2021

An Exploration of The Application of Spatial Network Screening Methods On Iowa Rural Road Crashes

Rosanna Maria Novellino Suarez
South Dakota State University

AN EXPLORATION OF THE APPLICATION OF SPATIAL NETWORK SCREENING METHODS ON IOWA RURAL ROAD CRASHES

BY

ROSANNA MARIA NOVELLINO SUAREZ

A thesis submitted in partial fulfillment of the requirements for the

Master of Science

Major in Civil Engineering

South Dakota State University

2021

THESIS ACCEPTANCE PAGE

Rosanna Novellino

This thesis is approved as a creditable and independent investigation by a candidate for the master's degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Michael Pawlovich
Advisor Date

Nadim Wehbe
Department Head Date

Nicole Lounsbery, PhD
Director, Graduate School Date

I dedicate this thesis to those who have unconditionally supported me and encouraged me to continue with my journey: my family and friends. I would also like to thank my advisor, Dr. Michael Pawlovich, for his guidance during graduate school and for this effort, Dr. Suzette Burckhard, for her advice during my university journey, and Dr. Christopher Saunders for his statistical analysis assistance. May the information within this thesis be useful for future evaluation of roadway safety and evaluation methods.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

AN EXPLORATION OF THE

APPLICATION OF SPATIAL NETWORK SCREENING METHODS

ON IOWA RURAL ROAD CRASHES

ROSANNA MARIA NOVELLINO SUAREZ

2021

Safety on the roadway system is important due to its usage on mobility and accessibility, especially on rural roads in the state of Iowa. Single vehicle run off road crashes have been increasing in the United States and studies and research has increased due to the concern with those. For this effort, a spatial-temporal method of traffic safety network screening is utilized in order to evaluate the concerning type of crashes in particular locations. The study of single vehicle run off road crashes using the proposed method is important since distributions and clusters of crashes along roadways can be observed and further evaluations can be performed.

CHAPTER 1. MOTIVATION AND OBJECTIVE

The roadway system is a valuable public resource that enables both mobility along highways and streets and access to public and private properties along the highways and streets (*1-2*). Regarding mobility, roadway professionals seek to provide maximum appropriate capacity and speeds to serve the needs of traffic seeking to progress from one location to another. Regarding accessibility, roadway professionals seek to provide appropriate access to properties. The appropriate levels of mobility and accessibility are generally based on the roadway functional classification, with the overall goal of safe and efficient operations on the roadway system maximize the value and serve users properly. However, due to the somewhat conflicting goals of mobility and accessibility, conflict often arises when inappropriate access is provided. These conflicts degrade safety and efficiency in the form of crashes, near misses, swerving, hard braking, increased delay and traffic congestion, lower speeds, as well as other results. Figures 1.1 and 1.2 include representations of the interactions between mobility and accessibility (*3*). Several studies and research have been conducted to evaluate the problem (*5-10*), but most have not been resolved.

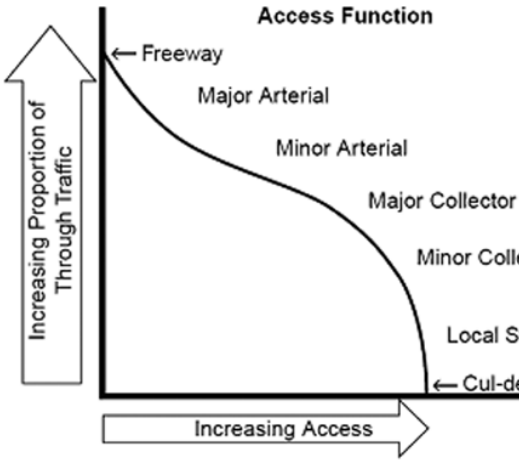

Figure 1.1: Roadway functional hierarchy *(2)*.

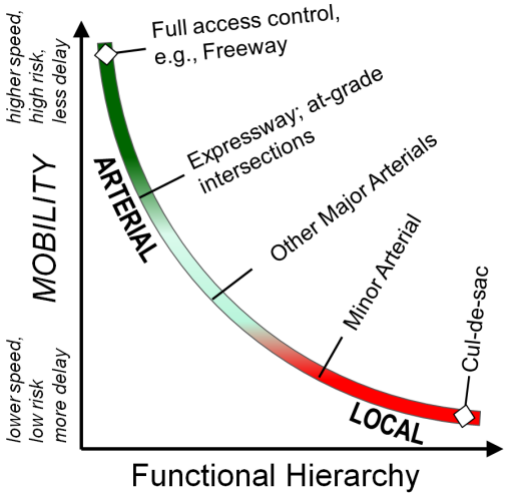

Figure 1.2: Access Control Hierarchy *(4)*.

When areas of concern are not properly evaluated, suitable solutions cannot be offered, and improvements do not occur for future traffic conditions (*5, 11-14*). Identifying roadway needs is essential when improving safety measurements (*12-14*). Roadway needs should be evaluated on existing networks but also prior to construction of network changes or additions to anticipate safety impacts. Implementing safety planning and measurements in transportation have become an essential step internationally and required in the United States (*14*).

Network screening is an essential tool in this process since it allows the identification of sites where safety measures can be applied and improvements can be made (*5, 9, 12-17*). Transportation agencies might address identified locations through improvements to geometric, signage, access, or other options. Multiple network screening methods exist, each with benefits and deficiencies. Most existing methods rely on connecting crashes to the roadway network which can be problematic in some cases.

Prior to development of network screening methods, identification of roadway sections needing improvement was problematic. During the 1970s, states were required by federal statute to "develop a highway safety improvement program that would be able to reduce the number and severity of crashes". To ease implementation, a federal Hazard Elimination Program (HEP) was initiated within the official highway safety act of 1986. After that, states developed Hazard Evaluation Systems (HES) which were used to identify sites with potential safety problems. This was the first-time locations with concern were evaluated and where network screening was basically born (18).

Also, Geographic Information Systems (GIS) and spatial and temporal analysis has been developed and used to identify crashes (*15, 17, 19, 21, 22, 26*). This is possible due to GIS's capabilities to collect information, integrated, and offer its visualization (*20*). The use of GIS has offered simplifications when managing and evaluating crash data (*21-23*). Several studies have used this application to develop models based on methodology and technology in order to obtain reliable risk estimates (*22*).

Using GIS, studies have applied clustering to crashes (*17, 19*). The method includes all crashes located in the area where the cluster is created as shown in Figure 1.3. In the figure, there are multiple segments and crashes shown as blue dots. The gray segments are representing local roads while the red segment represents US 18. In this case, only US 18 was intended to be evaluated. Due to the circular selection region, many unrelated crashes are included along the segment to be evaluated.

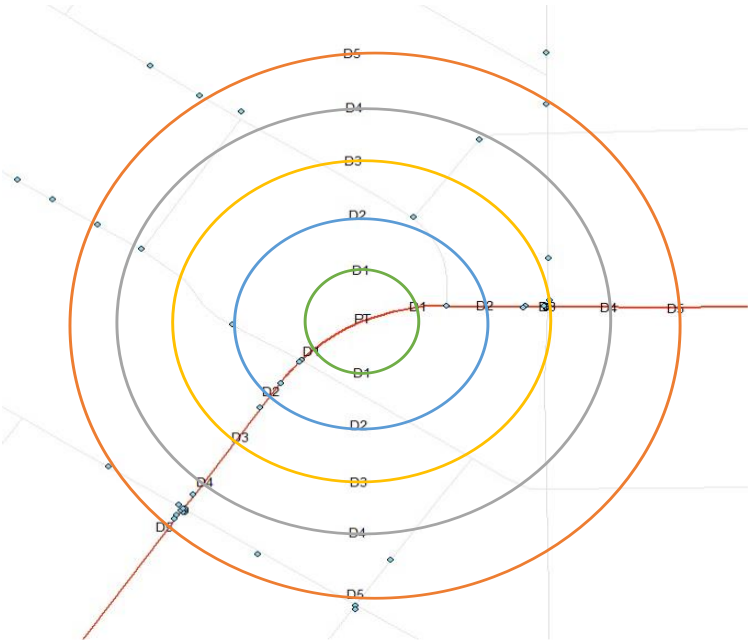

Figure 1.3: Evaluated Area Using General Clustering Method.

A modification of that approach is to include only the crashes located along US 18 by selecting only crashes along the roadway section as shown in Figure 1.4. As it can be observed, those crashes that belong to different roads are not included in the selection. While most existing methods initiate with the roadway network, an alternate method is to initiate with the crash locations and relate these to the network links (segments) and nodes (intersections) where they are located.

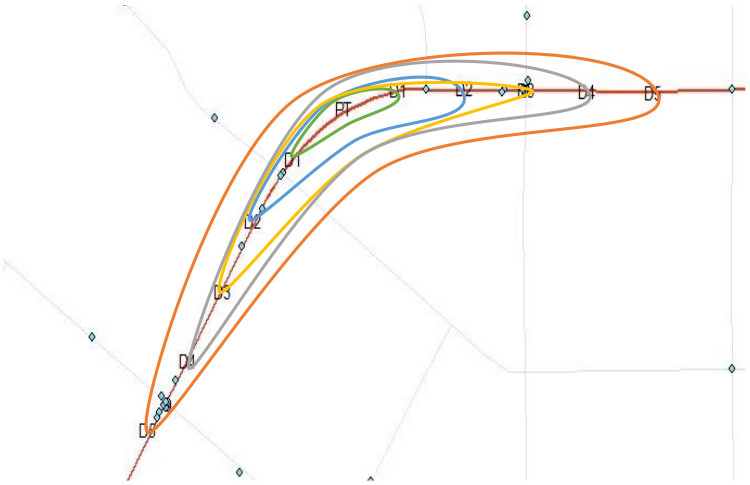

Figure 1.4: Evaluated Area Using Proposed Clustering Method.

