Technical Report Documentation Page

1. Report No. CTS 17-04	2.	3. Recipients Accession No.	
4. Title and Subtitle Alcohol-Related Hot-Spot Analysis and Prediction		5. Report Date May 2017	
		6.	
7. Author(s) William H. Schneider IV, Brandon Stakleff, Lauren Buser		8. Performing Organization Report No.	
9. Performing Organization Name and Address Department of Civil Engineering 209E Auburn Science and Engineering Center Akron, OH 44325		10. Project/Task/Work Unit No. CTS #2015040	
		11. Contract (C) or Grant (G) No. DTRT13-G-UTC35	
12. Sponsoring Organization Name and Address Roadway Safety Institute Center for Transportation Studies University of Minnesota 200 Transportation and Safety Building 511 Washington Ave. SE Minneapolis, MN 55455		13. Type of Report and Period Covered	
		14. Sponsoring Agency Code	
15. Supplementary Notes http://www.roadwaysafety.umn.edu/publications/			

16. Abstract (Limit: 250 words)

This project developed methods to more accurately identify alcohol-related crash hot spots, ultimately allowing for more effective and efficient enforcement and safety campaigns. Advancements in accuracy came from improving the calculation of spatial autocorrelation and interpolation, the identification of spatio-temporal patterns, and the influence of geographical patterns on the spatial distribution of crashes. The project then used the location-based hot-spot maps created using these improved methods to develop a new method of patrolling for intoxicated drivers. This method guides officers to statistically significant locations where intoxicated drivers are most likely to be, allowing officers to be more accurate while patrolling. Additionally, this method allows officers to pass through more alcohol-related crash locations per minute and mile than current patrolling practices. By improving how officers patrol, individuals may be deterred from driving while intoxicated, and alcohol-related crashes may be ultimately reduced.

17. Document Analysis/Descriptors alcohol use, drunk driving, high risk locations, spatial analysis, police patrol		18. Availability Statement No restrictions. Document available from: National Technical Information Services, Alexandria, Virginia 22312	
19. Security Class (this report) Unclassified	20. Security Class (this page) Unclassified	21. No. of Pages 140	22. Price

LIST OF FIGURES

Figure 3.1. Comparison of *Gi** z-scores obtained by Euclidean and network-based analysis for Cuyahoga County. ...17

Figure 3.2. Comparison of hot spot areas between Euclidean and network-based analysis. ...18

Figure 3.3. Identification of hazardous roadways. ...19

Figure 4.1. Hot Spots of Alcohol-Related Crashes. ...34

Figure 4.2. Spatio-Temporal Analysis of Alcohol-Related Crashes in Cuyahoga County by Time of Day. ...37

Figure 4.3. Comparison of Single and Multi-Vehicle Hot Spots by Time of Day. ...39

Figure 4.4. Spatio-Temporal Analysis of Alcohol-Related Crashes by Day of Week. ...41

Figure 4.5. Comparison of Single and Multi-Vehicle Hot Spots by Day of Week. ...43

Figure 5.1. County Population Density. ...51

Figure 5.2. Hot Spot Maps of Urban Counties. ...53

Figure 5.3. Hot Spot Maps of Rural Counties. ...56

Figure 6.3.1: Hot Spot Maps of Franklin (left), Summit (middle), and Ross (Right) Counties ...65

Figure 6.3.2: Significant Network Locations for Franklin County for 90% confidence levels (left), 95% confidence levels (middle), and 99% confidence levels (right). ...66

Figure 6.3.3: Significant Network Locations for Summit County for 90% confidence levels (left), 95% confidence levels (middle), and 99% confidence levels (right). ...66

Figure 6.3.4: Significant Network Locations for Ross County for 90% confidence levels (left), 95% confidence levels (middle), and 99% confidence levels (right). ...67

Figure 6.3.5: Significant Network Locations for Franklin County for 90% confidence levels (left), 95% confidence levels (middle), and 99% confidence levels (right), zoomed in. ...69

Figure 6.3.6: Significant Network Locations for Summit County for 90% confidence levels (left), 95% confidence levels (middle), and 99% confidence levels (right), zoomed in. ...70

Figure 6.3.7: Significant Network Locations for Ross County for 90% confidence levels (left), 95% confidence levels (middle), and 99% confidence levels (right), zoomed in. ...70

Figure 7.3.1: Corridor & Hot Spot Patrol Areas for Franklin, Summit, and Ross Counties 78

Figure 7.4.1: Optimized Corridor Routes (top) and Hot Spot Routes (bottom) for Franklin County for 3, 5, and 7 patrol cars respectively. 80

Figure 7.4.2: Optimized Corridor Routes (top) and Hot Spot Routes (bottom) for Summit County for 3, 5, and 7 patrol cars, respectively. 82

Figure 7.4.3: Optimized Corridor Routes (top) and Hot Spot Routes (bottom) for Ross County for 3, 5, and 7 patrol cars, respectively. 84

Figure 8.4.1: Results of First Failure Mode for Franklin County. 97

Figure 8.4.2: Results of First Failure Mode for Summit County. 98

Figure 8.4.3: Results of First Failure Mode for Ross County. 99

EXECUTIVE SUMMARY

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 spots 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.

The use of hot spots may be beneficial in improving campaigns to reduce alcohol-related crashes. Once the hot spot maps are created, this research uses the results to develop a new method of patrolling for intoxicated drivers. The hot spot maps are broken down into local indicators of spatial association, which show statistically significant locations where intoxicated drivers are likely to be present. Route optimization models are then used to guide officers to these locations. These models are compared with traditional methods of corridor patrolling through a series of performance metrics. Failure probability models are then created to further compare the two methods of patrolling, as well as aiding captains of jurisdictions in decision-making processes.

By utilizing location-based hot spots, new methodologies of patrolling may be developed in order to reduce the amount of alcohol-related crashes. This new method of patrolling will guide officers to statistically significant locations, allowing them to be more accurate while patrolling for intoxicated drivers. Additionally, this method proves to pass through more alcohol-related crash locations per minute and mile, indicating it may be more efficient than current practices of patrolling. By improving how officers patrol, people may more accurately be deterred from driving intoxicated and alcohol-related crashes may be ultimately reduced.

CHAPTER 1: 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. Using these relationships, safety campaigns may be improved by implementing new methodologies of patrolling with the goal of reducing the amount of alcohol-related crashes.

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, as well as examines new methodologies for the implementation of safety campaigns. 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 reduce the amount of alcohol-related crashes in the state of Ohio by creating a geospatial means to analyze motor vehicle crashes, then improving overtime patrols to be implemented in safety campaigns. 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. The analyses may then be used to develop new methods of patrolling, which may aid in reducing the number and severity of crashes.

The overall objective of this research, to reduce the amount of alcohol-related crashes in the state of Ohio by creating a geospatial means to analyze motor vehicle crashes, then improving overtime patrols to be implemented in safety campaigns, is achieved through six 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. The fourth analysis begins the development of new methods of safety campaigns. The spatial relationships previously developed are further broken down to locating statistically significant areas, which define exact areas where officers may be able to patrol. The fifth analysis uses a model to create routes for officers to follow which will guide them to the specific areas defined in the fourth analysis. This analysis will also compare this new method of patrolling to a traditional method of corridor patrolling. The sixth analysis creates a failure probability model that will justify the use of the newly developed method of patrolling while potentially helping administrator in the decision-making process. 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 THIS RESEARCH

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.

1.2.5 Chapter VI: Using Local Indicators of Spatial Association to Improve Patrols and Reduce Alcohol-Related Crashes

The purpose of this chapter is to locate significant areas for officers to patrol for intoxicated drivers. The significant areas are determined from the output of hot spot analyses in three counties in Ohio. The output of the hot spots is a series of points that may be broken down into local indicators of spatial association that identify a confidence interval for each output of the hot spot map. By using these points, officers will have more specific targets, backed up by statistical significance, while patrolling for intoxicated drivers. These points are defined by 90%, 95%, and 99% confidence, or no significance. This research found that utilizing the 95% confident network locations may be best in guiding officers patrolling for intoxicated drivers.

1.2.6 Chapter VII: Comparison of Traditional Corridor Based Enforcement with Route Optimization of Hot Spot Analysis

The goal of Chapter IV is to compare traditional corridor enforcement practices with the newly proposed hot spot route optimization (HSRO). The HSRO method of patrolling optimizes route to each of the locations provided in Chapter III. These routes are compared with the traditional method of corridor patrolling to determine the most efficient method of patrolling for intoxicated drivers. Comparing these two methods will help to determine if the HSRO method of patrolling emulates traditional methods of patrolling. The average amount of alcohol-related crash locations passed per mile and minute are use as performance metric for each method of patrolling and compared. Ultimately the HSRO method of patrolling is able to pass through more alcohol-related crash locations per mile and time, indicating that this method may be the most efficient in patrolling for intoxicated drivers.

1.2.7 Chapter VIII: Use of Failure Probability Models to Justify New Methods of Patrolling

This chapter uses two failure probability models to further compare the efficiencies between patrolling through corridors and the HSRO method of patrolling. Failure probability is used to determine the failure of scenarios given a number of variables. The goal of the first model is to determine the most amount of consecutive cycles that may be completed by a given fleet size, while the goal of the second model is to determine the cost-effectiveness of patrolling and the cost of potential pullovers. These models may be beneficial in determining which method of patrolling may be the most efficient to use. These models may also be useful to captains in determining the desired fleet size for patrolling for intoxicated drivers on a given night.

1.2.8 Chapter VIII: 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, as well as the development of new methods of patrolling using the spatial analyses. Additionally, this chapter reviews future recommendations of the research, including implementation practices and future studies of alcohol-related crash patterns. These future recommendations build upon the techniques used within this research.

CHAPTER 2: 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 (Songchitruksa 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-m^2 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 3: 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.

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 Euclidian 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

Injury Severity	Number of Crashes
Property Damage Only	1900
Injury	1400
Fatal	65
Total	**3,365**

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 *Gi** 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 *Gi** statistic is calculated using the following equation:

$$G_i^*(d) = \frac{\sum_j w_{ij}(d)x_j - W_i^*\bar{x}}{s\left[\frac{W_i^*(n-W_i^*)}{(n-1)}\right]^{1/2}} \tag{3.1}$$

where:

$$W_i^* = \sum_j w_{ij}(d) \tag{3.2}$$

$$s^2 = \frac{\sum_j x_j^2}{n} - \bar{x}^2 \tag{3.3}$$

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 *Gi** 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:

$$z_0 = \frac{\sum_{i=1}^{s} z_i \frac{1}{d_i^k}}{\sum_{i=1}^{s} \frac{1}{d_i^k}} \qquad (3.4)$$

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:

$$PAI = \frac{\frac{c}{C} \times 100}{\frac{l}{L} \times 100} \qquad (3.5)$$

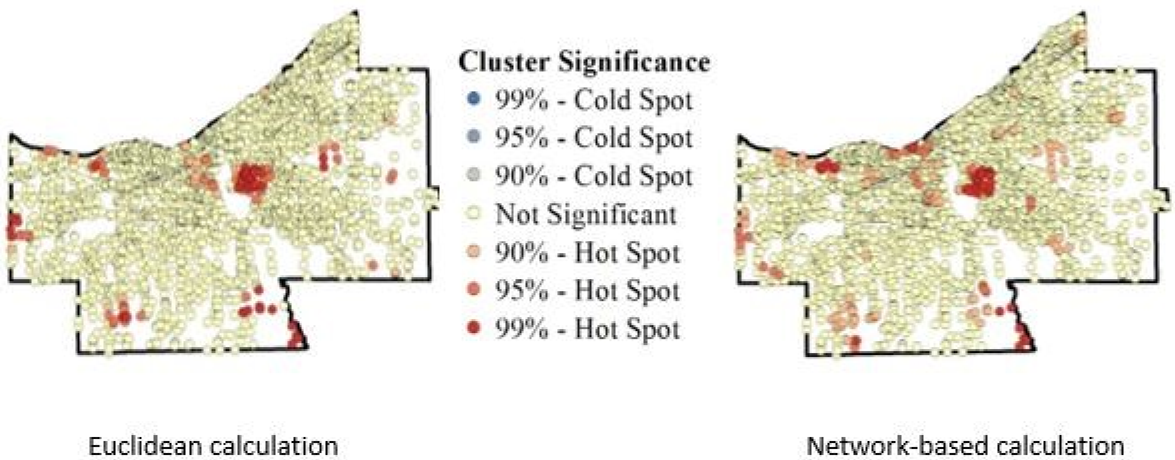
where, *c* is the societal crash cost of crashes in hot spots, *C* is the total societal crash cost of all crashes within the study area, *l* is the length of roadways identified as being located in the hot spot, and *L* is the total length of roadways within the entire study area. Thakali et al. (2015) 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 my 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.

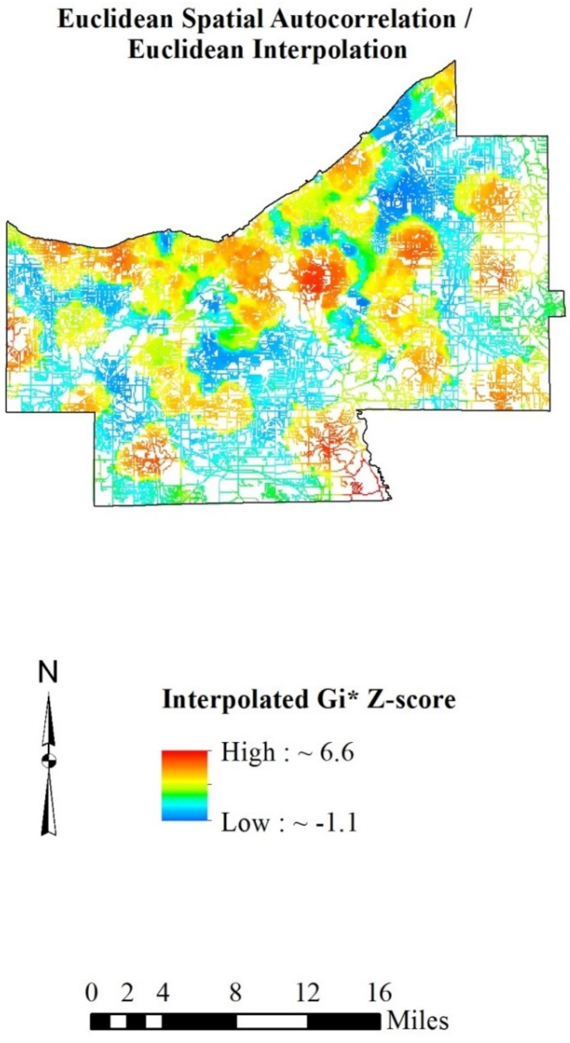


Note: Hot spots represent locations where high injury severity crashes are close in distance to other high severity crashes. Cold spots represent locations where low severity crashes are close in distance to other low severity crashes.

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.

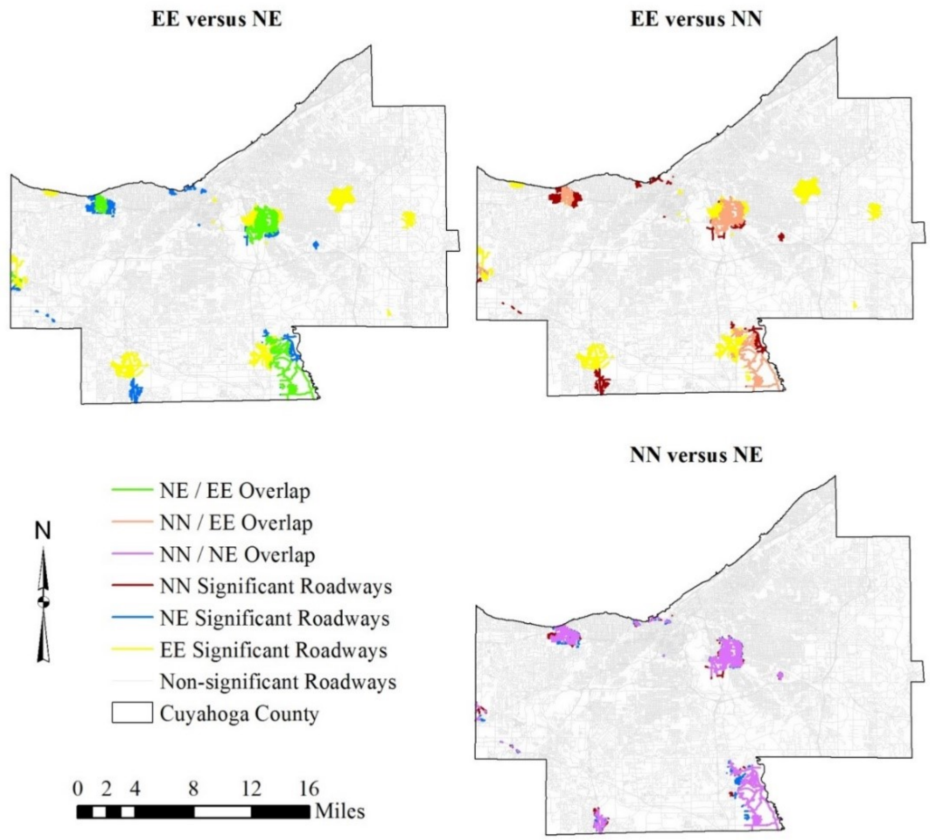


Note: Network interpolation completed with the use of SANET ver. 4.1.

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.

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

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.*

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 *Gi** have larger PAI values (3.50 and 3.55) as opposed to the value where the *Gi** 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.

	NE	NN
EE	44.88%	44.70%
NE		0.19%

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.

	NE	NN
EE	38.01%	39.62%
NE		1.67%

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 4: 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 campaign enacted by law enforcement aim is 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

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 *Gi** 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:

$$G_i^*(d) = \frac{\sum_j w_{ij}(d)x_j - W_i^*\bar{x}}{s\left[\frac{W_i^*(n-W_i^*)}{(n-1)}\right]^{1/2}} \tag{4.1}$$

where:

$$W_i^* = \sum_j w_{ij}(d) \tag{4.2}$$

$$s^2 = \frac{\sum_j x_j^2}{n} - \bar{x}^2 \tag{4.3}$$

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:

$$z_0 = \frac{\sum_{i=1}^{s} z_i \frac{1}{d_i^k}}{\sum_{i=1}^{s} \frac{1}{d_i^k}} \qquad (4.4)$$

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 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.

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

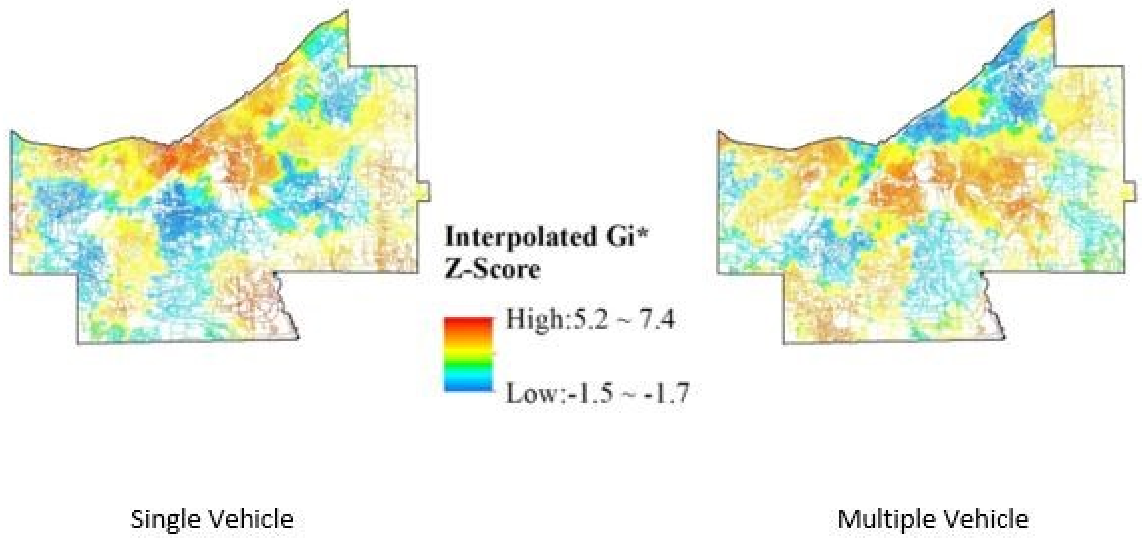


Figure 4.1. Hot Spots of Alcohol-Related Crashes.

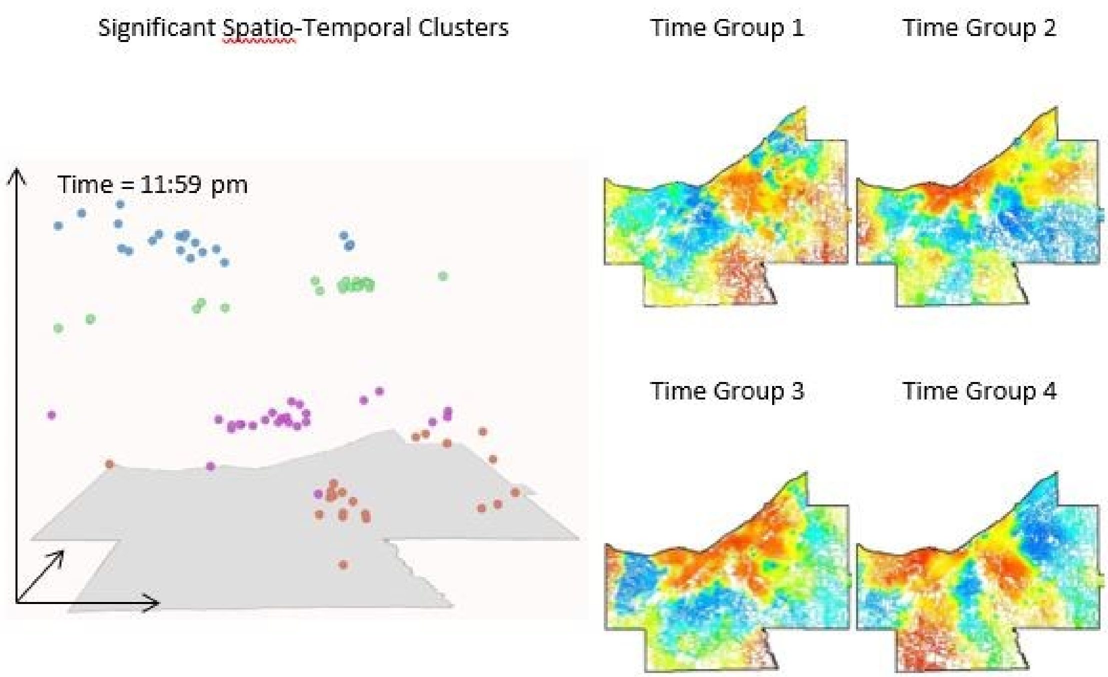
There are several locations within Figure 4.1 where significant clusters of high severity crashes occur. These significant clusters are identified when the z-score is greater than or equal to 1.96, which correlates to a 95% level of significance. The positively significant clusters related to those clusters 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.

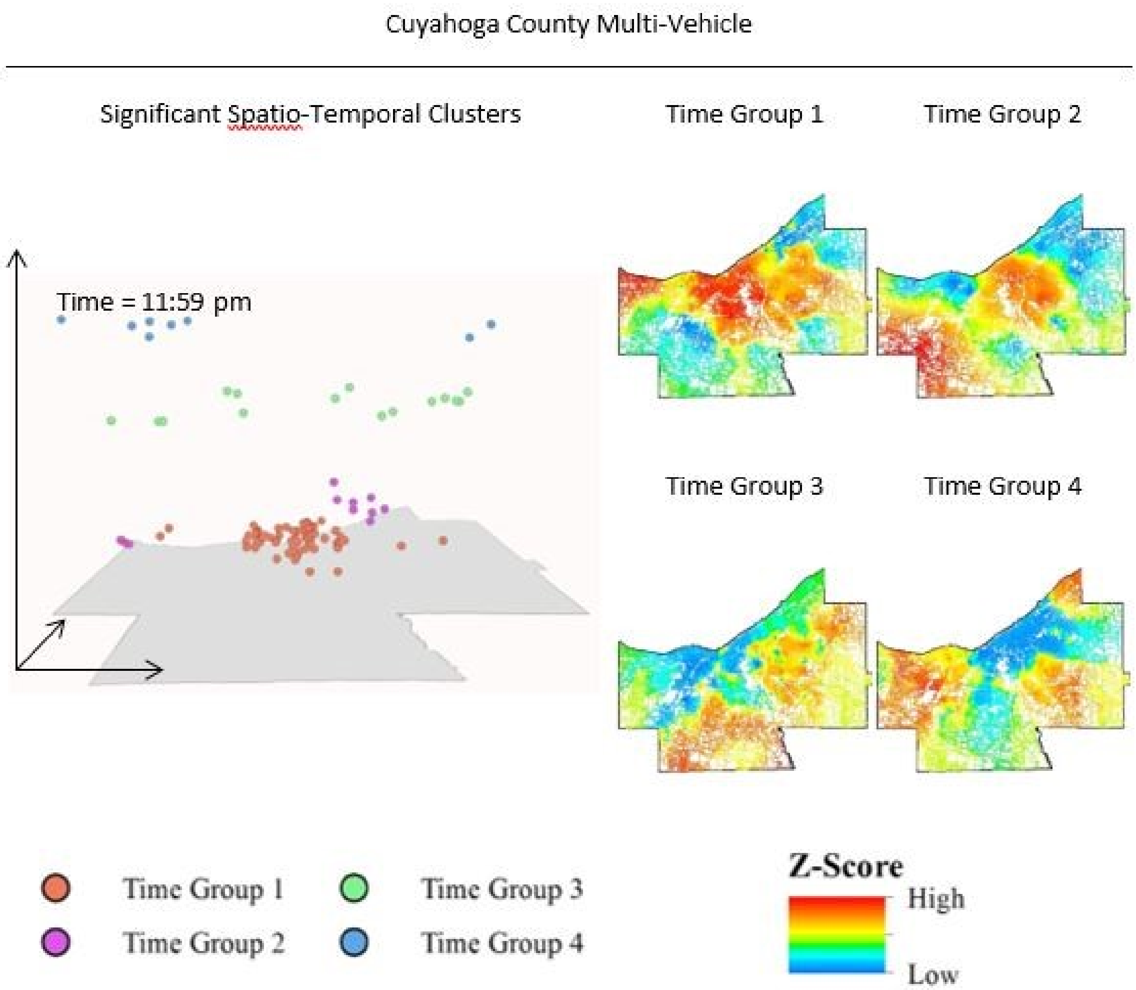
4.4.2 Spatio-Temporal Analysis

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





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.

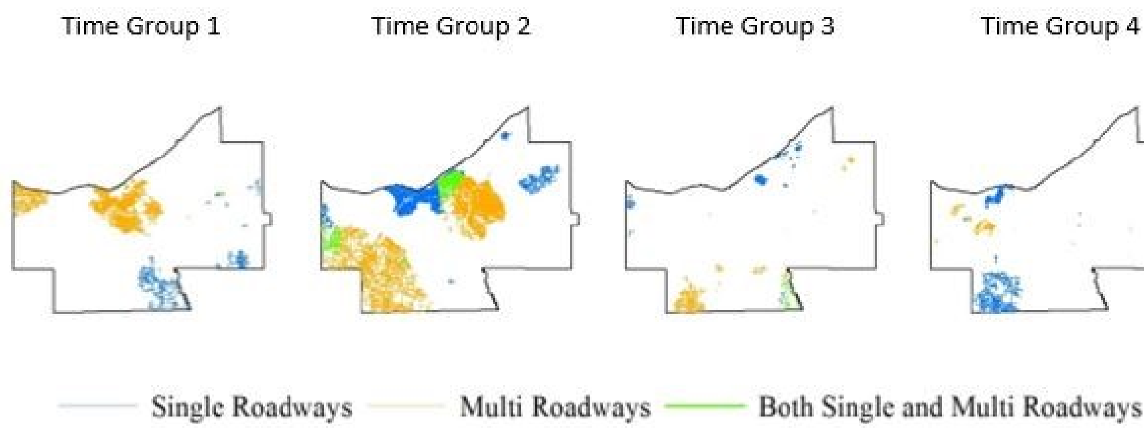
Grouped Cluster	Single Vehicle	Multi-Vehicle	Combined Timeframe
Time Group 1	12:00 am – 3:52 am	12:00 am – 1:59 am	12:00 am – 4:00 am
Time Group 2	4:15 am – 8:36 am	4:55 am – 6:30 am	4:15 am – 8:45 am
Time Group 3	4:05 pm – 8:37 pm	3:14 pm – 5:59 pm	3:00 pm – 8:45 pm
Time Group 4	10:08 pm – 11:46 pm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

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 locations 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.

