
<td>$24,996,400</td>
<td>230.1 miles</td>
<td>1.51</td>
</tr>
<tr>
<td>NE</td>
<td>$39,461,900</td>
<td>156.6 miles</td>
<td>3.50</td>
</tr>
<tr>
<td>NN</td>
<td>$39,386,700</td>
<td>154.0 miles</td>
<td>3.55</td>
</tr>
</tbody>
</table>

Note: The abbreviations EE, NE, and NN represents the type of distance used within the calculation of the $G_i^*$ and interpolation. EE indicates Euclidean based $G_i^*$ and Euclidean based interpolation. NE indicates network based $G_i^*$ and Euclidean based interpolation. NN indicates network based $G_i^*$ and network based interpolation.

The difference between the PAI values obtained for the Euclidean and network-based analyses may be seen in Table 3.2. The analyses that use a network-based $G_i^*$ have larger PAI values (3.50 and 3.55) as opposed to the value where the $G_i^*$ was calculated using a Euclidian approach (1.51), indicating the ability of the network-based analysis to identify a more highly concentrated societal crash cost than the Euclidean analysis. The increased concentration of high severity crashes allows for a larger impact to be realized when using the same law enforcement resources to cover each area, as more locations that contribute to the high severity crashes will be patrolled.

The percent difference in the societal costs is shown in Table 3.3.
Table 3.3. Percent difference in societal crash costs.

<table>
<thead>
<tr>
<th></th>
<th>NE</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>EE</td>
<td>44.88%</td>
<td>44.70%</td>
</tr>
<tr>
<td>NE</td>
<td>0.19%</td>
<td></td>
</tr>
</tbody>
</table>

Note: The abbreviations EE, NE, and NN represents the type of distance used within the calculation of the Gi* and interpolation. EE indicates Euclidean based Gi* and Euclidean based interpolation. NE indicates network based Gi* and Euclidean based interpolation. NN indicates network based Gi* and network based interpolation.

From Table 3.3 it may be seen that the largest variation between each of the analyses is the use of Euclidean based distances in the calculation of the Gi*. The use of Euclidean or network-based distances within the interpolation of the hot spots has a very minimal impact. The difference in the societal crash costs for the crashes when the Gi* calculation in the analysis is calculated using a network-based distance rather than a Euclidean distance is approximately $14,400,000. This indicates that in an effort to have the largest economic impact in crash reduction, using hot spots based on network based spatial autocorrelation is necessary.

In a similar fashion to those differences described for the societal crash costs, the percent difference for each of the three types of analyses may be seen in Table 3.4.

Table 3.4. Percent difference in length of roadway.

<table>
<thead>
<tr>
<th></th>
<th>NE</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>EE</td>
<td>38.01%</td>
<td>39.62%</td>
</tr>
<tr>
<td>NE</td>
<td>1.67%</td>
<td></td>
</tr>
</tbody>
</table>

Note: The abbreviations EE, NE, and NN represents the type of distance used within the calculation of the Gi* and interpolation. EE indicates Euclidean based Gi* and Euclidean based interpolation. NE indicates network based Gi* and Euclidean based interpolation. NN indicates network based Gi* and network based interpolation.

From Table 3.4, it may again be seen that the largest variation between each of the analyses is the use of Euclidean distances in the calculation of the Gi*. The difference in
the length of roadway between the analyses in which the $Gi^*$ calculation is either Euclidean or network-based is 76 miles. This indicates that using the Euclidean based spatial autocorrelation includes much more roadways than the network counterparts and would allow law enforcement to more flexibility in the areas that they patrol.

The use of Euclidean or network-based distances within the interpolation of the hot spots again has a very minimal impact. From examining the difference in percentages for both the societal crash cost and the length of roadway, it may be seen that there is a large difference within the length of roadway with respect to the societal crash costs. This difference is a major contributor to the variance in the PAI value obtained from the two analyses. The analysis containing the larger PAI value is the one that would be most beneficial for implementation. This would allow law enforcement agencies to use the least amount of resources (funds, manpower, etc.) to realize the highest economic impact (societal cost savings).

3.5 Conclusions

Hot spots provide a great opportunity to identify problem locations. The ability to accurately locate hot spots is pivotal in the use of such maps for focusing OVI enforcement patrols. The maps for developing patrol areas must be legally sound, and the legality of the maps comes from appropriately identifying roads in which to patrol. A large opposition to a traffic violation could be that the driver was targeted on a road that was inaccurately identified as hazardous. The appropriate roads to patrol are those that significantly contribute to hazardous conditions. Through statistically identifying a risk associated with roadways, bias may be removed from the development of patrol routes.
While advances have been made to identify hot spots for vehicle crashes, a discrepancy has been noted between the approaches used for the calculation of distances separating crashes. Previous research efforts to identify these hot spots have used two different approaches: using either Euclidean distances or using network-based distances. A Euclidean analysis examines the spatial distribution of crashes in a straight line distance from one crash to another, irrespective of the presence of buildings, water, fields, or other features. In contrast, a network-based analysis examines the spatial distribution of crashes along the path of roadways. In the latter approach, the distance separating one crash from another may only be calculated over a path that vehicles are capable of traveling.

Because both analysis approaches are currently in use and are important within the identification of high-risk areas for public safety campaigns, an investigation comparing the Euclidean analysis versus a network-based analysis was conducted. This comparison examined the relationship of vehicle crashes and identified high-risk roadways using each approach. The results indicate that using network-based distances in the calculation of spatial autocorrelation will produce a higher PAI than a spatial autocorrelation employing Euclidean distances. This signifies a greater societal crash cost per mile for high-risk roads, which would aid in more efficiently and accurately identifying hot spots for law enforcement purposes. The results of the comparisons between selected combinations of analysis approaches indicate that the NE and NN analyses return very similar results. However, the results for the NE and NN analyses differ greatly from the EE combination, where a Euclidean distance is used to calculate the $Gi^*$ spatial autocorrelation. These relationships are indicated by the NE and NN
analyses containing much larger societal crash costs while having the hot spots contained within a much smaller roadway length. Law enforcement would benefit from using either the NE or NN rather than EE combination, as these analyses would result in increased deployment efficiency for patrol efforts. From a standpoint of the legality of OVI stops, the network-based analysis provides a more compact area that does not unnecessarily identify additional roadways to be patrolled. The removal of unnecessary roadways reduces the potential for a traffic stop to be challenged due to targeting a driver on a roadway that may not be hazardous. Having an analysis that is more legally sound will reduce the ability of a suspected driver OVI to claim that they were illegally targeted. Additionally, identifying hot spots that require fewer roads to be patrolled while still targeting areas with high societal crash costs may effectively increase the efficiency of law enforcement efforts.

Overall, the effect of using network distance over Euclidean distances in the interpolation of crash spatial autocorrelation is minimal. While the network-based distances provide slightly better results, those analysts who either lack access to appropriate software or have computers with limited processing capacity may be more suited to interpolate hot spots using Euclidean distances, which are more readily obtained. However, the same is not the case for the calculation of the $Gi^*$ statistic, in which large differences are realized, and the use network-based distances is able to identify high-risk areas more effectively.
CHAPTER IV

A SPATIO-TEMPORAL HOT SPOT EXAMINATION OF ALCOHOL-RELATED SINGLE AND MULTIPLE VEHICLE CRASHES

4.1 Introduction

In 2012, there were 10,322 people killed in crashes throughout the United States where a vehicle operator had a blood alcohol concentration (BAC) of 0.08% or greater (NHTSA, 2015), accounting for 31 percent of all traffic related fatalities. This trend has continued at the same rate for the 15-year span between 1997 and 2012. The influence of alcohol on decision making and on the maneuvering skills of a driver have been well documented and researched, as indicated through studies by Holloway (1995), Mitchell (1985), and Ogden and Moskowitz (2004). The implications of alcohol extend across various types of motor vehicles, from motorcyclists doubling their chance of a fatality (Schneider and Savolainen, 2011) to the drivers of passenger vehicles being involved in higher severity crashes (Zhu and Srinivasan, 2011).

Many tactics are being applied to reduce the number of alcohol-related crashes. These tactics may range from informational outreach programs presented by educators to presence related target enforcement implemented by law enforcement officers. Educational programs allow for drivers to realize the impacts their actions will have upon themselves and other motorists. These programs may reflect upon the relative risk
associated with increased alcohol consumption (Zador, 1991) or the increased likelihood of injuries and death due to alcohol use (Hingson and Winter, 2003). The safety campaigns enacted by law enforcement aim to stop an intoxicated driver prior to a crash occurring. The performance of two tools used within these safety campaigns, such as saturation patrol and corridor patrol, has been examined by Maistros et al. (2014). The outcome of enforcement campaigns rely on the locations where the campaigns are implemented.

Spatial analyses are used in the determination of locations in which there are high alcohol-related crash rates. The identified locations may then be ideal for the implementation of target enforcement. The spatial analyses often investigate crashes based on multiple years of data combined together. The locations of interest are then determined purely on the spatial aspect of the crashes. Meliker et al. (2004) analyzed a little over two years of crash locations to spatial analysis to identify the presence of clustering in alcohol-related crashes. Meanwhile, Huang et al. (2010) examined five years of data, linking spatial autocorrelation to socioeconomic factors such as age and income. The identification of spatial patterns provides a location that may be targeted towards reducing crashes and injury severity; however, the optimal time to target these areas is unknown.

While the results of spatial investigations are very important and beneficial, there may be trends that go unnoticed due to changes in temporal periods. Temporal changes in spatial patterns of alcohol-related crashes are very important to investigate, as the presence of events or holidays may have an influence on drinking-driver occurrence. Farmer and Williams (2005) examined average deaths per day and average deaths per
hour in order to identify high death rates and alcohol involvement on holidays such as Independence Day and New Year’s Day. While dates such as this are useful, it is difficult to know the location in which such crashes occur.