In the United States, run-off-road (ROR) crashes are frequent in rural areas (*24-26*). About 30 people die every day due to rural ROR crashes (*26*). ROR crashes result in one-third of traffic fatalities on the rural road network making them a major problem in transportation safety (*25*). From 2016 to 2018, fifty-one percent of traffic fatalities in the United States were composed of roadway departures or ROR crashes (*26*).

Countermeasures to ROR crashes involve helping drivers maintain their lanes, reducing the potential of crashes due to lane departure, and minimizing the severity of crash occurrence (*26*). For example, shoulder or center rumble strips assist drivers to maintain their lanes, removal of roadside objects reduces the potential of crash occurrence, and the addition of breakaway features for signposts reduces the severity of crashes. However, first, in order to apply any countermeasures, the locations where this type of crashes are occurring must be identified.

The objective of the current exploration is to use spatial temporal techniques to evaluate rural and run-off-road crashes along coincidental network connections. By applying that method, the primary goal is to identify clusters and evaluate crash

distributions along the segments. The objective will be achieved by using ArcGIS, Microsoft Excel, R, and Google Earth.

ArcGIS will enable the combination of the crash and road data to visualize and classified the crashes. Crash locations, including all types of segments and intersections, and crash densities for all crash types will be presented on the map. The inclusion of those in the program, allows categorization and crash selection by including different fields, allowing the creation of new data sets. The data sets can include a particular type or manner of crash that requires further evaluation. Using Python, the crash points located along the segments can then be related to each other and crash counts can be generated.

From a selected point along the segment, crash counts can be obtained using the appropriate sight distance for the selected segment, and the location for those are also detected. With the crash count and the location of each selected area, different graphs were created with Microsoft Excel, R, and Google Earth. The intend behind the creation of the graphs was to allocate clusters among the evaluated segments. The three programs were used to compare the outputs of each, and to confirm the existence of clusters after applying the proposed method. With crash clusters available, traffic and roadway characteristics can be included in the analysis, and sites can be prioritized for civil engineering purposes. The efficient completion of the current exploration could contribute with practitioners who intent to apply network screening for crash evaluation.

CHAPTER 2. LITERATURE REVIEW

Traffic volume growth has been relatively consistent over the past several decades. As shown in Figure 2.1, except for a recession-related decline during the 2007-2012 timeframe and the recent COVID-related decline of 2020, national traffic volumes have shown a relentless upward trend.

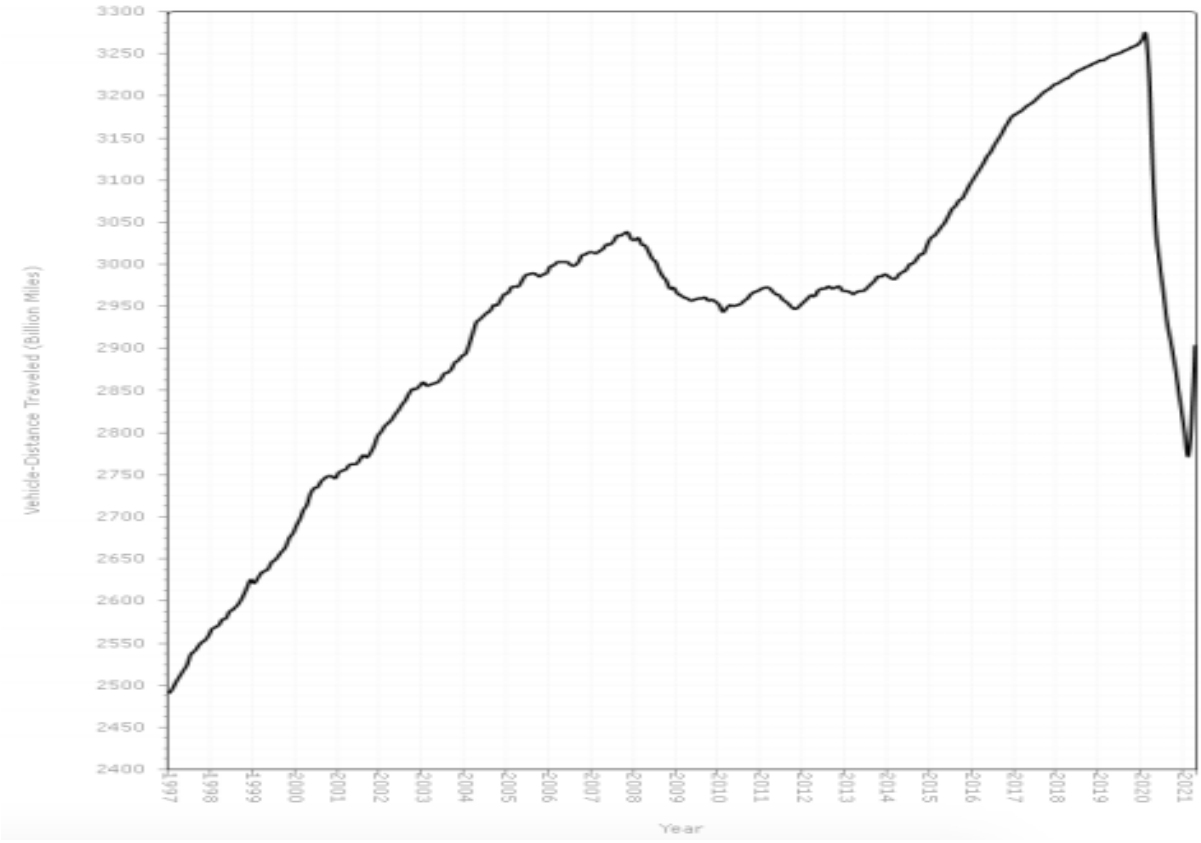

Figure 2.1: General Growth of Traffic in the United States (*27)*.

For Iowa, as shown in Figure 2.2, traffic growth has also increased steadily, not just for highway traffic but also other modes. The Highway VMT category has shown the most significant percentage increase, and it includes automobiles, pickup trucks, and motorcycles.

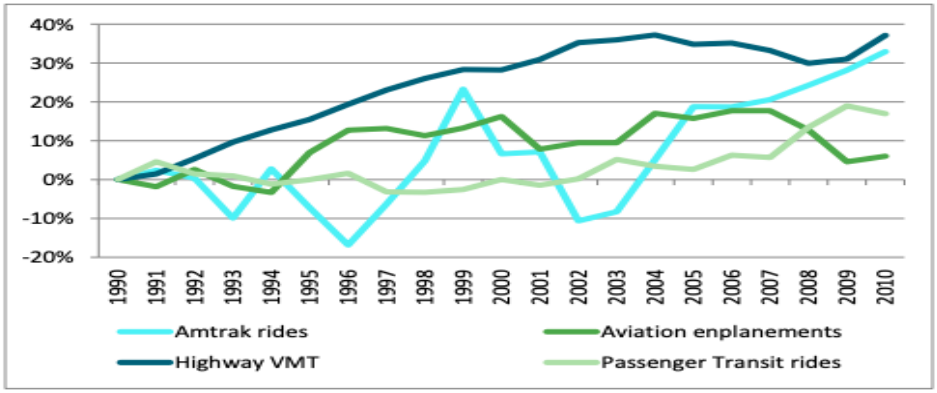

Figure 2.2: General Growth of Transportation in Iowa *(28)*.

A negative aspect of traffic volume growth is a corresponding increase of road traffic crashes, which have been shown to be significantly related to traffic volume (*6, 20, 23, 24, 29-46, 49*). Additionally, many of these crashes are severe, involving either fatalities or major injuries (*8, 19, 20, 29*). Approximately 1.2 million lives lost worldwide and roughly 38,000 annual fatalities in the United States due to road crashes (*20, 24, 25, 46*). However, fatalities on Iowa roadways have steadily decreased over the past several decades despite an increase in traffic volume (*28*). Thus, as crash frequency has been increasing with the growth in traffic, traffic safety is a public and socioeconomic concern (*6, 20, 46*), with Iowa having some success reducing fatalities.

To address this concern, much research has been devoted to evaluating road traffic safety (*5-10, 20, 23, 29-45, 48, 49*). Traffic safety research has addressed diverse topics that include various crash factors and types, countermeasure implementation and effectiveness, statistical methods, and approaches, and evaluating approaches for identifying crash clusters. The evaluation of crash factors has been directly related to drivers in several studies. Those studies include but are not limited to driver's speed, age, experience, substance involvement, distractions, and others (*6, 7, 31, 38, 39, 41, 50-54*).

Also, crash factors evaluations involve different road conditions analysis and weather conditions (*31, 50, 55, 56-58).* In general, drivers under the age of 24 are more involved in crashes when compared to other aged drivers (*7, 38, 39, 41, 52, 54*), fatal injuries are more likely to occur to older drivers when compared to other drivers age categories (*38, 50, 53*), the use of drugs and alcohol could be a cause leading to more crashes and worse injuries (*31, 50, 51*), male drivers are more likely to be involved in a crash but female drivers are more likely to have serious injuries (*31, 38, 39, 41, 54*). Weather conditions have also presented a relationship with crashes since an impact on crash occurrence since to persist when weather conditions are present (*31, 55, 56, 50, 57, 58*).