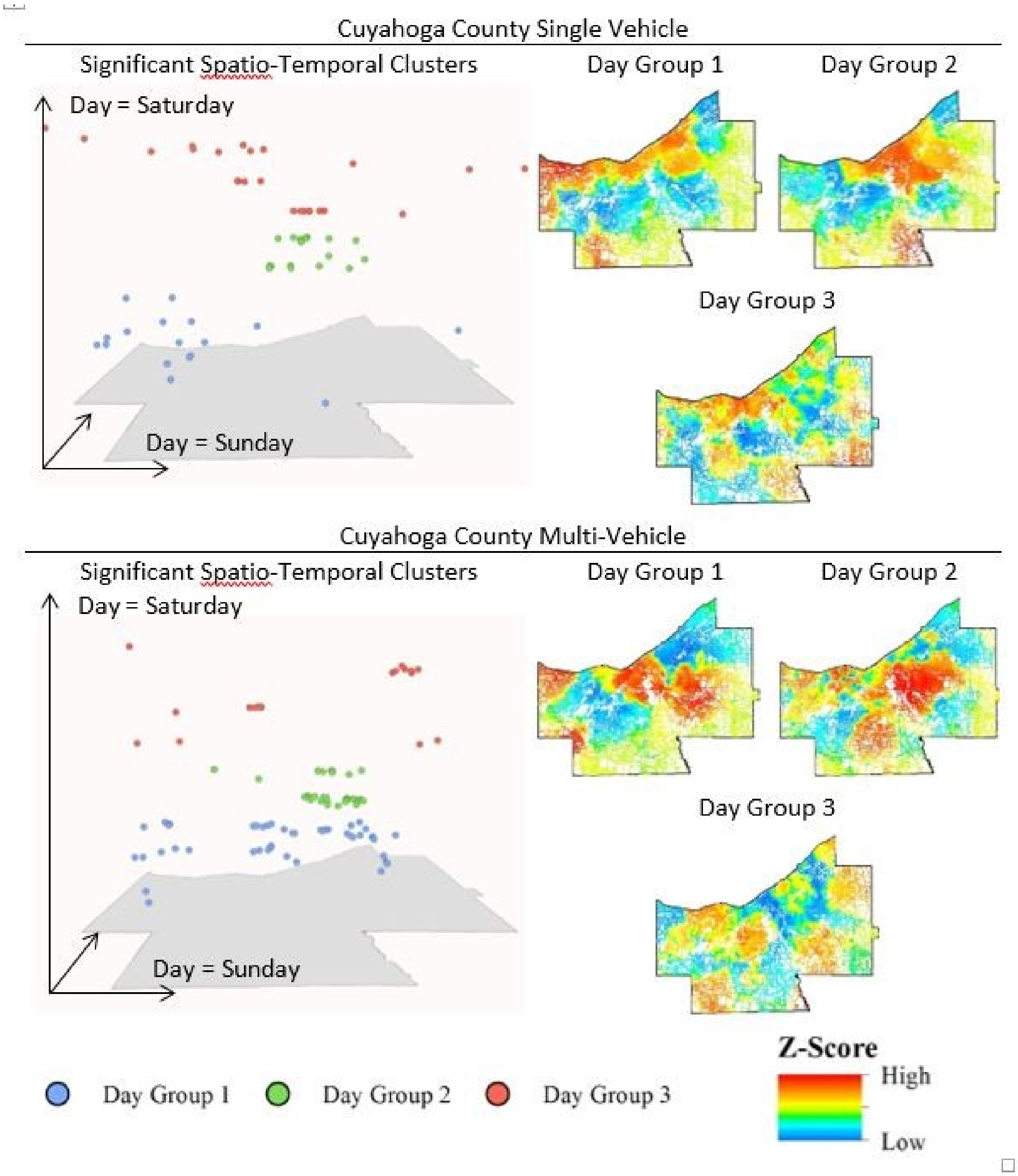


Note: Network interpolation completed with the use of SANET ver. 4.1.

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 there is a decrease in the number of vehicles, which in turn shifts the statistically significant multi-vehicle clusters to the more predominate statistically significant 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.



Note:

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

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; 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.

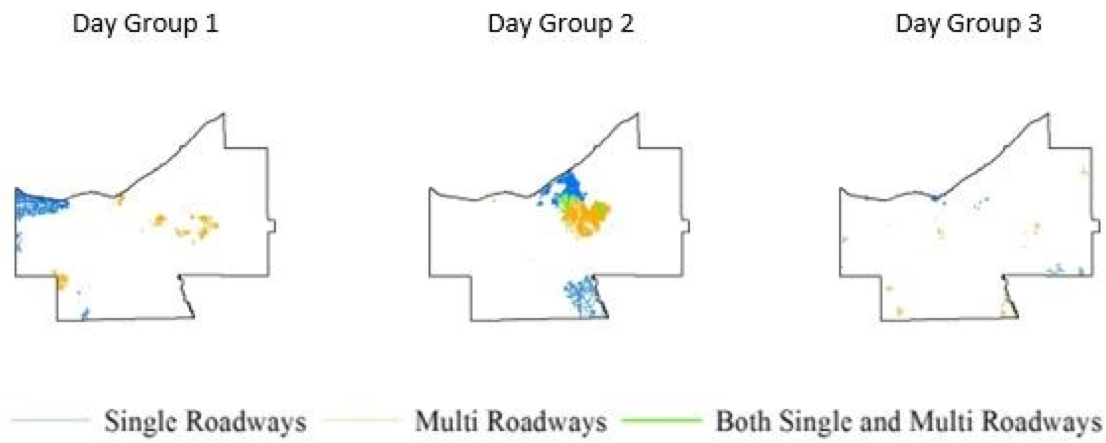
Grouped Cluster	Single Vehicle	Multi-Vehicle	Combined Timeframe
Day Group 1	Sunday – Monday	Sunday – Monday	Sunday – Monday
Day Group 2	Tuesday - Wednesday	Tuesday - Wednesday	Tuesday - Wednesday
Day Group 3	Thursday - Saturday	Thursday - Saturday	Thursday - Saturday

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.



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 5: 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 effect 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.

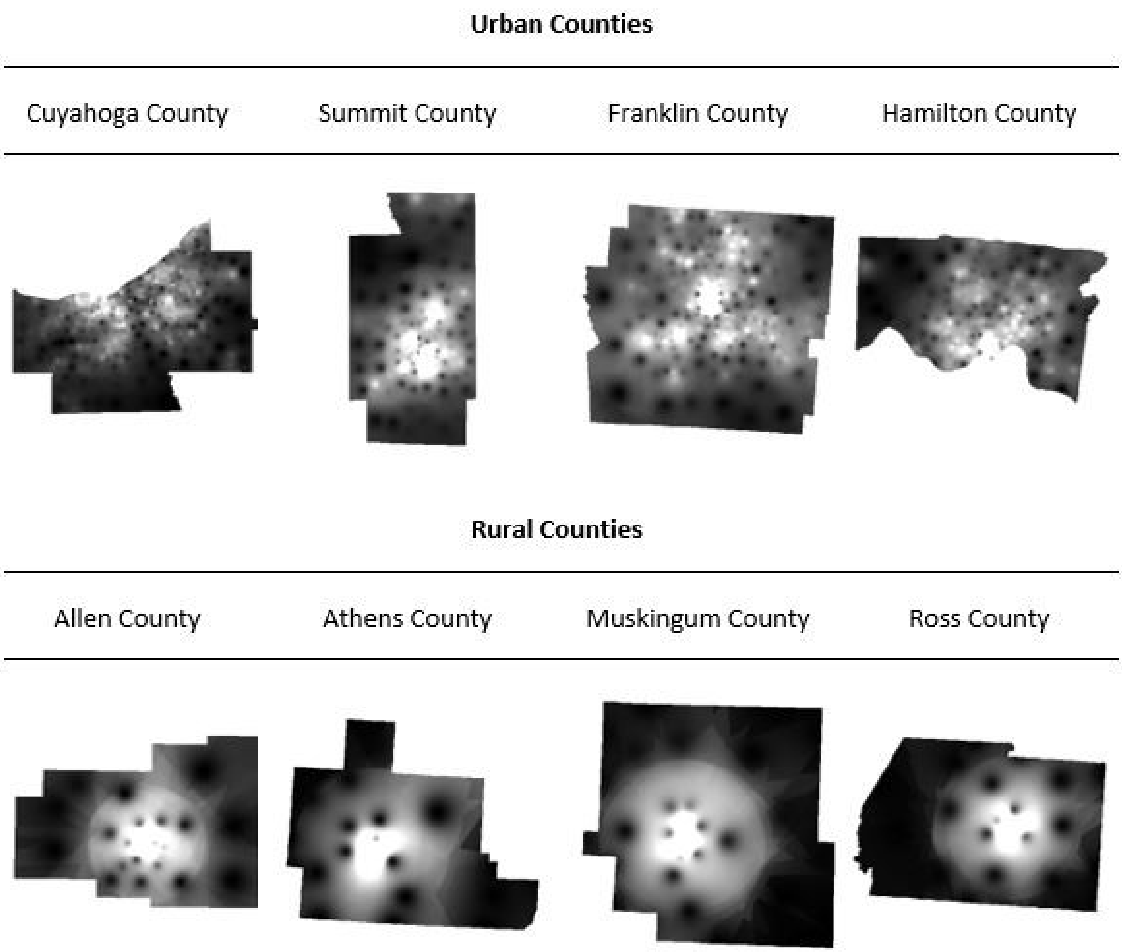
As hot spots are shown to indicate where law enforcement officers may patrol, the identification of statistically significant areas is important to determine. Hot spots of crashes are determined based on the relationship between a value pertinent to a crash location and the distance separating each crash 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.

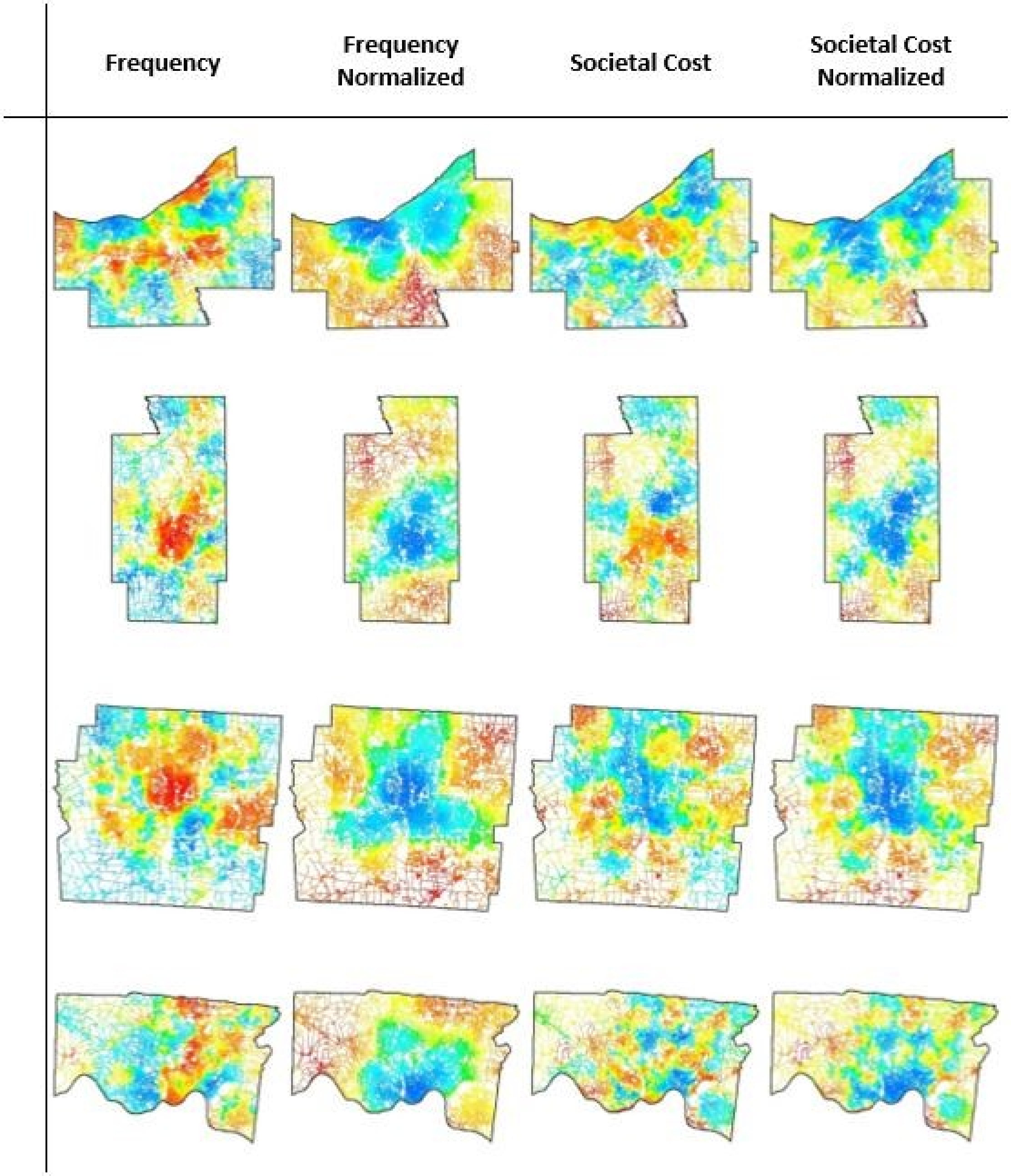


Note: The color ramp is based on the population density (persons per square mile). The lighter areas correlate to higher population densities. Meanwhile, the darker areas correlate to lower population densities.

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 cross covariance 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.



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.

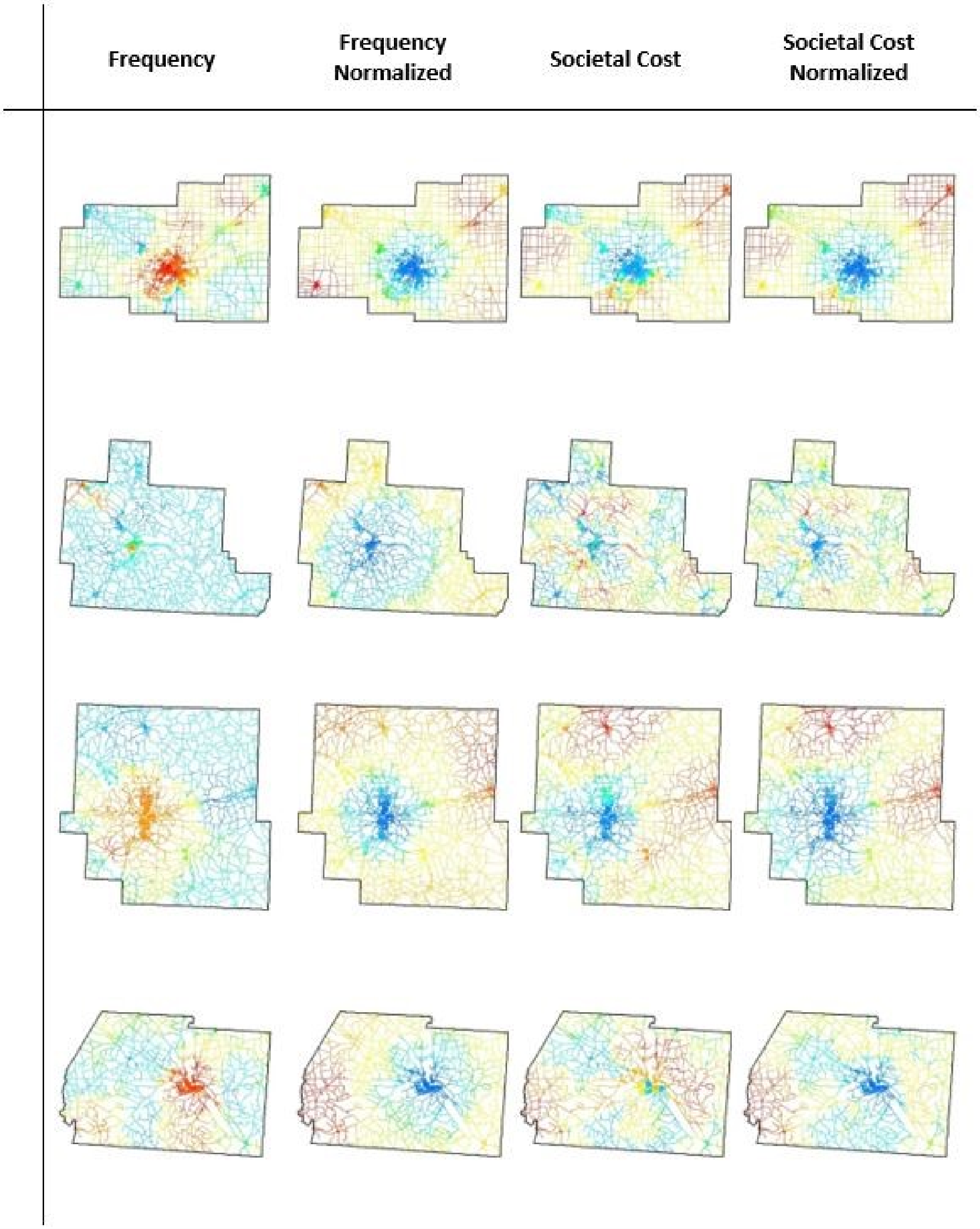
Figure 5.2 identifies the z-score relating to each roadway within the studied counties. Those roadways that are indicated in a red color are more significant towards the clustering of high values. On the other hand, those roadways that are blue in color are more significant towards clusters of low values. 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 area 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.

Urban Counties	Percent Urban	Percent Rural	Rural Counties	Percent Urban	Percent Rural
Cuyahoga			**Allen**		
Cost	70%	30%	Cost	12%	88%
Normalized Cost	79%	21%	Normalized Cost	11%	89%
Frequency	100%	0%	Frequency	100%	0%
Normalized Frequency	94%	6%	Normalized Frequency	9%	91%
Summit			**Athens**		
Cost	63%	37%	Cost	4%	96%
Normalized Cost	46%	54%	Normalized Cost	0%	100%
Frequency	100%	0%	Frequency	44%	56%
Normalized Frequency	70%	30%	Normalized Frequency	69%	31%
Franklin			**Muskingum**		
Cost	88%	12%	Cost	10%	90%
Normalized Cost	73%	27%	Normalized Cost	7%	93%
Frequency	100%	0%	Frequency	54%	46%

Normalized Frequency	84%	16%	Normalized Frequency	3%	97%
Hamilton			**Ross**		
Cost	92%	8%	Cost	0%	100%
Normalized Cost	81%	19%	Normalized Cost	0%	100%
Frequency	100%	0%	Frequency	52%	48%
Normalized Frequency	78%	22%	Normalized Frequency	0%	100%

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 6: USING LOCAL INDICATORS OF SPATIAL ASSOCIATION FROM HOT SPOT ANALYSES TO IMPROVE PATROLS AND REDUCE ALCOHOL-RELATED CRASHES

6.1 INTRODUCTION

Alcohol-related crashes have been a consistent problem in the United States. From 1999 through 2014 alcohol-related fatalities accounted for over 30% of total vehicle fatalities in the United States (NHTSA, 2014). The use of enforcement strategies, such as sobriety checkpoints and saturation patrols, helps to control the amount of intoxicated driving (Sanem et al., 2015), however due to the consistency of alcohol-related fatalities in recent years, there is still room for improvement. These enforcement strategies are also costly, creating a need for increased funding or efficiency in the current practices of patrolling.

The National Highway Traffic Safety Administration (NHTSA) utilizes the strategy of Data-Driven Approaches to Crime and Traffic Safety (DDACTS). DDACTS determines the most effective ways to reduce crimes and crashes through the use of location-based data collection. Temporal and environmental factors, as well as hot spot maps are used to identify significant locations of concern. These significant areas are then used to determine strategies to resolve the crime and crash problems.

Driving while intoxicated has been an area of concern since 1903 when the *Quarterly Journal of Inebriety* expressed concern about intoxicated operators of "motorized wagons" (Shepard, 1903). Moving forward in time to 2013-2014, the *National Roadside Survey of Alcohol and Drug Use by Drivers* reported that 5.2% of drivers were under the influence of alcohol (reduced from 1973 when 22.3% of drivers were intoxicated) (Berning et al., 2015). The reasons behind intoxicated driving and the dangers of it have been widely studied, including how the perception of one's own level of intoxication or how one views others in society influence their own decisions (Gellar and Smith, 2014; Meesmann et al., 2015; Christoforou et al., 2012; Turner and Georggi, 2001; Timmerman et al., 2003; Harrison and Fillmore, 2005). Though the amount of intoxicated drivers have decreased from when it first became a noticeable issue, they are still very prevalent, giving research and officers determination to discover which methods best deter people from driving intoxicated. Sanem, et al. (2015) found that the combined use of multiple enforcement strategies, such as sobriety checkpoints, saturation patrols, and open container laws, decrease the amount of intoxicated driving. The amount of enforcement in areas has also been proven to show reductions in the amount of people willing to drive intoxicated (Fell et al., 2014). Additional research shows the effects that enforcement strategies may have on the reduction of alcohol-related crashes (Fell et al., 2008; Jai et al., 2016; Fell et al., 2014; Blais et al., 2015; Elder et al., 2004; Vollrath et al, 2005).

Despite the research indicating that impaired driving is a problem and enforcement strategies may reduce the amount of alcohol-related crashes, efforts must continue to improve these enforcement practices. Hot spot methodologies, as explained by Songchitruska and Zeng (2010), have been proven to identify spatial relationships between high-impact crashes. Songchitruska and Zeng ultimately found the use of hot spot analyses to be an effective tool that may be used for decision making processes and incident detection. The use of hot spot analyses was also used by Maistros, et al. (2014) to locate significant areas of alcohol-related crashes in Stark County, Ohio. A similar application of hot spot analysis was conducted by Kuo, Lord, and Walden (2013) who took the hot spots of crimes and accidents and routed officers to the top five and top ten hot spots in College Station, Texas with hopes of reducing police dispatch times. These researchers, along with many others (Ratcliffe and McCullagh, 2001; Truong and Somenahalli, 2011; Carrick et al., 2014; Prasannakumar, et al., 2011; Khan et al., 2009; Cheng and Washington, 2008) have studied the benefits of hot spot analysis as a whole.

However further research has shown more in depth studies of these hot spots, analyzing local indicators of spatial association (LISA). Luc Anselin (1995) describes how the Gi and Gi* statistics may be used to identify these local indicators of spatial association, which may prove to be very beneficial in further studies of spatial data. De Vlack, et al. (2016) used LISA to show the substitutability of recreation areas in Belgium, showing the differences between hot and cold spots and how they relate to the Gi* statistics. Since the Anselin study of 1995, LISA has frequently been used to identify significant areas of interest (Johnston and Ramachandran, 2014; Nelson and Boots, 2008; Ratcliffe and McCullagh, 2001).

This research focuses on reducing the amount of alcohol-related crashes in the state of Ohio by developing a new method of patrolling through statistically significant hot spots. The DDACTS approach is utilized by taking hot spot maps created based on alcohol-related crash locations to show significant locations of these alcohol-related crashes. These hot spot maps are broken down into three confidence levels, 90% confidence, 95% confidence, and 99% confidence. Each county has a different amount of significant locations for each confidence level. Oftentimes there may be hundreds of significant locations that cover a concentrated area in the county, and as a result, patrolling each significant location may not necessarily be the most efficient way to locate intoxicated drivers. This research will determine if it is acceptable to patrol only the 99% confidence level locations, or the 95% and 99% confidence level locations. With the limited funding provided by the State, the amount of officers needed to patrol the network locations in a given shift time will be examined, as well as the officers' ability to cover the network locations. The goal of this chapter is to guide officers to more significant areas where intoxicated drivers are likely to be. This is found by using LISA from hot spot maps created off of frequency based alcohol-related crashes. The results of this research will be used to send officers to significant locations with hopes of reducing the amount of alcohol-related crashes.

6.2 DATA

The data sources for this study include crash records populated from the state of Ohio OH-1 crash reports (ODPS, 2015), the Ohio Department of Transportation (ODOT) Geographic Information Systems (GIS) roads layer (ODOT, 2016), and United States Census estimates information (United States Census Bureau, 2015). Using these three databases, the research team selected Franklin, Summit and Ross counties for analysis. All alcohol-related crashes in each county from January 1, 2012 through April 9, 2015 are included in the analysis.

Franklin County was selected due to its high population (greater than one million people), it encompasses the large metropolitan area of the City of Columbus, and its high number of alcohol-related crashes. Summit County was selected due to its large urban areas with a population greater than 500,000 people and a significantly high number of alcohol-related crashes. In contrast, Ross County is a predominantly rural county with a total population less than 100,000 people, and has a road network a quarter the size of Franklin County. Historically, Franklin County and Summit County are both in the top ten counties statewide with the highest number of alcohol-related fatalities per year. A summary the general characteristics of each of the three counties may be seen in Table 6.2.1.

Table 6.2.1: Comparison of Franklin, Summit, and Ross Counties

County Comparison

	County Population	Population Density (per sq. mi.)	Lane Miles	Total Alcohol-Related Crashes
Franklin	1,251,772	2,186	5,670	4,051
Summit	541,968	1,313	3,608	1,805
Ross	77,170	113	1,429	334

Note: County Population and Population Density determined from United States Census Bureau Quick Facts.
Lane Miles determined from Ohio Department of Transportation ArcGIS Roads Layer.
Total Alcohol-Related Crashes determined from Ohio Department of Public Safety OH-1 Crash Reports.

As shown in Table 6.2.1, there is a wide range of demographics, road networks and alcohol-related crashes between these three counties. This paper will use these counties as case studies as a demonstration for a new methodology that law enforcement agencies may use to help curb alcohol-related crashes.

6.3 METHODOLOGY

The hot spot maps of Franklin, Summit and Ross counties are used to help improve the efficiency of officers patrolling for intoxicated drivers. The output of these hot spot maps, explained by Songchitruska and Zeng (2010) and De Vlack et al. (2016), shows hot (red) and cold (blue) spots for each county. This may be seen in Figure 6.3.1. Each individual point on the map is considered a local indicator of spatial association (LISA), which has a Gi* statistic that shows each cluster's significance (De Vlack et al., 2016). The types of significance for these clusters may be broken down into seven categories, 90%, 95%, or

99% confidence level for hot and cold spots, or no significance. The focus of this research will remain on the hot spots, since they represent frequency-based alcohol-related crashes.