The next step is to consider the spatial-temporal realm, which combines the aspects of both the spatial and temporal analyses together. Spatio-temporal analyses have been categorized into three different types, including map animation, isosurfaces, and comaps (Brunsdon et al., 2007; Plug et al., 2011). Benefits and drawbacks for each of these methods have been described by Plug et al. (2011). The benefits include map animation’s use of clear visualizations, isosurface’s examination in three-dimensions, and comap’s display of consecutive maps. The drawbacks from using these methods include map animation’s need to be replayed multiple times for understanding and isosurface’s computational requirements. Prasannakumar et al. (2011) used a basic version of comaps, breaking the temporal time span into two different groups, monsoon season and non-monsoon season. Li et al. (2007) dove deeper into the use of comaps by comparing morning versus evening peak hours of travel and weekday versus Friday, Saturday, and Sunday.

This research compares the movement of hot spots by examining isosurfaces created from the Getis-Ord $Gi^*$ statistic. The goal of this research is to identify the variation between single vehicle alcohol-related crashes and multiple vehicle alcohol-related crashes. The use of the $Gi^*$ statistic has shown to be a useful way to determine locations of clustered crashes (Getis and Ord, 1992; Khan et al., 2008; Kuo et al., 2013; Prasannakumar et al., 2011; Songchitruska and Zeng, 2010; Truong and Somenahalli, 2011). The application of the moving timeframe to the $Gi^*$ statistic allows for crashes to
be identified as spatially relevant as long as they occur during a similar time period. The result of this research provides a further understanding of alcohol-related crashes both in the relationships between single and multiple vehicles and how crash patters change over time. By identifying the movements of crash patterns, shifts in tactics to reduce the number and severity of alcohol-related crashes may occur. These shifts would move the target location of implementations such as saturation or corridor patrols as clusters of crashes appear and disappear throughout the course of time. If these shifts did not occur, a target location may continually be used after a cluster disappears or at inappropriate times.

4.2 Data

This study analyzes crash records from the OH-1 crash reports, maintained by the Ohio Department of Public Safety, dating from January 1, 2012, through April 9, 2015. Specifically, alcohol-related crashes are investigated within Cuyahoga County, which contains one of the largest numbers of alcohol-related crashes from counties within the state and annually records over 1,000 alcohol-related crashes per year. These crashes were then subdivided into single vehicle and multiple vehicle data sets, which related to a total of 1,432 and 1,933 crashes, respectively. Single and multi-vehicle crashes have routinely been identified as being related to different crash mechanisms (Ivan et al., 1999; Qin et al., 2004; Geedipally and Lord, 2010). Therefore, the examination of these two types of crashes provides great insight into crashes that may exhibit different characteristics spatio-temporally. These studied crashes are further described in the following table.
Table 4.1. Descriptive Statistics of Alcohol-Related Crashes in Cuyahoga County, Ohio.

<table>
<thead>
<tr>
<th></th>
<th>Single</th>
<th></th>
<th>Multiple</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Percent</td>
<td>Count</td>
<td>Percent</td>
</tr>
<tr>
<td>Total Crashes</td>
<td>1432</td>
<td>1933</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time of Day</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5:00AM-12:00PM</td>
<td>122</td>
<td>9%</td>
<td>168</td>
<td>9%</td>
</tr>
<tr>
<td>12:00PM-5:00PM</td>
<td>109</td>
<td>8%</td>
<td>164</td>
<td>8%</td>
</tr>
<tr>
<td>5:00PM-11:00PM</td>
<td>463</td>
<td>32%</td>
<td>627</td>
<td>32%</td>
</tr>
<tr>
<td>11:00PM-5:00AM</td>
<td>738</td>
<td>52%</td>
<td>974</td>
<td>50%</td>
</tr>
<tr>
<td><strong>Day of Week</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sunday</td>
<td>208</td>
<td>15%</td>
<td>236</td>
<td>12%</td>
</tr>
<tr>
<td>Monday</td>
<td>129</td>
<td>9%</td>
<td>198</td>
<td>10%</td>
</tr>
<tr>
<td>Tuesday</td>
<td>133</td>
<td>9%</td>
<td>189</td>
<td>10%</td>
</tr>
<tr>
<td>Wednesday</td>
<td>168</td>
<td>12%</td>
<td>214</td>
<td>11%</td>
</tr>
<tr>
<td>Thursday</td>
<td>164</td>
<td>11%</td>
<td>265</td>
<td>14%</td>
</tr>
<tr>
<td>Friday</td>
<td>292</td>
<td>20%</td>
<td>425</td>
<td>22%</td>
</tr>
<tr>
<td>Saturday</td>
<td>338</td>
<td>24%</td>
<td>406</td>
<td>21%</td>
</tr>
<tr>
<td><strong>Injury Severity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Property Damage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Only</td>
<td>815</td>
<td>57%</td>
<td>1085</td>
<td>56%</td>
</tr>
<tr>
<td>Injury</td>
<td>584</td>
<td>41%</td>
<td>816</td>
<td>42%</td>
</tr>
<tr>
<td>Fatal Injury</td>
<td>33</td>
<td>2%</td>
<td>32</td>
<td>2%</td>
</tr>
<tr>
<td><strong>Road Contour</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Straight Level</td>
<td>935</td>
<td>65%</td>
<td>1591</td>
<td>82%</td>
</tr>
<tr>
<td>Straight Grade</td>
<td>143</td>
<td>10%</td>
<td>209</td>
<td>11%</td>
</tr>
<tr>
<td>Curve Level</td>
<td>213</td>
<td>15%</td>
<td>78</td>
<td>4%</td>
</tr>
<tr>
<td>Curve Grade</td>
<td>136</td>
<td>9%</td>
<td>46</td>
<td>2%</td>
</tr>
<tr>
<td>Unknown</td>
<td>5</td>
<td>0%</td>
<td>9</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Speed Related</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>579</td>
<td>40%</td>
<td>522</td>
<td>27%</td>
</tr>
<tr>
<td>No</td>
<td>853</td>
<td>60%</td>
<td>1411</td>
<td>73%</td>
</tr>
</tbody>
</table>

Note: For the crashes by day of week, those crashes that occur prior to 5:00 am are attributed to the previous day’s count.

Similarities and differences between single and multi-vehicle alcohol-related crashes may be seen in Table 4.1. The time of day and injury severity are very similar, down to within about 1 percent of each other. Even the difference based on day of the week between the
two groups of crashes have very similar timelines. Differences may then be seen based on
the contour of the road or whether the crash was speed related. Single vehicle alcohol-
related crashes have a higher occurrence of being located on curved roadways and being
speed related than their multi-vehicle counterparts. The differences in the road contour
and use excessive use of speed indicate that there may be variation in the location of
where these crashes are occurring.

Through an examination of the temporal spectrum of Table 4.1, it may
unsurprisingly be seen that at least half of the crashes occur between in the late night
hours, between 11:00 pm and 5:00 am. Those people who are drinking alcohol and
operating a vehicle at this time are usually doing so as a continuation of activities from
the previous night. To account for those people who may have begun drinking from the
night before, pre-midnight, those crashes that occur prior to 5:00 am were attributed to
the previous day. For example, a person goes out to a bar on a Saturday night and leaves
when the bar closes and crashes his or her vehicle on Sunday morning. Within this
example situation, the crash would be attributed to Saturday. This plays a large role into
the identification of crash occurrence by the day of the week. From Table 4.1, it may be
seen that as the days progress through the week, a peak number of crashes occur at the
end of the week, with at least 40 percent of the crashes represented on Friday and
Saturday. These temporal trends follow common thoughts that alcohol-related crashes
occur at night and on the weekend.

In order to analyze the crashes based on a temporal aspect of the spatio-temporal
analysis, the crash data was reorganized based on two different time scales, time of day
and day of week. All crashes from the three years of data were then condensed into one
complete cycle (either twenty-four hours or seven days) on both time scales (time of day or day of week). In order to identify clusters that may appear at either end of a cycle, the crashes at the beginning of the cycle were repeated at the end of one complete time period. For example, after all of the crashes were condensed into a single 24-hour time period, the first two hours of crashes were repeated onto the following day’s time. In total, the analysis then covers 26 hours and allows for clustering to be identified at the beginning and end of the day. The use of the first two hours again as the last two hours allows, for example, a crash at 12:30 am to be related to ones that occur at both 11:30 pm and 1:30 am.

4.3 Methodology

This research examines the spatio-temporal distribution of single and multi-vehicle alcohol-related crashes. The spatial patterns, while important, only paint one part of an overall picture. Analyzing crashes purely on a spatial analysis only gives an indication of where crashes are occurring if they were to occur at the same time. The idea of examining crashes solely on an individual basis misses some key relationships that have been exposed through temporal examination.

As described within the data section, the commonly believed temporal pattern is that alcohol-related crashes occur at night and during the weekend. While this study investigates the influence of temporal components to alcohol-related crashes, the objective of this study is not to reaffirm this belief. The objective is to identify the movement of clustered crashes as time progresses throughout the day or week. While many crashes may occur at these known times, there may be clusters of high severity crashes that occur in a wide variety of locations throughout the day or week. The
identification of these multiple locations and their shift in movement throughout time is
the objective of this research. It would be inappropriate to maintain a target location at
one site throughout an entire day or week, as the pattern would be likely to move
throughout the county.

The location of clusters throughout time is identified by examining the spatial
autocorrelation of crashes as time progresses. The examination of spatial autocorrelation
is identified through the use of the Getis-Ord $G_i^*$ statistic. The ability to identify the
spatial autocorrelation as time progresses is accomplished by implementing a moving
timeframe that determines which crashes are neighbors with one another. Those crashes
that are considered to be neighbors occur within a specified time period and distance
from one another. The determination of the time period and distance are further explained
in the spatial weights matrix section. As time continues, crashes are either included or
excluded from spatial autocorrelation analysis. Multiple iterations of spatial
autocorrelation are examined through this use of the moving timeframe.