Crash types are highly involved in most safety analysis studies. In most cases, crash types are widely related to severity and crash counts. In several studies, crash types are directly involved with modeling creation for different purposes (*32, 35, 59-62*). The reason for that is that similar crash types are likely to occur when there are deficiencies on the road design, but different crash types occur when different crash factors such as driver and environmental conditions are involved. Most of the time, crash types are associated to different conditions at certain locations (*32, 35, 59-62*). Because of that, generating models depending on the crash type is not always successful, however, it could be used as a complementary option. Angle, head-on, rear-end, sideswipe, same direction and opposite direction, pedestrian-involved, and single vehicle crash types are the most evaluated in studies (*32, 35, 59-62*).

Most traffic safety studies have the common objective of identifying or developing effective countermeasures to prevent and reduce severe crashes. Depending on different circumstances of why the crashes are happening, different countermeasures

can be applied (*31, 54, 63-67*). Some studies have identified general countermeasures including hazard warnings, control on speed limits, improving on light conditions, roundabouts incorporation, and many others (*31, 54, 63-67*). It has been proven that the application of effective countermeasures offers a potential to reduce crashes (*67*). Previous studies have also focused on the evaluation of countermeasures for run-off-road crashes. Some are the improvement of curves design, high friction surface treatments, the increase of safety campaigns addressed towards drivers, rumble strips or guardrails, increasing the separation between opposing lanes, using traversable roadside slope designs, relocating, or shielding fixed roadside objects, and others (*54, 64*). Another method to evaluate and identify crashes, is the development of crash clusters.

Much of this research addresses steps within the Roadway Safety Management Process, which involves: network screening, diagnosis, countermeasure selection, economic appraisal, project prioritization, and safety effectiveness evaluation (*11*). These steps are often combined into four distinct steps: network screening, diagnosis and countermeasure selection, economic appraisal and project prioritization, and effectiveness evaluation (HSM). Though the process is cyclical, as shown in Figure 2.3, network screening is often viewed as the initial step where the identification of potential sites for safety countermeasure application occurs (*5, 9, 12-17*). Diagnosis and countermeasure selection is subsequently applied to the identified sites with economic appraisal and project prioritization comparing the selected countermeasure projects for implementation. Effectiveness evaluation occurs after some time has passed since evaluation and this leads to improvements in the cyclical process over time. For this thesis, the focus is primarily on the network screening aspect but with impacts on subsequent steps in mind.

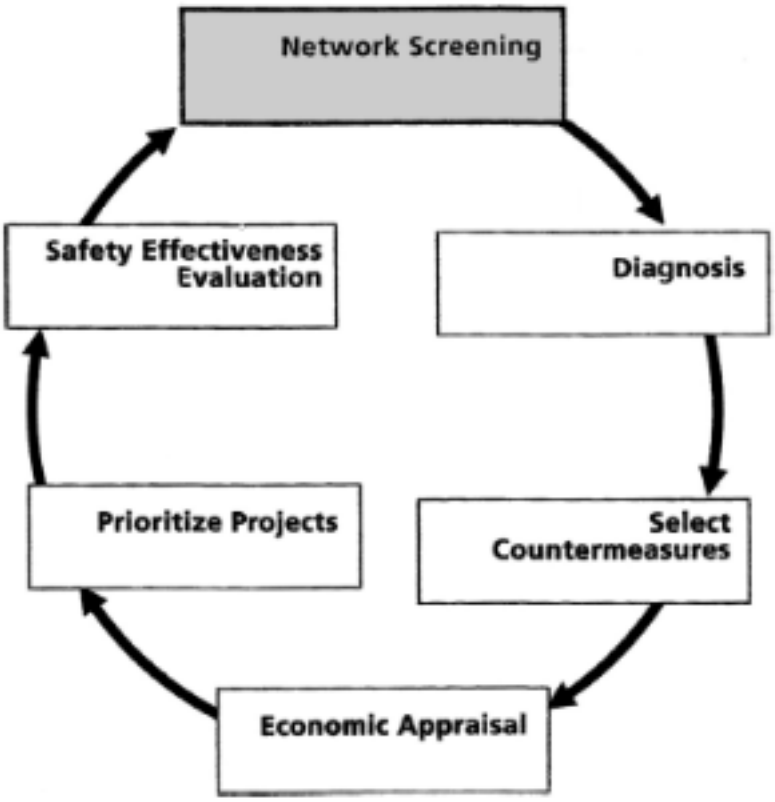

Figure 2.3: Roadway Safety Management Process *(HSM)*.

With the focus on network screening and, more specifically, network screening of single vehicle run-off-road (SVROR) crashes, a literature review of related topics follows: network screening, crash cluster analysis), and run-off-road crash research.

Network Screening

In the U.S., federal mandates during the 1970s required states to each develop and implement a highway safety improvement program. Network screening, or identification of hazardous locations and elements, became an obvious important first step. Following the network screening process, per the four-step Roadway Safety Management Process, countermeasures and projects were developed with subsequent effectiveness evaluation intended. To provide guidance and ease implementation, a federal Hazard Elimination

Program (HEP) was initiated in 1978. With the impetus of the HEP, many states developed Hazard Evaluation Systems (HES) which included network screening to identify likely hazardous locations and develop a list of safety improvement candidate locations. (*18*)

Network screening is an analytical process to quantify and assess safety performance of the roadway network, comparing individual locations (sites) to one another and ranking the sites on relevant criteria. The process is a crucial initial step that identifies sites of concern by first determining sites with greater tendency for crashes, whether total crashes or some subset thereof (e.g., severity or some other crash characteristic such as run-off-road crashes). Subsequently, these sites are assessed to whether they are outside some norm. Historically, the assessment of sites based on roadway geometrics (e.g., segments and intersections) has been based on density, rate, or severity comparisons but more recently assessed via more rigorous statistical means.

Some studies have based the evaluation of crashes depending on driver's speed, age, experience, substance involvement, and distractions (*6, 7, 31, 38, 39, 41, 50-54*). Others have involved different road conditions analysis and weather conditions in order to establish relations among crashes (*31, 50, 55-58*). Another very commonly used subset to determine sites with greater tendency for crashes is to define crash types and crash severity (*32, 35, 59-62*). The identified sites are then assessed for potential countermeasures.

Related to the determination of sites with greater tendency for crashes, analysts have historically summarized data based on geometric features of the roadway network, namely intersections and segments (*5, 12-14, 62, 63*). Whereas this certainly makes

logical sense as engineering countermeasures would generally be applied across the network, basing the identification of safety concerns on factors on network elements that may disperse impacts over lengths and obfuscate actual safety concerns. For example, if a length of roadway has an intense grouping of crashes along a short stretch but the crash impacts are dispersed along the entire length, this may not be obvious and may not be flagged for review or mitigation. Conversely, if the crashes are spaced reasonably evenly, which has been noted as being the case for run-off-road (*31, 54, 63-67*), then the section of roadway again may not be flagged even though a more systemic countermeasure might be appropriate. Furthermore, with regard to intersections, the length issue is less obvious; however, the assigning of crashes to intersections for subsequent summarization often ignored the crashes in proximity to the intersection that may have resulted from operations at the intersection. Developing a method to assess the distribution of crashes around a location both to identify concentration and spread is thus important, perhaps by ignoring the roadway geometrics initially and basing the identification on crash clustering initially.

Due to this importance, network screening methods have been often researched for potential improvements both to initial site identification (*9, 10, 16, 20, 32, 34, 44, 52, 53, 64, 66-75*) and to subsequent assessment (*31, 54, 63-67*). Recent methods have taken advantage of statistical methodology, including multivariate and spatial statistic (as well as other) improvements over time (*7, 29, 32, 33, 35, 36, 43, 48, 56, 58, 59, 61-63, 69, 77-83*). These improvements are intended to produce viable results for the subsequent steps. Viable, valid network screening methods return results where locations with a significant frequency of collisions are identified and subsequently evaluated (*70*). When areas of

concern where crashes are constantly taking place at are not properly evaluated, suitable solutions for those crashes cannot be offered, hence improvements and countermeasures are not established and applied for future conditions (*5, 11-14*).

Over the past few decades, several different network screening methods have been used, ranging from reasonably simplistic, easily applied to more sophisticated. These methods have evolved over time and development continues. Some of the methods include: crash frequency, crash rate, crash severity, rate quality control, index methods, spatial location and proximity, Bayesian methods, and multivariate modeling techniques (*6*).

The crash frequency method consists of the summarization of a number of crashes for a particular spot location; this method is related to the spot map method. The spot locations are compared and ranked depending on their crash frequency and those with high crash locations are further evaluated for statistical significance (89).

The crash rate method evaluates the risk of exposure depending on high crash locations. Similar to the crash frequency method, this method also classifies locations with higher rates as high crash locations. In this method, the basis for ranking crashes is the number of found crashes divided by the number of vehicles present in the location (*89*).

Crash severity methods are those which incorporate the measure of severity into the crash analysis as their name indicate. This method includes frequency, rate, and ratio of more severe crashes. Severe crashes are those which are given a more relative weight when compared to those considered as less severe. Including the severity of crashes in the

analysis helps agencies to offer more attention to those sites in which more fatal crashes are occurring when compared to crashes with a lower severity (*89*).