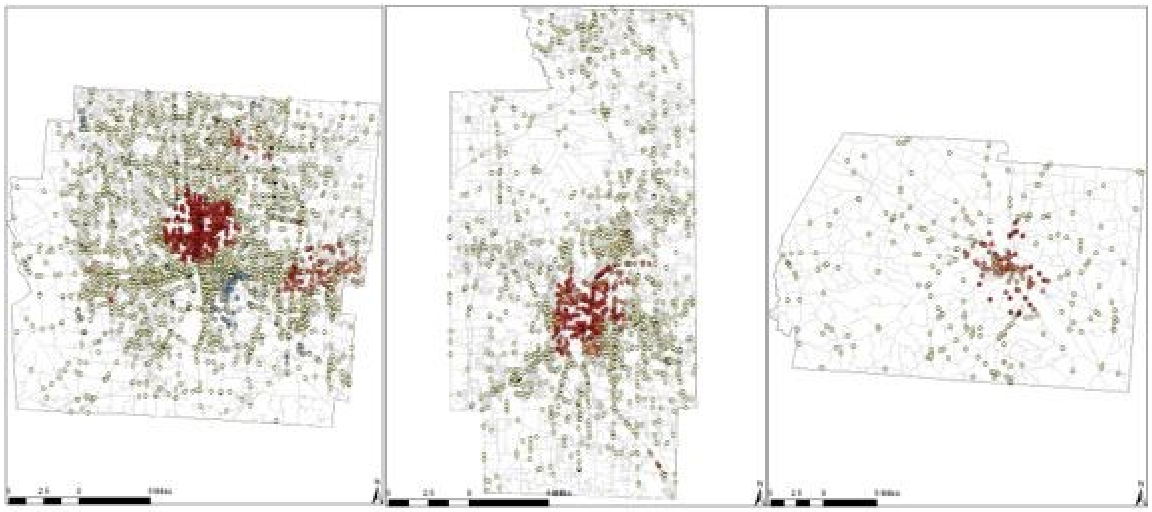


Figure 6.3.1: Hot Spot Maps of Franklin (left), Summit (middle), and Ross (Right) Counties

Network locations, which are points that are used in ArcGIS to guide patrol officers to each hot spot, are overlaid onto the 90%, 95%, and 99% confident level hot spots for each map. Locations identified as 99% confident are the most significant, indicating that there is a 1% chance that any location in this category is not actually significant. This pattern is similar for locations in the 95% and 90% confidence levels. The network locations are all located on road networks, allowing officers to eventually be routed to them. The significant network locations for Franklin, Summit and Ross counties may be seen in Figure 6.3.2, Figure 6.3.3, and Figure 6.3.4.

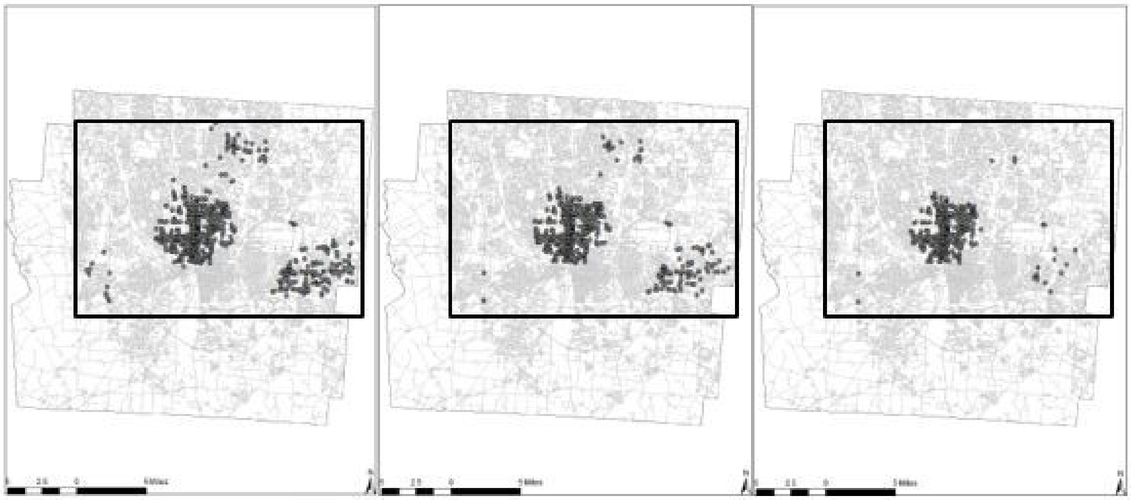


Figure 6.3.2: Significant Network Locations for Franklin County for 90% confidence levels (left), 95% confidence levels (middle), and 99% confidence levels (right).

Figure 6.3.3: Significant Network Locations for Summit County for 90% confidence levels (left), 95% confidence levels (middle), and 99% confidence levels (right).

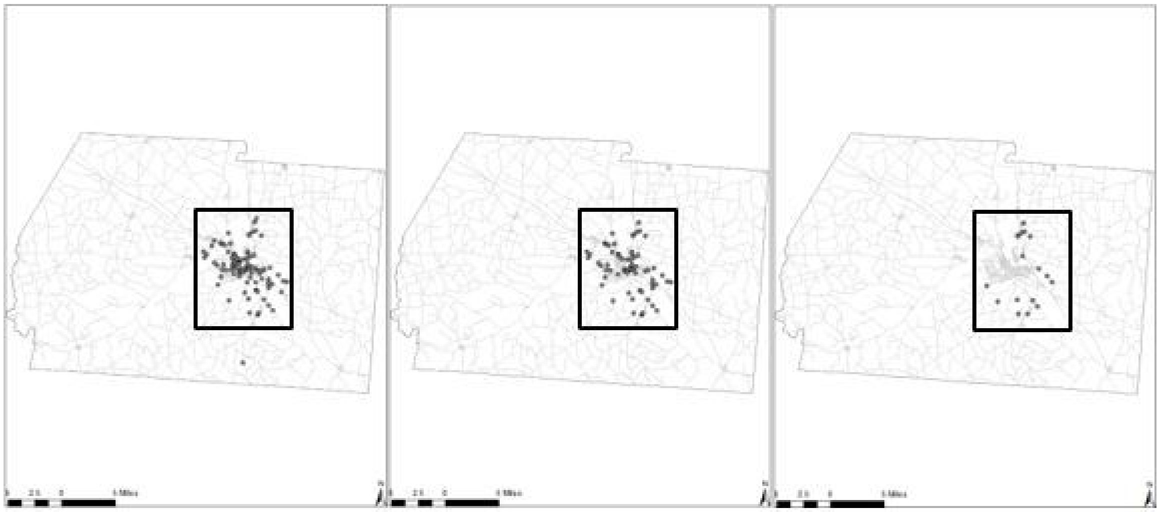


Figure 6.3.4: Significant Network Locations for Ross County for 90% confidence levels (left), 95% confidence levels (middle), and 99% confidence levels (right).

Each column in Figure 6.3.2, Figure 6.3.3, and Figure 6.3.4 represents a different confidence level. It should be noted that as the significance increases, the amount of network locations decreases. This may be verified through Table 6.3.1, which shows the exact of amount of network locations for each county.

Table 6.3.1: Count of Network Locations for each Confidence Level for Franklin, Summit, and Ross Counties

Count of Network Locations per Confidence Levels

County	90%	95%	99%
Franklin	857	728	566
Summit	398	329	184
Ross	111	79	18

Note: The count of 95% network locations includes the 99% network locations. Similarly, the count of 90% network locations includes the count of 95% and 99% network

Table 6.3.1 shows the decrease in network locations as the confidence level increases. This indicates, that the points at the 99% confidence level may be more significant; however there may be less of them. Officers patrolling these locations may have fewer locations to pass through; however they may be most successful patrolling these locations. Officers patrolling the network locations at the 90% confidence level will have many locations to patrol, however officers may not be as efficient when patrolling these locations. The confidence level of each location represents that location's ability to reject the null hypothesis. The network locations with a higher confidence level indicate locations where there is a higher chance of intoxicated drivers being present. As a result, network locations with a higher chance of presence of intoxicated drivers often occurs less than network locations with a lower chance of presence of intoxicated drivers. This methodology is similar to how Johnston & Ramachandran (2014) identified statistically significant hot and cold spots from LISA analyses, and will be used to improve the efficiency of officers being able to locate intoxicated drivers.

One solution in determining where officers should patrol for intoxicated drivers would be to send them to all significant locations in each county, in order to locate as many intoxicated drivers as possible. However funding is often limited for state agencies, therefore it may be more efficient for officers to patrol only the more significant locations. By having fewer locations to patrol, less manpower may be required to cover these locations in a shift. In order to justify officers patrolling only the network locations in the 95% or 99% confidence levels, relationships between the three levels of network locations are created.

The first relationship observed is the amount network locations at the 95% and 99% confidence level compared to the amount network locations at the 90% confidence level. This relationship may be seen in Table 6.3.2. If for example, the percent of 99% confident network locations was very large, it may be acceptable to allow patrol officers to patrol only the 99% confident network locations, since they will be going to most locations and have the highest chances of stopping intoxicated drivers.

Table 6.3.2: Percent of 95% and 99% Confident Network Locations Compared to the Amount of 90% Confident Network Locations

Count and Percent of Network Locations per Confidence Levels

County	90%	95%	Percent of 90	99%	Percent of 90
Franklin	857	728	85	566	66
Summit	398	329	83	184	46
Ross	111	79	71	18	16

Note: This table shows the number of network locations in each confidence level as well as the proportion of 95% and 99% confident network locations that are included in the 90% confident network

The 99% confident network locations account for less than half of the total significant network locations for Ross and Summit counties. If officers were to patrol only the 99% confident network locations they will fail to patrol over half of the total significant locations where intoxicated drivers are likely to be present, however the locations they do patrol will have the highest chance of a presence of intoxicated drivers. With less network locations to patrol, less officers may be needed to patrol these locations during a given shift time. Using the 99% confident network locations may be the best economic practice for police stations in some counties, spending less money on manpower (less man-hours worked as a result of less officers needed to patrol the network locations) and having the most accuracy on locating the intoxicated drivers. However in some counties the 99% confident network locations may not cover enough area for officers to be efficient in patrolling for intoxicated drivers, whereas the 95% or 90% confident network locations may provide more coverage for officers to patrol.

The 95% confident network locations account for over 70% of the total network locations for all three counties. With more locations to patrol, officers will be more accurate in covering the significant locations. Since these locations are 95% confident, there remains only a 5% chance that any location is not significant, so the chances of locating intoxicated drivers remains high. However, since there are more significant network locations, more officers may be needed to patrol these locations in a given shift time. As a result, the cost of patrolling these locations may be increased.

The second relationship observed to determine if it may be acceptable to use the 95% or 99% confident network locations is the coverage of each group of network locations. As seen in Figure 6.3.5, Figure 6.3.6, and Figure 6.3.7, the network locations are mainly dispersed around very few central locations. If the 99% or 95% confident network locations appear to be covering most of the significant locations, it may be acceptable to allow officers to patrol only those locations, as opposed to patrolling all significant locations. However, if the network locations of higher confidence levels fail to cover a large portion of the significant areas, it may be more useful for officers to patrol a lower confidence level for the purpose of stopping more intoxicated drivers. Figure 6.3.2, Figure 6.3.3, and Figure 6.3.4 shows the network locations for each confidence level for Franklin, Summit, and Ross counties. The identified section of each county in Figure 6.3.2, Figure 6.3.3, and Figure 6.3.4 is shown in Figure 6.3.5, Figure 6.3.6, and Figure 6.3.7.

Figure 6.3.5: Significant Network Locations for Franklin County for 90% confidence levels (left), 95% confidence levels (middle), and 99% confidence levels (right), zoomed in.

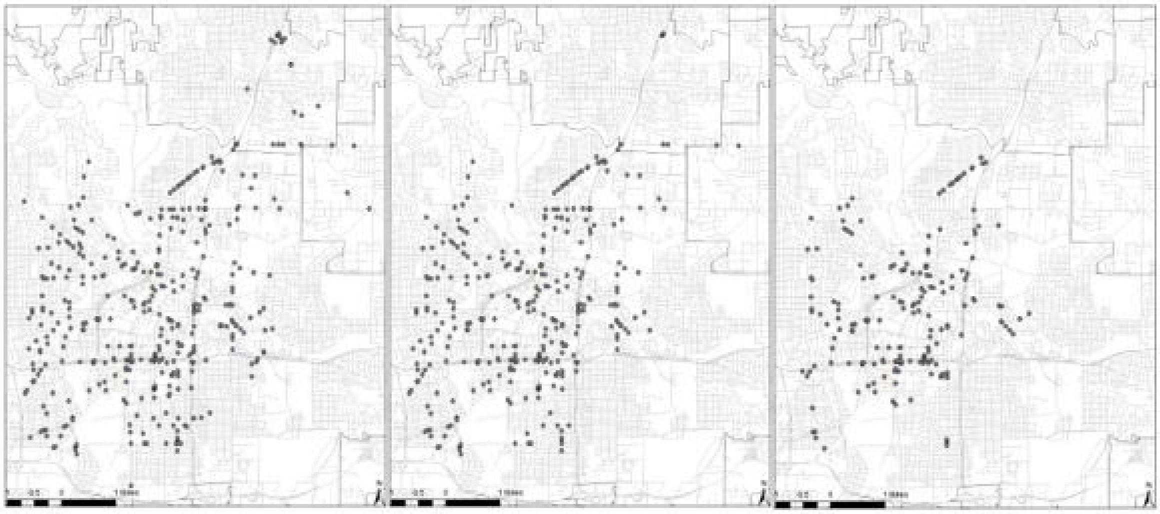


Figure 6.3.6: Significant Network Locations for Summit County for 90% confidence levels (left), 95% confidence levels (middle), and 99% confidence levels (right), zoomed in.

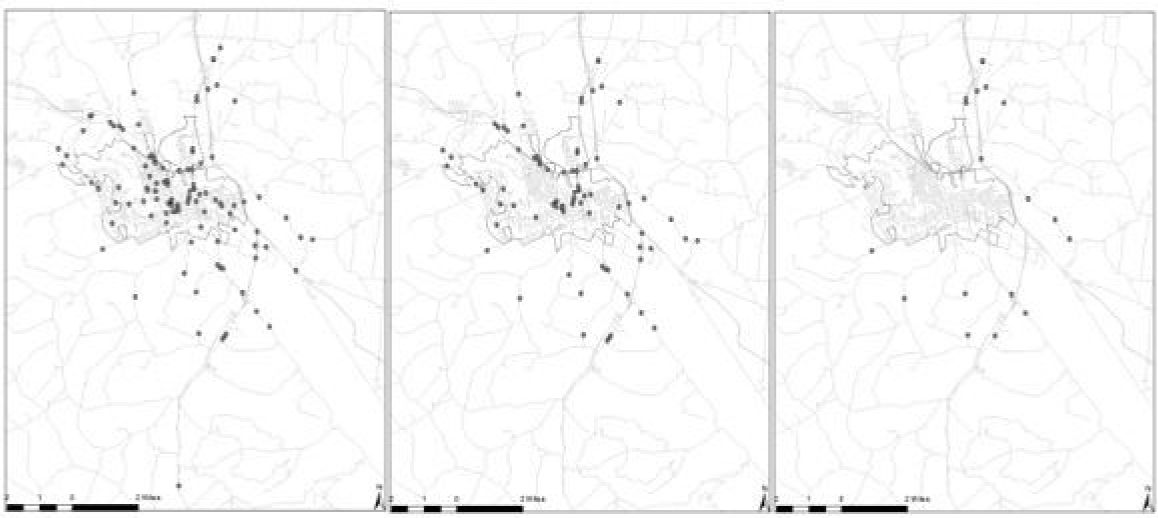


Figure 6.3.7: Significant Network Locations for Ross County for 90% confidence levels (left), 95% confidence levels (middle), and 99% confidence levels (right), zoomed in.

The 99% confident network locations for Franklin, Ross, and Summit counties represent 66%, 16%, and 46% of the total significant network locations, respectively (as seen in Table 6.3.2). The 99% confident network locations in Franklin County are greatly reduced, especially when noticing the top of area 2, as well as all of area 3. The 99% confident network locations for Summit County are located only in the City of Akron, instead of in surrounding suburbs of Tallmadge and Cuyahoga Falls. Ross County reflects a similar trend to Franklin County noticing that network locations are not present at all within the City of Chillicothe. Patrolling the only 99% confident network locations for these three counties may reduce chances of locating intoxicated drivers where their presence is still significant. The network locations for

Franklin, Summit, and Ross counties with a 95% confidence level represent 85%, 71% and 83% of the total significant network locations, respectively (as seen in Table 6.3.2). This means that officers patrolling the 95% confident network locations may cover most of the majority of the total significant areas, allowing officers increased effectiveness when patrolling for intoxicated drivers. The 95% confident network locations for Franklin County remain largely present in all identified areas. These trends are similar for both Summit and Ross Counties. The number of 95% confident network locations is reduced, however officers patrolling these locations will continue to patrol each city as well as surrounding suburbs.

Patrolling the 95% and 99% confident network locations for each county may have a reduced cost when comparing to patrolling for all three confidence levels. There are fewer network locations in the 95% confidence level than the 90% confidence level, indicating fewer officers may be required to patrol these locations in a given shift time. The decrease in the amount of officers may reduce the cost of patrolling; however, chances of stopping intoxicated drivers at these locations remain high since there is either a 5% or 1% chance of each location not being significant. Patrolling the 99% confident network locations may yield the most reduced costs. The 99% confidence level has the least amount of network locations for each county, meaning there may be less officers required to patrol the significant areas for a given shift time, as well as less equipment required to cover the patrols. The decrease in officers and equipment indicate less man-hours and reduced costs. Despite the reduced amount network locations, officers have the greatest chance of locating intoxicated drivers at the 99% confidence level.

The third relationship to determine the confidence level for officers to patrol uses radii of different lengths around the 95% and 99% confident network locations. If a large amount of network locations in the 95% or 99% confidence levels are within a specific distance from the 90% confident level network locations, it may be appropriate to justify using the 95% or 99% confident network locations for officers to patrol. The different lengths of radii used may be seen in Table 6.3.3, which represent the interests of different professionals using this application. For example, someone who models crashes may need to be very detailed in locating each crash. Miller and Karr (1998) express the concern of the location of crashes, and how these locations are important in the modeling after accidents. As a result, the end user may want a shorter radius to compare crashes, such as 0.01 or 0.05 miles. However, a police officer who is looking to patrol significant locations within their jurisdiction may not necessarily need to go through every back and side road to locate the intoxicated drivers. Giving patrol officers a radius of 0.1 to 0.2 miles may be more forgiving for the application of locating intoxicated drivers. ArcGIS is used to locate all network locations for each radius with the “select by location” tool. By using this tool with the selection method set at “within a distance,” a buffer is created around the selected layer (95% or 99% confidence level network locations) at the specified radius to select all the 90% confident network locations within the specified radius. The proportion of 90% confident network locations that are located within each radius of the 95% and 99% confident network locations may be seen in Table 6.3.3.

Table 6.3.3: Proportion of 90%Confident Network Locations in Each Radius of 95% and 99% Confident Network Locations

Proportion of 90% Confident Network Locations in Each Radius around the 99% Confident Network Locations							
County	0.9	Radius Distance (miles)					
		0.01	0.05	0.1	0.15	0.2	0.25
Franklin	857	66%	67%	69%	71%	76%	76%
Summit	398	46%	48%	51%	55%	58%	63%
Ross	111	16%	16%	17%	17%	18%	18%
Proportion of 90% Confident Network Locations in Each Radius around the 95% Confident Network Locations							
County	0.9	Radius Distance (miles)					
		0.01	0.05	0.1	0.15	0.2	0.25
Franklin	857	85%	86%	88%	89%	91%	92%
Summit	398	83%	84%	85%	88%	90%	92%
Ross	111	71%	72%	73%	73%	74%	74%

Note: This table shows the proportion of 90% confident network locations within each radius of the 95% and 95% confident network locations.

Again, these tables represent the proportion of 90% confident network locations within each radius of the 95% and 99% confident network locations. The proportions for the 99% confident network locations are much lower than 95% confident network locations. This is a result of the lesser amount of 99% confident network locations, however this may also be indicative of which confidence level to use. Ross County shows the lowest relationships for both confidence levels, with a maximum proportion of 90% confident network locations located anywhere near the 99% confident network locations equal to 18%. Meaning if officers were to patrol locations within 0.2 miles of the 99% confident network locations, they will only pass through 18% of the total significant areas of Ross County. Although the 99% confident network locations are the most significant, with the highest chances of a presence of intoxicated drivers, the 90% and 95% confident network locations are still significant. Essentially 82% of the total significant area in Ross County will be “missed” by patrol officers, leaving room for people to drive intoxicated without getting caught. Similarly, Summit County has a fairly low relationship of 90% confident network locations to 99% confident network locations. Officers responsible for locating intoxicated drivers in Summit County will patrol 58% of the total significant areas when accounting for a 0.2 mile radius, leaving 42% of the total significant areas without patrols.

The comparison of 90% confident network locations to 95% confident network locations in Table 6.3.3 appears to be much more reliable in patrolling for intoxicated drivers. Ross County officers patrolling the 95% confident network locations have a minimum of 71% coverage of the total significant area (compared to the 16% for the 99% confident network locations), which then increases up to 74% for a larger radii. This allows officers to patrol much more of the significant area, resulting in greater chances of locating intoxicated drivers. Franklin and Summit County have very similar results with a minimum of 85% and 83% coverage, respectively, of the total significant area. As the radius increases, the coverage increases up to 92% for both counties, leaving only 8% of the total significant area untouched by patrol officers. Officers patrolling the 95% confident network locations for Franklin and Summit Counties may be much more accurate in locating intoxicated drivers, while keeping the amount of network locations reduced from the 90% confidence level. Sending officers to patrol the lesser locations of the 95% confidence level, which are also more significant than the 90% confidence level, may significantly reduce the cost of manpower and time it takes to patrol these locations.

6.4 CONCLUSION

This research shows statistically significant locations that may be used by patrol officers to reduce the amount of alcohol-related crashes by locating areas where intoxicated drivers are likely to drive. Officers may improve the effectiveness in their saturation patrols by patrolling these locations. Given local indicators of spatial association with three different levels of significance, individual points are located that show significant areas where intoxicated drivers are likely to be present. This research was able to narrow down the total significant locations to provide officers with the most significant areas to patrol while accounting for limited budgets.

The 95% confident network locations appear to be the most reliable network locations to use when patrolling for intoxicated drivers. Every point in these locations has either a 5% or 1% chance of not being a significant location where intoxicated drivers are present. Although the amount of these locations is reduced, they are present throughout most significant areas in each county. Since the number of 95% confident network locations is lower, the amount of manpower required to patrol these locations in a given shift time may be less than the amount required covering the total amount of significant locations. These locations also account for the two most significant confidence levels, indicating that the chance of locating intoxicated drivers is higher. Sending officers to patrol the 95% confident network locations may yield increased effectiveness in the ability to locate more intoxicated drivers, as well as increased efficiency in spending the funding provided by the State of Ohio.

It has been proven that the use of enforcement strategies, specifically high visibility and saturation patrols, produce reductions in alcohol-related crashes. This research shows that sending officers to patrol 95% confident network locations may have the most effective results in locating intoxicated drivers. Additionally, these network locations remain significant while considering the limited funding of jurisdictions.

CHAPTER 7: COMPARISON OF TRADITIONAL CORRIDOR BASED ENFORCEMENT WITH ROUTE OPTIMIZATION OF HOT SPOT ANALYSIS

7.1 INTRODUCTION

In 2014, 31% of the all fatal crashes in the state of Ohio involved at least one alcohol-impaired driver (NHTSA, 2016). The trend of alcohol-related crashes and alcohol-related fatalities is a problem not only in the state of Ohio, but one which faces the nation as a whole. From the late 1980's through the early 1990's, there was a steady reduction in the percentage of alcohol-related fatalities. Since 1994, the rate has steadily remained above roughly 30%. While a large scale study completed in 1988 by Moskowitz and Robinson (1987) found the effects of alcohol on driving related skills, improvements in the coming years were a result of joint efforts between whole communities, engineers, and law enforcement, and many studies outlining advances in methods to reduce alcohol-related crashes and fatalities.

Studies on the geospatial patterns and trends of alcohol-related crashes are relevant in determining methods to reduce the amount of alcohol-related crashes. In 1983, Colón and Cutter used multiple regression to determine relationships between motor vehicle accidents and alcohol availability (Colón and Cutter, 1983). Several studies followed in determining spatial relationships between alcohol-related crashes to help refine national and regional trends in alcohol-related crashes (Gary et al., 2003; Ponicki et al, 2013).

In addition to geospatial work, there have been several studies which focus on law enforcement and improving their effectiveness as crash countermeasures. Dedicated impaired driving saturation patrols are identified by NHTSA as highly effective, easily implementable proven countermeasures (Goodwin et al., 2015). While the effectiveness of this practice has been proven, there have been several recent studies focused on improving the efficiency and implementation of these patrols in varying capacities and situations (Fell et al., 2014; Sanem et al., 2015; Maistros and Schneider IV, 2016, accepted, not yet published). Another important aspect of impaired driving patrols is the specific type of patrol that is conducted. Different patrols options include corridor enforcement, where officers stick to specific roads that are over represented in alcohol-related crashes, and saturation patrols, during which officers work specific areas that are over represented in alcohol-related crashes. Of the several methods for defining high crash areas for saturation patrol, Data Driven Approaches to Crime and Traffic Safety (DDACTS) is a method developed by NHTSA which focuses on the use of spatial clustering. Location based crashes may be used in order to find spatial relationships, commonly known as hot spots, which show areas of significance. Other papers have gone on to use spatial clustering as a means to define route points (Maistros and Schneider IV, 2016).

The objective of this chapter is to compare traditional corridor enforcement patrols to new route optimized segments using points defined through hot spot analyses. This is completed through the use of Esri's ArcGIS Vehicle Routing Problem. Once the routes for each method are modeled, they will be compared by locating the total amount of alcohol-related crash locations passed by each route. Officers passing through more alcohol-related crash locations per mile or time may be more effective in ultimately reducing the amount of alcohol-related crashes (Fell et al., 2014).

7.2 DATA

The data sources for this study include crash records populated from the state of Ohio OH-1 crash reports (ODPS, 2015), the Ohio Department of Transportation (ODOT) Geographic Information Systems (GIS) roads layer (ODOT, 2016), and United States Census estimates information (United States Census Bureau, 2015). Using these three databases, the research team selected Franklin, Summit and Ross counties for analysis. All alcohol-related crashes in each county from January 1, 2012 through April 9, 2015 are included in the analysis.

Franklin County was selected due to its high population (greater than 1 million people), it encompasses the large metropolitan area of the City of Columbus, and its high number of alcohol-related crashes. Summit County was selected due to its large urban areas with a population greater than 500,000 people and a significantly high number of alcohol-related crashes. In contrast, Ross County is a predominantly rural county with a total population less than 100,000 people, and has a road network a quarter the size of Franklin County. Historically, Franklin County and Summit County are both in the top 10 counties statewide with the highest number of alcohol-related fatalities per year. A summary the general characteristics of each of the three counties may be seen in Table 7.2.1.

Table 7.2.1: Comparison of Franklin, Summit, and Ross Counties

County Comparison

	County Population	Population Density (per sq. mi.)	Lane Miles	Total Alcohol-Related Crashes
Franklin	1,251,772	2,186	5,670	4,051
Summit	541,968	1,313	3,608	1,805
Ross	77,170	113	1,429	334

Note: County Population and Population Density determined from United States Census Bureau Quick Facts.
Lane Miles determined from Ohio Department of Transportation ArcGIS Roads Layer.
Total Alcohol-Related Crashes determined from Ohio Department of Public Safety OH-1 Crash Reports.

As shown in Table 7.2.1, there is a wide range of demographics, road networks and alcohol-related crashes between these three counties. This chapter will use these counties as case studies as a demonstration for a new methodology that law enforcement agencies may use to help curb alcohol-related crashes.

7.3 METHODOLOGY

The DDACTS initiative that was developed by NHTSA encourages local law enforcement to develop data driven approaches to address high crime and crash areas. Within traffic enforcement, agencies may use various strategies which may include sobriety checkpoints, saturation patrols, and multi-jurisdiction, which are multi-agency short-term high visibility details.

This chapter focuses specifically on the improvement of saturation patrols which may be used during high visibility enforcement overtime (HVEO) patrols. In Ohio, Franklin and Summit counties qualify for additional state funding for impaired driving HVEO programs due to their high numbers of alcohol-related fatalities. Traditionally in Ohio, saturation patrol allows the local agency the maximum flexibility as to when and where to implement their enforcement detail. While the flexibility of generic saturation patrols allows officers to work a variety of areas, it does not always provide the direction and guidance to address the greatest occurrences of impaired driving.

7.3.1 Methodology One: Corridor Enforcement

One approach to improve saturation patrol is through the development of corridor specific routes. The fundamental difference between saturation and corridor patrols is corridor enforcement requires the officer to patrol one specific segment of road. This segment is typically defined by either the number of alcohol-related arrests or the number of alcohol-related crashes. Typically an agency will select the top 5, 10 or 15 roads with the highest numbers of alcohol-related crashes and will only patrol these segments. In this chapter, the corridors for the three counties are defined as the 15 road segments with the highest number of alcohol-related crashes within each county.