The spatial analysis and the spatio-temporal analysis are conducted in a very
similar manner. The only difference is that the temporal components are removed for the
spatial analysis. This temporal component is present within the spatial weights matrix and
the cluster grouping analysis. In order to aid in the identification of spatial distribution
within the spatial analysis, the significance of the clustering values is interpolated using
inverse distance weighting (IDW). All distances that are used within the calculation of
these spatial and spatio-temporal analyses are network based distances that follow along
the path of the roadway system. An in-depth explanation of the processes used within the
spatial and spatio-temporal analyses is described in the remainder of this chapter.
4.3.1 Crash Weighting

The spatial autocorrelation between one crash and another is determined based on the injury severity of the crash, similar to that conducted by Truong and Somenahalli (2011). Within this research, the highest injury severity of all parties involved in each crash is used as the record’s overall weight. The recorded injury severities pertain to three levels of severity: fatal injury (K), injury (A/B/C), and property damage only (O). These injury severity levels then correlate directly to the societal cost of crashes identified in the Highway Safety Manual (AASHTO, 2010). These crash cost guidelines attribute a higher weight to crashes that contain higher injury severities.

4.3.2 Spatial Weights Matrix

Within both the spatial and spatio-temporal analyses, the spatial weights matrix designates which crashes are deemed as neighbors with one another based on the distance of separation of two given crashes. A binary system is used in the creation of the matrix to identify which crashes are neighbors with one another. Those crashes that are neighbors receive a value of 1; those crashes that are not neighbors receive a value of 0. Through the spatial analysis, all crashes that are within the threshold distance are deemed to be neighbors with one another. This differs from the spatio-temporal analysis, which also takes a moving window timeframe into account. Not only do the crashes need to be within the threshold distance, but they must also occur within one unit of time either before or after a crash to be considered a neighbor. The unit of time examined within this research is either 1 hour or 1 day depending on the investigation completed throughout the results.
The threshold distance is calculated along the path of the roadway and is determined based on the ability of crashes to have at least one neighboring crash. Such a distance may be overestimated during a time when crashes are less frequent and underestimated when crashes are more frequent. In order to determine an adequate threshold, the distance required for each crash to have one neighbor is calculated. This returned a total of 1,432 distances for single vehicle crashes and 1,933 distances for multi-vehicle crashes. The average of these values, for each the single and multi-vehicle crashes, is used as the threshold distance. This average is calculated to remove over- or underestimation.

4.3.3 Cluster Identification

The cluster identification determines the spatial autocorrelation among crashes based on the comprehensive cost of each injury severity level and the spatial weights matrix. The spatial autocorrelation is calculated using the Getis-Ord \( Gi^* \). The calculation of the \( Gi^* \) statistic may be seen in the following equations:

\[
G^*_i(d) = \frac{\sum_j w_{ij}(d)x_j - W^*_i \bar{x}}{s\left[\frac{W^*_i(n-W^*_i)}{(n-1)}\right]^{1/2}}
\]  \hspace{1cm} (4.1)

where:

\[
W^*_i = \sum_j w_{ij}(d)
\]  \hspace{1cm} (4.2)

\[
s^2 = \frac{\sum_j x_j^2}{n} - \bar{x}^2
\]  \hspace{1cm} (4.3)

In Equations 4.1 and 4.2, \( w_{ij}(d) \) is the spatial weight matrix consisting of binary weights with a value of 1 assigned to all locations within distance \( d \), \( x_j \) is the value of the
comprehensive cost based on the crash injury severity, $\bar{x}$ is the average cost of all crashes, and $n$ is the total number of crashes (Prasannakumar et al., 2011).

The $Gi^*$ statistic identifies the level of dispersion among crashes based on the weighted injury severity level. The result of this statistic is a z-score indicating the dispersion at each crash location. The z-score relates to the null hypothesis that all of the crashes are randomly distributed. Z-scores that are positive and statistically significant represent locations where high injury severity weights are clustered together. Those locations that are negative and statistically significant represent locations where low injury severity weights are clustered together. All other locations that are not statistically significant are considered to be randomly distributed.

4.3.4 Spatial-Temporal Cluster Groupings

Cluster locations that are deemed to be statistically significant through the calculation of the $Gi^*$ are then selected to determine if there is grouping present within both the spatial and temporal components. The process of identifying groupings of significantly clustered crashes begins by analyzing only those crashes that are considered to be significantly clustered, based on their z-score. The clusters with a z-score of 1.96 or greater, which relates to a 95% level of significance, are deemed to be significantly clustered. In order to accurately group all of the significantly clustered crashes, the k-means clustering algorithm was implemented, as seen in Anderson (2009), Oltedal and Rundmo (2007), Vlahogianni et al. (2010), and Xu et al. (2012), which has the ability to specify within what group each crash should be contained. Golob and Recker (2004) describe the k-means process as one that minimizes the variability of crash attributes within a cluster while at the same time maximizing the variability between different
clusters of crashes. The crash attribute used to divide the crashes into multiple groups is the time/date in which the crash occurred.

4.3.5 Hot Spot Interpolation

Once the spatial autocorrelation has been determined at each crash location, the level of clustering at all points along the roadway is able to be identified. This is accomplished by interpolating the z-score throughout the entire roadway network. By identifying the z-score at all locations, a smooth transition between significantly clustered and non-clustered locations is determinable. Only those locations that are significantly clustered may then be used as areas in which law enforcement may patrol for alcohol enforcement.

The interpolation of the z-scores is accomplished using inverse distance weighting (IDW). The ability of IDW to determine unknown values at all locations based on the separation distance from known values is described by Mehdi et al. (2011). The unknown z-scores are calculated from IDW through the following equation:

\[
z_0 = \frac{\sum_{i=1}^{s} \frac{z_i}{d_i^k}}{\sum_{i=1}^{s} \frac{1}{d_i^k}}
\]

(4.4)

where, \(z_0\) is the z-score being estimated at point 0, \(z_i\) is the known z-score value at point \(i\), \(s\) is the total number of crash locations used to estimate the unknown z-score, \(d_i\) is the distance separating point \(i\) from point \(0\), and \(k\) identifies the level of influence based distance between points (Ansari and Kale, 2014).
4.4 Results

The results of this study examine hot spots determined through both spatial and spatio-temporal analyses. The results of these two types of analyses are also compared to temporal descriptive statistics, identified in the data section.

4.4.1 Spatial Analysis

The spatial distribution considered in this research is identified from the $Gi^*$ statistic for both single and multi-vehicle crashes. These $Gi^*$ z-scores were interpolated in an effort to show the clustering relationship throughout all roadways within the study area and not specific crash locations. The IDW interpolation, which was conducted along the roadway network using SANET (ver. 4.1), identifies the cluster significance of all crashes. These interpolated values may be seen in the following figure.

![Interpolated Gi* Z-Score](image)

Note: Network interpolation completed with the use of SANET ver. 4.1.

Figure 4.1. Hot Spots of Alcohol-Related Crashes.

There are several locations within Figure 4.1 where significant clusters of high severity crashes occur. These significant clusters are identified when the z-score is greater than or equal to 1.96, which correlates to a 95% level of significance. The positively significant clusters related to those clusters that contain high severity crashes in close proximity to
one another. Negatively significant crashes show locations where low severity crashes are clustered in close proximity to one another; however, there are no negatively significant clusters present with either type of crash. The highly clustered areas occur, for both the single and multi-vehicle crashes, around the city of Cleveland and several other smaller areas along the outer perimeter of the county. The significant areas for both types of data are identified at similar locations, with minor differences in the region covered for each type of crash. The differences in these locations bring to the forefront the basic idea that single and multi-vehicle crashes do not occur at exactly the same place. This requires each type of cluster to have a campaign tailored to the type of crash by which it is analyzed. For instance, single vehicle clusters may need more of a focus on those drivers speeding around curved sections of roadway. The pure spatial analysis provides a great general idea of where safety implementations may originate. However, there is no sense on when would be the optimal time to provide these implementations, as a reference to any temporal aspect is not present for this purely spatial investigation. For instance, it is unknown whether 2:00 am, 10:00 pm, or another interval is the optimal time to implement a safety campaign in a specific location. Without this consideration of time, clusters of crashes may or may not be present at an identified location.

4.4.2 Spatio-Temporal Analysis

While the spatial analysis provides an idea of the spatial distribution and the temporal analysis provides insight into when crashes are occurring, neither of these analyses overlap and tell a complete story. For example, it may be known that a specific area contains clustered crashes, as identified through spatial analysis. Additionally, the time of day or day of week when most crashes occur may be known. However, it is not
known whether those clusters identified through the spatial analysis will be present at the time the temporal analysis designates. It would not be beneficial to assume that crashes are always clustered in the same location, set up a safety campaign at that location, and not have a cluster occur. Therefore, the ability to merge the two capabilities into a single analysis is necessary. Within the spatio-temporal analysis, the crashes are analyzed not only based on their spatial distribution but also on the time at which they occurred. This allows crashes that occur at a similar time frame to be considered as clustered. Crashes that occur in a similar location but outside of this timeframe may then not necessarily register a cluster at the same location but at a different time. The spatio-temporal analysis allows for an examination of both the distribution of crashes and a temporal aspect to be investigated together.