The rate quality control method is used to evaluate those sites in which crash rates are greater when compared to similar areas. Similar to the crash rate method, this method includes statistical control test in order to determine what the critical crash rate is in similar places to determine whether the observed site is significant or not (89).

Lastly, index methods are those that intend to include the severity indices with other factors as the ones previously mentioned. The most common of these include the weighted rank method, the crash probability index method, and the Iowa method. The weighted rank method consists of the inclusion of some of the previously mentioned methods in order to calculate a single index value for each site, and then, rank them as in previous methods to assign significance. The crash probability index method creates a combination of the results of crash frequency, crash rate, and crash density, in order to then rank the sites and determine which sites should be prioritized. The Iowa method generates three ranking lists involving frequency rank, a rate rank, and a severity rank, which are then combined into a single rank to also determine significance of sites (*64, 89*).

For spatial location and proximity, often this involves use of Geographic Information Systems (GIS) (*12, 19-23*) for application; however, this need not be the case given valid coordinates within the database. Different methods that have been developed to assess spatial location and proximity for network screening include: short/sliding window (*6, 11, 14, 16, 29*), Continuous Risk Profiles (CRP) (*6, 11, 14, 20*), Point Pattern Analysis (PPA) (*5, 10, 16, 32, 81*), and Kernel Density Estimation (KDE) (*17, 19, 20, 48,*

69). The network screening method developed as part of the research related to this thesis essentially improves upon KDE by travelling along the network to determine proximity and crash densities rather than a radial proximity. These crash clustering methods, and in particular, KDE are discussed more thoroughly in the following Crash Clustering Methods section.

Bayesian methods that have been applied to network screening include two primarily distinct subsets, Empirical Bayesian (*16, 32, 68, 70, 71, 84, 85*) and Full Bayesian (*29, 68, 32-34, 79, 84*) or Bayesian hierarchical model (*22, 31, 70*), and Bayesian spatial model (*7, 77*). However, the focus of this thesis is not on the statistical methodology but on the development of crash clustering.

Generally, when applying some of the previously mentioned methods, the data is first classified into different characteristics and factors which can include variation in accidents and injuries factors (*77*), spatial characteristics or considerations (*29, 33, 79, 84, 86*), driver behaviors depending on different elements (*6*), crash frequency, crash rate, crash risk (*7, 9, 20, 84*), historical crash data (*17*) and others. Overall, roadway crashes are typically classified depending on the interaction of the driver, the vehicle, the roadway characteristics, and environmental factors (*35*).

The next section revisits the topic of spatial location and proximity and focuses further on the crash clustering methods, including additional details regarding network screening. Though Bayesian statistical methods have been used to enhance network screening, again that is not the focus for this thesis.

Crash Clustering Methods

Hotspots identification, also known as black spots and sites with promise, is the process in which locations composed by a higher number of crashes usually caused by local risks factors are recognized (*16, 32, 67, 69-71*). The location of hotspots in intersections and along different types of segments is very important since it allows improvements to be made on the identified affected areas (*68, 72*). One of the most important aspects when identifying hotspots is to make sure that only true hotspots are being recognized and no errors are being included on the hotspot identification (*10, 16, 17, 19, 20, 23, 34, 36, 68, 70, 72, 73*).

Different studies have been focused on the comparison and analyses of methods in order to determine which method is found to be the most efficient when identifying hotspots depending on their given characteristics (*23, 33, 67, 69, 73*). Even though some methods have been proven to perform better than other in different studies, their performance is dependent on the locations of the crashes and their circumstances (*69*). It can be said that the methods that have been compared the most have included spatial modeling and temporal methods. In some studies, they have been found to perform better when compared to the other methods (*35, 36, 73*). Other methods that have been helpful for hotpot identification in different studies have been GIS (*19-23*) as well as Empirical Bayes (*16, 34, 67, 68, 70, 71, 85*)

Clustering has been another recognized method to identify and classify hotspots (*12, 13, 16, 19, 23, 24, 31, 48*). The main objective of clustering is to reclassify the data and arrange it in the most suitable way in order to evaluate the crashes effectively (*23, 48*). When clusters are implemented, hidden relationships and patterns can be found for

crashes and high risks locations are identified (*16, 12, 13, 24*). In some studies, this method has been combined with hierarchical Bayesian models (*16, 24, 31*), with the kernel density estimation method (24), by applying K-functions (*12, 13, 19, 23*) and others.

Clusters allow the identification of contributing factors and relevant properties that allow the development of more accurate test scenarios (*24, 31*). Clusters are mostly defined and classified depending on roadway conditions and design, crashes characteristics and kinematics or environmental conditions (*24, 31*). They can also be subdivided later based on the variance of their variables and characteristics to simplify the evaluation of the data (*36*). Cluster analysis is commonly applied to crash segmentation applying temporal or spatial distribution and to pattern identification in order to recognize factors and variations in crashes (*24*).

Research have shown that identifying crashes in order to reduce them is as important as understanding the relationship between road network risk and prediction variables (*48*). In order to reduce the frequency of traffic accidents, those much be identified withing the road segments (*5*). Safety network screening have been widely used to improve road locations (*31*). In order to provide roadway safety, locations with high collisions must be identified (*14*). For that reason, identifying hotspots have become a priority in different studies (*7, 13, 15, 17, 18, 21, 68, 77*). In order to do this important step, a variety of steps have been used and developed (*68*).

There are various methods that are used to estimate crash frequencies or severities in different locations (*15*). Most of the locations where those are evaluated are through

segments and intersections (*15*). Crash screening methods are very popular and used to analyze crashes in large areas and identify hotspots (*15*). The following paragraphs contain a short explanation of the most common methods.

Sliding Window Method (SWM)

The sliding window method consist of the estimation of potential for safety improvement at the start of a segment and then the estimation procedure is repeated until the window approaches the end of the segment that is being evaluated (*6, 16*). For the sliding window method, the hot spot window length, and the minimum number of crashes per hotspot are used as defining parameters. Under this method, critical threshold is used to determine whether hotspots are present or not and it consist of the minimum number of crashes per hotspot (*16*).

Previous studies have used a sliding window to calculate both crash count by severity and the predicted crash count on the basis of the developed SPFs and to calculate additional performance indices on the basis of both the link and the window data (*29*). In most studies, SWM has been compared to other methods (*6, 11, 14, 16, 29*). In some studies, SWM has presented an undesired number of false negatives (*11, 16*). Although SWM has been useful to identify individual locations that needed improvement, other methods have demonstrated better performance overall (*6, 11, 14, 16, 29*).

Peak Searching Method (PS)

The PS method is performed when the segment is subdivided to small windows and with the data, potential for safety improvement (PSI) is estimated in order to test them later using coefficient of variation until a maximum PSI is found, or the window reaches the end of the segment (*6, 15)*. In other words, the peak searching method divides

the segment into windows with the same length and the coefficient of variation is determined in order to obtain the statistical significance of the crashes (*16*). In most studies, the peak searching method has been compared to different methods (*6*, *11*, *15*, *16*). Similar to the sliding window method, the peak searching method has presented an undesired number of false negatives and false positives (*11*). Because of the lack of accuracy of the peak searching method, some studies have found superior methods when they are compared, and those have been selected instead (*6*, *11*, *16*).

Continuous Risk Profile

The network screening method of continuous risk profile (CRP) was developed with a web-based application also called California Safety Analyst (CASA) (*11*). CRP first reduces the data to only significant collisions to then obtain the predicted collision frequency from the corresponding SPF (*11*, *20*). The CRP method produces a measure of risk which identify the density of collisions per unit distance of roadway (*14*). It is able to monitor the risk changes over the years and to quantify effective countermeasures (*14*). One of the advantages of this method, is that it does not require SPFs estimations or segmentation of roadways since it uses spatial correlation (*11*, *14*).

In some studies, the CRP approach has been compared with SWM, PS, other methods, and it has been found that false positives and false negatives are reduced when using CRP while a better general performance have been observed (*6*, *11*, *14*, *20*). The CRP approach has successfully detected high collision concentration locations in previous studies (*6*, *11*, *14*, *20*).

Point Pattern Analysis

Point pattern analysis is based on the study of spatial arrangements of points in a determined space. The method is used in multiple areas in order to relate different points and its easier application is usually on maps. The modeling of this method occurs when it is inferred that an arrangement is the result of some process. the arrangement is typically on simple circles and ellipses, depending on given distances. At the end of the process, all points have the same location (*5, 10, 16, 32, 81*).

The point pattern analysis method has also been used on hotspots studies of crashes. It can be said that the point pattern analysis is divided into density-based methods and distance-based methods. On the density-based methods, the kernel density estimation (KDE) and quadrat analysis are used (the former being the most popular), and for the distance-based methods K-functions and Moran's I analysis are used. Results for this method have been successful, but there are some other methods that have outperform it due to the lack of accuracy that it can present in some cases (*5, 10, 16, 32, 81*).

Clustering

Clustering analysis consists of grouping a set of variables associating their similitudes in order to create groups. Clustering as a segmentation method is intended to organize and group large data into a small number of groups by associating the characteristics of the found crashes. This method has been widely used in multiple studies in order to analyze data (*7, 8, 12, 13, 15, 18, 48, 71*).