7.3.2 Methodology Two: Route Optimized Hot Spot Analysis

The second methodology used within this chapter is the development of new route optimized segments using hot spot analysis (HSA). HSA is a robust method of cluster analysis which aggregates crashes based on location and statistical significance. Crashes that lie within a defined distance of one another are combined so to for a new aggregate point halfway between the two original crashes with a count value representing the two original crashes. This process is repeated until there are no more crashes within the re-defined distance of the original crash point. Once all crash locations are analyzed for aggregation, local and global cluster analysis is performed. Local clustering analysis determines the significance of the clustering at an individual point as described by Ord and Getis (1995), Wulder and Boots (1998), Prasannakumar et al. (2011), and Lees (2006). Global clustering is performed to determine if the aggregate points are statistically clustered in relation to one another within the study area. For this study, the area is defined by the three counties for which the analysis is performed. HSA has proven to be effective in the crash identification and in directing the efforts of law enforcement (Carrick et al., 2014; Maistros et al., 2014; Ratcliffe and McCullagh, 2011).

With the improvement of geospatial crash locations, HSA continues to be refined and accepted by practitioners and researchers. One problem with HSA is it only defines the problem area, it does not provide guidance on how to best patrol within that area. To help address this limitation, this chapter will utilize the HSA analysis and will develop a route optimization model within the HSA. The hot spot route optimization (HSRO) model is developed first by defining the network locations or desired areas for patrol. For this chapter, the network locations are determined by the output of the HSA. Anselin's 1995 study (Anselin, 1995), explains how the Gi and Gi* statistics from the results of HSA may be used to identify individual points, or network locations, within the output of the HSA known as local indicators of spatial association. These individual network locations are able to give a statistical significance to each output on the hot spot map. Various levels of significance are often identified to show the confidence that any output may reject the null hypothesis, which is that the hot spot is not significant. These levels include 99% confidence, 95% confidence, and 90% confidence, or showing no significance.

Since the network locations in this chapter are based off alcohol-related crash locations, a point with a 95% confidence indicates that there is a 5% chance that location does not represent a location where alcohol-related crashes may typically occur significant. Similar relationships may be seen from De Vlack et al. (2016), who explain the significance of hot and cold spots when analyzing the substitutability of recreation areas in Belgium. Additionally, Johnston & Ramachandran (2014) show how the use of local indicators of spatial association are able to reject the null hypothesis of homogeneity, while also utilizing these data for their focus of stated preference welfare estimates. For the purposes of this chapter, local indicators of spatial association that are at or above 95% confidence are identified and utilized as network locations to be used in the route optimization model. These points may be seen as the black dots in Figure 7.3.1. When patrolling the 95% confident network locations, each of those points have a 95% chance that the alcohol-related crash location is significant, giving officers a greater chance of locating intoxicated drivers.

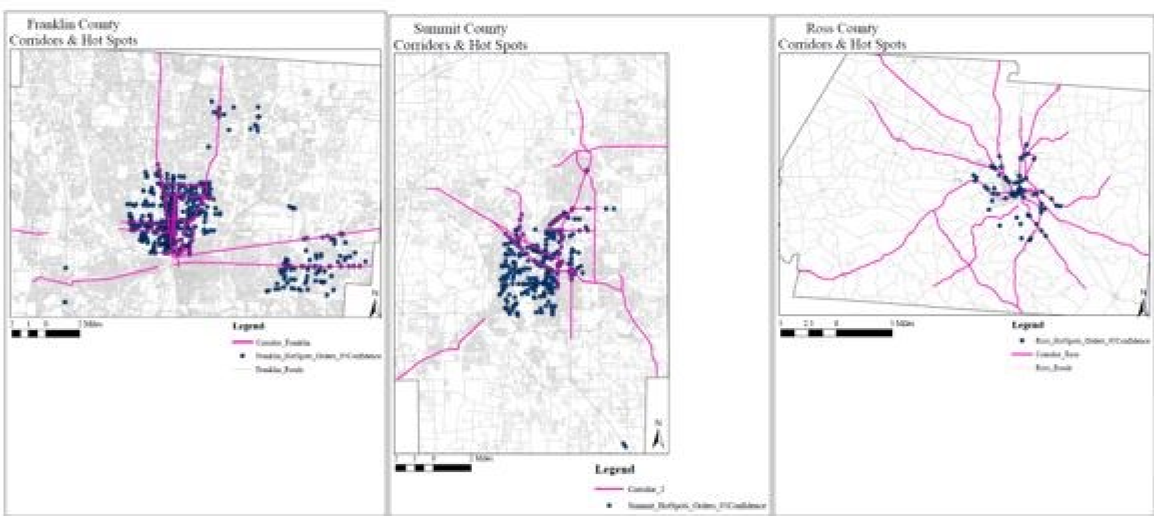
Once the network locations, seen as the black dots in Figure 7.3.1 are determined Esri's ArcGIS Vehicle Routing Problem may be used to create a HSRO model based on crash locations. In this study, the 95% confident local indicators of spatial association are utilized as the network locations. These are the areas that the patrol routes for officers will be guided to while patrolling for intoxicated drivers. At these locations, there is a 5% chance that each area is not significant, allowing for a 95% chance that the point is significant and officers may be more successful in locating intoxicated drivers. Additionally by patrolling these locations, officers may be able to show more of a presence in areas where intoxicated drivers may be more likely to drive. Fell, et al. (2015) explains how less people are likely to drive intoxicated in communities where enforcement of intoxicated driving is more prevalent. Patrol cars are also added into the system. The patrol cars are modeled so they are able to travel through any part of the county.

To determine the results, the model must first route the patrol cars through the identified corridors. The total time that is calculated for each individual route is then used as a restriction for the HSRO model. By restricting the total time for the HSRO model, each individual patrol car will travel the same amount or less time than when patrolling the corridors. Restricting the total time of each route allows for a similar comparison between the current method of corridor enforcement and the new method of HSRO.

7.3.3 Comparison of Corridor Enforcement and HSRO Analysis Parameters

As stated previously, the main goal of this chapter is to compare two methodologies for improving alcohol enforcement. In order to achieve this main goal, the two methodologies, as best as possible, must use similar parameter constraints such as total travel time, total time, and number of patrol cars. In this chapter the corridor enforcement model, methodology one, is implemented first. The results, travel time, total time, and number of patrol cars from the corridor enforcement methodology are then used as parameter inputs for the hot spot route optimization model. The parameters for both models are determined by the corridors, and they must match. This is to allow for a similar comparison between the current method of corridor enforcement and the new method of hot spot route optimization.

Figure 7.3.1, is developed using similar parameter constraints between the two methodologies. This figure shows the top 15 corridors per county as well as the 95% hot spots network locations.



Note: The dots are based on the output of the HAS, showing 95% confident local indicators of spatial association. Additionally, the lines represent the top 15 corridors in each county, as previously identified.

Figure 7.3.1: Corridor & Hot Spot Patrol Areas for Franklin, Summit, and Ross Counties

Figure 7.3.1 shows the base relationship of the HSRO (dots) versus the top 15 corridors representing roads with the most alcohol-related crashes for each county. The corridors will first be modeled in order to determine what officers may currently be patrolling. The results from the corridor routes are then used to model the statistically significant HSRO. The final product is a map of the corridor and HSRO methods showing the routes for when three, five, and seven patrol cars are deployed, each route depicted from a different color gradient. The corridors are shown first in order to view what officers may currently be patrolling, followed by the HSRO to show the new method of what officers may patrol to further reduce the amount of alcohol-related crashes.

7.4 RESULTS

The final results of the map include total time, total travel time, and total miles. It may be noted that there is a clear difference between the total time and total travel time. The total travel time indicates how long the officers are patrolling each route without any stops. Similarly, the total miles show the amount of miles each officer drives on their given route. The total time incorporates a stop time, which simulates the pullovers officers make. In order to simulate this, a service time is added onto each network location, the summation of which represents the amount of time officers are pulled over per hour. As a result, the total cycle time for each patrol route is a realistic vision of what may occur in the field. Assuming that patrol officers generally have 1.5 stops per hour and an average stop time for each pullover may be set at 15 minutes. The service time is averaged over the total amount of network locations in each county, and may be seen in equation 7.1:

$$t_s = \frac{(\lambda_p)*(t_p)}{N_l} * N_p \tag{7.1}$$

where t_s represents the service time, λ_p represents the rate of pullovers, t_p represents the average time of pullovers, N_l represents the number of network locations per county, and N_p represents the number of patrol cars in given scenario. For example, if one officer were to patrol their route and it took one hour to patrol, the total service time may represent about 22.5 minutes (for 1.5 stops at exactly 15 minutes each). However, this time may vary since stops are not always exactly 15 minutes each and officers may not always pullover exactly 1.5 people per hour. The total travel time and total miles are used to determine the total amount of alcohol-related crash locations passed per mile of travel as well as the amount of alcohol-related crash locations passed per total time of travel. The final corridor and hot spot patrol maps of Franklin County may be seen in Figure 7.4.1. Additionally, the total times, total travel times, and total miles for each patrol in each scenario may be seen in Table 7.4.1.

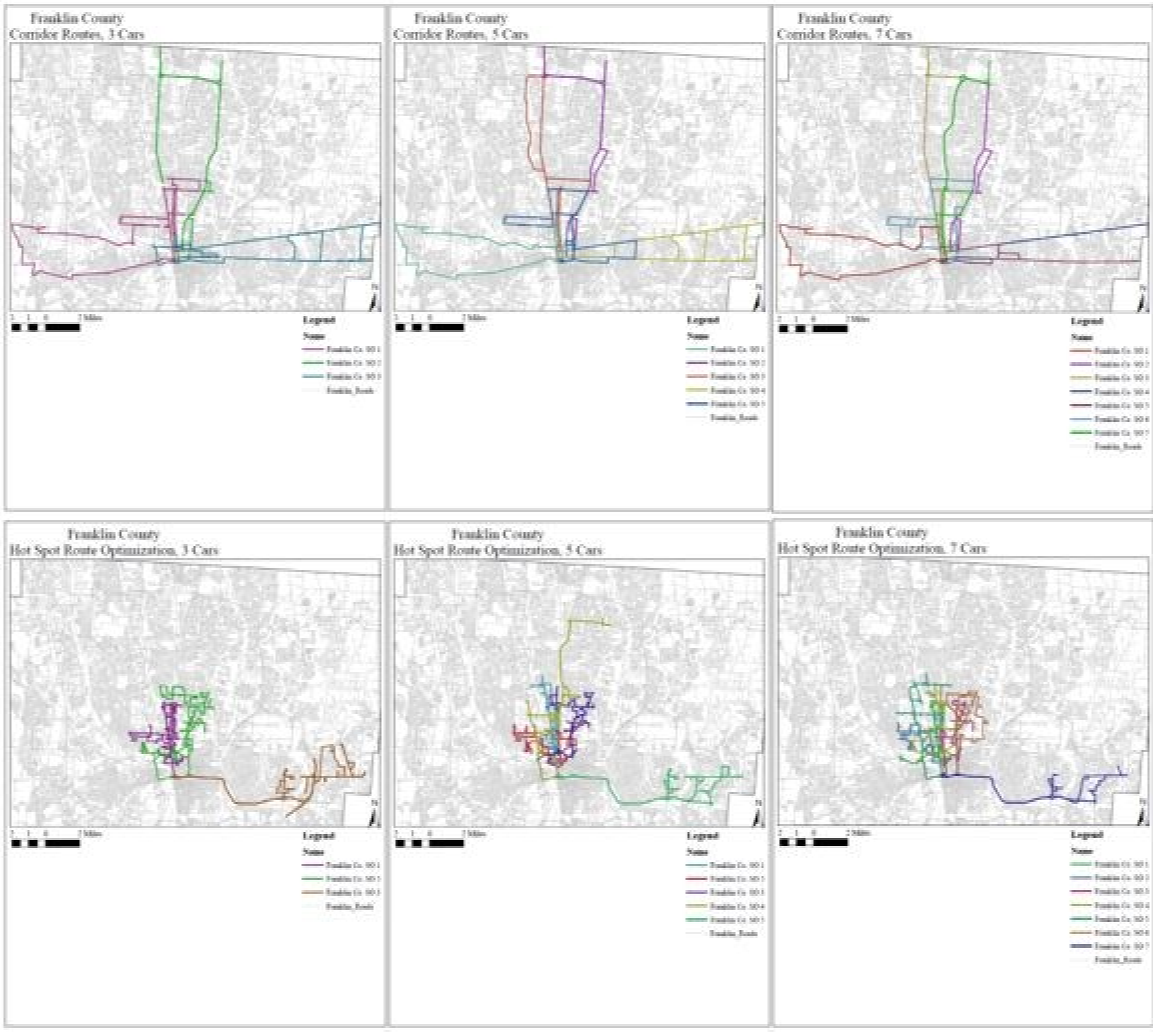


Figure 7.4.1: Optimized Corridor Routes (top) and Hot Spot Routes (bottom) for Franklin County for 3, 5, and 7 patrol cars respectively.

Table 7.4.1: Total Miles, Total Time and Total Travel Time for Franklin County when 3, 5, and 7 Patrol Cars are Deployed

		3 Cars			5 Cars			7 Cars		
		Miles	Total Time	Travel Time	Miles	Total Time	Travel Time	Miles	Total Time	Travel Time
Franklin County Corridor Route Results	Car 1	50	106	83	29	80	60	28	69	43
	Car 2	51	98	78	38	72	47	25	69	45
	Car 3	48	110	83	30	80	60	30	62	44
	Car 4	-	-	-	36	76	49	27	65	44
	Car 5	-	-	-	31	65	44	27	58	34
	Car 6	-	-	-	-	-	-	22	68	46
	Car 7	-	-	-	-	-	-	34	66	43
Franklin County HSRO Results	Car 1	43	109	82	28	80	54	23	64	36
	Car 2	48	104	86	32	80	58	25	63	42
	Car 3	55	93	85	34	80	60	27	67	46
	Car 4	-	-	-	41	76	61	28	69	46
	Car 5	-	-	-	41	70	60	29	64	46
	Car 6	-	-	-	-	-	-	28	68	47
	Car 7	-	-	-	-	-	-	34	57	47

Note: The miles and speed of travel used to find the Miles of each route and Travel Time are calculated based on the ArcGIS Ohio Roads Layer used in the model. The Total Time incorporates an assumed stopping time at each network location in order to account for the rate of 1.5 pullovers per hour.

The optimized corridor routes have a maximum total time of 110, 80, and 69 minutes for three cars, five cars, and seven cars, respectively. These travel time inputs were used as the maximum time restrictions for the HSRO model. The final HSRO model requires the patrol to go through as many network locations as possible in the given time restrictions from the corridor routes. Franklin County has similar results between the corridor and HSRO method. Since the corridor times are restricting the HSRO model, the HSRO model utilizes as much of the time as possible in order to pass through the most network locations. This ensures that officers have the optimum amount of presence in traveling through significant locations when patrolling for intoxicated drivers.

The overall comparison between the two methodologies show that the corridor enforcement is spread out throughout the entire county while the HSRO model enforcement detail is more localized through the metropolitan areas. It may also be noted that the location of the routes in the HSRO vary for each amount of patrol cars modeled. This is due to the time restrictions from the corridor routes not allowing the patrol cars in the HSRO model to travel through every area that the hot spot locations are present. This may be beneficial as it does not require patrol officers to drive throughout the entire county. Instead, officers may stay in few central locations while continuing to have the same effect on presence and effectiveness of patrolling for intoxicated drivers.

The same methodology is repeated for Summit County (Figure 7.4.2) and Ross County (Figure 7.4.3). Similar trends may be seen in both counties where the corridors often span throughout the county and the hot spot locations, based off the statistically significant areas of hot spots, are located in one general area. The optimized corridor and hot spot patrol routes for Summit County may be seen in Figure 7.4.2. Additionally, the total times, total travel times, and total miles for each patrol in each scenario may be seen in Table 7.4.2.

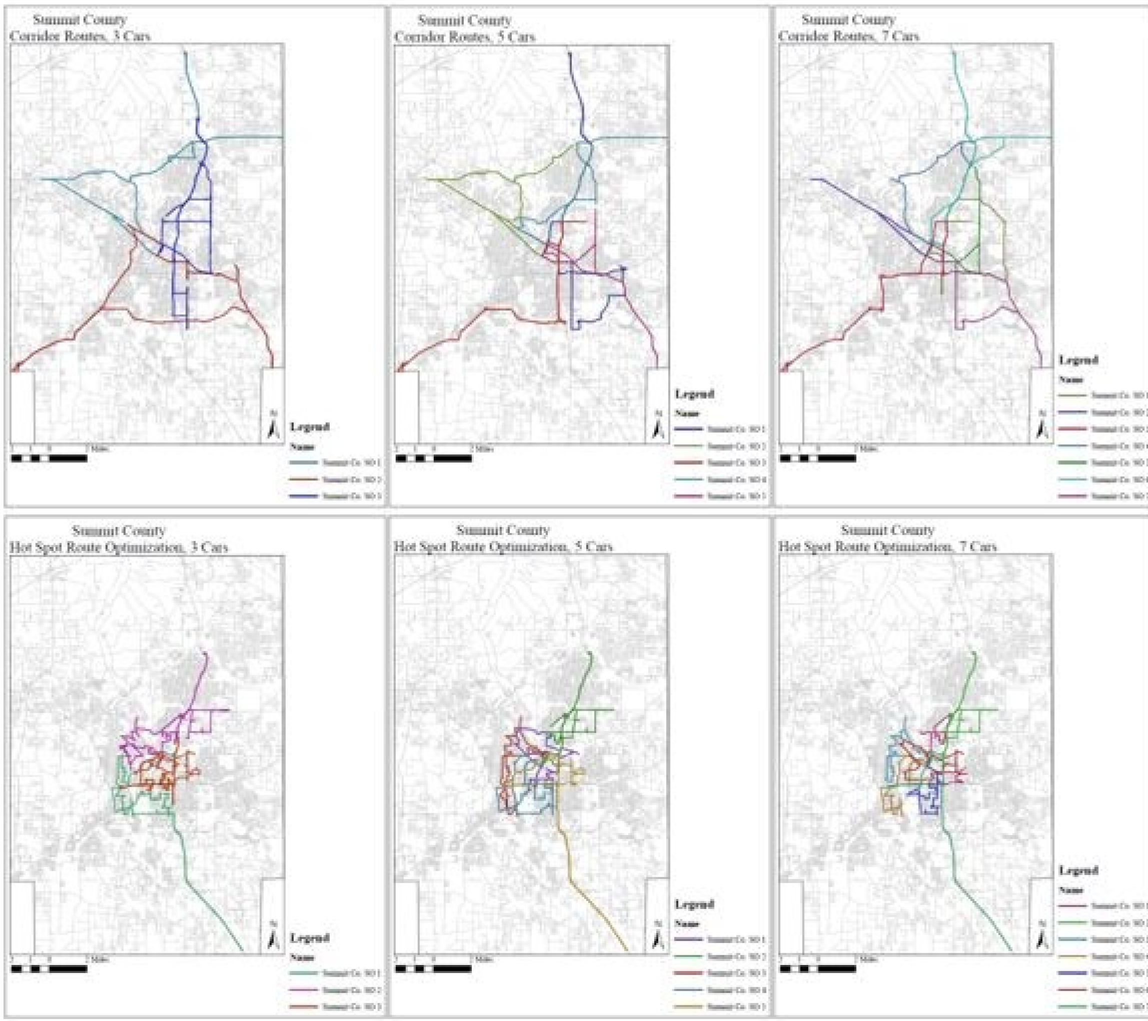


Figure 7.4.2: Optimized Corridor Routes (top) and Hot Spot Routes (bottom) for Summit County for 3, 5, and 7 patrol cars, respectively.

Table 7.4.2: Total Miles, Total Time and Total Travel Time for Summit County when 3, 5, and 7 Patrol Cars are Deployed

		3 Cars			5 Cars			7 Cars		
		Miles	Total Time	Travel Time	Miles	Total Time	Travel Time	Miles	Total Time	Travel Time
Summit County Corridor Route Results	Car 1	48	96	76	40	70	48	27	59	38
	Car 2	48	96	75	28	71	48	16	54	27
	Car 3	48	93	67	36	71	45	30	55	32
	Car 4	-	-	-	28	70	44	20	56	30
	Car 5	-	-	-	26	64	49	17	61	44
	Car 6	-	-	-	-	-	-	34	62	37
	Car 7	-	-	-	-	-	-	24	61	42
Summit County HSRO Results	Car 1	49	96	75	19	59	36	19	49	30
	Car 2	44	96	73	32	69	47	24	47	35
	Car 3	40	92	66	22	70	47	15	47	30
	Car 4	-	-	-	24	66	42	16	43	27
	Car 5	-	-	-	38	68	47	15	49	29
	Car 6	-	-	-	-	-	-	16	49	32
	Car 7	-	-	-	-	-	-	32	49	38

Note: The miles and speed of travel used to find the Miles of each route and Travel Time are calculated based on the ArcGIS Ohio Roads Layer used in the model. The Total Time incorporates an assumed stopping time at each network location in order to account for the rate of 1.5 pullovers per hour.

For Summit County, the total amount of time required for officers to patrol the HSRO model is less than what is required for the corridor routes. This occurs because the corridors span much more throughout the county than the hot spots, requiring more time to patrol the corridors than the hot spot locations. As a result, officers patrolling the routes for the HSRO model travel consistently less time and miles than the when patrolling the corridors. This may prove to be an interesting result if officers are able to pass through more alcohol-related crash locations per mile and per time in the HSRO model than the corridor model. This may be a significant result because it may indicate that officers are able to be more efficient in patrolling for intoxicated drivers.

As seen in Table 7.4.2, the maximum total time utilized in Summit County from the corridor routes is 96, 71, and 62 minutes for three, five, and seven cars respectively. Again the travel time and number of cars patrolling in the corridors are used as maximum travel time and number of cars patrolling for the HSRO model. This is applicable for when three patrol cars are deployed, however when five patrol cars and seven patrol cars are deployed, the maximum amount of time needed for the hot spot patrols is less than the maximum amount of time allowed by the corridor patrols. This means that all statistically significant areas are able to be patrolled in the hot spot patrol in the same amount of time or less than the corridor patrols. If officers are able to patrol in the HSRO areas in less time than the corridor areas, while also traveling through more alcohol-related crash locations, officers may be more effective in showing a presence and stopping intoxicated drivers in less time than when patrolling corridors.

Similar results are seen for Ross County in Figure 7.4.3 while the total times, total travel times, and total miles for each patrol in each scenario may be seen in Table 7.4.3. Ross County is similar to Summit County in that the corridors tend to span throughout the whole county, while the hot spot areas are located in one central location. This again allows the hot spot patrols to consistently require less time to travel through all areas than the corridor patrols. If the hot spot patrols are able to pass through more alcohol-related crash locations per mile than the corridor patrols, it may be more beneficial for officers to conduct patrols through HSRO.

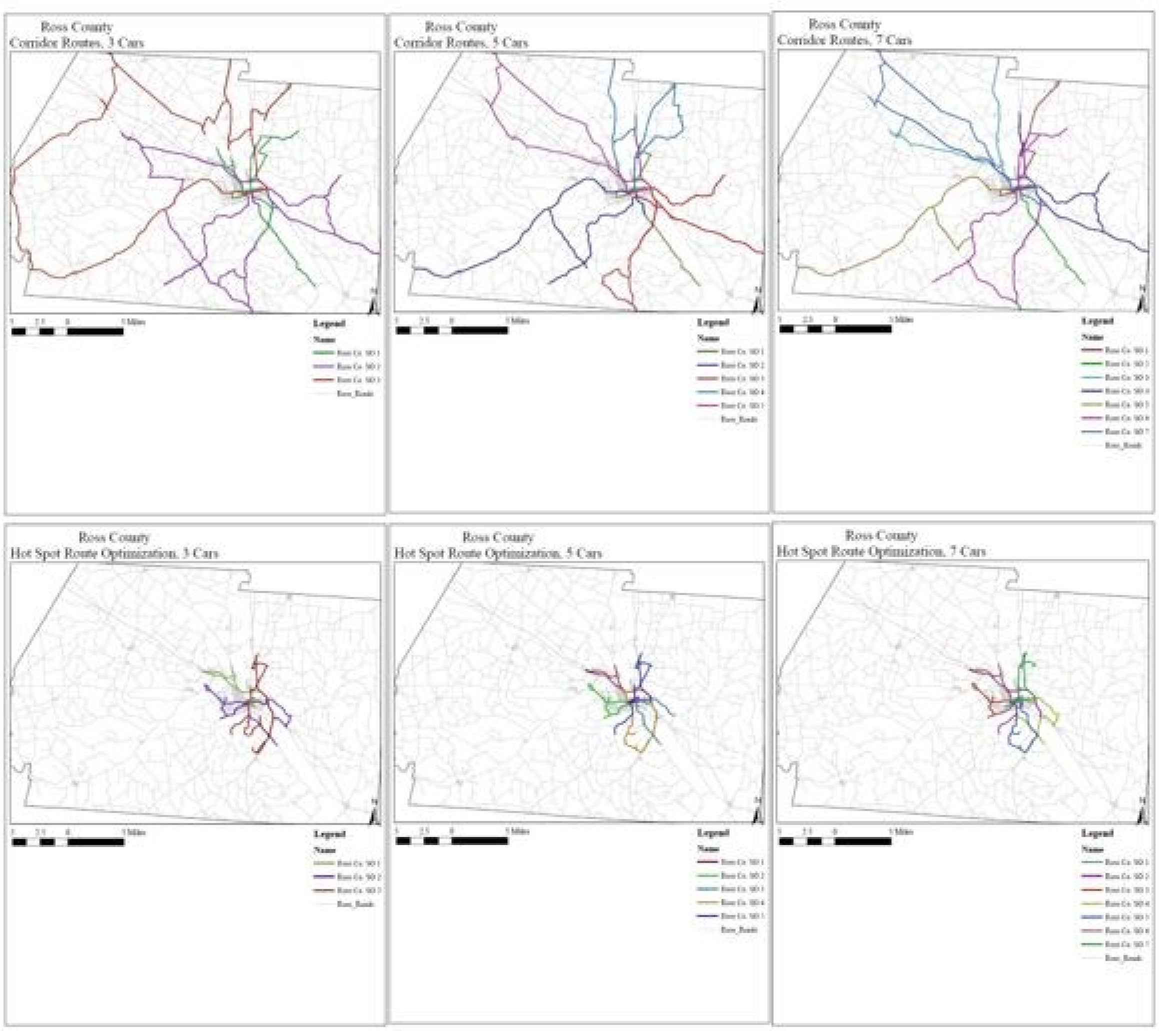


Figure 7.4.3: Optimized Corridor Routes (top) and Hot Spot Routes (bottom) for Ross County for 3, 5, and 7 patrol cars, respectively.

Table 7.4.3: Total Miles, Total Time and Total Travel Time for Ross County when three, five, and seven cars Patrol Cars are Deployed

		3 Cars			5 Cars			7 Cars		
		Miles	Total Time	Travel Time	Miles	Total Time	Travel Time	Miles	Total Time	Travel Time
Ross County Corridor Route Results	Car 1	77	162	134	43	102	73	32	104	87
	Car 2	112	204	182	68	129	105	47	99	84
	Car 3	109	205	188	74	127	105	50	104	80
	Car 4	-	-	-	52	110	91	52	101	79
	Car 5	-	-	-	56	120	102	52	97	73
	Car 6	-	-	-	-	-	-	54	91	66
	Car 7	-	-	-	-	-	-	50	87	59
Ross County HSRO Results	Car 1	29	74	41	18	56	29	5.1	48	31
	Car 2	30	57	57	14	51	29	19	50	30
	Car 3	35	59	59	21	57	34	14	49	29
	Car 4	-	-	-	19	55	33	16	50	28
	Car 5	-	-	-	22	58	38	15	48	26
	Car 6	-	-	-	-	-	-	15	48	22
	Car 7	-	-	-	-	-	-	16	41	11

Note: The miles and speed of travel used to find the Miles of each route and Travel Time are calculated based on the ArcGIS Ohio Roads Layer used in the model. The Total Time incorporates an assumed stopping time at each network location in order to account for the rate of 1.5 pullovers per hour.