The result of the spatio temporal analysis is a four-dimensional map. These four dimensions are longitude, latitude, time, and z-score. There are a couple different options to comprehend the results of the analysis. First, to make the multi-dimensional map easier to understand, only significant clusters, with a z-score greater than or equal to 1.96, are shown. This reduces the map to three-dimensions and allows for the identification of when and where clusters are occurring. Different trends in the clustering of crashes may also be noticed, such as: movements through the progression of time, groupings of clusters, or temporal or spatial gaps. In order to better quantify these movements and groupings of clusters, the k-means algorithm is used. The use of this algorithm removes arbitrary grouping of clusters by the analyst. The z-scores within each group may then be interpolated along roadways to identify the locations where law enforcement may patrol while implementing safety campaigns. Additionally, with hot spot maps created for each
grouped time period, multiple maps may be compared to one another. This analysis for single and multi-vehicle alcohol-related crashes by time of day may be seen in the following figure.
Cuyahoga County Single Vehicle

Significant Spatio-Temporal Clusters

Time Group 1  Time Group 2

Time Group 3  Time Group 4

Cuyahoga County Multi-Vehicle

Significant Spatio-Temporal Clusters

Time Group 1  Time Group 2

Time Group 3  Time Group 4

Note: Network interpolation completed with the use of SANET ver. 4.1. The hot spot maps of the time groups (1-4) relate to the grouped clusters shown in the Significant Spatio-Temporal Clusters map.

Figure 4.2. Spatio-Temporal Analysis of Alcohol-Related Crashes in Cuyahoga County by Time of Day.
In Figure 4.2, both spatio-temporal clusters and spatio-temporal hot spot maps, based on k-means groupings, may be seen. The multi-dimensional plots of clustered crashes depict both the location and time throughout the day in which the clusters occur. The spatial location is spread out in relation to where the correlating crashes occurred within the county. The temporal depiction is identified as those crashes closest to the surface of the county (depicted in Figure 4.2) are at the beginning of the day, 12:00 am, and those farther away from the surface are later in the day, 11:59 pm. The groupings of clusters and their associated time spans within each Time Group is not user specified. It is calculated, however, using the k-means clustering algorithm for both the single and multi-vehicle crashes. The timeframe relating to each time group of clusters may be seen in the following table.

Table 4.2. Time Groupings for Clusters by Time of Day.

<table>
<thead>
<tr>
<th>Grouped Cluster</th>
<th>Single Vehicle</th>
<th>Multi-Vehicle</th>
<th>Combined Timeframe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Group 1</td>
<td>12:00 am – 3:52 am</td>
<td>12:00 am – 1:59 am</td>
<td>12:00 am – 4:00 am</td>
</tr>
<tr>
<td>Time Group 2</td>
<td>4:15 am – 8:36 am</td>
<td>4:55 am – 6:30 am</td>
<td>4:15 am – 8:45 am</td>
</tr>
<tr>
<td>Time Group 3</td>
<td>4:05 pm – 8:37 pm</td>
<td>3:14 pm – 5:59 pm</td>
<td>3:00 pm – 8:45 pm</td>
</tr>
<tr>
<td>Time Group 4</td>
<td>10:08 pm – 11:46 pm</td>
<td>10:04 pm – 11:43 pm</td>
<td>10:00 pm – 12:00 am</td>
</tr>
</tbody>
</table>

Note: Time groups for the single and multi-vehicle clusters are determined through the use of the k-means clustering algorithm.

The k-means clustering separates the clusters into four separate groups. The time for the first and last cluster included in each group may be seen in Table 4.2. These time groups do not overlap for consecutive groups for both the single and multi-vehicle clusters. Therefore, a combined timeframe was created that encompasses both the single and multi-vehicle crashes for comparison. The closest 15-minute interval that encompasses
both the single and multi-vehicle crashes within each time group was used for ease of understanding.

Significant clusters of high severity crashes seen early in the day for both the single and multi-vehicle crashes in Figure 4.2, are located in a similar area as the significant hot spots found in Figure 4.1. While this may lead one to think that an overall spatial analysis is sufficient, the location of significantly clustered crashes for the remaining times of the day differ. As time progresses, there is then a lack of crash clustering in the same location, as identified in Figure 4.1, for the remainder of the day. Specifically for the single vehicle crashes, clusters may be seen towards the southeastern portion of the county. As time continues through the day, the clusters move towards the north-central portion of the county and move towards the western side of the county at the end of the day. For multi-vehicle crashes, clusters begin in the early hours in the north-central portion of the county. As the day progresses, these clusters then spread out in all directions towards the edges of the county.

Not only are the individual movements of hot spots important to determine for either the single or multi-vehicle crashes, it is imperative to identify their interaction with each other. The location of statistically significant clusters of single and multi-vehicle crashes, along with the portions of significant roadways that overlap, may be seen in the following figure.
Note: Network interpolation completed with the use of SANET ver. 4.1.

Figure 4.3. Comparison of Single and Multi-Vehicle Hot Spots by Time of Day.

From Figure 4.3, it may be seen that the location of significant clusters for both the single and multi-vehicle crashes are fairly separate. Some larger areas of overlap may be seen in Time Groups 2 and 3, and extremely small amounts of overlap are identified in Time Groups 1 and 4. The lack of overlapping significantly clustered roadways further contributes to the notion that single and multi-vehicle crashes occur due to differing circumstances. Changes in the specific location of significant clusters may be seen between Times Groups 1 and 2. In Time Group 1, significant clusters of multi-vehicle crashes are seen to be located in the north central portion of the county. This differs from the significantly clustered single vehicle crashes located towards the southeastern portion of the county. When progressing to Time Group 2, the significantly clustered multi-vehicle crashes begin shifting away from their original location and significant clusters of single vehicle crashes then appear. These shifts between clusters of single and multi-vehicle crashes may then rise from a reduction of vehicle on the roadway. In Time Group 1, when more vehicles are present, clusters of multi-vehicle crashes may be seen. In Time Group 2 once there is a decrease in the number of vehicles, the once multi-vehicle clusters turn into single vehicle crash clusters. The shifts in clusters between single and multi-vehicle crashes imply that if a law enforcement tactic were to be used within the north-central location. The campaign in this area would have to switch from targeting multi-vehicle crashes to targeting single vehicle crashes. Very few to no significant clusters appear to be located in the same area throughout the entire day. This further
identifies the need for law enforcement to alter the location of safety campaigns to adjust to spatio-temporal patterns.

While the analysis of the time of day provides a description of when and where clusters of crashes are occurring throughout the day, it is still necessary to ascertain an idea of which day in the week the crashes occur. As commonly thought, and seen from the temporal portion of the descriptive statistics in Table 4.1, the ideal times to target alcohol intoxicated drivers is on Thursday, Friday, and Saturday. However, without identifying clusters of crashes throughout the week, the accuracy of this spatio-temporal trend may be unknown. To resolve this lingering question, a plot of the spatio-temporal clustering, depicted in the same manner as Figure 4.2, for both single and multi-vehicle crashes may be seen in the following figure.
Figure 4.4. Spatio-Temporal Analysis of Alcohol-Related Crashes by Day of Week.

In Figure 4.4, similar to composition of Figure 4.2, both spatio-temporal clusters and hot spot maps based on k-means groupings, may be seen. The multi-dimensional plots again depict the location of significantly clustered high severity crashes throughout the county;

Note: Network interpolation completed with the use of SANET ver. 4.1.

In Figure 4.4, similar to composition of Figure 4.2, both spatio-temporal clusters and hot spot maps based on k-means groupings, may be seen. The multi-dimensional plots again depict the location of significantly clustered high severity crashes throughout the county;
however, the temporal component now indicates the day of the week in which the cluster is present. The timeframe for the week starts off on Sunday, where depicted clusters are close to the surface of the county. As the week progresses through to Saturday, the clusters raise higher and higher from the surface of the county. Similar to the establishment of the Time Groups, the grouping of clusters into Day Groups is not user specified. The groups are again determined using the k-means clustering algorithm for both the single and multi-vehicle crashes. The timeframe relating to each day group may be seen in the following table.

Table 4.3. Time Groupings for Clusters by Day of Week.

<table>
<thead>
<tr>
<th>Grouped Cluster</th>
<th>Single Vehicle</th>
<th>Multi-Vehicle</th>
<th>Combined Timeframe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day Group 1</td>
<td>Sunday – Monday</td>
<td>Sunday – Monday</td>
<td>Sunday – Monday</td>
</tr>
<tr>
<td>Day Group 2</td>
<td>Tuesday - Wednesday</td>
<td>Tuesday - Wednesday</td>
<td>Tuesday - Wednesday</td>
</tr>
<tr>
<td>Day Group 3</td>
<td>Thursday - Saturday</td>
<td>Thursday - Saturday</td>
<td>Thursday - Saturday</td>
</tr>
</tbody>
</table>

Note: Day groups for the single and multi-vehicle clusters are determined through the use of the k-means clustering algorithm.

The k-means clustering is now separated into three groups for the day of the week, as may be seen in Table 4.3. The days for both the single and multi-vehicle crashes fell on the same intervals. Therefore, when examining both sets of crashes together, the day groups align to be exactly the same.

As may be seen in Figure 4.4, the significant clusters of single vehicle crashes shift extensively throughout the county. These clusters originate in the north western part of the county during the beginning of the week. Through the middle of the week, the single vehicle clusters may be found in the north-central portion of the county. Finally, at the end of the week, the single vehicle clusters disperse widely throughout the county. The significant clusters of multi-vehicle crashes also vary in location throughout the
week. The multi-vehicle clusters are fairly spread-out throughout the county at the beginning of the week. By the middle of the week, there is a large significant cluster located just east of the center of the county. At the end of the week, the clusters are dispersed throughout the entire county. The large condensed areas of significantly clustered crashes seen in the early parts of the week may require a regional effort to provide a reduction in crash severity and occurrence. In contrast, the more dispersed condition of clusters may require local agencies in the area of specific clusters to address the problem of alcohol-related crashes.