Crash Prediction Model

In some studies model crash frequencies that accounts for spatial correlations are evaluated. One method that has been previously used is the site-level CPM which uses the same mean and variance of the crash frequency and assumes that the observations are mutually independent. The zone-level CPM has been also used on county, states, traffic analysis zone, local health area, and others (*86*). Macro CPMs have been used to incorporate safety into traffic planning and traffic safety estimation. Micro CPMs requires the mean and variance of crash frequencies to be equal to analyze observations (*48*). After evaluating the CPM method, it has been found that the spatial modeling in CPM includes spatial correlation among the observations and generates a better safety assessment (*48, 86*).

Kernel Density Estimation (KDE)

Density estimation is an important concept in statistics. Density estimation offers valuable information about a particular data set involving different features such as skewness and multimodality. For a random quantity of X with a probability function of *f,* the function offers a natural description of the distribution of X. Also, with the function of *f,* probabilities associated with X can be observed (*47*).

The function *f* can be applied as (*47*):

$$P(a < X < b) = \int_a^b f(x)dx \qquad \text{for all } a < b.$$

Under the density-based methods, the kernel density estimation (KDE) is used to estimate crash frequencies or severities (*5, 7, 12, 13, 15, 17, 21, 30, 42-44, 49, 74-76, 80-82*). KDE is a non-parametric method in which a density estimation technique is used. The identification of hotspots is a systematic process in which road sections that suffer from high-risk crashes are identified. The KDE method can be used to estimate the density of traffic accidents on a roadway network (*5, 7, 12, 13, 15, 17, 21, 30, 42-44, 49, 74-76, 80-82*). This method enables the evaluation of local probability accident occurrence and the probable dangerousness of particular area (*81*).

KDE has been used to detect highly crash risk sections, to observed temporal variation of hotspots across the road network along with the Moran's Index method, to investigate spatial variations and several more. KDE have been adopted in previous studies and expected crash rates have been identified (*5, 7, 12, 13, 15, 17, 21, 30*). All of the studies in which KDE have been implemented, have shown potential for this method and satisfying results (*49*). The kernel density estimation tool has become one of the most promising spatial tools since it determines the risk of an accident by also making comparisons (*5, 7, 10, 12, 13, 15, 17, 21, 30*).

In a general view of the kernel density estimation tool, each point has a symmetrical surface and the distance between the points and the reference locations is determined. That is done in order to estimate the distribution of accident points. Once the hotspots are identified, they are classified by homogenous types based on environmental characteristics, and clusters are created (*7, 12*).

In more detailed, on the KDE method, a continuous crash density surface is created by using a circular search area by the kernel function which is applied on each crash. After that, grid cells are overlaid over the study area for each of which the density is estimated by the addition of the overlapping density surface of each crash point (*49*). In other words, this method places a symmetrical surface over the evaluated points and analyses the distance from the selected points to the location of reference based on the mathematical function that it includes. After that, the values for all of the surfaces per point are added for each of the selected locations of reference. The kernels are created over a point in order to generate a smooth and continuous surface. The estimated density is calculated by adding these values at each observation withing the bandwidth (*43, 76, 83*).

The kernel density function can be applied as (*76*):

$$f(x,y) = \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{d_i}{h}\right)$$

In this equation, $f(x,y)$ represents the density estimate at the location of (x,y). n represents the number of observations, h is the bandwidth or the kernel size, K is the kernel function, and d_i represents the distance between the selected location of (x,y) (*76*). For a better understanding of the variables involved in this function, please refer to Figure 2.3.

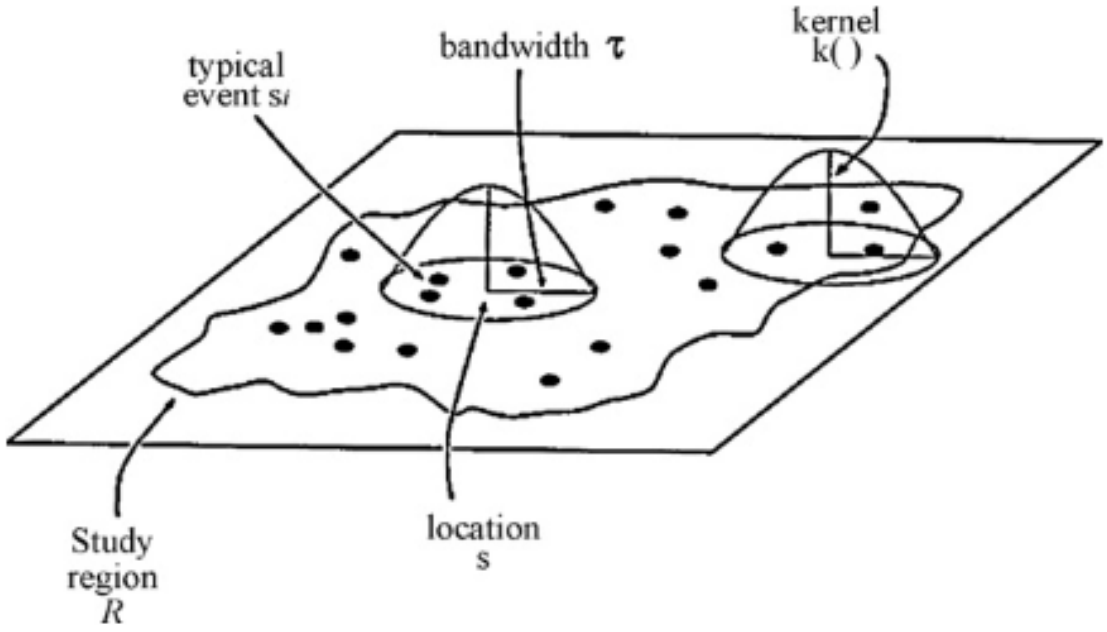

Figure 2.4: Diagram of how the kernel density method works (64).

On the other hand, the kernel function is applied as (*49*):

$$f(x,y) = \sum_{i=1}^{n} \frac{1}{n \times 2 \times \pi h^2} \times W_i \times K(\frac{d_i}{h})$$

In this equation, $f(x,y)$ represents the density estimate at the location of (x,y). n represents the number of observations, h is the bandwidth or the kernel size, K is the kernel function, and d_i represents the distance between the selected location of (x,y), W_i is the intensity of the observation (*49*, *76*). For a better understanding of the variables involved in this function, please refer to Figure 2.4.

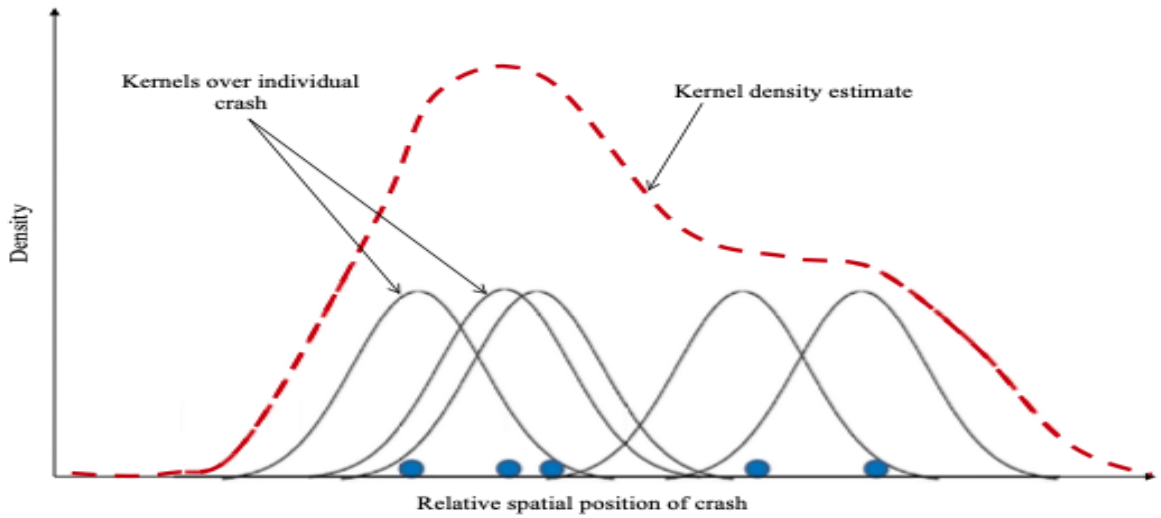

Figure 2.5: Kernel density method with kernels (*49*).

Kernel functions can be normal, uniform, quartic, epanicnikov, and triangular. Among all of those options, the most popular are normal for quartic functions and ArcGIS (*49, 74, 83*). The normal kernel function is used more often since it provides convenient mathematical properties and when it was compared to the others in previous studies, it resulted to have a smaller efficiency loss. On the other hand, the epanicnikov function is better when it comes to mean square error (*74*). The selection of the kernel function depends on what the study area and the data include and what results are expected from the particular study.

Even though KDE have several parameters, some are more influential than others. The two main parameters which affect KDE the most are bandwidth and cell size. The bandwidth is highly important since it determines the extent of search area (*43, 49, 75, 80*). In all cases, it is important to experiment with different intervals of bandwidths in order to determine which one fits the data more accurately. Generally, the range goes

from 20 to 1,000 meters but since every data set is different, the ranges might change depending on the area to be study. In particular studies, the bandwidth has been selected to be equal and double the cell size in order to follow the hotspots pattern (*49, 75*). If the bandwidth is large, a smooth density pattern is created which makes the separation of local hotspots more challenging. If the bandwidth is small, then only individual hotspots' locations are highlighted and evaluated creating a sharp density pattern (*43*).