As seen in Table 7.4.3, the maximum total time utilized by the corridors in Ross County is 205, 129, and 104 miles for three, five, and seven cars, respectively. When considering the HSRO model, the maximum total time needed to patrol through all network locations for three, five, and seven cars is 74, 58, and 50 minutes. The major difference comes when noticing that the HSRO model is located in the one central location while the corridors span throughout the county. When using the HSRO model to patrol, officers may be able to create much more of a presence in the significant areas of the county, as opposed to traveling through the whole county and potentially only being able to travel their respected route once or twice during a given three or four hour shift.

After the maps have been optimized for the corridor and HSRO patrols, the total miles traveled by each car, as well as the total time each patrol car took to complete its respective route is recorded. Additionally, the total amount of alcohol-related crash locations passed through by each route system is recorded. This ensures that if two patrol cars are driving along the same road for a period of time, each alcohol-related crash location passed through by each car is recorded one time only.

In each county, the total time and total miles taken to patrol the corridors and HSRO models differ. In order to determine which method may be more efficient, the total crashes per mile, as well as total crashes per total time (in minutes) are calculated. The amount of alcohol-related crash locations is currently used as a reference for patrolling for intoxicated drivers. In this case, the patrol route that is able to pass through more crash locations per mile or more crash locations per minute may be seen as the most efficient route. These results may be seen in Table 7.4.4.

Table 7.4.4: Total Crashes per Mile and Total Crashes per Total Time Comparison between Corridor and Hot Spot Patrols for Franklin, Summit and Ross Counties

		3 Cars		5 Cars		7 Cars	
		Corridor	HSRO	Corridor	HSRO	Corridor	HSRO
Franklin County	Total Crashes per Mile	5.79	5.98	5.51	5.51	4.70	5.02
	Total Crashes per Total Time (minutes)	2.75	2.85	2.42	2.51	1.98	2.15
Summit County	Total Crashes per Mile	3.07	3.48	2.61	3.44	2.74	3.42
	Total Crashes per Total Time (minutes)	1.55	1.63	1.19	1.40	1.13	1.41
Ross County	Total Crashes per Mile	0.56	1.02	0.55	1.03	0.49	1.02
	Total Crashes per Total Time (minutes)	0.29	0.51	0.27	0.35	0.24	0.31

Note: Total Crashes per Mile is found by dividing the total amount of alcohol-related crash locations passed in each fleet by the total miles traveled in each fleet. Total Crashes per Total Time is found by dividing the total amount of alcohol-related crash locations passed in each fleet by the Total Time recorded for each fleet.

Franklin County, the most urban county of the three case studies also has the most crashes in the studied time frame with 4051 located alcohol-related crashes. While patrolling the HSRO model, more alcohol-related crash locations per mile were passed through with three and seven cars than the corridor patrols, however when five cars were patrolling, the amount of alcohol-related crash locations passed is the same as the corridor patrols. Though there is no difference with five patrol cars, the amount of alcohol-related crash locations passed through per time is greater for the HSRO patrols than the corridor patrols. Additionally, HSRO patrols passed through more alcohol-related crash locations per minute for three and seven patrol cars than the corridor patrols.

These results are very similar for Summit and Ross Counties. Summit County has a total of 1,805 alcohol-related crashes. The total alcohol-related crash locations passed per mile is significantly higher for the HSRO as well as the total crash locations passed per total time. With this knowledge and the knowledge that the amount of the routes for the HSRO model are shorter for five and seven cars in Summit County, it may be viable to conclude that officers may be more efficient in patrolling using the HSRO model.

Ross County, with 331 total alcohol-related crashes again shows similar results to Summit County. The total amount of alcohol-related crash locations passed per mile in the HSRO model is at least doubled for each amount of officers patrolling than the corridors. Additionally the total amount of alcohol-related crash locations passed per time is greater for the HSRO model than the corridor in each situation. Ross County is very similar to Summit County in that each result for the HSRO model shows either less time patrolling or more alcohol-related crash locations passed by the patrols. Given these results, patrol officers are able to travel less distance and less time for HSRO patrols while covering not only more alcohol-related crashes, but also areas that have been found to be statistically significant from the hot spots.

When studying the results of Franklin County, the results are in favor of HSRO patrols in eleven out of twelve study areas. Similarly, the results of Summit and Ross Counties are in favor of the HSRO patrols in all twelve of the study areas. Fewer miles are traveled, and more alcohol-related crash locations are passed per mile and minute of patrolling. Using Franklin County as a representative of urban counties, and Ross County as a representative of rural counties, it may be possible to conclude that HSRO patrolling may be the most effective method when patrolling for intoxicated drivers.

Ultimately, the results for the HSRO method of patrolling are favored. When considering the amount of crashes passed per mile, this means that officers patrolling the HSRO model will pass more crash locations in fewer miles than the corridor routes. Additionally, the results show that officers will pass more alcohol-related crash locations in less amount of time for the HSRO method than the corridor method. By not only traveling through more alcohol-related crash locations, but also the statistically significant locations, as defined by the hot spots, officers may be more available to stop intoxicated drivers. Using the theory that a presence of enforcement reduces the amount of intoxicated drivers, as explained by Fell et al.(2014), Sanem et al. (2015), Kenkel (1993), and Tay (2005), the hot spot methodology proves to be the most effective, as well as most promising for ultimately reducing the amount of alcohol-related crashes.

7.5 CONCLUSIONS

With continued efforts going towards reducing the amount of alcohol-related crashes comes new research. The goal of this chapter is to determine whether patrolling for intoxicated drivers is more effective when conducting HSRO patrols or corridor patrols. Corridors are found by locating the top 15 road segments with the most alcohol-related crash locations during the studied time period for Franklin, Summit, and Ross counties. HSRO patrol locations are found using statistically significant hot spot locations. By using Esri's ArcGIS program to optimize routes for the three counties for corridor patrols first, then using the results to compare HSRO patrols, it is possible to determine which method may be more efficient to use. Table 5 shows the results when comparing the amount of alcohol-related crash locations passed per mile, as well as amount of alcohol-related crash locations passed per time. This allows for an equal comparison when the corridor models have more travel time and travel more miles than the HSRO models. The model that is able to travel through more alcohol-related crash locations per time and per mile may be seen as the most effective. By passing through more alcohol-related crash locations per mile and per time, officers may be able to locate more intoxicated drivers and create more of a presence to reduce the amount of alcohol-related crashes. Overall, hot spot patrols produced better outcomes, often allowing the patrols to drive less miles and less time but pass through more alcohol-related crash locations. Not only are the HSRO patrols passing through more alcohol-related crash locations per mile traveled, but they are also patrolling the statistically significant locations produced from hot spot analyses. This research shows that when patrolling for intoxicated drivers, it may be overall more beneficial to use the HSRO method. The HSRO method not only passes through more alcohol-related crash locations, but it also travels through the statistically significant areas as identified through HSA. This may yield to locating more intoxicated drivers, and ultimately reducing the amount of alcohol-related crashes.

CHAPTER 8: USE OF FAILURE PROBABILITY MODELS TO JUSTIFY NEW METHODS OF PATROLLING

8.1 INTRODUCTION

Driving while intoxicated continues to be a problem in the United States despite the efforts of researchers and law enforcement officers. Though the amount of alcohol-related crashes has remained consistent since 1999 (NHTSA, 2014), studies have continued in order to find some form of deterrence for drinking and driving (Carrick and Washburn, 2011; Fell et al., 2015); Sanem et al., 2015). Studies have continued to find some form of deterrence for drinking and driving, such as increased enforcement and mass media campaigns (Tay, 2005; Kenkel, 1993; Blais et al., 2015; Elder et al., 2004), however the amount of alcohol-related crashes has remained consistent since 1999.

So far, this research has analyzed proposed methodologies to determine ways to reduce the amount of alcohol-related crashes considering limited resources. Traditional methods of corridor enforcement have been compared with hot spot route optimization (HSRO) models, which are developed from hot spot analyses. In order to determine which method of patrolling is most efficient, the amount of alcohol-related crash locations passed per minute and per mile is used as a performance metric. Providing time and cost restrictions through the use of failure probability models may help to further justify the application of these methods. Failure probability is a proven method to determine the chance of failure of different scenarios and is widely used, whether it be for structural applications, such as pipeline failures (Yuhua and Datao, 2005; Dundulis et al., 2016), or chances that a specific treatment of asphalt will fail after a given time (Dong and Huang, 2015).

This research will utilize failure probability models to compare the traditional method of corridor patrolling to the new method of patrolling through HSRO. The failure probability presented will simulate a realistic application of these methods of patrolling in order to determine which may be the most efficient method to reduce the amount of intoxicated drivers. Two types of failure modes are used to compare the different methods of patrolling. The first type of failure will be the chance that an officer is unable to complete each consecutive cycle of their patrol, while the second type of failure will indicate the chance that officers patrolling through each consecutive cycle are more costly than the chance of pullovers themselves. The method of patrolling that is able to complete more cycles or has lower chances of failure for each consecutive cycle may be the more efficient method to use when patrolling for intoxicated drivers.

While the use of failure probability may immediately determine which method of patrolling is most efficient, it may also be used in the future to help law enforcement officials and captains in decision making processes by providing a guide of the amount of resources that may be used in patrolling on a given night. These models may be especially beneficial during the Drive Sober or Get Pulled Over (DSOGPO) campaigns. DSOGPO campaigns typically occur between mid-August through Labor Day, and again through the late November and December (NHTSA, 2016). During this time, high visibility enforcement officers (HVEO) are sent to patrol for intoxicated drivers through saturation patrols and sobriety check points. By using failure probability models, captains may be able to more effectively guide their officers to areas where intoxicated drivers are more likely to be present.

As useful as these failure probability models may be for law enforcement officials and captains, they may be seen equally as useful on the larger scale of communities. The DSOGPO campaign is not restricted to people involved with law enforcement. Mass media campaigns are also utilized during this time to advertise the dangers of impaired driving through funding supplied by the United States Department of Transportation (NHTSA, 2013). The results of the failure probability models may help captains and volunteers determine better locations to promote campaigns against intoxicated driving. For example, if the models show one method of patrolling is more effective than the other, these organizations may use the locations of the more effective model for their media outlets.

8.2 DATA

The data sources for this study include of crash records populated from the state of Ohio OH-1 crash reports (ODPS, 2015), the Ohio Department of Transportation (ODOT) Geographic Information Systems (GIS) roads layer (ODOT, 2016), and United States Census estimates information (United States Census Bureau, 2015). Using these three databases, the research team selected Franklin, Summit and Ross counties for analysis. All alcohol related crashes in each county from January 1, 2012 through April 9, 2015 are included in the analysis.

Franklin County was selected due to its high population (greater than 1 million people), it encompasses the large metropolitan area of the City of Columbus, and its high number of alcohol related crashes. Summit County was selected due to its large urban areas with a population greater than 500,000 people and still has a high number of alcohol related crashes. In contrast, Ross County is a predominantly rural county with a total population less than 100,000 people, and has a road network quarter the size of Franklin County. Historically, Franklin County and Summit County are both in the top 10 counties statewide with the highest number of alcohol related fatalities per year. A summary the general characteristics of each of the three counties may be seen in Table 8.2.1

Table 8.2.1: Comparison of Franklin, Summit, and Ross Counties

County Comparison

	County Population	Population Density (per sq. mi.)	Lane Miles	Total Alcohol-Related Crashes
Franklin	1,251,772	2,186	5,670	4,051
Summit	541,968	1,313	3,608	1,805
Ross	77,170	113	1,429	334

Note: County Population and Population Density determined from United States Census Bureau Quick Facts.
Lane Miles determined from Ohio Department of Transportation ArcGIS Roads Layer.
Total Alcohol-Related Crashes determined from Ohio Department of Public Safety OH-1 Crash Reports.

As shown in Table 8.2.1, there is a wide range of demographics, road networks and alcohol related crashes between these three counties. This chapter will use these counties as case studies as a demonstration for a new methodology that law enforcement agencies may use to help curb alcohol related crashes.

Additional data for this chapter are based on the HSRO and corridor results from the previous chapters. The total travel time of the individual patrol cars in each fleet and method of patrolling may be seen in Tables Table 7.4.1, Table 7.4.2, and Table 7.4.3. These results, as well as the number of network locations each patrol passes through on their respected routes will be directly used in the failure probability models.

8.3 METHODOLOGY

The main goal of this research is to reduce the amount of alcohol-related crashes in the state of Ohio through the improvement of overtime patrols. So far, this research has used hot spot maps created from alcohol-related crash locations to find local indicators of spatial association (LISA). It was determined that the LISA's that represent a 95% confidence may be used as network locations in a vehicle routing problem for proposed method of HSRO. The proposed method of HSRO and traditional method of corridor patrolling are compared by finding the amount of alcohol-related crash locations passed per time and distance for when fleets of three, five, and seven patrol cars are routed through the counties. These results show one performance metric of which method of patrolling may be the most effective in reducing the amount of alcohol-related crashes. The remainder of this chapter presents a second performance metric to help determine which method of patrolling may be most optimal in reducing the amount of alcohol-related crashes.

8.3.1 Failure Mode 1

Failure probability is widely used to help determine the chances of system failure for a number of applications (Leon and Macías, 2005; Ahammed and Melchers, 1997). This research uses failure probability for two different applications. The first application determines the chance that officers are unable to complete patrolling their route on each consecutive cycle that they patrol. This stems from the theory presented by Sanem, et al. (2015) where saturation patrols, among other forms of deterrence, are associated with less driving under the influence (DUI) violations. As a result, this research views officers who are able to complete more cycles of their patrol route as more effective in the overall cause of reducing the amount of alcohol-related crashes.

In failure probability models, a limit state function (LSF) is used to define the failure of the system. The LSF may be generally defined by Equation 8.1:

$$l = C - D \qquad (8.1)$$

where l represents the LSF of the model, C represents the capacity of the model, and D represents the demand of the model. The demand for this first failure mode is represented by the patrol time, which is defined as the total time the officers take to patrol their respected routes. The capacity of this first failure mode is the working shift time, which will remain at a constant three hours for the purposes of this chapter.

As the capacity model, or time of working shift, stays consistent throughout this first failure mode, the demand model varies and depends on a number of variables. A rate of pullovers has been incorporated throughout this research, and has a constant value of 1.5 pullovers per hour for each patrol car. Using this occurrence of pullovers, a Poisson distribution is completed over 100,000 simulations to simulate a different amount of pullovers in each scenario. By incorporating the Poisson distribution, variation is included in the system, allowing for a more realistic failure probability model. A time is then associated with each pullover that an officer makes, which is determined by assuming an average of 15 minutes and generating random numbers from an exponential distribution. Given the total amount of pullovers and their associated time for each simulation, the total time of pullovers for each mission is calculated as seen in Equation 8.2:

$$\mathrm{t_{tp}} = \sum t_p \qquad (8.2)$$

where $\mathrm{t_{tp}}$ is the total time of pullovers in each patrol, and t_p is the time of each individual pullover. Equation 8.2 is set up as a summation in order to account for the total time of pullovers in each individual simulation, or patrol. Once the total time of pullovers is determined for each mission, the total time of each patrol may be calculated as seen in Equation 8.3:

$$\mathrm{T_m} = \mathrm{n_c} * \mathrm{t_c} + \mathrm{t_{tp}} \qquad (8.3)$$

where $\mathrm{T_m}$ is the total patrol time for each simulation, $\mathrm{n_c}$ is the number of cycles the patrol car is able to fully complete throughout the duration of the shift, the $\mathrm{t_c}$ is the amount of travel time the individual patrol car takes to complete its respected cycle, and the $\mathrm{t_{tp}}$ is the total time the officer spent on pullovers for each simulated patrol.

Once the total time of each mission is calculated, the failure may then be calculated for each simulation using the equation of the LSF as a base equation for Equation 8.4:

$$f = (l \leq 0) = (T_{sh} - T_m \leq 0) \qquad (8.4)$$

where f is the failure of each simulation, l represents the LSF, T_{sh} is the shift time (three hours), and T_m is the patrol time for each simulation. When Equation 8.4 is a negative value, f will be denoted as a one, indicating failure of the system. If the system does not fail, Equation 4 will be a positive value and f will be denoted as a zero. In order to determine the total failure of the system, the sum of all failures for each simulation is averaged over the total number of simulations, as seen in Equation 8.5:

$$P_f = \frac{\sum f}{n_s} \tag{8.5}$$

where P_f represents the total probability of failure of the system, $\sum f$ represents the summation of failure for all simulations, and n_s represents the total number of simulations in the system.

The probability of failure for fleets of three, five, and seven patrol cars in each county are calculated as mentioned above. The output for each fleet is presented on a graph with a different failure for each individual patrol car in the fleet. The method of patrolling that allows officers to complete more cycles of patrolling within their given shift time is more desirable. This allows officers to have more of a presence in their county.

With these results, the total failure of the system may then be determined for each fleet. It should be noted that the patrol cars in each fleet are independent of each other (the success/failure of one patrol car is not related to the success/failure of any other patrol car). However, despite the fact that each patrol car is independent, the success of the fleet as a whole system is dependent on the success all individual patrol cars. The success of the system for this first probability mode may be defined as all patrol cars being able to complete each consecutive cycle of patrolling. Knowing the probability of failure of the cars in each fleet, the probability of success may be defined by Equation 8.6:

$$P_{ss} = 1 - P_f \tag{8.6}$$

where P_{ss} is the probability of success, and P_f is the probability of failure of each individual patrol car. Using the probability of success of each patrol car in a fleet, the probability of failure of the system may be found by Equation 8.7:

$$P_{fs} = 1 - \prod P_{ss} \tag{8.7}$$

where P_{fs} represents the probability of failure of the system, and P_s represents the probability of success of the system. The results of the system failure will show one value of failure for each fleet. For example, there will only be one failure result for a whole fleet of three cars instead of three individual results for each car in the fleet. Additionally, the system will then show a separate result for the fleet of five patrol cars and the fleet of seven patrol cars. Typically the system failure occurs when the first patrol car in a fleet fails. This is because an entire fleet may not succeed unless all individual cars in the fleet succeed for each cycle.

This first failure mode determines the maximum number of cycles each patrol car in a fleet of three, five, and seven patrol cars is capable of completing in the three hour shift time. The failure for each fleet depends on the total time it takes for each patrol to travel through their respected route and the total time of simulated pullovers for each patrol. This failure mode may not only be used to help determine how many officers are needed in a fleet on a given night, but it is also used to compare the HSRO method of patrolling versus the traditionally used method of corridor patrolling. If one method is able to patrol more cycles of their respected routes in a given shift time, that allows the officers to create more of a presence in the statistically significant areas. Additionally, if one method is able to patrol through their routes with less officers than the other, less money will be spent while patrolling for intoxicated drivers.

8.3.2 Failure Mode 2

The second application of failure probability compares the chance that officers patrolling on a given shift are more costly than the cost of potential pullovers for the county. Law enforcement agencies often see a strict amount of budgeting and manpower restrictions. This second failure mode may help to justify the current restrictions or it may justify an increase in funds to help improve the cause of reducing the amount of alcohol-related crashes.

The second failure mode begins with a LSF defined as the system failing when the cost of sending officers to patrol (demand model) is greater than the cost of potential pullovers (capacity model). Both, the capacity and demand model, are dependent on a number of variables from the previous chapter. The capacity model is dependent on the number of pullovers for each individual patrol car in each fleet. The occurrence of pullovers is determined using a Poisson distribution, similar to the first failure mode and the model is run through total of 100,000 simulations. A cost is associated with each pullover, based on fines and penalties the driver is responsible for, and determined using a triangular distribution with a low cost of \$250, a mean cost of \$630, and a high cost of \$1000 (NOLO, 2016). These costs are randomly assigned to each pullover in each patrol car in the 100,000 simulations. The total cost of the pullovers for each simulated patrol make up the capacity model for the LSF, and may be seen in Equation 8.8:

$$C_p = \sum C_{up} \qquad (8.8)$$

where C_p is the total cost of pullovers for each simulated patrol, and $\sum C_{up}$ is the summation of the cost of each individual pullover for each simulation.

In order to calculate the demand model the cost of miles driven and the cost of manpower is determined. For the purposes of this research, the cost of miles is based off the Internal Revenue Service (IRS) standard mileage rates from 2013 at 56.5 cents per mile, while the cost of manpower is randomly determined again using a triangular distribution with a low rate of \$21 per hour, a mean rate of \$28 per hour and a maximum rate of \$37 per hour. These costs are provided through the United States Bureau of Labor Statistics (2015) and represent the 25, 50, and 75 percentile of hourly police officer wages in the country. The cost of miles is found using Equation 8.9:

$$\mathrm{C_{mi}} = 0.01 * \mathrm{c_{mi}} * \mathrm{T_{mi}} \qquad (8.9)$$

where $\mathrm{C_{mi}}$ represents the cost of miles for each patrol car in the simulated fleets (in dollars), $\mathrm{c_{mi}}$ represents the IRS rate of mileage cost, and $\mathrm{T_{mi}}$ represents the total miles traveled by each officer in a given shift, based on the results from the HSRO and corridor patrols in the previous chapter. Next, the cost of manpower is calculated by Equation 8.10:

$$\mathrm{C_{mp}} = \mathrm{c_{man}} * \mathrm{T_{sh}} \qquad (8.10)$$

where $\mathrm{C_{mp}}$ is the cost of manpower associated with each car in the simulated fleet, $\mathrm{c_{man}}$ is the hourly wage an officer makes, which is randomly determined from the previously defined triangle distribution, and $\mathrm{T_{sh}}$ is the total shift time, also previously defined as three hours. The total cost of miles and manpower are then used to determine the total cost of patrolling for the cars in each simulation, as described in Equation 8.11:

$$C_{pa} = C_{mp} + C_{mi} \qquad (8.11)$$

where C_{pa} represents the cost of the individual patrols for each simulation, C_{mp} represents the cost of manpower for each patrol, and C_{mi} represents the mileage cost for each patrol. The total cost of patrolling for each simulated shift is the final variable that makes up the demand model for the second mode failure probability.

Using the cost of patrolling and cost of pullovers, failure may be calculated using the previously defined LSF, as seen in Equation 8.12:

$$f = (C_p - C_m \leq 0) \qquad (8.12)$$

where f represents the failure of each patrol car for each simulation, C_p represents the cost of pullovers from each patrol car, and C_m represents the cost of each patrol's mission. The failure in each mission is similar to the first failure mode. When the value of the failure is negative, the model fails, and the failure is denoted as a one. When the value of failure is positive, the model is successful, indicating that the cost of pullovers is more than the cost of patrolling, and the failure is denoted as a zero. Finally, the probability of failure for each individual patrol car in each fleet is determined by averaging the failure for each simulation over the total number of simulations, as seen in Equation 8.13:

$$P_{fc} = \frac{\sum f}{n_s} \qquad (8.13)$$

where P_{fc} represents the total probability of failure of the system, $\sum f$ represents the summation of failure for all simulations, and n_s represents the total number of simulations in the system. Equation 8.13 represents the probability of failure for each individual patrol car in a fleet. This method is repeated for fleets of three, five, and seven patrol cars in each county.

Once the probability of failure is determined for each car in each fleet, the probability of failure for the system is determined. The system failure is determined by how often the total cost of patrolling for a whole fleet in a given shift is greater than the total cost of pullovers over the 100,000 simulations. This may be seen through Equation 8.14:

$$f_{all} = (C_{pua} - C_{ma} \leq 0) \tag{8.14}$$

where f_{all} represents the total failure for each simulation, C_{pua} represents the total cost of pullovers in each simulation, and C_{ma} represents the total cost of missions of the fleets in each simulation. The denotations are similar to the failures in Equations 8.4 and 8.12 where a negative value represents a failure and is denoted as a one, and a positive value represents a success and is denoted by a zero. Finally, the probability of failure of the whole system for each fleet is found by averaging out the total number of failures over the total number of simulations, as seen in Equation 8.15:

$$P_{fa} = \frac{\sum f_{all}}{n_s} \tag{8.15}$$

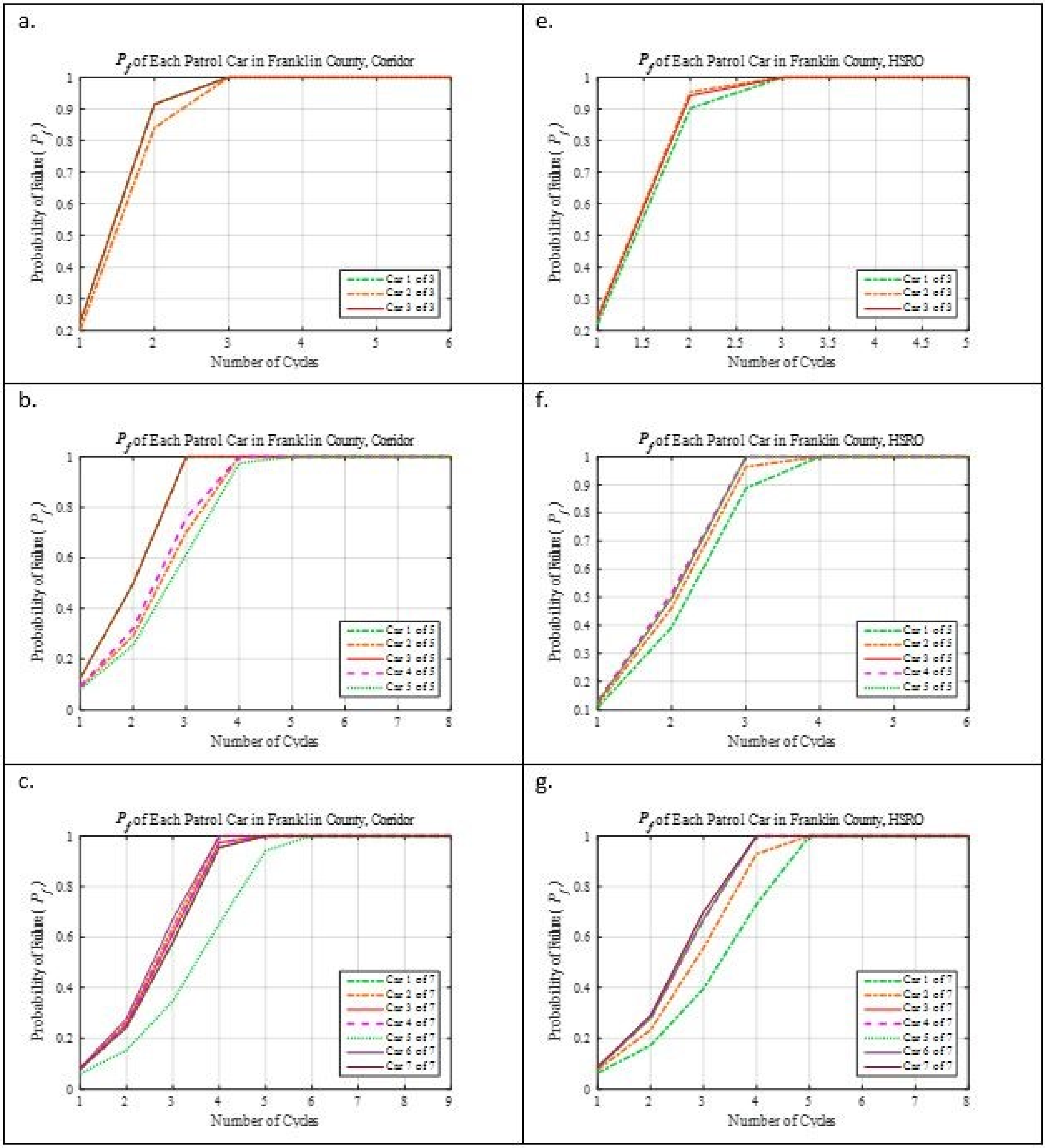
where P_{fa} represents the probability of failure of the system, $\sum f_{all}$ represents the total sum of failures in all simulations, and n_s represents the total number of simulations.

The second failure mode is evaluated for all counties at three, five, and seven patrol cars for the HSRO method of patrolling and patrolling through corridors. Using this mode of failure probability, it may be possible to determine if one method of patrolling is more cost effective than the other. If the system of one method of patrolling shows a lesser chance that the cost of manpower is greater than the cost of potential pullovers, captains may consider utilizing that method of patrolling.