As the individual movements of both the single and multi-vehicle clusters have been identified, the combined interaction of the two types of crashes must again be investigated. The roadways pertaining to statistically significant clusters of both single and multi-vehicle crashes may be seen in the following figure.

![Comparison of Single and Multi-Vehicle Hot Spots by Day of Week](image)

Note: Network interpolation completed with the use of SANET ver. 4.1.

Figure 4.5. Comparison of Single and Multi-Vehicle Hot Spots by Day of Week.

As seen in Figure 4.5, there are again very limited occurrences of the single and multi-vehicle clusters appearing in the same location during the same time period. The largest combined area of both single and multi-vehicle clusters may be seen in Day Group 2. All
other overlapping roadways are very small in Day Groups 1 and 3. Within Day Group 2, besides the overlapping portions of roadway, the significant clusters of single and multi-vehicle crashes occur in a very similar area. This does not occur throughout either the beginning or end of the week, however. The shifts in the location of significantly clustered crashes may readily be seen, as generally no hot spot covers the same location twice. This has a large influence on safety campaigns and would require multiple shifts in the locations patrolled by law enforcement. In comparison to the overall spatial analysis shown in Figure 1, only a small portion of the spatio-temporal hot spots occur in the same location as those determined without the influence of time.

4.5 Conclusion

Investigating the occurrence of crashes where an operator was under the influence of alcohol is important to both understanding the mechanics behind such crashes and identifying a campaign to reduce their number. Each aspect of the spatial, temporal, and spatio-temporal analysis tells a different story. While individual pieces may come from the spatial analysis and the temporal analysis, their marriage allows for the proper targeting of areas where alcohol intoxicated drivers may be traveling. The spatio-temporal analysis not only implements a similar procedure to that of the purely spatial analysis, but also includes a moving timeframe to capture a temporal movement of the identified clusters. The use of the spatial weights matrix is a key ingredient into linking the spatial separation of crashes along a roadway network to a varying window of time. By providing an in-depth analysis into the crashes, relationships that are not recognized by either spatial or temporal analyses alone may be noticed, which may contribute to a deeper understanding of how to effectively reduce the occurrence of the crashes.
The results of this study identified movements of hot spots both throughout the time of day and day of week. These movements are very important in the determination of a location to implement a safety campaign. For example, it is seen that within the day of week analysis, barely any of the hot spots reoccurred in the same location between the three time/day groups. If a safety campaign were to have been implemented in one location without adapting to the temporal movement of crashes, large significant clusters of crashes would remain unaddressed. Similar to the time of day analysis, if a safety campaign were to be implemented only in locations identified through Time Groups 1 and 2, valuable resources may be wasted as hot spots in those areas dissolve into Time Groups 3 and 4.

Different strategies may be needed at various locations and times to address the issue of operating a vehicle while intoxicated, and these strategies may be related to the overall size or location of the identified hot spot. Large condensed hot spots may require a regional effort to reduce the severity and occurrence of crashes. Meanwhile, multiple small dispersed hot spots may require the effort of many local agencies in specific areas. Overall, this spatio-temporal analysis allows for an identification of when and where to stage safety implementations that spatial or temporal analyses alone may miss. By only investigating the relationship as to when or where crashes are occurring using a single form of analysis, an inefficient safety campaign may be implemented.
CHAPTER V
EXAMINING THE USE OF NORMALIZATION IN MAPPING OF ALCOHOL-
RELATED HOT SPOTS

5.1 Introduction

A total of 33,561 traffic related fatalities occurred in 2012 (FHWA, 2015), the latest year of available data. Of these crashes, nearly one-third of the crashes resulted from an operator having a blood alcohol concentration (BAC) level of 0.08 or greater. This trend of having approximately 31% has been a continuing trend for at least the past 20 years. Studies investigating the effects of alcohol and the habits of drivers who drink have provided a wide breadth of knowledge. For instance, Kennedy et al. (1996) identified the high-risk involved with young drivers and alcohol use, stating that 70% of male drivers involved in alcohol-related fatal crashes were between the ages of 20 and 39. Voas, Tippetts, and Fell (2003) continued the investigation of young age and drinking through a study relating to the effects of minimum legal drinking age, which identified that the establishment of a zero tolerance BAC reduced alcohol involved crashes. Naimi et al. (2003) further studied the habits of drinkers, determining an increased likelihood of binge drinkers to drive impaired. The effects of drinking on driving-related skills has additionally been investigated by Moskowitz and Florentino (2000) at low BAC levels in an effort to determine the most effective legal limits.
All of the previously listed research provides a great indication of the actions and habits of alcohol impaired drivers. While this information is important to know, a major contributor to reducing the number of alcohol-related crashes is the use of law enforcement. There are a number of strategies that are used to aid in this reduction that involve a high presence of law enforcement officers in specific areas. These types of strategies provide high visibility enforcement, which informs drivers that preventing driving under the influence of alcohol is a top priority. The presence of law enforcement is often in the form of saturation patrol or corridor patrol. Through corridor patrols, officers patrol the roadways known to contain the highest number of alcohol-related crashes. Saturation patrol performs in a similar manner; however, instead of being restricted to a few specific roads, a defined area is covered. Maistros et al. (2014) investigated a case study of both saturation and corridor patrol in which hot spots were used to identify the locations that law enforcement could cover. This case study identified that within hot spots, there is a statistically significant difference in average number stops per hour versus the number of stops per arrest of a person operating a vehicle while under the influence.

As hot spots are shown to indicate where law enforcement officers may patrol, the identification of statistically significant areas is important to determine. Hot spots of crashes are determined based on the relationship between a value pertinent to a crash location and the distance separating each crash location from one another. There are a couple of different options for the value used within the calculation of the \( Gi^* \); it may either be based on the frequency of crashes or the severity of crashes which have previously occurred. Hot spots usually identify locations where high values are in close
relation to one another. A few methods may be employed to identify the spatial relationship of crashes. These methods include, but are not limited to, the use of kernel density estimation (KDE), Moran’s $I$, and the Getis-Ord $Gi^*$ statistic. KDE identifies the magnitude of the value in question per an area unit (Erdogan et al., 2008). Moran’s $I$ identifies the relationship of similar or dissimilar values in relation to each other and allows for the determination of outliers (Erdogan, 2009). Meanwhile, the $Gi^*$ statistic determines the location of concentrated high or low values (Getis and Ord, 1992).

Songchitruska and Zeng (2010) further explain the similarities and differences between some of these spatial statistics and the importance of using the $Gi^*$ statistic for identifying hot spots. Kuo et al. (2013) used the frequency of crashes to calculate the Getis-Ord $Gi^*$ statistic. This allowed clusters of crashes and crimes to be identified for police patrol routes. On the other hand, Truong and Somenahalli (2011) showed the ability to use injury severity as a weighting system for the calculation of the $Gi^*$ statistic. The resulting significant clusters of high severity crashes were then used to identify unsafe bus stops.

While the use of hot spot analyses allows for specific areas of concern to be identified, there are often concerns raised when the hot spots are concentrated towards major cities or city centers. The general statement that is brought to the forefront is that due to high population densities there will, of course, be clusters of crashes in those locations. Comments have traditionally been raised that the relationship between crashes and population density should be addressed. Therefore, this research is directed towards tackling the issue of normalizing hot spots of crashes by population density.
5.2 Data

Alcohol-related crashes from January 1, 2012 through April 9, 2015, are investigated in this study. These crashes were obtained from the Ohio Department of Public Safety’s OH-1 crash reports. The crashes were then divided and analyzed based on eight different counties. A breakdown of each county and their respective geographical description may be seen in the following table.

Table 5.1. County Geographical Makeup.

<table>
<thead>
<tr>
<th>County</th>
<th>Major City</th>
<th>Percent Urban</th>
<th>Percent Rural</th>
<th>Population</th>
<th>Total Area (sq. mi.)</th>
<th>Alcohol-Related Crashes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuyahoga</td>
<td>Cleveland</td>
<td>91%</td>
<td>9%</td>
<td>1,280,122</td>
<td>459.0</td>
<td>3,366</td>
</tr>
<tr>
<td>Summit</td>
<td>Akron</td>
<td>71%</td>
<td>29%</td>
<td>541,781</td>
<td>420.0</td>
<td>1,809</td>
</tr>
<tr>
<td>Franklin</td>
<td>Columbus</td>
<td>72%</td>
<td>28%</td>
<td>1,163,414</td>
<td>544.0</td>
<td>4,040</td>
</tr>
<tr>
<td>Hamilton</td>
<td>Cincinnati</td>
<td>75%</td>
<td>25%</td>
<td>802,374</td>
<td>412.5</td>
<td>2,711</td>
</tr>
<tr>
<td>Allen</td>
<td>Lima</td>
<td>14%</td>
<td>86%</td>
<td>106,331</td>
<td>406.8</td>
<td>401</td>
</tr>
<tr>
<td>Athens</td>
<td>Athens</td>
<td>3%</td>
<td>97%</td>
<td>64,757</td>
<td>508.5</td>
<td>242</td>
</tr>
<tr>
<td>Muskingum</td>
<td>Zanesville</td>
<td>5%</td>
<td>95%</td>
<td>86,074</td>
<td>672.5</td>
<td>391</td>
</tr>
<tr>
<td>Ross</td>
<td>Chillicothe</td>
<td>3%</td>
<td>97%</td>
<td>78,064</td>
<td>692.8</td>
<td>334</td>
</tr>
</tbody>
</table>

Note: Overall county population obtained from United States Census Bureau (2013).

The counties selected for this analysis, which are listed in Table 5.1, were chosen due to the fact that they cover a wide range of geographies and crash occurrences. This allows the study to be more robust and not focus on only one type of study area. Four of the counties are comprised of at least 70% urban areas. In order to balance the urban counties, four additional counties were then selected that contain a similar percentage of rural areas. The mainly urban counties contain the four largest amounts of crashes, ranging from 1,809 to 4,040 alcohol-related crashes during the time period studied.