After using the network KDE method to identify the potential hotspots, in some studies, potential hotspots are extracted onto density maps to help with visualization (*5, 12*). In some cases, once the network KDE method is applied, segments show zero crashes and those can be discarded (*5*). Similarly, to other studies, the KDE method has also been evaluated and compared with other methods (*5, 30, 76*). For some studies, after comparing methods, the network KDE results to identify potential hotspots performs better than results based on aggregated crash data (*5, 49*) and in other cases, it just meets with the requirements of the study (*5, 7, 12, 13, 15, 17, 21, 30, 42-44, 49, 74-76, 80-82*).

As it has been previously stated, several studies have used the Kernel Density Estimation tool as a method to identify hotspots. The geographic information system (GIS) has played an important role in the process of identifying crashes and preventing them. GIS is a tool on expansion, always proving new analysis opportunities and tools for research (56). This tool offers efficiency when providing screening analysis in order to screen and diagnosticate for crashes (*42-44 75, 76, 82, 83*).

The Kernel Density Estimation has been used in the frame of the spatial analysis along network (SANET) (*42-44 75, 76, 82,*). This method can be applied to data located

in GIS since a toolbox was included for the general KDE process under the spatial analysis network option (*42*). Previous studies have used the SANET tool, and the results have shown a precise definition of crashes proving the accuracy of the tool.

Kernel density estimation has two different forms. Planar Kernel Density Estimation (PKDE) and Network Kernel Density Estimation (NKDE). Both of these methods have been widely used when identifying clusters. (*13, 15, 42*). Those two methods have been compared in other studies, and results have shown that NKDE is less depending on input parameters which is advantageous when determining clusters in some cases. Because of the advantages that the NKDE method offers, it has been chosen as the primary method for studies (*13*). Also, in previous studies the NKDE method has identified continuous local hotspots which is advantageous when identifying hotspots (*15*).

The Planar Kernel Density Estimation is widely used when studying crashes since the density is calculated within a homogenous 2D space. The PKDE analyses events withing a 2-D homogenous space in which the longer the distance between a point and the location, the lower the weight of the selected point. This method uses the Euclidean distance between the located crashes alternating the result (*42*). The two key parameters that are involved in this method are the kernel function k and the bandwidth r. PKDE has outperformed NKDE in previous studies since PKDE is less dependent on the given input parameters (*42*). On the other hand, one of the disadvantages of this method, is that it uses Euclidean distance between the events which is not accurate since crashes occur along a road network (*42*).

The planar kernel density function can be applied as (*42*):

$$\lambda(s) = \sum_{i=1}^{n} \frac{1}{\mu r^2} k\left(\frac{d_{is}}{r}\right)$$

In this equation, $\lambda(s)$ represents the density in the locations s, r represents the radius (bandwidth) of KDE, k is the weight of point i, k is the weight of point i at a distance d_i of the location s (*42*). This function can also be referred as the kernel function of the relationship between d_{is} and r.

The Network Kernel Density Estimation follows the PKDE process by including a calculation of the density of point type events on a linear unit instead of using the 2-D homogenous area unit that the PKDE employs. Previous studies have involved this type of KDE instead of the planar alternative when events were occurring on a road network.

The network kernel density function can be applied as (*42*):

$$\lambda(s) = \sum_{i=1}^{n} \frac{1}{r} k\left(\frac{d_{is}}{r}\right)$$

In this equation, $\lambda(s)$ represents the density in the locations s, r represents the search radius (bandwidth) of KDE, k is the weight of point i, k is the weight of point i at a distance d_i of the location s (*42*). This function can also be referred as the kernel function of the relationship between d_{is} and r.

Advantages

As expected, due to its highly use for different studies, the KDE method has various advantages when compared to the other clustering methods. One advantage is the considerations of spatial autocorrelation of crashes when preparing clusters (*49*).

Another advantage of this method is its efficiency when obtaining scores for exiting samples and also providing new samples that reflect the underlying data structure. That is performed by providing the new total log probability density under the model and generating models for a data set (*74*).

This method has also been found to be superior and advantageous to multiple methods that involve statistical hotspot and clustering techniques (*43*). The major and most important advantage of this method is how accurate this method has been proven to be when taken into different studies and compared with others. Also, this method allows the user to determine the spread of risk of an accident or crash. The spread of risk is the area around an existing cluster in which accident are more likely to occur based on the spatial dependency of the selected place (*76*).

For those and more reasons, KDE has always been considered when creating clusters and evaluating crashes along roadways (*49*). After the evaluation of KDE and completed comparisons, it can be said that in general, this is a simple method to implemented to data, and it is also an easy method that still offers accurate results.

Challenges

As expected with any method, some challenges can be found. The main challenge when implementing KDE is selecting the appropriate value for some of its variables. As stated earlier, the selection of the bandwidths plays a very important role in the kernel density estimation. If the variable is assigned incorrectly, then the results are not going to be as accurate and other methods could be able to outperform KDE when the results are compared. In order to overcome this disadvantage, a series of bandwidths should be tested in order to find the one that fits the study better.

Also, some studies have brough the concern that KDE considers discrete phenomenon as a continuous area (*44*). This can also be overcome by carefully evaluating the area in which this method will be applied and correcting any found or suspicious discrepancies.

As it was mentioned earlier, the use of clusters becomes crucial when identifying hotspots and when evaluating crashes in an effective manner (*16, 19, 12, 13, 31, 48, 23, 24*). The most common methods are hierarchical clustering, density-based clustering, and K-means clustering. Hierarchical clustering allows you to identify clusters by breaking up the dataset into different groups while also allowing you to evaluate additional groups inside the previous groups (*16, 31, 24*). Density-bases clustering methods tend to focus only on points that are tightly packed instead of assuming that every single point is part of a cluster (*48*). Lastly, K-means clustering considers every single point from the dataset and evolve the clustering over a series of iterations (*12, 13, 19, 23*).

Agglomerative hierarchical cluster analysis has been a one of the used methods in previous studies to evaluate run-off-road crashes. One of the benefits of using that approach, is that the main characteristics of the clusters are driven by the variables that were used in the cluster analysis. For this method, each crash in the dataset is considered its own cluster and then similarities are found among them, and a new cluster is created by merging the currently existing clusters. The clustering procedure was composed by three main steps. The first step included the measurement of the distance between two clusters and the determination of their location. The second step included the agglomeration criterion in which the previously selected clusters were merged depending on their characteristics, and a new cluster was created. The third and last step of the clustering procedure included the evaluation of the existing clusters in order to determine the number of clusters that were going to be used for the study (*24*).

Kernel density estimation for the spatial clustering investigation of crashes have been used in several studies. Different studies have considered this method to be the most successful and the most promising to describe and analyze spatial patterns and crashes (*17, 19, 48*). One alteration that could be made, is adding another value to the kernel density method. Adding another value could improve the sites selection that could need further investigation by identifying high density accident areas (*19*). A different way to proceed with the kernel density method, could be to apply natural break cluster to the dataset. Natural break cluster can be used in this case to determine the most appropriate arrangement that results in clusters of density. From the kernel density results, the areas could be then divided into different classes and the differences among them can be compared in order to minimize the deviation within each class (*48*).

Other methods have been implemented by different studies. An example of those other methods is the multidimensional clustering algorithm. This method includes a graph theoretical framework which is advantageous for segmentation and grouping multidimensional data which can be applied to multivariate data that evaluate crashes. The segmentation process includes a graph in which each edge contains a particular weight that includes the similarity of two different crashes. The graph is then partitioned into two sets in order to minimize the similarities between the existing crashes on the dataset and continue to work only with the most relevant clusters (*37*). Even though there are multiple clustering methods, the data should be evaluated in order to choose the appropriate method and obtain the most appropriate clustering results.

Run-off-Road Crashes

Run-of-road crashes, also known as roadway departures, are defined as crashes that occur when a single vehicle crosses an edge line or a center line, or when it leaves the traveled way (*24-26*). Not only do they constitute a significant portion of crashes in the United States but many result in fatalities (*24, 78*). Run-of-road crashes have been associated and reported as one-third of traffic fatalities on the rural road network making them a major problem in transportation safety (*25*). From 2016 to 2018, fifty-one percent of traffic fatalities in the United States were roadway departure or run-off-road crashes (*26*).

Evaluation of run-off-road crashes is needed in order to reduce their frequency. As crash locations are random but crash types are not, studies should be applied to run-off-road crashes (*87*). Studies have proposed to apply countermeasures to keep vehicles

on the roadway (*26, 87*) and to target high-risk locations (*87*) using systemic applications in order to provide roadway safety and prevent future crashes (*25, 26, 87*).

Run-off-road crashes have been evaluated with cluster methods in order to evaluate their factors and variations with the aim of improving future conditions. In previous studies, the two largest clusters for run-of-road crashes involved drift in daylight and drift at night with the characteristic that drivers were able to maintain control of the vehicle previous to the crash. Run-of-road crashes could be directly involved with drift events, low-friction curves, excessive curve speeds, and driver maneuvers (*24*).

Factors Affecting Run-off-Road Crashes

A very important aspect on reducing run-off-road crashes is to identify the key factors that are directly associated with the crash (*37, 39*). Previous studies have identified a variety of combinations of factors that have included but are not limited to:

- Human factors or driver characteristics
- Crash and vehicle characteristics
- Roadway related information
- Environmental factors

More specifically, different studies have analyzed data sets considering different key factors for run-off-road crashes. Those studies have included factors that are considered to be directly affecting the safety of the road and resulting in fatal run-off-road crashes (*38-41, 54*).