The results of each failure mode may help to determine any significant differences between the HSRO method of patrolling and patrolling through the traditional method of patrolling through corridors. Additionally, these results may justify current practices of patrolling for intoxicated drivers, or the benefit of implementing new practices of patrolling. The goal of these methods of failure probability is to determine which method may be most effective and efficient in patrolling for intoxicated drivers and reducing the amount of alcohol-related crashes.

8.4 RESULTS

The final results of the first failure mode show a comparison of the traditional method of corridor patrolling and the proposed method of HSRO patrolling. The results of the first failure mode compare the amount of cycles each fleet is able to complete in a given shift time for each method of patrolling. Fleets that are able to patrol through more cycles in a given shift may show more of a presence in the county as well as have a higher potential to reduce the amount of alcohol-related crashes. It should be reiterated that these results are based off the routing results from CHAPTER 7: where the results of the routes for the corridor model were used as restrictions for the routes in the HSRO model. Results for the first failure mode comparing corridor patrolling and patrolling through HSRO for Franklin, Summit, and Ross Counties may be seen in Figure 8.4.1, Figure 8.4.2, and Figure 8.4.3.


a.
P_f of Each Patrol Car in Franklin County, Corridor
Probability of Failure (P_f)
Number of Cycles
Car 1 of 3
Car 2 of 3
Car 3 of 3
e.
P_f of Each Patrol Car in Franklin County, HSRO
Probability of Failure (P_f)
Number of Cycles
Car 1 of 3
Car 2 of 3
Car 3 of 3
b.
P_f of Each Patrol Car in Franklin County, Corridor
Probability of Failure (P_f)
Number of Cycles
Car 1 of 5
Car 2 of 5
Car 3 of 5
Car 4 of 5
Car 5 of 5
f.
P_f of Each Patrol Car in Franklin County, HSRO
Probability of Failure (P_f)
Number of Cycles
Car 1 of 5
Car 2 of 5
Car 3 of 5
Car 4 of 5
Car 5 of 5
c.
P_f of Each Patrol Car in Franklin County, Corridor
Probability of Failure (P_f)
Number of Cycles
Car 1 of 7
Car 2 of 7
Car 3 of 7
Car 4 of 7
Car 5 of 7
Car 6 of 7
Car 7 of 7
g.
P_f of Each Patrol Car in Franklin County, HSRO
Probability of Failure (P_f)
Number of Cycles
Car 1 of 7
Car 2 of 7
Car 3 of 7
Car 4 of 7
Car 5 of 7
Car 6 of 7
Car 7 of 7

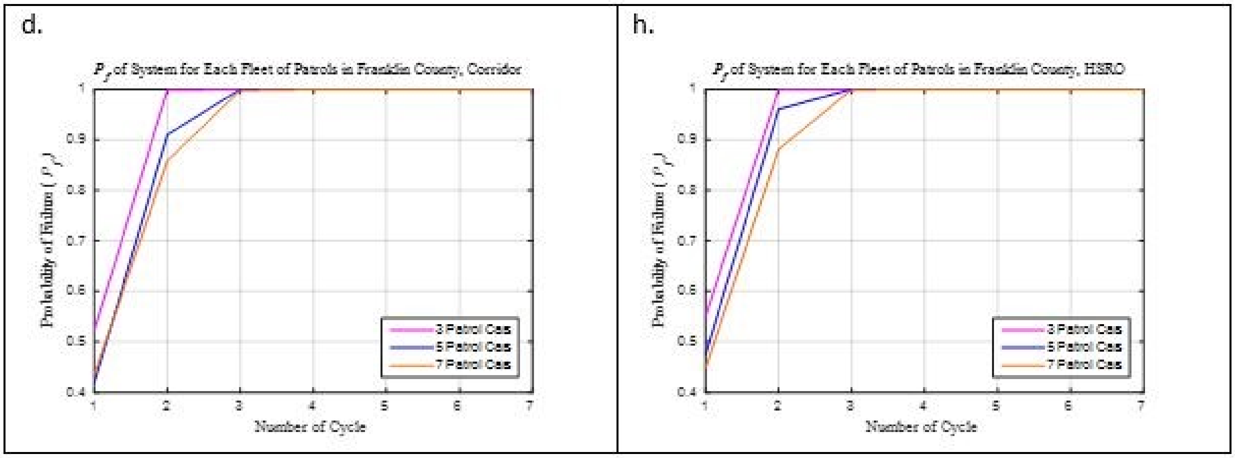
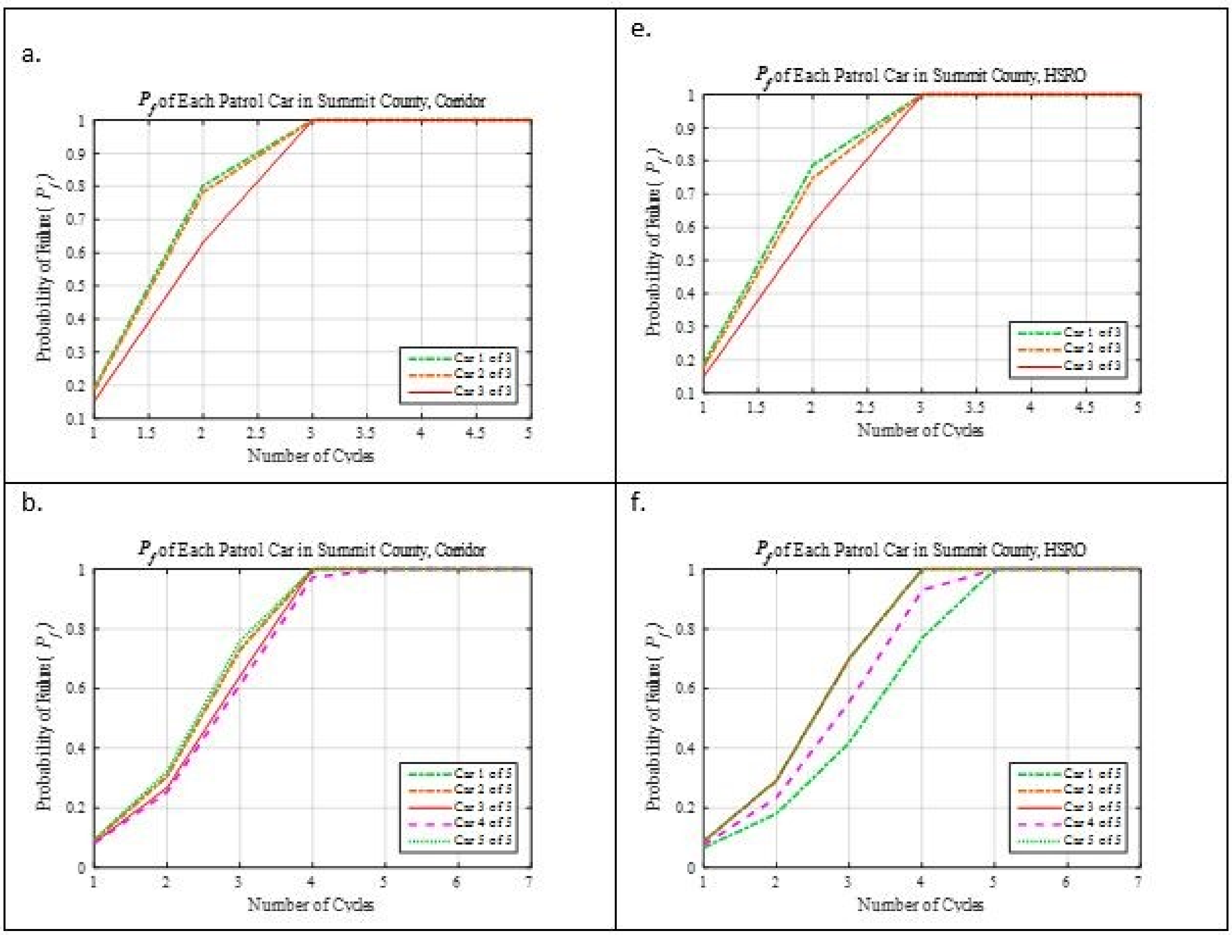


Figure 8.4.1: Results of First Failure Mode for Franklin County.



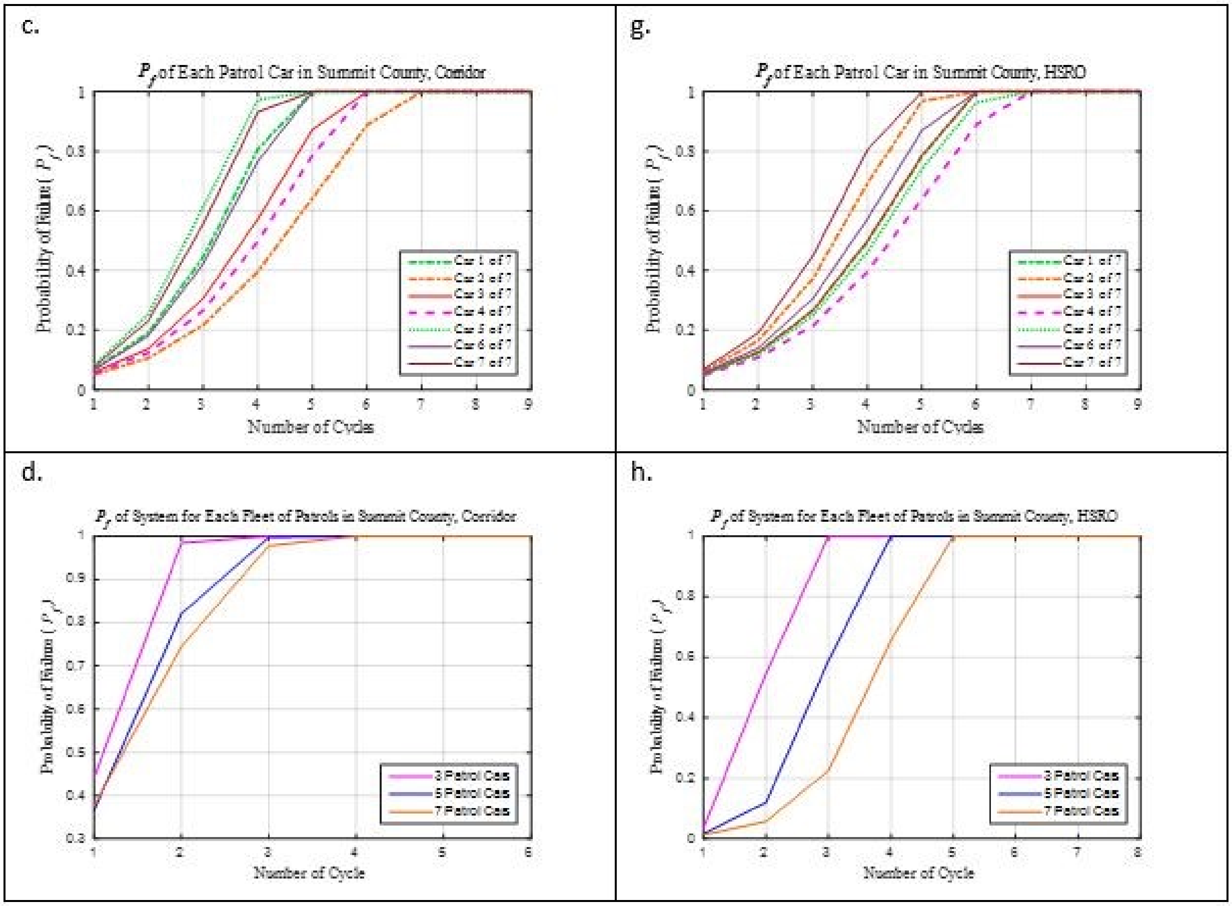
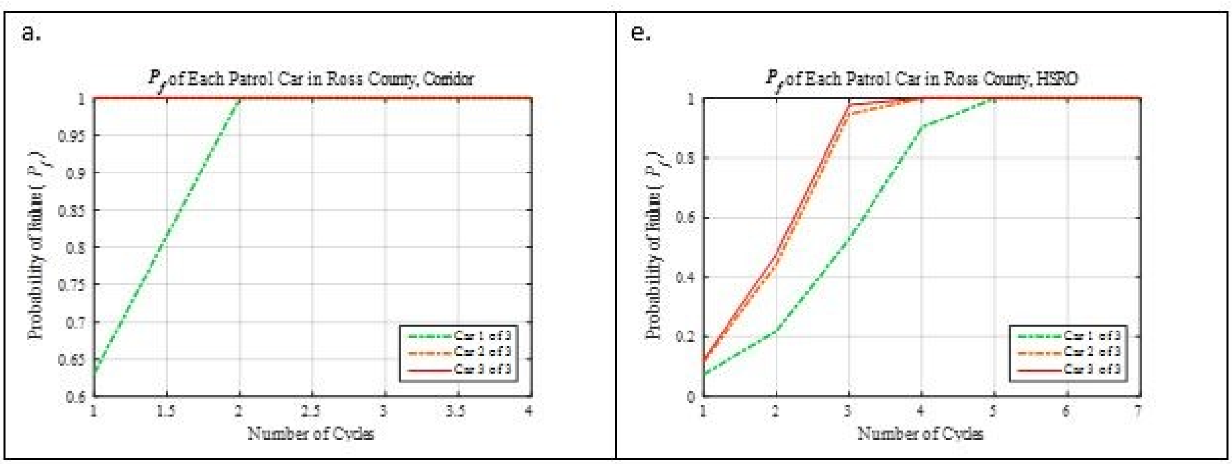


Figure 8.4.2: Results of First Failure Mode for Summit County.



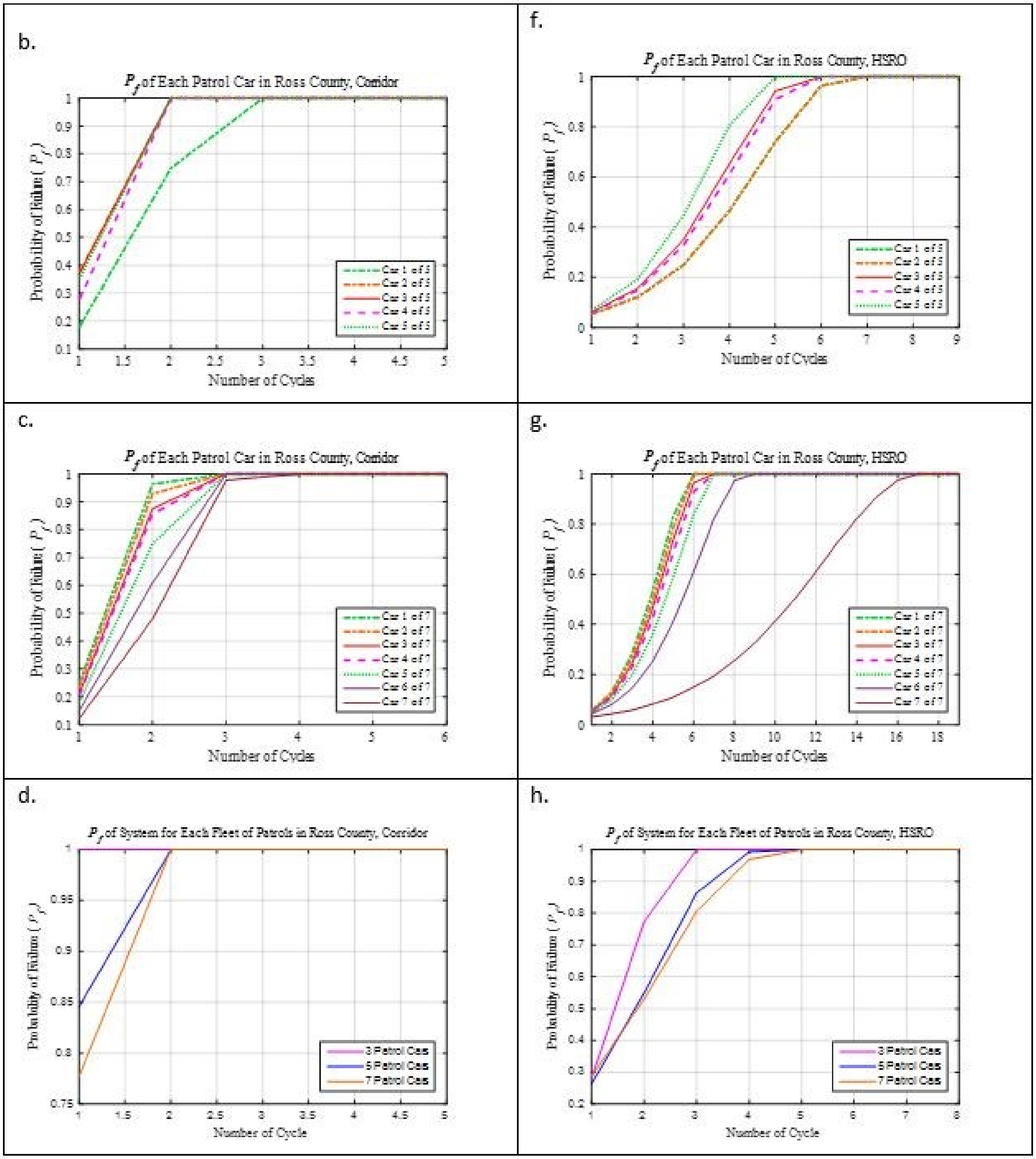


Figure 8.4.3: Results of First Failure Mode for Ross County.

Note: Figures 1-3 are developed using MATLAB and are based on the time results from CHAPTER 7: .

Figure 8.4.1, Figure 8.4.2, and Figure 8.4.3 show the comparison between corridor patrolling and patrolling through HSRO for Franklin, Summit, and Ross Counties, respectively. It is important to consider the amount of cycles each individual officer is able to complete because the goal is for each officer to complete the maximum amount of cycles before failing. This is to ensure that officers are able to show the maximum amount of presence on the roads to deter people from driving intoxicated. When a failure occurs, this means the officer is unable to completely patrol the next consecutive cycle without going over the allotted time in their shift, and the chance for deterring intoxicated drivers has ended.

Results for Franklin County are very similar between the two methods of patrolling. This is due to fact that the results for patrolling through corridors are used as restrictions for patrolling through the HSRO. For fleets of three, five, and seven patrols, both methods of patrolling allow officers to complete the same amount of cycles. However minor differences appear when considering they, the officers, are to complete each cycle. This may be seen in a fleet of seven patrol cars where the corridor method of patrolling allows three officers to complete four cycles of patrolling and one officer to complete five cycles of patrolling before failing, but the HSRO method of patrolling only allows two officers to complete four cycles of patrolling before failing.

When comparing the fleets as a system, again both methods of patrolling are very similar for Franklin County. Each fleet of patrol cars is able to complete the same number of cycles for both methods. It may be noted that for fleets of three and seven patrol cars, officers are able to complete two and three consecutive cycles before failing, however the system failure for each fleet appears to show a failure at two and three cycles, indicating officers are unable to complete these cycles. This is a result of how the probability of failure is calculated in a system with independent variables to determine the failure of the fleets. Since the chance of failure is so high for the individual patrol cars, the system approaches failure at that value.

Summit County is very similar to Franklin County, in that the results of the corridor patrolling were used as restrictions for patrolling through HSRO. Each individual fleet of vehicles is able to complete the same amount of patrols, however the individual cars in each fleet differ in results. The HSRO method of patrolling has more cars completing more cycles of patrolling for fleets of five and seven patrols. This is beneficial because it allows for more officers to have an increased presence while patrolling for intoxicated drivers, indicating a greater effect of deterrence. This is also beneficial when considering how the systems for each fleet are calculated.

The fleet systems in Summit County ultimately yield the same failures for each method of patrolling, however the chance of failure for each cycle is significantly different. For example, when patrolling through two cycles, the corridor method of patrolling has a much larger chance of failure than the HSRO method of patrolling. This is similar for when three and four cycles are patrolled through each method. These trends indicate that the HSRO method of patrolling may be more reliable since officers are less likely to be unable to complete patrolling their cycle in the given shift time for each consecutive cycle. Again, the system failure for Summit County appears to occur earlier than the individual fleets for five and seven patrols due to how the probability of failure is determined with a series of independent variables.

Finally, the comparison between the corridor method of patrolling and the HSRO method of patrolling are analyzed for Ross County. Since the time results for the corridor method of patrolling are much greater than the time results for the HSRO method of patrolling, as seen in Table 7.4.3, the failure probability results are also much different. For every fleet that is deployed, officers patrolling the HSRO method are consistently able to complete more cycles than while patrolling the corridor method. This allows for much more presence of officers, especially in the statistically significant locations of Ross County.

When considering the fleets of each method of patrolling in Ross County, results are consistent with the individual fleets. Officers patrolling through corridors are only able to complete one cycle of patrolling for fleets of five and seven cars. However, the HSRO method of patrolling allows officers to complete four and five cycles of patrolling for fleets of five and seven patrols. This again allows for much more of a presence of officers, as well as more time patrolling in the statistically significant locations to deter and reduce the amount of intoxicated drivers.

For the system of each fleet, it may be beneficial to consider the value of the probability of failure different fleet. This may help to determine the exact differences between officers patrolling through corridors and officers patrolling through HSRO. Table 8.4.1 shows the system failure results for Franklin, Summit, and Ross counties.

Table 8.4.1: Probability of Failure for Franklin, Ross, and Summit County Fleet Systems

Probability of Failure Values for Systems of Fleets in Each County

Cycles	Cars in Fleet (Corridor)			Cars in Fleet (HSRO)		
	3	5	7	3	5	7
Franklin County System Results						
1	0.52	0.42	0.43	0.55	0.48	0.45
2	**1.00**	0.91	0.86	**1.00**	0.96	0.88
3	1.00	1.00	**1.00**	1.00	1.00	**1.00**
Summit County System Results						
1	0.44	0.37	0.38	0.04	0.02	0.01
2	0.98	0.82	0.75	0.55	0.12	0.06
3	1.00	**1.00**	0.98	1.00	0.59	0.22
4	1.00	1.00	1.00	1.00	1.00	0.66
5	1.00	1.00	1.00	1.00	1.00	1.00
Ross County System Results						
1	1.00	0.85	0.78	0.28	0.26	0.29
2	1.00	1.00	1.00	0.77	0.55	0.53
3	1.00	1.00	1.00	**1.00**	0.86	0.81
4	1.00	1.00	1.00	1.00	0.99	0.97
5	1.00	1.00	1.00	1.00	1.00	**1.00**

Note: This table shows the probability of failure of fleets of three, five, and seven patrol cars in each county based on the travel time of each patrol and a rate of 1.5 pullovers per officer per hour.

Note: The bold numbers indicate failure probability values rounded to one, however the values themselves do not equal one.

The system failure results, described in Section 8.3.1 , may be seen in Table 8.4.1. These results show the probability of failure of each consecutive cycle for the different number of fleets in each county. When the value equals a one, this means the fleet has failed, or is unable to complete the next consecutive cycle. Once the system fails, it is unable to come back, explaining why some fleets show multiple failures.

The system failure results seen in Table 8.4.1 may help to determine the differences between the two methods of patrolling. These results match with graphs (d) and (h) in Figures 8.4.1-8.4.3. When considering Franklin County, the chances of failure for each consecutive cycle in the fleets are extremely close to each other. For example, when a fleet of five patrol cars are deployed, the corridor method of patrolling has a 42% and 91% chance of failure while completing one and two cycles, respectively, where the HSRO method of patrolling has a 48% and 96% chance of failure for completing one and two cycles, respectively. This means that nearly half the time, a fleet of five officers will only be able to complete one cycle of patrolling. Though the failure results are so close, the HSRO method of patrolling has officers travelling through the statistically significant locations identified through the hot spots. This allows for a greater chance of locating intoxicated drivers through the HSRO method of patrolling.

Again, similar results are found when comparing the methods of patrolling for Summit County. When three and five fleets of patrols are deployed, the HSRO and corridor methods of patrolling are able to complete the same amount of cycles. However, the chances of failure are much lower for the HSRO method of patrolling. An example of this is seen in the fleet of five patrols which yields a 2%, 12%, and 59% chance of failure for 1-3 cycles of patrolling through HSRO, where the corridor method of patrolling yields a 37%, 82%, and nearly 100% chance of failure. This means that officers are much more likely to complete up to three cycles of their route while patrolling for intoxicated drivers in the HSRO method of patrolling. This allows officers to be more visible, and have a much greater chance of stopping and deterring intoxicated drivers than while patrolling through corridors.

When comparing the systems of fleets for each method of patrolling in Ross County, the results reflect the individual results from each fleet. This may be seen in Figure 3, and also in the failure probability values in Table 8.4.1. When a fleet of seven cars is deployed for the corridor method of patrolling, the fleet has a 78% chance of failure at one cycle. This means that officers are fairly unlikely to complete even one cycle of patrolling given the rate of pullovers for Ross County. However, when a fleet of seven cars are deployed for patrolling through the HSRO method, officers have a 53% chance of failure at two cycles. This indicates that officers are more likely to complete more cycles while patrolling through HSRO than while patrolling through corridors. This allows officers to have a greater presence to deter intoxicated drivers while also having more of a chance of stopping intoxicated drivers. Additionally, officers patrolling through HSRO are present in the statistically significant areas as defined by the hot spot analyses.

When considering a captain’s perspective based on the results of the first failure probability model for each county, officer are able to patrol the same, or more cycles through the HSRO method of patrolling than the corridor method. Additionally, officers generally have less of a probability of failure while patrolling through the HSRO method, indicating greater chance they are able to complete each consecutive route, showing more of a presence to both deter and reduce the amount of intoxicated drivers. Additionally, this gives officers a presence through the statistically significant areas as identified through hot spot analyses. When considering captains in the decision making process, a captain may be able to use these results to determine the fleet size in a given night. These decisions may be dependent on funding available or the level of concern in a given night for intoxicated drivers, such as when patrolling through DSOGPO campaigns.

The results of the second failure probability mode compare the chance that the cost of sending a fleet of officers to patrol for intoxicated drivers is less than the cost of the potential pullovers. The chance of potential pullovers is based on the previously defined 1.5 pullovers per hour for a patrol car, and the time and distance results of the routes for each county in CHAPTER 7: are used to develop this failure mode. The goal of this failure mode is to determine if it is more costly to fund patrols when considering the chance of pullovers for each county. The results for each individual fleet in all counties are compiled and may be seen in the appendices. The final results are then compiled from each system of fleets for each county, which may also be seen in the appendices. It may be noted that the chance of failure for the individual patrols in each county are much higher than the fleets as a system results for each county. This is explained in the methodology, section 8.3.2 , where the total cost of pullovers for all patrol cars is compared to the total cost of the fleet. In most cases, the total cost of pullovers far exceeds the cost to run the fleet, resulting in a lower chance, or no chance, of failure for the systems. The failure rates for each county seen in Table 8.4.2.

Table 8.4.2: Failure Mode Two System Failure Results for Each County

Failure Mode 2 System Results

Cycles	Cars in Fleet (Corridor)			Cars in Fleet (HSRO)		
	3	5	7	3	5	7
Franklin	0.000	0.000	0.000	0.000	0.000	0.000
Summit	0.000	0.000	0.000	0.000	0.000	0.000
Ross	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table is based on the randomly distributed cost of pullovers and cost of manpower, given a rate of 1.5 pullovers per officer per hour.

Table 8.4.2 shows the results of the second failure mode for each fleet in Franklin, Summit, and Ross counties. Since the rate of pullovers in each county is set at 1.5 pullovers per hour, and the designated shift time is 3 hours, cost of total potential pullovers in for each fleet in each county is constantly greater than the cost it would take to patrol for intoxicated drivers. The results of failure mode two indicate that patrolling for intoxicated drivers is always worth the cost since there is nearly zero chance of not having enough pullovers to outweigh the cost of manpower and equipment.

The rate of 1.5 pullovers per hour has been a constant variable used throughout this research. Maistros et al. (2016) found this average rate of pullovers through research conducted in Stark County, Ohio. Though this research uses a rate of 1.5 pullovers per hour, it is not guaranteed that officers throughout the whole state, or nation, have the same efficiency of stops.