Meanwhile, the counties that are comprised of mainly rural areas contain from 242 to 401 alcohol-related crashes. The wide range of populations may also be seen within Table
5.1. The urban counties all contain over 500,000 people, while the rural counties contain less than 110,000. The overall size of each county is similar, ranging from about 400 to 700 square miles in area. However, those counties that have the smaller population sizes generally cover a larger area, making their overall population density less than in the counties with higher population. These various types of counties are used to investigate a wide range of population distributions.

In order to determine the population distribution, such as the population density and urban/rural geographical information, decennial census information was obtained from the United States Census Bureau for the year 2010. The census data was obtained at two levels, the block and tract levels. The census blocks provided the information pertinent to an area being described as urban or rural. Meanwhile, the census tracts provided the population density values. Population densities could be obtained from census blocks; however, at the block level, many areas may be seen to contain populations of zero. When populations of zero occur, normalized crash values drastically spike in locations where there are no residents but a large presence of people.

5.3 Methodology

The influence of normalization of hot spots, which are calculated based on both the frequency and the injury severity of crashes, is being investigated in this research effort. Therefore, a total of four hot spots are being investigated for each county: two normalized hot spots and two non-normalized hot spots. The frequency of crashes is determined by those crashes that occur in the exact same location. Meanwhile, the injury severity is based on the greatest level of injury realized by all parties involved. The injury severity is also weighted based on the societal crash costs determined within the Highway
Safety Manual (AASHTO, 2010). The weighted injury severities relate to either a fatal injury (K), injury (A/B/C), or property damage only (O) crash severity level.

5.3.1 Population Density

Population density information is created using data from the latest decennial census. The latest census available for this study is obtained from the 2010 census. This data is aggregated into zones within each county. These zones provide a specific population and attributed area. However, due to the specific boundaries obtained from zonally based values available from the census, population densities could drastically change in a matter of feet when changing from one census tract to another. If using these values straight as they were obtained, the population density relating to two crashes within the same tract would be the same, no matter if they were two feet apart or 2,000. Similarly, two crashes that are only 20 feet apart but contained within two separate census tracts may relate to very different population densities.

In order to smooth the population densities to provide a gradual change, it is necessary to interpolate the values obtained from the decennial census. Inverse distance weighting (IDW) interpolation is used to accomplish this task. The use of IDW and its abilities to interpolate unknown values is further described by Mehdi et al. (2011). IDW interpolation is computed using the following equation:

\[
Z_0 = \frac{\sum_{i=1}^{s} \frac{z_i}{d_i^k}}{\sum_{i=1}^{s} \frac{1}{d_i^k}}
\]

where, \(Z_0\) is the estimated z-score at unknown location \(0\), \(z_i\) is the measured z-score at location \(i\), \(s\) is the number of crash clusters used to estimate the unknown z-score, \(d_i\) is the network based distance separating locations \(i\) and \(\theta\), and \(k\) is the power that smooths the
z-scores based on the influence of distance (Ansari and Kale, 2014). This allows each crash to have its own specific associated population density, even if it is within the same census tract as another crash. When the normalization by population density is applied, the value studied for spatial autocorrelation is either crashes per person per square mile or societal cost per person per square mile.

5.3.2 Spatial Autocorrelation

Spatial autocorrelation is calculated in this study using the $G_i^*$ statistic. This statistic functions on the null hypothesis that all crashes are randomly distributed. Using a statistical significance level of 0.05, associated with a z-score of $\pm 1.96$ for a crash, indicates that the null hypothesis should be rejected and the crash may be considered to be either a hot spot or a cold spot. Hot spots are those in which high values, either the frequency or the cost of crashes, are located in close proximity to other high values. Cold spots, on the other hand, are those in which low values are located in close proximity to other low values. The $G_i^*$ statistic is calculated using the following equation:

$$G_i^* (d) = \frac{\sum_j w_{ij}(d) x_j - W_i^* \bar{x}}{s \left[ \frac{W_i^* (n-W_i^*)}{(n-1)} \right]^{3/2}}$$  \hspace{1cm} (5.2)

where:

$$W_i^* = \sum_j w_{ij}(d)$$  \hspace{1cm} (5.3)

$$s^2 = \frac{\sum_j x_j^2}{n} - \bar{x}^2$$  \hspace{1cm} (5.4)

where $w_{ij}(d)$ is the spatial weight, $x_j$ is either the frequency or cost associated value, $\bar{x}$ is the average of all frequency or cost values, and $n$ is the total number of crashes (Prasannakumar et al., 2011).
The spatial weight used in the calculation of the $Gi^*$ is dependent upon the distance separating one crash from another in comparison to the threshold distance. The threshold distance is one such that all crashes have at least one neighbor. The spatial weight is a fixed value for those crashes that occur within the threshold distance. All crashes that occur within this distance retain a value of one, while all other crashes retain a value of zero. This allows crashes that are within the threshold distance to be included in the $Gi^*$ calculation. All distances to determine whether a crash is within the threshold distance or not are calculated using a network-based distance. This type of distance strictly follows along the path of the roadway.

5.3.3 Interpolation of Spatial Autocorrelation

The calculation of spatial autocorrelation provides a specific value to each studied crash location. In order to depict these statistically significant clustered locations, the clustering value of each crash must be interpolated along the roadways. This provides a clearly defined hot spot area in which law enforcement may operate safety measures. The interpolation used to display the crash data is different from that used within the demographic information. As the distance measurements for the calculation of the spatial autocorrelation follows along the roadway network, so do those of the interpolation. The demographic information does not necessarily follow a strict network, and patterns may smoothly transition over open fields, backyards, playgrounds, and other areas. Crash patterns, however, are restricted because they occur on a roadway network. The theory of interpolating values along a network may be common; however, the availability of software to complete this task is not. Therefore, SANET (Ver. 4.1) was retained for the
completion of this task. This software uses IDW to determine the interpolated value at all locations.

5.4 Results

There are a few different levels of census data that could be converted to population density. These levels range from the block, block group, tract, and county levels. The census block is the smallest level, which would work great for obtaining the best resolution of population data; however, at this level, there are many areas that contain values of zero population. This trend decreases as use of census data transfers from individual blocks to block groups, and finally to the level of a census tract. There still are some census tracts that contain populations of zero; however, the occurrence of these is very minimal in comparison to both the census block and block groups. When populations of zero occur, normalized crash values drastically spike in locations where there are no residents but there may be a large number of people traveling within the area. The population density for each census tract is calculated based on the population observed in a tract divided by the area in which each tract covers. Due to the boundaries obtained from zonally based values available from the census, population densities could drastically change in a matter of feet when changing from one census tract to another.

In an effort to reduce the effect of the boundaries and to smooth values over census tracts with zero population, the population densities of the census tracts are interpolated. IDW interpolation is used to accomplish this smooth transition throughout an entire counties area. Maps of these interpolated population densities may be seen in the following figure.
Note: The color ramp is based on the population density (persons per square mile). The lighter areas correlate to higher population densities. Meanwhile, the darker areas correlate to lower population densities.

Figure 5.1. County Population Density.

It may be seen from Figure 5.1 that there is typically one densely populated area within each county. These highlighted areas are the locations of concern when investigating the normalization of hot spots. The peak population densities for the two densest urban counties are 28,956 and 23,231 people per square mile, relating to the cities of Columbus in Franklin County and Cleveland in Cuyahoga County, respectively. The peak population densities for the two least dense rural counties are 3,525 and 5,445 people per square mile, relating to the cities of Zanesville in Muskingum County and Lima in Allen County, respectively. There is a visual difference in the interpolated population densities between the urban counties and the rural counties. The urban counties have more census tracts being interpolated and higher populations in the areas surrounding the central city.
in the county. This leads the densities depicted in Figure 5.1 to appear less intense and more spread out. Meanwhile, the rural counties have larger census tracts and the population density in the central city in the county has a higher influence. This leads the densities, depicted in Figure 5.1, in the location of these central cities to appear much more intense. The influence of the shape of the population densities has a direct relation to the normalization of hot spots. While population density is a good indication of where people are present, roadway density was also believed to have an impact on the normalization of clusters. The additional input of roadway density was examined for its impact; however, an investigation of the crosscovariance did not reveal any trends that would have improved the normalizing factor.

Four hot spot maps were created for each of the eight counties studied in this research effort. The hot spots are based on the frequency of crashes, frequency of crashes normalized by population density, societal cost of the crashes, and the societal cost of the crashes normalized by population density. Each of these hot spot maps for the heavily urban counties may be seen in the following figure.
<table>
<thead>
<tr>
<th></th>
<th>Frequency Normalized</th>
<th>Societal Cost</th>
<th>Societal Cost Normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuyahoga</td>
<td>![Map Cuyahoga]</td>
<td>![Map Cuyahoga Societal Cost]</td>
<td>![Map Cuyahoga Societal Cost Normalized]</td>
</tr>
<tr>
<td>Summit</td>
<td>![Map Summit]</td>
<td>![Map Summit Societal Cost]</td>
<td>![Map Summit Societal Cost Normalized]</td>
</tr>
<tr>
<td>Franklin</td>
<td>![Map Franklin]</td>
<td>![Map Franklin Societal Cost]</td>
<td>![Map Franklin Societal Cost Normalized]</td>
</tr>
<tr>
<td>Hamilton</td>
<td>![Map Hamilton]</td>
<td>![Map Hamilton Societal Cost]</td>
<td>![Map Hamilton Societal Cost Normalized]</td>
</tr>
</tbody>
</table>

Note: Network interpolation completed with the use of SANET ver. 4.1. The color ramp is based on $Gi^*$ z-score. The roadways in red are more significant towards clustering of high values. The roadways in blue are more significant towards clustering of low values.