In the human factors or driver characteristics category, information such as driver age, intoxication, condition of the driver, driver gender, driver injury, carelessness, speeding and driver distraction have been considered in previous studies (*38-41, 54*).

Under this category, speeding has been identified as one of the key factors that could lead to run-off-road crashes (*38, 39, 54*). Other factors that have been observed to contribute are distraction, inexperience, and reckless driving (*39, 54*). For gender characteristics, male passenger-car drivers at dawn seem to be vulnerable to fatal run-off-road crashes while females between the ages of 65 and 74 driving non-passenger cars have been observed to be higher risk (*38*). It was also found that compared to females, males have a greater propensity toward property damage only during a run-off-road crash. For the driver age, older drivers have shown to be more likely involved in fatal run-off-road crashes under partial access control zones and under general conditions (*38, 41*). Regarding intoxication of the drivers, it was found that drug or alcohol usage during driving increases the risk of being severely injured during a run-off-road crash (*41*, *54*)

The crash characteristics have included crash year, crash time, and collision type (*38, 40, 41, 54*) while vehicle related information has been limited to include the vehicle condition and the vehicle type (*38, 39, 54*). Previous studies have also found that 69% of run-off-road crashes were overturned and 31% occurred due to hitting objects outside the carriageway (*38, 39*).

For the roadway related category, data access control, alignment, speed limit, lighting condition, road condition, road type, intersection, surface condition, road width, lane width, shoulder width, road-side objects and highway type have been included (*38-41, 54*). Previous studies have found that the majority of run-off-road crashes take place on rural roads and straight road segments (*39, 54*). It was also found that roads with a small safety zone tend to have more run-off-road crashes when compared to others. Also,

roadways with strong curvatures have been found to add 3 times more run-off-road crashes when compared to straight roads (*40*).

Regarding environmental factors, weather, time of day, and day of week have been evaluated (*38-41*). Studies have found that the majority of run-off-road crashes have occurred at clear weather conditions with enough lighting and on dry road surface (*39, 41*). Even though most crashes occur on dry road surfaces, it was found that the presence of ice, snow or wet road can worsen category of the run-off-road crash when compared to dry surfaces (*41, 54*). Also, it can be noted that 38% of the run-off-road crashes took place during the evening time (*39, 54*) and more than half occurred during working days. (*39*)

CHAPTER 3. DATA METHODOLOGY

The data used for development of the methodology and the corresponding analyses was obtained from the Iowa Department of Transportation (Iowa DOT). The data included road, crash, and traffic data for the years from 2015 to 2019. The road data include many attributes across all road classifications, i.e., interstates, US routes, state routes, and local roads. The crash data include attribute tables with the "statistical" (i.e., non-personal) data for all reported crashes that meet the State reporting threshold. While 5 years of crash data are included, only 1 year of road data were included for this initial analysis. There are two primary reasons for this. First, while the crash data are updated and generally released within a few months of the end of the year, the road data releases are usually delayed. Second, for the purpose of initial methodology development, one year of road data proved simpler to manage.

To reduce the data to rural, single-vehicle, run-off-road (SVROR) crashes, ArcGIS (ESRI, 2021) software was used. Appropriate queries and relates were used to develop subsets of the data, using the relevant tables from the road and crash data. First, the data were reduced to rural roads and crashes. The crash data were then further reduced to run-off-road (ROR) then single-vehicle only. The results of the queries and relates for annual rural SVROR crashes are shown in Figures 3.1 through 3.5 as well as Table 3.1. Figures 3.1 through 3.5 have several similarities, with potential clusters of SVROR crashes visible at a statewide level, which may not result in tight clusters at a roadway level. However, overall, the SVROR crashes seem somewhat sparse which is not unexpected. Beyond that, summarizations of the most relevant tables among the data,

were obtained for descriptive statistic purposes after each table was related to the final subset of interest.

Figure 3.1: Crashes after querying data for the year of 2015.

Figure 3.2: Crashes after querying data for the year of 2016.

Figure 3.3: Crashes after querying data for the year of 2017.

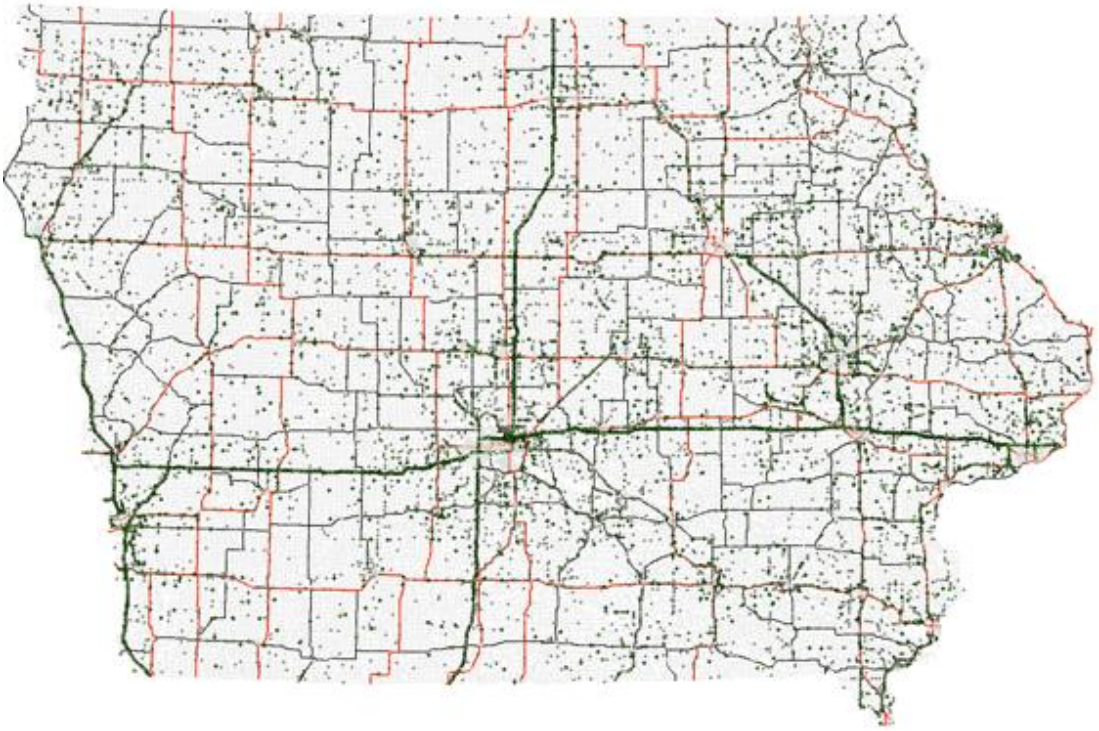

Figure 3.4. Crashes after querying data for the year of 2018.

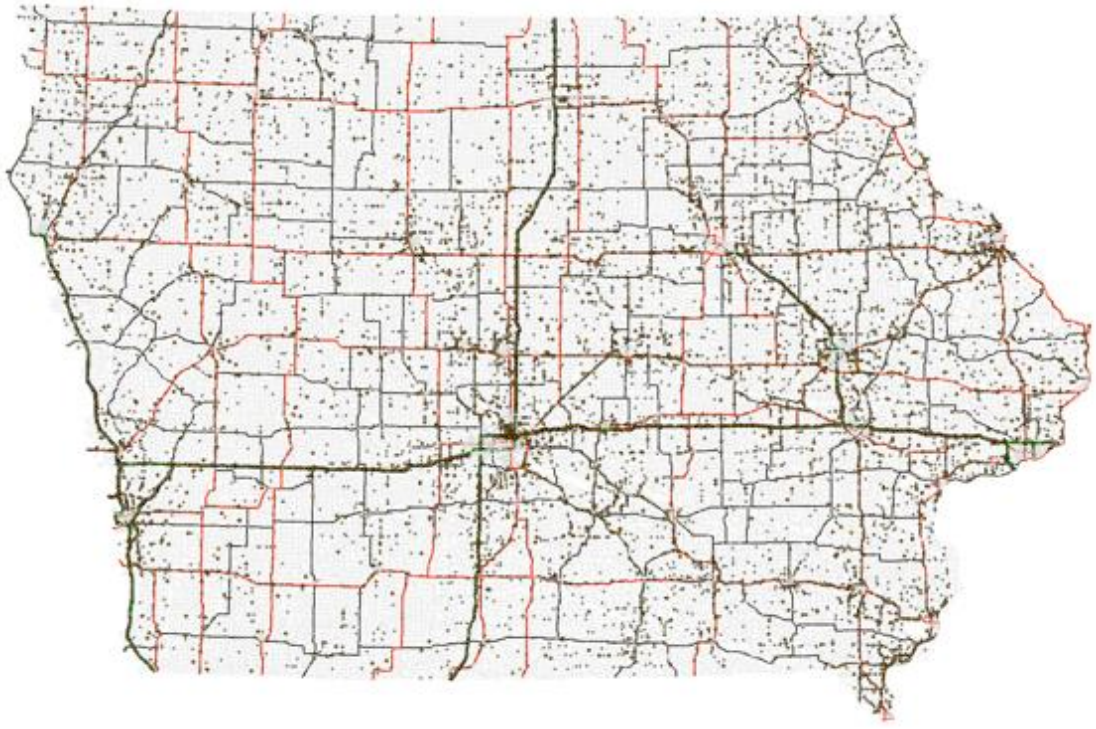

Figure 3.5. Crashes after querying data for the year of 2019.