Since failure mode two shows a probability of failure equal to zero for each scenario of patrolling, it may be beneficial to conduct the test with different rates of pullovers. Conducting failure mode two at different rates of pullovers will both account for different rates of pullovers for different jurisdictions, but also determine the rate of pullovers where a chance of failure is present. Since a rate of 1.5 pullovers per hour has virtually zero chance that the cost of patrolling for intoxicated drivers is more costly than the pullovers conducted, theoretically any higher rate of pullovers will also yield the same results. Because of this, the maximum rate of pullovers studied will remain at 1.5 pullovers per hour.

The systems of fleets for Franklin, Summit, and Ross counties is modeled for rates of 1.25, 1.0, 0.75, 0.5, and 0.25 pullovers per hour per officer. The results for the fleet systems may be seen in Table 8.4.3.

Table 8.4.3: Sensitivity Analysis of Failure Mode Two for Franklin, Ross and Summit Counties

Failure Mode 2 Sensitivity Analysis System Results

Pullover Rate	Cycles	Cars in Fleet (Corridor)			Cars in Fleet (HSRO)		
		3	5	7	3	5	7
1.5	Franklin	0.000	0.000	0.000	0.000	0.000	0.000
	Summit	0.000	0.000	0.000	0.000	0.000	0.000
	Ross	0.000	0.000	0.000	0.000	0.000	0.000
1.25	Franklin	0.000	0.000	0.000	0.000	0.000	0.000
	Summit	0.000	0.000	0.000	0.000	0.000	0.000
	Ross	0.000	0.000	0.000	0.000	0.000	0.000
1	Franklin	0.000	0.000	0.000	0.000	0.000	0.000
	Summit	0.000	0.000	0.000	0.000	0.000	0.000
	Ross	0.000	0.000	0.000	0.000	0.000	0.000
0.75	Franklin	0.002	0.000	0.000	0.002	0.000	0.000
	Summit	0.003	0.000	0.000	0.002	0.000	0.000
	Ross	0.002	0.000	0.000	0.002	0.000	0.000
0.5	Franklin	0.017	0.004	0.001	0.017	0.004	0.001
	Summit	0.019	0.004	0.001	0.018	0.004	0.001
	Ross	0.017	0.004	0.001	0.018	0.004	0.000
0.25	Franklin	0.138	0.098	0.050	0.138	0.095	0.049
	Summit	0.143	0.099	0.051	0.142	0.100	0.052
	Ross	0.136	0.093	0.052	0.135	0.092	0.053

Note: This table is based on the randomly distributed cost of pullovers and cost of manpower, given the different rates of pullovers.

Table 8.4.3 shows the chance that officers patrolling for intoxicated drivers are more costly than the pullovers themselves. As the rate of pullovers decreases, the chance of failure increases for both methods of patrolling. When anything rates that is greater than one pullover per hour shows a probability of failure of zero. This is a promising result, meaning that there is zero chance that the cost of patrolling will be greater than the cost of potential pullovers in every scenario. As the rate of pullovers are reduced to below one pullover per hour, a probability of failure becomes present, indicating that there is now a chance that the cost of patrolling may be greater than the cost of potential pullovers.

For example, if an officer has a rate of 0.25 pullovers per hour, the officer will on average speak to one driver in four hours. Since the rate of pullovers is an average and the amount of actual pullovers is randomly distributed over 100,000 simulations, there may be situations where officers have zero or one pullover per three hour shift. However, since the failure of the system combines the pullovers from the cars in each fleet, the chance of failure decreases as the amount of patrol cars in the fleet increases. Similarly, the rates of failure also increase as the rate of pullovers decreases. This is because fewer pullovers indicates less money going toward the cost of pullovers, and with the same amount of officers patrolling, the probability that the cost of patrolling is greater than the cost of pullovers increases. In each scenario, the chance of failure is extremely low, with the highest chance of failure being present when the pullover rate is 0.25 with a 13%-14% chance of failure with a fleet of three patrol cars. Given that these are the highest rate of failure, they remain significantly low.

The probability of failure for each rate of pullovers is very similar between the HSRO method of patrolling and patrolling through corridors. Table 8.4.3 shows a fraction of a difference between the two methods of patrolling, occasionally showing zero difference. However, it may be noted that the HSRO method of patrolling continues to have officers driving through the statistically significant areas, as defined through hot spot analyses, in each county. The use of failure mode two may be beneficial for captains in the decision making process when determining patrols for a given night. This failure mode may also be an indication that the cost of patrolling is not an issue considering the rate that officers may be able to stop intoxicated drivers.

With failure rates this low, captains may be confident in knowing that sending officers to patrol for intoxicated drivers will be less costly than the chances of pulling over potentially intoxicated drivers. Additionally, since there is very little chance that there will be zero pullovers in a given shift time, captains may be more confident in sending any size of fleet out to patrol. The results of the second failure mode ultimately show the chance that is it more costly to patrol for intoxicated drivers, than the cost of pullovers themselves. These results may assist captains in the decision making process when determining fleet sizes for a given night of patrolling.

8.5 CONCLUSION

With the ultimate goal of this research going towards reducing the amount of alcohol-related crashes, it is beneficial to use mathematical methods that help to justify all models that have been created. This chapter utilizes two modes of failure probability that may justify the use of different patrol methods and give captains a guide when determining fleet sizes. The first mode of failure probability determined the maximum number of cycles an officer may be able to patrol in a given shift time. For Franklin and Summit Counties, the results were very similar between the corridor and HSRO methods of patrolling. For Ross County, the HSRO method of patrolling consistently allowed officers to complete more cycles in the given shift for every fleet size. The results were similar between the two methods of patrolling for Franklin and Summit Counties, the HSRO method of patrolling guides officers through the statistically significant areas of alcohol-related crashes in each county. Since these areas are statistically significant, the chance of officers locating intoxicated drivers may be increased.

The second mode of failure probability determines the chances that patrolling for intoxicated drivers is more costly than the chance of potential pullovers. Again, the results between the two methods of patrolling were similar, in that the system of fleets for both methods of patrolling had zero chance that the cost of sending officers to patrolling for intoxicated drivers was more costly than the potential pullovers themselves. Through the use of failure probability, it may be possible to determine the method of patrolling for intoxicated drivers that is most efficient as guide captains in decision-making practices with the ultimate hopes of reducing the amount of alcohol-related crashes.

CHAPTER 9: CONCLUSION AND RECOMMENDATIONS

9.1 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. Using the examination of crash patterns, new methods of patrolling are developed for officers to reduce the amount of intoxicated drivers. The objectives of the new methodology are including identification of significant areas for officers to patrol, comparison of methods of patrolling, and creation of failure probability models.

9.2 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.

9.3 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.

9.4 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.

9.5 IDENTIFICATION OF SIGNIFICANT AREAS FOR OFFICERS TO PATROL

The first objective of identifying significant areas for officers to patrol initially uses hot spot analyses. The output of a hot spot analysis is comprised of local indicators of spatial association, with are location that each have a different value representing the statistical significance of that location. These locations, otherwise referred to as network locations, are compared at each significance level in each of the counties studied.

To determine which confidence level should be used in patrolling, the amount of network locations is compared at each confidence level, and compared between the 90% and 95%, as well as 90% and 99%. Theses comparisons are used to determine if "too much" of the significant areas will be missed if officers are sent to patrol only the 95% or 99% confident network locations compared to patrolling every significant network location. Typically as the confidence level increases, the number of network locations decreases. As a result, this research looks into potentially having officers patrol only a specific level of confident network locations as opposed to every significant network location. Additionally, a number of radii of different lengths are placed around each of the 95% and 99% confident network locations in order determine if the amount of 90% confident network locations that are within a the 95% and 99% confident network locations. This may help to justify using a set of higher confidence network locations, and officers having fewer locations to patrol.

Given the comparisons studied in this first objective, results show that it may be acceptable for officers to patrol only the 95% confident network locations. By patrolling only the 95% confident network locations, officers have fewer locations to patrol, while also patrolling the more significant locations, resulting in an increased efficiency. Though this objective is able to show the significant locations for officers to patrol, it does not show how officers should patrol these locations.

9.6 COMPARISON OF METHODS OF PATROLLING

Given the statistically significant hot spot locations of patrolling, the next step is to determine how officers should patrol these locations. This is completed through the use of Esri's ArcGIS Vehicle Routing Problem. However, in order to give credibility to the HSRO method of patrolling, it is compared with the traditional method of corridor patrolling. This is completed by taking the top 15 corridors with the most amounts of alcohol-related crashes and using the times and fleet sizes as restrictions while routing officers through the HSRO method of patrolling.

Once each method of patrolling is routed, the number of alcohol-related crash locations passed per time and per mile is compared for each county studied. Ultimately, the HSRO method of patrolling was able to pass through more alcohol-related crash locations per time and per mile for each fleet size and in each county. This indicates that not only are officers patrolling through the statistically significant locations in the HSRO method, but they are also able to pass through more alcohol-related crash locations when comparing to the corridor method of patrolling.

9.7 CREATION OF FAILURE PROBABILITY MODELS

The use of failure probability models is common in many applications of civil engineering. This research created failure probability models to again compare the differences between the HSRO method of patrolling and patrolling through corridors. Two failure probability modes are created in this research.

The first mode is used to determine the maximum amount of cycles each officer in a fleet is able to patrol, as well as the maximum amount of cycles a whole fleet is able to patrol before failing. Failure for this mode is defined as the chance that an officer has a being unable to complete each consecutive cycle. This model is built off the theory that increased presence of officers within an area results in a decrease of intoxicated drivers. Additionally, the increase in presence of officers allows them to have greater chances of locating intoxicated drivers. The first failure mode found that the HSRO method of patrolling is able to complete the same or more consecutive cycles in a given shift. This means that, again, not only are officers able to have more of a presence while patrolling, but they are also patrolling in the statistically significant areas, as defined by hot spot analyses.

The second failure mode is used to determine the chance that the cost of patrolling is greater than the cost of potential pullovers. This is to determine the cost-effectiveness of patrolling for intoxicated drivers for both method of patrolling. The second failure mode is based on a rate of 1.5 pullovers per hour and a shift time of three hours. Results for each of the fleets showed a failure rate of zero, indicating that there is a zero percent chance that the value of patrolling is ever greater than the value of potential pullovers.

A sensitivity analysis is then used to determine the rate of pullovers that will affect the failure rate of second failure mode. With a rate of pullovers equal to 0.5 pullovers per hour, the rates of failure for a fleet of three patrol cars is between one and two percent for both methods of patrolling. When the rate of pullovers is decreased to 0.25 pullovers per hour, the rates of failure for all three fleets are increased significantly. At a rate of 0.25 pullovers per hour, the chances that officers patrolling are unable to pullover enough people to outweigh the cost of patrolling are between 13% and 14% for a fleet of three cars, 9% and 10% for a fleet of five cars, and 4% and 5% for a fleet of seven cars. Despite the fact that the rates of failure for the second failure mode are extremely low, the HSRO method of patrolling may still be seen as more significant since it guides officers to significant locations as defined by hot spot analyses.

9.8 FUTURE RECOMMENDATIONS

This research shows new methods of patrolling based on the results of spatial analyses with the ultimate goal of reducing the amount of alcohol-related crashes. Future research may be utilized in a number of ways moving forward, beginning with the implementation of these methodologies, followed by a spatio-temporal analysis to examine how the patterns of crashes vary over time.

9.8.1 Implementation of HSRO

Though this research explains the methodologies of HSRO, implementation has not yet occurred. Contact with jurisdictions that recognize the DDACTS methodologies of improving patrols may be beneficial in sharing resources to implement this research. Through implementation, further studies on the amount of DUI's and alcohol-related crashes that occur on nights of patrolling are may be used to determine validate the research. After small scale studies, this research may be expanded through multiple jurisdictions within the county or state to determine any legitimate results in reducing the amount of alcohol-related crashes

9.8.2 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.

9.9 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. These analyses were used to develop new methods of patrolling for officers to reduce the amount of intoxicated drivers. 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.

REFERENCES

1. Aguero-Valverde, J., & Jovanis, P. P. (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis & Prevention, 38(3), 618-625.
2. Aguero-Valverde, J., & Jovanis, P. P. (2008). Analysis of road crash frequency with spatial models. Transportation Research Record: Journal of the Transportation Research Board, 2061(1), 55-63.
3. Aguero-Valverde, J. (2013). Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates. Accident Analysis & Prevention, 50, 289-297.
4. Ahammed, M., and R. E. Melchers. 1997. "Probabilistic Analysis of Underground Pipelines Subject to Combined Stresses and Corrosion." *Engineering Structures* 19: 988-994.
5. American Association of State Highway Transportation Officials. (2010). Highway Safety Manual, First Edition, Volume 1. Washington, D.C.: American Association of State Highway Transportation Officials.
6. Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. Accident Analysis & Prevention, 41(3), 359-364.
7. Ansari, S., & Kale, K. (2014). Methods for Crime Analysis Using GIS. International Journal of Scientific & Engineering Research, 5(12), 1330-1336.
8. Anselin, L. (1995). Local indicators of spatial association-LISA. Geographical analysis, 27(2), 93-115.
9. BAČKALIĆ, S. (2013). TEMPORAL ANALYSIS OF THE TRAFFIC ACCIDENTS OCCURENCE IN PROVINCE OF VOJVODINA. Transport Problems, 8:1
10. Berning, Amy, Richard Compton, and Kathryn Wochinger. February, 2015. *Results of the 2013-2014 National Roadside Survey of Alcohol and Drug Use by Drivers.* National Highway Traffic Safety Administration, Washington, D.C.: NHTSA's Office of Behavioral Safety Research.
11. Blais, Étienne, François Bellavance, Alexandra Marcil, and Laurent Carnis. 2015. "Effects of Introducing an Administrative .05% Blood Alcohol Concentration Limit on Law Enforcement Patterns and Alcohol-Related Collisions in Canada." *Accident Analysis & Prevention* 101-111.
12. Blazquez, C. A., & Celis, M. S. (2013). A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile. Accident Analysis & Prevention, 50, 304-311.
13. Boots, B., & Tiefelsdorf, M. (2000). Global and local spatial autocorrelation in bounded regular tessellations. Journal of Geographical Systems, 2(4), 319-348.
14. Brunsdon, C., Corcoran, J., & Higgs, G. (2007). Visualising space and time in crime patterns: A comparison of methods. Computers, Environment and Urban Systems, 31(1), 52-75.
15. Carrick, Grady, Ilir Bejleri, and Yiqiang Ouyang. 2014. "Methodological Approach to Spatiotemporal Optimization of Rural Freeway Enforcement in Florida." *Transportation Research Record: Journal of the Transportation Research Board* 2425: 1-9. doi:10.3141/2425-01.

16. Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. Security Journal, 21(1), 4-28.
17. Chen, Y., Berrocal, V. J., Raymond Bingham, C., & Song, P. X. (2014). Analysis of spatial variations in the effectiveness of graduated driver's licensing (GDL) program in the state of Michigan. Spatial and spatio-temporal epidemiology, 8, 11-22.
18. Cheng, Wen, and Simon Washington. 2008. "New Criteria for Evaluating Methods of Identifying Hot Spots." *Transportation Research Record: Journal of Transportation Research Board* 2083: 76-85. doi:10.3141/2083-09.
19. Christoforou, Zoi, Matthew G. Karlaftis, and George Yannis. 2012. "Effects of Alcohol on Speeding and Road Positioning of Yound Drivers- Driving Simulator Study." *Transportation Research Record: Journal of the Transportation Research Board* 2281: 32-42. doi:10.3141/2281-05.
20. Colón, I., & Cutter, H. S. (1983). The relationship of beer consumption and state alcohol and motor vehicle policies to fatal accidents. *Journal of Safety Research, 14*(2), 83-89.
21. Connor, J., Norton, R., Ameratunga, S., & Jackson, R. (2004). The contribution of alcohol to serious car crash injuries. Epidemiology, 15(3), 337-344.
22. De Vlack, Jeremy, Steven Broekx, Inge Liekens, Leo De Nocker, Jos Van Orshoven, and Liesbet Vranken. 2016. "Contrasting Collective Preferences for Outdoor Receation and Substitutability of Nature Areas using Hot Spot Mapping." *Landscape and Urban Planning* 151: 64-78.
23. Dong, Qiao, and Baoshan Huang. 2015. "Failure Probability of Resurfaced Preventative Maintenance Treatments: Investigation into long-Term Pavement Performance Program." *Transportation Research Record* 2481: 65-74.
24. Dundulis, Gintautas, Inga Žutautaitė, Remigijus Janulionis, Eugenijus Ušpuras, Sigitas Rimkevičius, and Mohamed Eid. 2016. "Integrated Failure Probability Estimation Based on Structural Integrety Analysis and Failure Data: Natural Gas Pipeline Case." *Reliability Engineering and System Safety* 156: 195-202.
25. El-Basyouny, K., & Sayed, T. (2009). Accident prediction models with random corridor parameters. Accident Analysis & Prevention, 41(5), 1118-1123.
26. Elder, Randy W., Ruth A. Shults, David A. Sleet, James L. Nichols, Robert S. Thompson, Warda Rajab, and Task Force on Community Preventative Services. 2004. "Effectiveness of Mass Media Campaigns for Reducing Drinking and Driving and Alcohol-Involved Crashes." *American Journal of Preventivbe Medicine* 27 (1): 57-65.
27. Erdogan, S., Yilmaz, I., Baybura, T., & Gullu, M. (2008). Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar. Accident Analysis & Prevention, 40(1), 174-181.
28. Erdogan, S. (2009). Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. Journal of safety research, 40(5), 341-351.
29. Evans, L. (1990). The fraction of traffic fatalities attributable to alcohol.Accident Analysis & Prevention, 22(6), 587-602.

30. Farmer, C. M., & Williams, A. F. (2005). Temporal factors in motor vehicle crash deaths. Injury Prevention, 11(1), 18-23.
31. Fell, James C., Elizabeth A. Langston, John H. Lacey, A. Scott Tippets, and Ray Cotton. 2008. *Evaluation of Seven Publicised Enforcement Demonstration Programs to Reduce Impaired Driving: Georgia, Louisiana, Pennsylvania, Tennessee, Texas, Indiana, and Michigan.* NHTSA.
32. Fell, James C., Geetha Waehrer, Robert B. Voas, Amy Auld-Owens, Katie Carr, and Karen Pell. 2014. "Effects of Enforcement Intensity on Alcohol Impaired Driving Crashes." *Accident Analysis & Prevention* 181-196.
33. Gary, S. L., Aultman-Hall, L., McCourt, M., & Stamatiadis, N. (2003, September). Consideration of driver home county prohibition and alcohol-related vehicle crashes. *Accident Analysis & Prevention, 35*(5), 641-648.
34. Geller, E. Scott, and Ryan C. Smith. 2014. "Field Investigation of College Student Alcohol Intoxication and Return Transportation from At-Risk Drinking Locations." *Transportation Research Record: Journal of the Transportation Research Board* 2425: 67-73. doi:10.3141/2425-09.
35. Geedipally, S., & Lord, D. (2010). Identifying hot spots by modeling single-vehicle and multivehicle crashes separately. Transportation Research Record: Journal of the Transportation Research Board, (2147), 97-104.
36. Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. Geographical analysis, 24(3), 189-206.
37. Golob, T. F., & Recker, W. W. (2004). A method for relating type of crash to traffic flow characteristics on urban freeways. Transportation Research Part A: Policy and Practice, 38(1), 53-80.
38. Goodwin, A., Thomas, L., Kirley, B., Hall, W., O'Brien, N., & Hill, K. (2015). *Countermeasures that work: A highway safety countermeasure guide for state highway safety officers, Eighth Edition.* Washington, D.C.: National Highway Traffic Safety Administration.
39. Govenors Highway Safety Association (GHSA). 2016. *Impaired Driving*. Accessed October 11, 2016
40. Harrison, Emily L.R., and Mark T. Fillmore. 2005. "Are bad Drivers More Impaired by Alcohol? Sober Driving Precision Predicts Impairment from Alcohol in a Simulated Driving Task." *Accident Analysis & Prevention* 37: 882-889.
41. Huang, H., Abdel-Aty, M., & Darwiche, A. (2010). County-level crash risk analysis in Florida: Bayesian spatial modeling. Transportation Research Record: Journal of the Transportation Research Board, (2148), 27-37.
42. Imprialou, M. I. M., Quddus, M., & Pitfield, D. E. (2014). High accuracy crash mapping using fuzzy logic. Transportation Research Part C: Emerging Technologies, 42, 107-120.
43. Ivan, J. N., Pasupathy, R. K., & Ossenbruggen, P. J. (1999). Differences in causality factors for single and multi-vehicle crashes on two-lane roads. Accident Analysis & Prevention, 31(6), 695-704.

44. Jia, Keqin, Judy Fleiter, Mark King, Wenjun Ma, Jing Lei, and Jianzhen Zhang. 2016. "Alcohol-Related Driving in China: Countermeasure Implications of Research Conducted in Two Cities." *Accident Analysis & Prevention.*
45. Johnston, Robert J., and Mahesh Ramachandran. 2014. "Modeling Spatial Patchiness and Hot Spots in Stated Preference Willingness to Pay." *Environ Resource Econ* 59: 363-387. doi:10.1007/s10640-013-9731-2.
46. Kang, S., Spiller, M., Jang, K., Bigham, J. M., & Seo, J. (2012). Spatiotemporal Analysis of Macroscopic Patterns of Urbanization and Traffic Safety. Transportation Research Record: Journal of the Transportation Research Board, 2318(1), 45-51.
47. Karim, M. A., Wahba, M. M., & Sayed, T. (2013). Spatial Effects on Zone-Level Collision Prediction Models. Transportation Research Record: Journal of the Transportation Research Board, 2398(1), 50-59.
48. Kenkel, D. S. (1993). Drinking, Driving and Deterrence: The Effectiveness and Social Costs of Alternative Policies. *The Journal of Law & Economics, 36*, 877-913.
49. Kennedy, B. P., Isaac, N. E., & Graham, J. D. (1996). The role of heavy drinking in the risk of traffic fatalities. Risk analysis, 16(4), 565-569.
50. Khan, G., Qin, X., & Noyce, D. A. (2008). Spatial analysis of weather crash patterns. Journal of Transportation Engineering, 134(5), 191-202.
51. Khan, G., Santiago-Chaparro, K. R., Qin, X., & Noyce, D. A. (2009). Application and integration of lattice data analysis, network K-functions, and geographic information system software to study ice-related crashes. Transportation Research Record: Journal of the Transportation Research Board, 2136(1), 67-76.
52. Kim, K., & Levine, N. (1996). Using GIS to improve highway safety. Computers, environment and urban systems, 20(4), 289-302.
53. Kim, K., Pant, P., & Yamashita, E. (2010). Accidents and accessibility. Transportation Research Record: Journal of the Transportation Research Board, 2147(1), 9-17.
54. Kuo, P. F., Lord, D., & Walden, T. D. (2013). Using geographical information systems to organize police patrol routes effectively by grouping hotspots of crash and crime data. Journal of Transport Geography, 30, 138-148.
55. Lee, J., Abdel-Aty, M., & Jiang, X. (2014). Development of zone system for macro-level traffic safety analysis. Journal of Transport Geography, 38, 13-21.
56. Lees, B. (2006). “The spatial analysis of spectral data: Extracting the neglected data.” *Applied GIS, 2*(2), 14.1-14.13.
57. Leon, David De, and Oscar Flores Macías. 2005. "Effect of Spaial Correlation on the Failure Probability of Pipelines Under Corrosion." *Pressure Vessels and Piping* 82: 123-128.
58. Levine, N., Kim, K. E., & Nitz, L. H. (1995). Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns. Accident Analysis & Prevention, 27(5), 663-674.
59. Li, L., Zhu, L., & Sui, D. Z. (2007). A GIS-based Bayesian approach for analyzing spatial–temporal patterns of intra-city motor vehicle crashes. Journal of Transport Geography, 15(4), 274-285.

60. Loo, B. P., & Yao, S. (2013). The identification of traffic crash hot zones under the link-attribute and event-based approaches in a network-constrained environment. Computers, Environment and Urban Systems, 41, 249-261.
61. Loukaitou-Sideris, A., Liggett, R., & Sung, H. G. (2007). Death on the crosswalk A study of pedestrian-automobile collisions in los angeles. Journal of Planning Education and Research, 26(3), 338-351.
62. Maistros, A. R., Schneider, W. H., Beverley, R., & Stakleff, B. A. (2014). Case Study Evaluation of Saturation and Corridor Enforcement Patrols.Transportation Research Record: Journal of the Transportation Research Board, 2425(1), 10-16.
63. Maistros, A. R., & Schneider IV, W. H. (2016). A Comparison of Overtime Patrol Stops Made Inside and Out Of Cluster Identified Hotspots. *Journal of the Transportation Research Board, accepted not yet published*.
64. Meesmann, Uta, Heike Martensen, and Emmanuelle Dupont. 2015. "Impact of Alcohol Checks and Social Norm on Driving Under the Influence of Alcohol (DUI)." *Accident Analysis and Prevention* 251-261.
65. Mehdi, M. R., Kim, M., Seong, J. C., & Arsalan, M. H. (2011). Spatio-temporal patterns of road traffic noise pollution in Karachi, Pakistan. Environment international, 37(1), 97-104.
66. Meliker, J. R., Maio, R. F., Zimmerman, M. A., Kim, H. M., Smith, S. C., & Wilson, M. L. (2004). Spatial analysis of alcohol-related motor vehicle crash injuries in southeastern Michigan. Accident Analysis & Prevention, 36(6), 1129-1135.
67. Miller, John S., and Duane Karr. 1998. "Experimental Application of Global Positioning System to Locate Motor Vehicle Crashes. Impact on Time and Accuracy." *Transportation Research Record* 1625: 41-49.
68. Mitra, S. (2009). Spatial autocorrelation and Bayesian spatial statistical method for analyzing intersections prone to injury crashes. Transportation Research Record: Journal of the Transportation Research Board, 2136(1), 92-100.
69. Moskowitz, H., & Fiorentino, D. (2000). A review of the literature on the effects of low doses of alcohol on driving-related skills (No. HS-809 028,).
70. Moskowitz, H., & Robinson, C. (1987). Driving-Related Skills in Impairment at Low Blood Alcohol Levels. *Alcohol Drugs Traffic Safety*, 79-86.
71. Naimi, T. S., Brewer, R. D., Mokdad, A., Denny, C., Serdula, M. K., & Marks, J. S. (2003). Binge drinking among US adults. Jama, 289(1), 70-75.
72. National Highway Traffic Safety Administration [NHTSA]. 2014. Fatality Analysis Reporting System: Query FARS Data, Available from http://www-fars.nhtsa.dot.gov/Main/index.aspx (accessed October 13, 2014)
73. National Highway Traffic Safety Administration (NHTSA), (2015). Fatality Analysis Reporting System: Fatal Crash Trends; 2015. http://www-fars.nhtsa.dot.gov/Main/index.aspx. Accessed April 11, 2015.