Figure 5.2. Hot Spot Maps of Urban Counties.

Figure 5.2 identifies the z-score relating to each roadway within the studied counties.

Those roadways that are indicated in a red color are more significant towards the clustering of high values. On the other hand, those roadways that are blue in color are more significant towards clusters of low values. The frequency clusters are calculated based on the number of crashes in the same location, while the cost based maps are
calculated based on the societal costs of crashes in the same location. The normalized maps are calculated using either the frequency or cost of crashes divided by the population density, in persons per square mile. Trends, such as those presented in Figure 5.2, may be depicted for each type of hot spot map. For those maps based on the frequency of crashes, hot spots are generally found towards the largest city within the county. The demographics of these cities are also the location of the highest population densities. This similarity in location indicates the influence of population density on the frequency based maps. These maps also contain a more consolidated hot spot in the high populous areas than the hot spots identified from the remaining types of maps. The influence of a safety campaign in such an area would provide a target of letting the population know that alcohol-related crashes are of concern. These locations may be best suited for educational campaigns due to the high influence of population or for high visibility campaigns, where large numbers of motorists would see the presence of enforcement.

The second column of maps is similar to the first, with the aspect that they are both determined based on the frequency of crashes; however, this set of maps is normalized based on the population density of the surrounding area. Within the second column of hot spot maps, almost the reverse of the hot spots based purely on the frequency may be seen. In other words, there is a tendency towards cold spots, or locations of low values in close proximity to other low values, at locations of high population density. The hot spots in the second column of maps then shifts towards the outer edges of the counties. The inclusion of cold spot in the same area as the hot spots from the maps in the first column does not remove the influence of population density. It
in turn identifies a significantly clustered area in the same location and identifies additional hot spots in the outer edges of the county that must then be included in safety campaigns. This would thus require an even larger effort by educators, enforcement, and engineers to eliminate hazards.

The third column of maps represents those that are clustered based on the societal cost of the highest injury severity involved in the crash. Within these maps, the hot spots return back towards the major metropolitan areas. However, the hot spots are not necessarily located at the highest population areas, as seen from the first column of maps. The cost-based hot spots tend to have a higher presence in the areas surrounding the high population areas, when compared to the frequency based maps, but they are not as dispersed to the outer portions of the counties, as seen in the normalized frequency based maps. Thus, the influence of high population areas is not as great as those seen from the frequency based maps. In turn, safety campaigns implemented in locations identified by the societal cost based maps would have a higher impact on the crashes it may reduce. A safety campaign in these identified areas would be best suited for lowering the overall severity of crashes.

The maps of societal costs normalized by population density are similar to those maps of crash frequency normalized by population density. There are, however, some small differences in the hot and cold spots. The cold spots, again, tend to appear near the highly dense population area, and the hot spots appear towards the outer edges of the county. These similarities slightly differ in the aspect that the cold spots are not as vast or are constrained by the presence of a nearby hot spot. Similar to the effect caused by the normalized frequency maps, the inclusion of both cold spots and hot spots would create
the need for a larger effort by educators, enforcement, and engineers to eliminate or reduce alcohol-related crashes.

In an effort to determine if the effects seen in the highly urban counties are specific to those population conditions, four counties that are comprised of mostly rural areas are also examined. These maps cover both the normalized and non-normalized analyses based on either the frequency or cost of crashes. The maps for these additional four counties may be seen in the following figure.
Note: Network interpolation completed with the use of SANET ver. 4.1. The color ramp is based on z-score of the $Gi*$. The roadways colored in red are more significant towards clustering of high values. The roadways in blue are more significant towards clustering of low values.

Figure 5.3. Hot Spot Maps of Rural Counties.

The maps that may be seen in Figure 5.3 contain similar trends to those described for Figure 5.2, where the non-normalized maps form hot spots around the highly dense populations, in contrast to the normalized maps that form cold spots in the same area.

Even though the non-normalized maps exhibit hot spots in similar areas, around the
presence of these dense populations, those resulting from urban counties tend to be larger and more apparent than those in rural counties. The mostly rural counties have smaller and less dense population demographics, resulting in hot and cold spots that are generally smoother and less interrupted by one another. The areas around the major metropolitan cities seen within Figure 5.2 may be seen to more rapidly change between being a hot spot and a cold spot. This effect is less noticeable in Figure 5.3, where the change is often more gradual. Additional differences between Figures 5.2 and 5.3 are the intensity of the color ramps depicting the z-score along the roadways. These color ramps appear to be different in a visual sense, but the only variation is due to the density of roads within urban versus rural counties.

One aspect that may be gleaned from both Figures 5.2 and 5.3 is that the normalization of the spatial autocorrelation generally takes the hot spot out of the densely populated areas and moves them towards the outer edges of the counties. Meanwhile, cold spots develop in areas similar to those of the hot spot that was just normalized. Additionally, both the frequency and cost-related hot spots are identified in similar areas; however, there are some differences. The frequency-based hot spots seem to be highly related to the location of densely populated areas. The cost-based maps, however, seem to be less discretionary about the population density of the area in which they are located.

Some more telling information about the demographics of where the hot spots are located may come from an examination of the urban and rural areas within each county. Even though each county contains more than 70% of either urban or rural areas, the composition of which locations the hot spots relate to changes from map to map. The
amount of roadway that each hot spot covers in both urban and rural environments in each county may be seen in the following table.

Table 5.2. Geographical Coverage of Significant Hot Spots.

<table>
<thead>
<tr>
<th>Urban Counties</th>
<th>Percent</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Urban</td>
<td>Rural</td>
</tr>
<tr>
<td>Cuyahoga</td>
<td>70%</td>
<td>30%</td>
</tr>
<tr>
<td>Normalized Cost</td>
<td>79%</td>
<td>21%</td>
</tr>
<tr>
<td>Frequency</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>Normalized</td>
<td>94%</td>
<td>6%</td>
</tr>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summit</td>
<td>63%</td>
<td>37%</td>
</tr>
<tr>
<td>Normalized Cost</td>
<td>46%</td>
<td>54%</td>
</tr>
<tr>
<td>Frequency</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>Normalized</td>
<td>70%</td>
<td>30%</td>
</tr>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Franklin</td>
<td>88%</td>
<td>12%</td>
</tr>
<tr>
<td>Normalized Cost</td>
<td>73%</td>
<td>27%</td>
</tr>
<tr>
<td>Frequency</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>Normalized</td>
<td>84%</td>
<td>16%</td>
</tr>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hamilton</td>
<td>92%</td>
<td>8%</td>
</tr>
<tr>
<td>Normalized Cost</td>
<td>81%</td>
<td>19%</td>
</tr>
<tr>
<td>Frequency</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>Normalized</td>
<td>78%</td>
<td>22%</td>
</tr>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural Counties</td>
<td>Percent</td>
<td>Percent</td>
</tr>
<tr>
<td></td>
<td>Urban</td>
<td>Rural</td>
</tr>
<tr>
<td>Allen</td>
<td>12%</td>
<td>88%</td>
</tr>
<tr>
<td>Normalized Cost</td>
<td>11%</td>
<td>89%</td>
</tr>
<tr>
<td>Frequency</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>Normalized</td>
<td>9%</td>
<td>91%</td>
</tr>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Athens</td>
<td>4%</td>
<td>96%</td>
</tr>
<tr>
<td>Normalized Cost</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>Frequency</td>
<td>44%</td>
<td>56%</td>
</tr>
<tr>
<td>Normalized</td>
<td>69%</td>
<td>31%</td>
</tr>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muskingum</td>
<td>10%</td>
<td>90%</td>
</tr>
<tr>
<td>Normalized Cost</td>
<td>7%</td>
<td>93%</td>
</tr>
<tr>
<td>Frequency</td>
<td>54%</td>
<td>46%</td>
</tr>
<tr>
<td>Normalized</td>
<td>3%</td>
<td>97%</td>
</tr>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ross</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>Normalized Cost</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>Frequency</td>
<td>52%</td>
<td>48%</td>
</tr>
<tr>
<td>Normalized</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The percent of roadway is based on the length of road, of a statistically significant cluster of high values, which passes through either urban or rural land types.

The change in the demographics associated with the hot spots may be seen in Table 5.2.

These percentages are based on the amount of roadway, which is part of a statistically significant cluster of high values, in either urban or rural land types. For example, if half of the roadways identified as being a statistically significant cluster of high values fall
within an urban area, it would be attributed to being 50% urban. The hot spots calculated through the frequency of crashes are seen to relate to the highest percentage of urban roadways. This follows the explanation described earlier for Figures 5.2 and 5.3, where the frequency based hot spots correlate to the most densely populated areas. One thought regarding normalization is that even when weighting crashes by injury severity, the hot spots tend to lean towards densely populated areas. It may be seen, however, that this is not always the case, and that oftentimes hot spots of crash costs relate to a higher percentage of rural roadways than their normalized counterparts. There is a greater tendency for the percentage of urban and rural roadways identified within cost based hot spots to relate to the overall percent of urban and rural roadways within each county. This identifies that the cost based maps relate the best to the overall demographics of the county and have the least bias of population density present of the four types of hot spot maps analyzed.

5.5 Conclusion

A past concern with hot spots is their tendency to occur in highly populous areas. Many suggestions have risen through past research that population density should be accounted for within the calculation of hot spots. In attempt to implement such variables, the act of normalizing hot spots by population density was investigated through this study. A wide range of geographies were studied in attempt to investigate the reaction of normalization in areas of both high and low populations. In total, four counties that contain at least 70% urban areas and four counties that contain at least 70% rural areas were considered.
With the census population being obtained at the tract level, the calculated population densities were bound by zonal boundaries. This created the possibility for drastic changes in population density when moving from one census tract to another. In order to remove this aggregated trend, the population density was interpolated over entire counties. The use of IDW created a smooth transition of values from one crash to another. From the interpolated population densities, the locations to be accounted for through normalization are able to be identified. The peak population density for all of the counties examined ranged from almost 29,000 down to about 3,500 people per square mile. This allowed for the effects of a wide range of geographies to be examined.

Hot spots were identified through the calculation of the $Gi^*$ statistic. This statistic was examined using two main variables of concern, frequency of crashes and the cost of injury severity. Additionally, both of these variables were normalized for population density. Similarities and differences were able to be seen when comparing the non-normalized and normalized maps. The non-normalized maps tended to have hot spots closer to the highly populated areas, as was the concern giving reason to conduct this study. The normalized maps removed the hot spots from these same areas, and forced the clustering of high values to be indicated in remote areas around the edges of each county. This created hot spots in locations where crashes rarely occurred, which may make the implementation of safety tactics less effective. Additionally, with the movement of hot spots away from dense populations came the inclusion of large cold spots. These cold spots turned up in the locations of the densely populated areas, which effectively reduced the purpose of normalizing the maps, by creating a new cluster in the location of dense populations. When comparing the location of hot spots within the non-normalized maps,
variations in their geographical makeup are able to be identified. These variations relate to the cluster maps based on the frequency of crashes to be centrally located in dense urban environments; meanwhile, the maps based on the societal crash costs contained hot spots covering much larger rural geographies. The implementation of safety campaigns in dense population areas may make the efforts of law enforcement more widely known to the public. On the other hand, covering a variety of geographies and not being heavily persuaded by population density may ultimately reduce the injury severity of alcohol-related crashes. This study showed that while the cost-based hot spots are directed towards locations of higher populations, it is not a strictly confounding relationship. The cost-based hot spots routinely addressed less dense, rural locations.

Overall, the appropriate hot spot analysis methodology to use depends on the application of the study. The normalized maps, while reducing the presence of hot spots in densely populated areas, negates its purpose by introducing cold spots in the same location. Thus, the non-normalized hot spot maps still have relevance. The frequency-based hot spot maps contain the highest proclivity to target densely populated areas. For the use of reducing alcohol-related crashes, this procedure would be most applicable to the implementation of high visibility enforcement campaigns. This in turn may send a signal to all drivers that there is a high presence of law enforcement interested in stopping alcohol intoxicated drivers before they crash. The cost-based hot spot maps contain the ability to address both urban and rural communities. This procedure provides the best opportunity for reducing alcohol-related crashes, while at the same time not specifically targeting densely populated areas. The best opportunity for cost-based hot spot maps is
the implementation of saturation or corridor patrols, which may have an emphasis on reducing high severity crashes.
CHAPTER VI
CONCLUSION AND RECOMMENDATIONS

6.1 Introduction

The identification of crash locations is important to educators, enforcement, and engineers alike. Knowing where crashes are likely to occur provides a basis of where to implement safety plans. A scatter plot of crash locations may provide a general idea of where the crashes occur; however, it is difficult to draw any forthcoming results. In order to determine the distribution of crashes, an examination using spatial analysis must occur. While there are many spatial analysis options available, this research examines several improvements to advance the examination of crash patterns. These advancements pertain to: 1) the calculation of spatial autocorrelation and interpolation, 2) the identification of spatio-temporal patterns, and 3) the influence of geographical patterns on the spatial distribution of crashes.

6.2 Calculation of Spatial Autocorrelation and Interpolation

Hot spot analysis allows for the identification of roadways that may be patrolled by law enforcement in an effort to reduce alcohol-related crashes. The roadways identified through a hot spot analysis provide a defined location where law enforcement may search for drivers who may be operating their vehicles while intoxicated. The use of a statistically backed analysis reduces the bias involved in determining roadways that law enforcement is assigned to patrol. Increased bias and the patrol of roadways that do not
effectively address the problem of alcohol-related crashes may raise issues with the legality of a stop performed on a suspected driver.

Through a comparison of the Euclidean and network distances, a large variance in the prediction accuracy index was identified for the calculation of the Getis-Ord \( Gi^* \) statistic. The variations, however, are minimal within interpolation calculations of hot spots when using Euclidean distances and network-based distances. Thus, while the use of network-based distances in the interpolation of hot spots is only slightly beneficial, the use of network-based distances within the calculation of the \( Gi^* \) is crucial. By using network-based distances within the calculation of the \( Gi^* \) and either measurement for the interpolation of hot spots, law enforcement would benefit from a more compact and efficient analysis. These benefits rise from the reduction of unnecessarily patrolled roadways and increases in societal crash costs; thus, improving the legality of roadways that are patrolled for impaired driving enforcement campaigns.

6.3 Identification of Spatio-Temporal Patterns

This research investigated both single and multiple vehicle alcohol-related crashes. While alcohol consumption is mainly a social behavior, spatio-temporal changes have a large effect in the distribution of crashes. A strong understanding of this distribution is essential to direct the efforts of educators and law enforcement, who attempt to reduce the overall occurrence of alcohol-related crashes. The examination of these crashes delves into the aspects of where and when these crashes occur and identifies differences between both types of crashes. By identifying shifts in the spatial patterns throughout time, the effects of implementations made to ensure safer roadways may be more pronounced.
The movement of clusters separates the spatial analysis from the spatio-temporal analysis. The results indicate that hot spots may move widely throughout a given time span. Given these shifts in hot spot locations, law enforcement must also alter the location of safety campaigns designed to reduce the number and severity of alcohol-related crashes. If the location of safety campaigns does not change as time progresses, there exists a risk of implementing a safety campaign in a non-hot spot location. Additionally, due to changes in the size of hot spots, the type of patrol may need to be altered to address large, condensed hot spots rather than small, dispersed hot spots.

6.4 Influence of Geographical Patterns on the Spatial Distribution of Crashes

In an effort to reduce alcohol-related crashes, the use of high visibility campaigns, saturation patrols, and corridor patrols are important tools utilized by law enforcement. The ability to identify the location to implement these tools relies on spatial analyses. Through spatial analysis, hot spots of crashes are able to be identified, and these hot spots statistically identify locations where law enforcement agencies should focus their efforts. In the creation of these maps, there is often concern that hot spot maps only target high population areas. In an effort to address this issue, this study examined the usefulness of normalizing these maps based on population density.

The comparison of normalized to non-normalized hot spot maps returned a total of four different types of maps examined over eight counties. Variations are found between each of the four types of maps. These variations are directly related to the type of geographies that included the statistically significant hot spots. By analyzing these variations, it is discovered that normalizing the hot spots is not necessary. Differences between the examination of frequency and societal cost hot spot maps indicate a
separation in the demographics being targeted. Those hot spots targeting high populations are found to result from hot spots based on the frequency of crashes. By examining hot spots based on the injury severity of crashes, the focus of high population areas was removed and the hot spots were dispersed among both urban and rural geographies.

6.5 Future Recommendations

While the results of this research advance the use of spatial analyses in crash examinations and the implementation of safety campaigns, future advancement may be possible. This future research may further link the use of hot spots to the implementations designed by educators, law enforcement, and engineers. Increased use of spatial analyses allows campaigns used by professionals in these fields to become ever more beneficial and efficient.

6.5.1 Linking Safety Campaigns to Hot Spots

Hot spots are shown to identify the distribution and location of motor-vehicle crashes. The implementation of safety campaigns relies heavily on the determination of locations where crashes are occurring. Improvements may continually be made towards identifying areas where crashes are likely to occur; however, without linking the success or failure of spatial analysis results to safety campaigns, improvements in reducing crashes may not progress as rapidly as needed. In order to accomplish this task, the strict use of hot spots should be used in a safety campaign. The use of a campaign may need to be implemented over multiple years to identify any significant changes in driver behavior. However, through an analysis of crashes in the implemented region, the influences should be identified. By identifying the influences that reduce crashes, the
effective implementation and advancement of using hot spots to locate campaign sites may be achieved.

6.5.2 Predicting Future Hot Spot Locations

Hot spot analyses of crashes are continually used in the investigation of past crashes in order to identify the locations where crashes are occurring. This information is useful; however, by only looking for the locations of where crashes have occurred in the past, the analyses are being reactive instead of proactive. Such research would first have to apply spatio-temporal techniques to identify patterns of movement. These movements would then have to be related to changes occurring within the environment surrounding the crashes. The ability to use past crash data to predict the movement of crashes in the upcoming year would give safety campaigns a leading edge.

6.6 Conclusions

This research investigated and applied new techniques to analyze motor-vehicle crashes. This research aids in the advanced identification of hot spots for motor-vehicle crashes. This research examined the current state of the practice. In building upon this current state, the most up-to-date crash data and geographic information was examined. This data was analyzed using new techniques that improved the accuracy of identified hot spots, determined the movement of hot spots through time, and identified the relationship of spatial autocorrelation to geographic attributes. The results of these analyses allow for increased efficiency of educational, enforcement, and engineering campaigns aimed at reducing the severity and occurrence of crashes. The efficiency is raised due to removal of unnecessarily patrolled roadways from enforcement campaigns, the identification of
when and where safety campaigns should be located, and how the ideal location for different types of safety campaign may be identified by studying various aspects of crashes. Additionally, future research is needed that may build upon the results found from this study.


 BAČKALIĆ, S. (2013). TEMPORAL ANALYSIS OF THE TRAFFIC ACCIDENTS OCCURRENCE IN PROVINCE OF VOJVODINA. Transport Problems, 8:1


Evans, L. (1990). The fraction of traffic fatalities attributable to alcohol. Accident Analysis & Prevention, 22(6), 587-602.


SANET, Spatial Analysis along Networks (Ver.4.1). Atsu Okabe, Kei-ichi Okunuki and SANET Team, Tokyo, Japan.