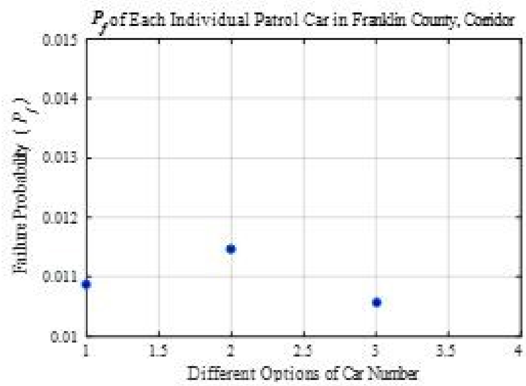
74. Nelson, Trisalyn A., and Barry Boots. 2008. "Detecting Spatial Hot Spots in Landscape Ecology." *Ecography* 31: 556-566. doi:10.1111/j.0906-7590.2008.05548.x.
75. Ng, K. S., Hung, W. T., & Wong, W. G. (2002). An algorithm for assessing the risk of traffic accident. Journal of safety research, 33(3), 387-410.
76. Nicholson, A. (1998). Analysis of spatial distributions of accidents. Safety science, 31(1), 71-91.
77. NOLO. 2016. *Ohio OVI Laws.* http://dui.drivinglaws.org/ohio.php.
78. Ohio Department of Transportation (ODOT). 2016. Transportation Mapping System: Data Downloads. Ohio Department of Transportation
79. Ohio Department of Public Safety (ODPS). 2015. Crash Extracts 2012-2015. Ohio Department of Public Safety
80. Oltedal, S., & Rundmo, T. (2007). Using cluster analysis to test the cultural theory of risk perception. Transportation research part F: traffic psychology and behaviour, 10(3), 254-262.
81. Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis, 27*(4), 286-306.
82. Peck, R. C., Gebers, M. A., Voas, R. B., & Romano, E. (2008). The relationship between blood alcohol concentration (BAC), age, and crash risk.Journal of Safety Research, 39(3), 311-319.
83. Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., & Wets, G. (2012). Application of different exposure measures in development of planning-level zonal crash prediction models. Transportation Research Record: Journal of the Transportation Research Board, 2280(1), 145-153.
84. Plug, C., Xia, J. C., & Caulfield, C. (2011). Spatial and temporal visualisation techniques for crash analysis. Accident Analysis & Prevention, 43(6), 1937-1946.
85. Ponicki, W. R., Gruenewald, L. G., & Remer, L. (2013, June). Spatial panel analyses of alcohol outlets and motor vehicle crashs in California: 1999-2008. *Accident Analysis & Prevention, 55*, 135-143.
86. Prasannakumar, V., Vijith, H., Charutha, R., & Geetha, N. (2011). Spatio-temporal clustering of road accidents: GIS based analysis and assessment. Procedia-Social and Behavioral Sciences, 21, 317-325.
87. Pulugurtha, S. S., Krishnakumar, V. K., & Nambisan, S. S. (2007). New methods to identify and rank high pedestrian crash zones: An illustration. Accident Analysis & Prevention, 39(4), 800-811.
88. Pulugurtha, S. S., Duddu, V. R., & Kotagiri, Y. (2013). Traffic analysis zone level crash estimation models based on land use characteristics. Accident Analysis & Prevention, 50, 678-687.
89. Qin, X., Ivan, J. N., & Ravishanker, N. (2004). Selecting exposure measures in crash rate prediction for two-lane highway segments. Accident Analysis & Prevention, 36(2), 183-191.
90. Quddus, M. A. (2008). Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accident Analysis & Prevention, 40(4), 1486-1497.

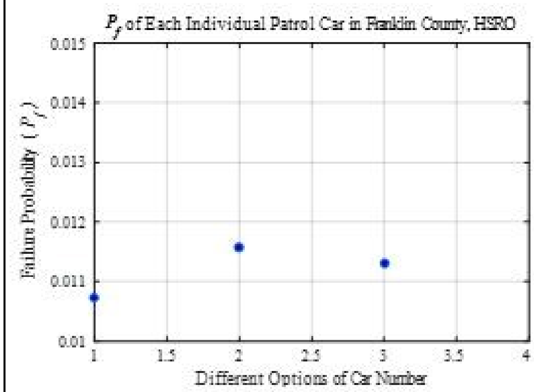
91. Ratcliffe, J.H., and M.J. McCullagh. 2001. "Chasing Ghosts? Police Perception of High Crime Areas." *Brit. J. Criminol* 41, 330-341.
92. Sanem, Julia R., Darin J. Erickson, Patricia C. Rutledge, Kathleen M. Lenk, Toben F. Nelson, Rhonda Jones-Webb, and Traci L. Toomey. 2015. "Association Between Alcohol-Impaired Driving Enforcement-Related Strategies and Alcohol-Impaired Driving." *Accident Analysis and Prevention* 104-109.
93. SANET, Spatial Analysis along Networks (Ver.4.1). Atsu Okabe, Kei-ichi Okunuki and SANET Team, Tokyo, Japan.
94. Scheiner, J., & Holz-Rau, C. (2011). A residential location approach to traffic safety: Two case studies from Germany. Accident Analysis & Prevention, 43(1), 307-322.
95. Schneider, R. J., Khattak, A. J., & Zegeer, C. V. (2001). A proactive method of improving pedestrian safety proactively with geographic information systems: example from a college campus. Transportation Research Record: Journal of the Transportation Research Board, 1773(1), 97-107.
96. Schneider, R. J., Ryznar, R. M., & Khattak, A. J. (2004). An accident waiting to happen: a spatial approach to proactive pedestrian planning. Accident Analysis & Prevention, 36(2), 193-211.
97. Schneider IV, W., & Savolainen, P. (2011). Comparison of severity of motorcyclist injury by crash types. Transportation Research Record: Journal of the Transportation Research Board, (2265), 70-80.
98. Shepard, M.D., C.H. 1903. *Quarterly Journal of Inebriety* 25: 407.
99. Songchitruksa, P., & Zeng, X. (2010). Getis-Ord spatial statistics to identify hot spots by using incident management data. Transportation Research Record: Journal of the Transportation Research Board, 2165(1), 42-51.
100. Sukhai, A., & Jones, A. P. (2013). Understanding geographical variations in road traffic fatalities in South Africa. South African Geographical Journal, 95(2), 187-204.
101. Tay, R. (2005). General and Specific Deterrent Effects of Traffic Enforcement. *Journal of Transport Economics and Policy, 39*, 209-223.
102. Thakali, L., Kwon, T. J., & Fu, L. Identification of crash hotspots using kernel density estimation and kriging methods: a comparison. Journal of Modern Transportation, 1-14.
103. Timmerman, Mary Ann, E. Scott Geller, Kent E. Glindemann, and Angela K. Fournier. 2003. "Do the Designated Drivers of College Students Stay Sober?" *Journal of Safety Research* 34: 127-133.
104. Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic geography, 234-240.
105. Tompson, L., & Townsley, M. (2010). (Looking) Back to the Future: using space-time patterns to better predict the location of street crime. International journal of police science & management, 12(1), 23-40.
106. Treno, A. J., Johnson, F. W., Remer, L. G., & Gruenewald, P. J. (2007). The impact of outlet densities on alcohol-related crashes: a spatial panel approach. Accident Analysis & Prevention, 39(5), 894-901.

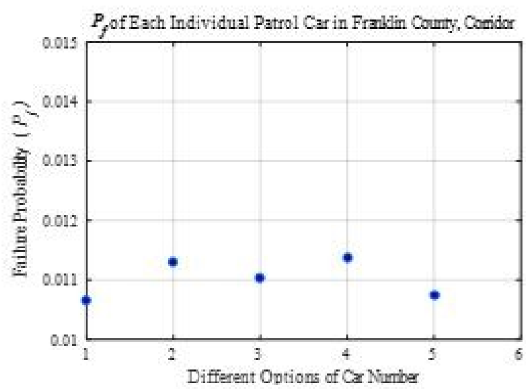
107. Truong, L. T., & Somenahalli, S. V. (2011). Using GIS to identify pedestrian-vehicle crash hot spots and unsafe bus stops. Journal of Public Transportation, 14(1), 99-114.
108. Turner, Patricia A, and Nevine Georggi. 2001. "Analysis of Alcohol-Related Motorcycle Crashes in Florida and Recommended Countermeasures." *Transportation Research Record: Transportation Research Board* 1779: 189-196.
109. U.S. Bureau of Labor Statistics. 2015. "Occupational Employment and Wages, May 2015- 33-3051 Police and Sherrif's Patrol Officers."
110. United States Census Bureau. 2013. "State & County Quick Facts." Available from http://quickfacts.census.gov/qfd/states/39/39151.html (accessed July 29, 2013)
111. U.S. Department of Transportation, National Highway Traffic Safety Administration. (2009). Data driven approaches to crime and traffic safety (DDACTS) Operational Guidelines (DOT HS 811 185).
112. Vandenbulcke, G., Thomas, I., & Int Panis, L. (2014). Predicting cycling accident risk in Brussels: A spatial case–control approach. Accident Analysis & Prevention, 62, 341-357.
113. Vlahogianni, E. I., Karlaftis, M. G., Golias, J. C., & Halkias, B. M. (2010). Freeway Operations, Spatiotemporal-Incident Characteristics, and Secondary-Crash Occurrence. Transportation Research Record: Journal of the Transportation Research Board, 2178(1), 1-9.
114. Voas, R. B., Tippetts, A. S., & Fell, J. C. (2003). Assessing the effectiveness of minimum legal drinking age and zero tolerance laws in the United States. Accident Analysis & Prevention, 35(4), 579-587.
115. Vollrath, M., H.-P. Krüger, and R. Löbmann. 2005. "Driving Under the Influence of Alcohol in Germany and the Effect of Relaxing BAC Law." *Transportation Research Part E* 41: 377-393.
116. Wang, X. C., Kockelman, K. M., & Lemp, J. D. (2012). The dynamic spatial multinomial probit model: analysis of land use change using parcel-level data. Journal of Transport Geography, 24, 77-88.
117. Wang, Y., & Kockelman, K. M. (2013). A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. Accident Analysis & Prevention, 60, 71-84.
118. Wong, D. W. (1999). Geostatistics as measures of spatial segregation. Urban Geography, 20(7), 635-647.
119. Wulder, M., & Boots, B. (1998). Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic. *International Journal of Remote Sensing, 19*(11), 2223-2231.
120. Xu, C., Wang, W., Liu, P., Guo, R., & Li, Z. (2014). Using the Bayesian updating approach to improve the spatial and temporal transferability of real-time crash risk prediction models. Transportation Research Part C: Emerging Technologies, 38, 167-176.
121. Yamada, I., & Thill, J. C. (2004). Comparison of planar and network K-functions in traffic accident analysis. Journal of Transport Geography, 12(2), 149-158.

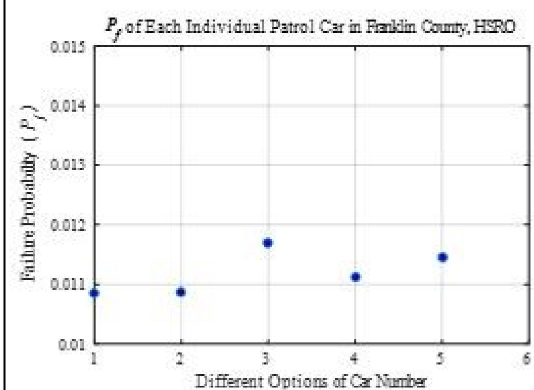
122. Yiannakoulias, N., Bennet, S. A., & Scott, D. M. (2012). Mapping commuter cycling risk in urban areas. Accident Analysis & Prevention, 45, 164-172.
123. Young, J., & Park, P. Y. (2014). Hotzone identification with GIS-based post-network screening analysis. Journal of Transport Geography, 34, 106-120.
124. Yu, R., & Abdel-Aty, M. (2013). Utilizing support vector machine in real-time crash risk evaluation. Accident Analysis & Prevention, 51, 252-259.
125. Yuhua, Dong, and Yu Datao. 2005. "Estimation of Failure Probability of Oil and Gas Transmission Pipelines by Fuzzy Fault Tree Analysis." *Journal of Loss Prevention in the Process Industries* 18: 83-88.

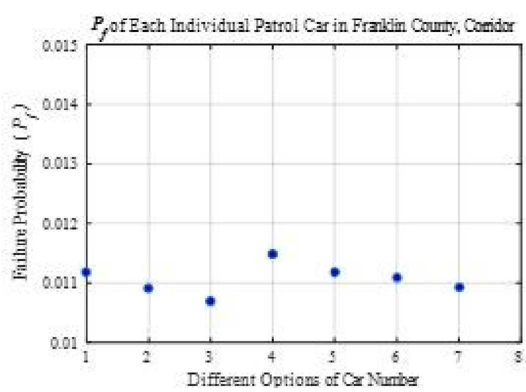
APPENDIX A. FAILURE PROBABILITY MODE 2 RESULTS

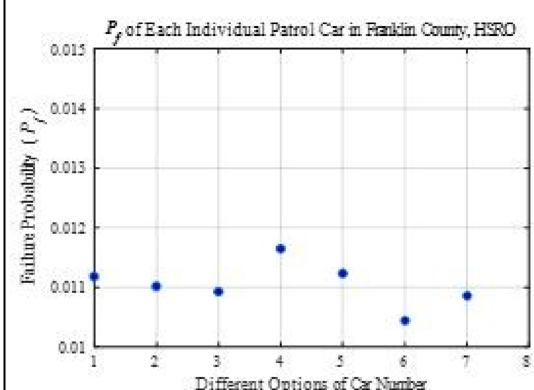

P_f of Each Individual Patrol Car in Franklin County, Corridor
Failure Probability (P_f)
0.015
0.014
0.013
0.012
0.011
0.01
1
1.5
2
2.5
3
3.5
4
Different Options of Car Number


P_f of Each Individual Patrol Car in Franklin County, HSRO
Failure Probability (P_f)
0.015
0.014
0.013
0.012
0.011
0.01
1
1.5
2
2.5
3
3.5
4
Different Options of Car Number


P_f of Each Individual Patrol Car in Franklin County, Corridor
Failure Probability (P_f)
0.015
0.014
0.013
0.012
0.011
0.01
1
2
3
4
5
6
Different Options of Car Number


P_f of Each Individual Patrol Car in Franklin County, HSRO
Failure Probability (P_f)
0.015
0.014
0.013
0.012
0.011
0.01
1
2
3
4
5
6
Different Options of Car Number


P_f of Each Individual Patrol Car in Franklin County, Corridor
Failure Probability (P_f)
0.015
0.014
0.013
0.012
0.011
0.01
1
2
3
4
5
6
7
8
Different Options of Car Number


P_f of Each Individual Patrol Car in Franklin County, HSRO
Failure Probability (P_f)
0.015
0.014
0.013
0.012
0.011
0.01
1
2
3
4
5
6
7
8
Different Options of Car Number

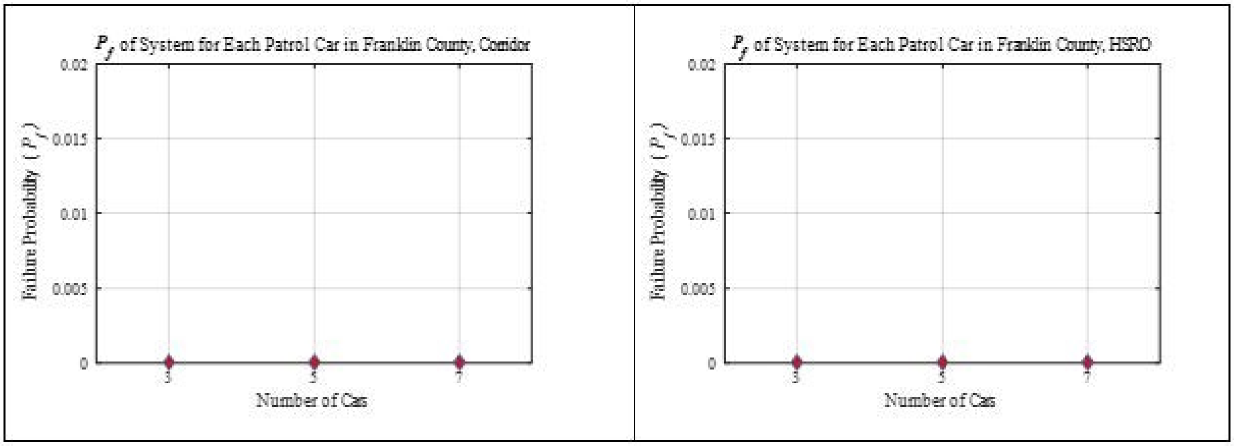
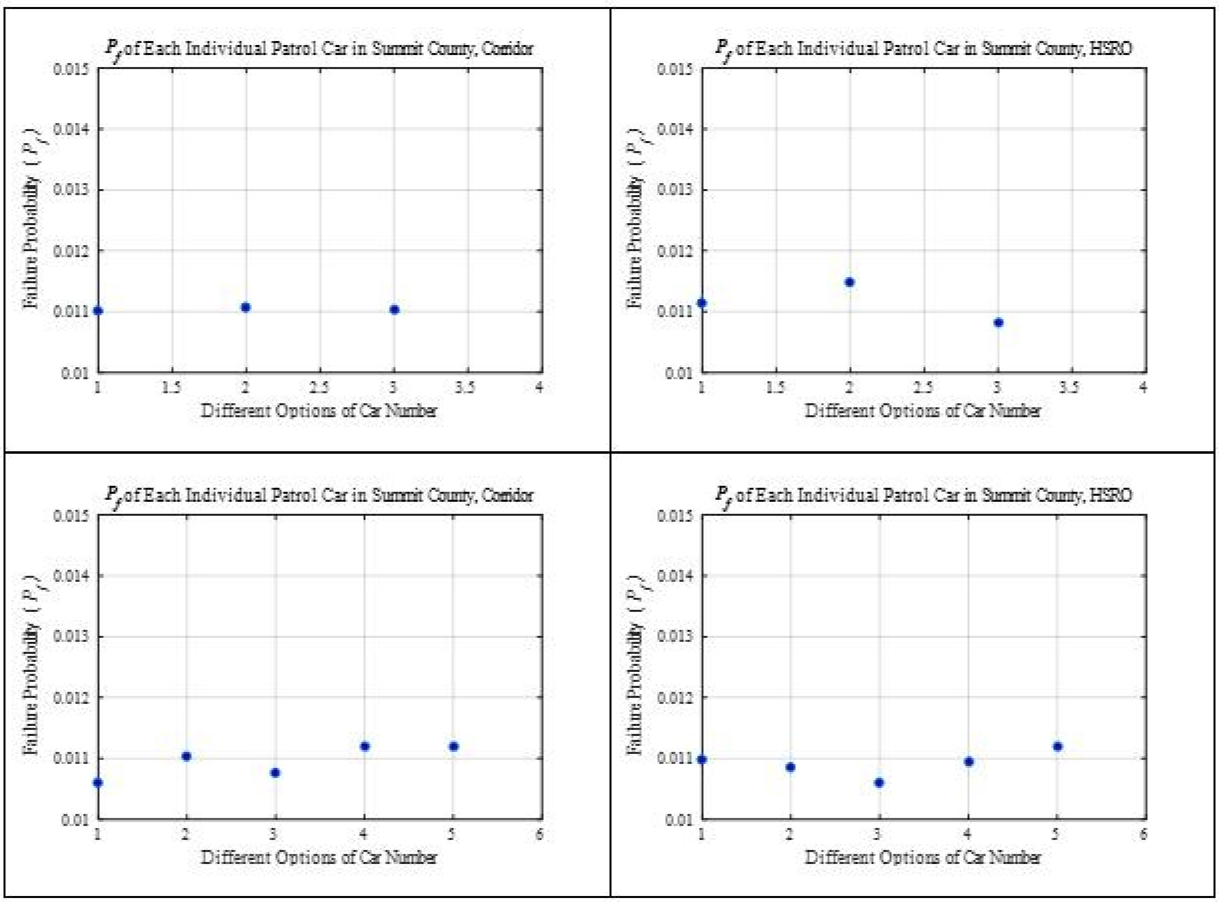


Figure A.1: Results of Second Failure Mode for Franklin County.



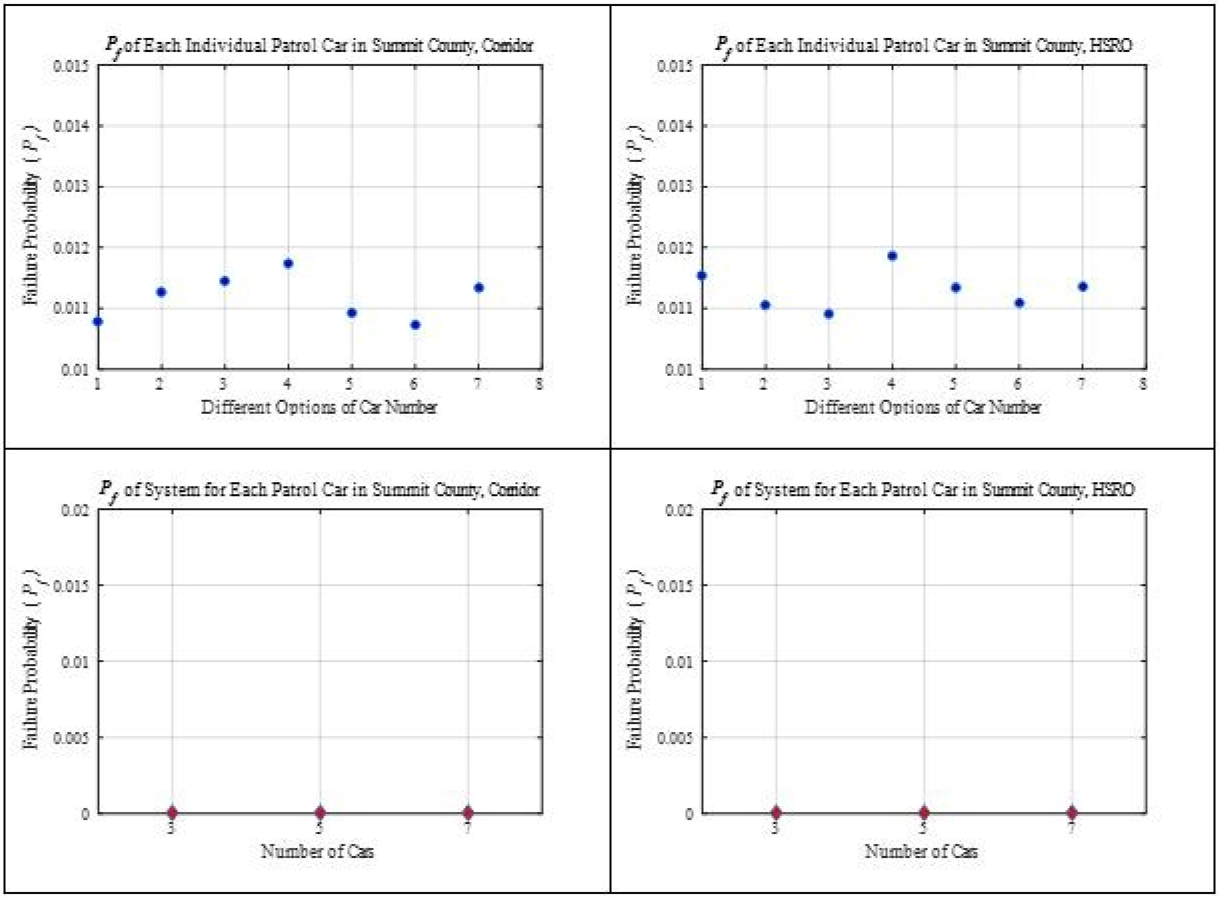
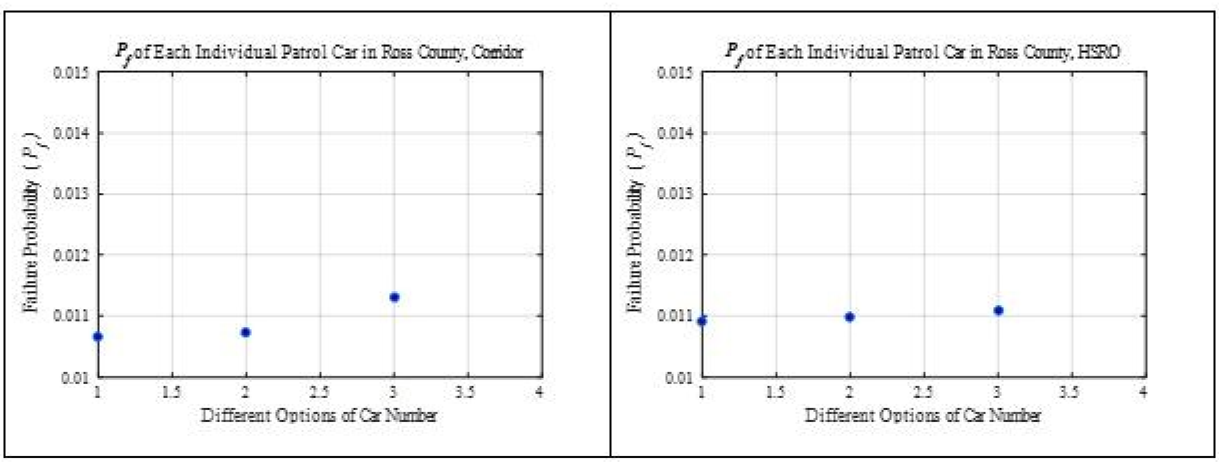


Figure A.2: Results of Second Failure Mode for Summit County.



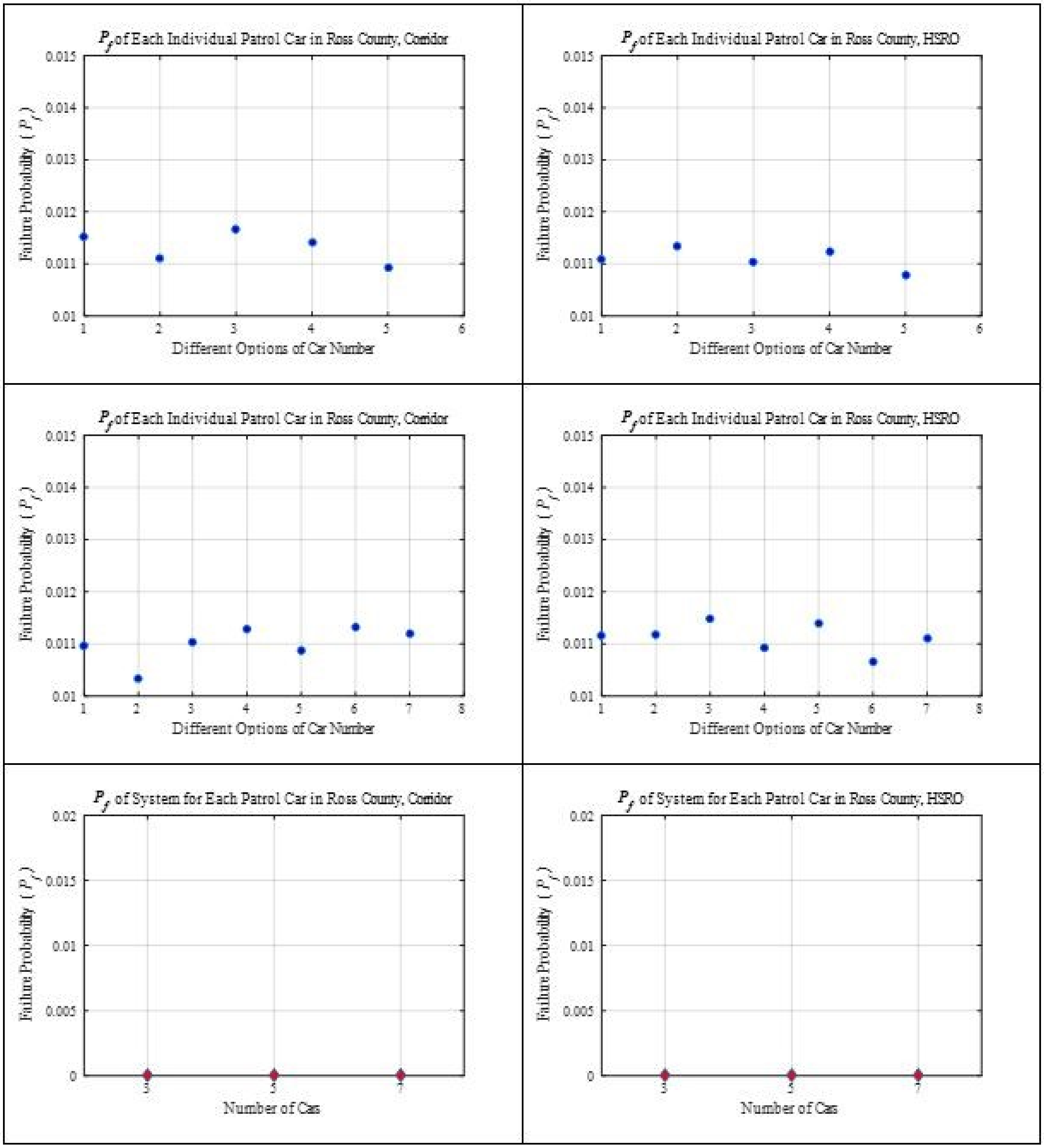


Figure A.3: Results of Second Failure Mode for Ross County.