As noted, from Figures 3.1 through 3.5 rural, SVROR crashes appear sparse though with potential concentrations at a statewide viewing level. This matches expectation from prior studies (*31, 54 63-67*). However, first, even at this map scale, there are some apparent regional concentrations that repeat annually. This might be due to the possibility of design improvements on certain areas if that is the cause for the repetition of those crashes. Second, the map scale may bely the actual concentrations, either showing more concentration than actuality or the reverse. That could be avoided by observing the crashes in a different perspective or by applying a statistical method for more precise evaluations. Third, the crash points shown may not depict the actual crash frequency as several crashes could have been located at the same point(s). The developed methodology attempts to identify clusters of crashes more precisely, specifically rural SVROR crashes for this study.

Table 3.1: Annual crashes for rural, rural ROR, and rural SVROR with percentages.

	2015	*2016*	*2017*	*2018*	*2019*
Rural	17,127	17,389	17,510	18,313	18,622
Rural & ROR	6,390	6,413	6,256	6,361	6,262
Rural, ROR & Single Vehicle	5,355	5,360	5,290	5,444	5,263
Rural & ROR (%)	37.3	36.9	35.7	34.7	33.6
Rural, ROR & Single Vehicle (%)	31.3	30.8	30.2	29.7	28.3

Within Table 3.1, results from the tiered sub-queries (rural; rural, ROR; and rural, SVROR) are shown for the 5 evaluation years. Generally, the trend seems to be

increasing crashes with 2019 having the highest frequency for both rural and rural, ROR crashes and only slightly less than 2018 for rural, SVROR. Additionally, percentages of rural crashes that were ROR for the analysis years are shown. The percentages appear reasonably consistent, from 37.3% to 33.6%, with perhaps some decline during the period. Beyond that, the percentages of rural, ROR crashes that involved only a single vehicle (rural, SVROR) are shown, with similar observations regarding consistency with a slightly lower percentage in the final year.

Figure 3.6 presents the previously discussed results in a graph. Run-of-road crashes have been associated and reported to one-third of traffic fatalities on the rural road network making them a major problem in transportation safety (*25*). In this case, they represent over 30% of rural crashes which is one of the motives of this study. Even though there has been a slight reduction during the evaluation years for this particular data, the problem with SVROR crashes does persist in the United States (*26*). The remainder of the data exploration and description is based on the rural, SVROR crash subset.

SVROR Crash Severity

Often safety analysis begins with an assessment of crash severity. The annual crash severity summary for rural, SVROR crashes is shown in Table 3.2. Crash severity is generally denoted by the KABCO scale which subsets severity into fatal, major injury, minor injury, and possible/unknown injury as well as property damage only (PDO). Fatal crashes involve at least one fatality but could also involve additional, non-fatal injuries. Similarly, major injury crashes

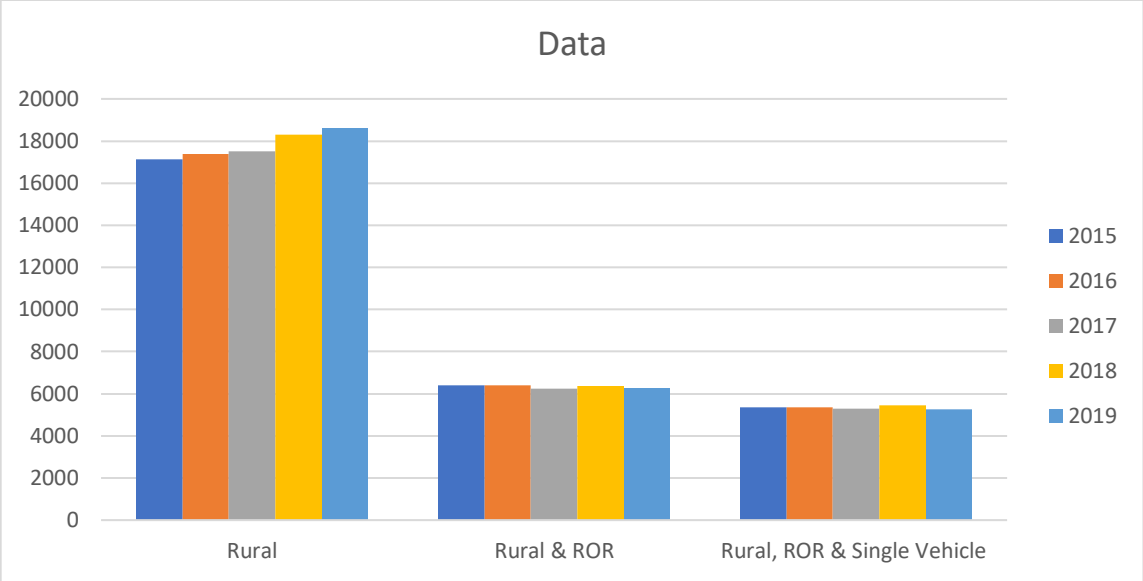

Figure 3.6: Annual crashes for rural, rural ROR, and rural SVROR.

involve at least one major injury but no fatalities though, again, these could include less severe injuries. Clearly, the number of fatal crashes (and, similarly for other severity levels) does not

necessarily equate to the number of fatalities, i.e., there could be more fatalities than fatal crashes and additional non-fatal injuries. PDO crashes involve only damage to property.

Table 3.2: Annual SVROR crashes by crash severity.

	2015	*2016*	*2017*	*2018*	*2019*
Fatal	76	99	89	68	81
Major Injury	356	353	342	290	265
Minor Injury	1,003	944	1,046	958	898
Possible/Unknown	1,001	985	1,026	971	932
Property Damage Only	2,919	2,984	2,787	3,157	3,087

As shown in Table 3.2, no clear increasing or decreasing trend is apparent over these 5 years for fatal crashes. The same can largely be said for minor injury crashes. However, major injury crash frequencies may indicate some decline; however, this may be due to some short-term variation. A reverse trend with a similar caveat can be observed for possible/unknown crashes. Finally, the number of PDO crashes seems to have increased during the 5 years but, again, though the trend seems more apparent, there is one aberrative year that inserts uncertainty. Figure 3.7 displays these results graphically. ROR crashes constitute a significant portion of crashes in the United States and also it is known that most of those crashes result in fatalities (*24*, *78*). For the current data, most crashes did not result in fatalities but instead, in property damage only. This could be due to the location in which these crashes are occurring or variations on the roadway design that might prevent fatal crashes. The reasoning for that could be further evaluated in a different study. Perhaps the inclusion of clusters for severe crashes in future evaluations would offer a clearer reason of why less fatal crashes are occurring when compared to the other crash severities.

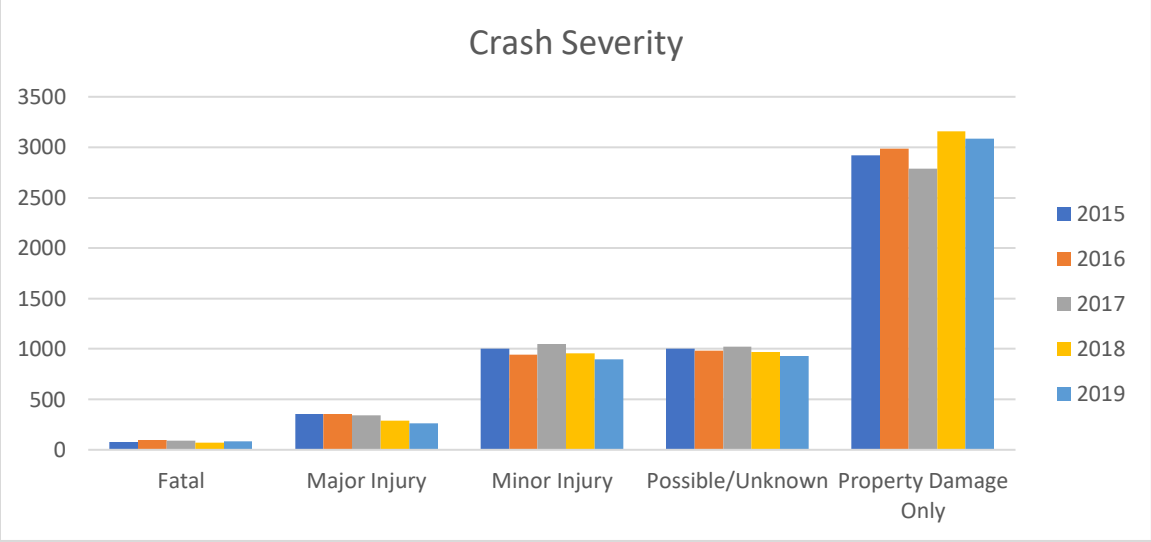

Figure 3.7: Annual SVROR crashes by crash severity.

SVROR Time of Day

Another frequent safety consideration is the time of day during which crashes occur. Typically, most crashes occur during the daytime hours, particularly the morning and evening commute, as most traffic occurs during that timeframe. However, the poorer sight distance brought about by darkness can influence the normal expectation, particularly in rural areas where artificial lighting is not available. To assess the time of day for crashes, the crash data field that uses 2-hour time bins (e.g., midnight to 1:59am, 2:00am to 3:59am, etc.) was used. Table 3.3 displays the summary of the time of the day for rural, SVROR crashes.

Table 3.3: Annual SVROR crashes by time of day.

	2015	*2016*	*2017*	*2018*	*2019*
Midnight to 1:59 AM	317	340	377	347	329
2:00 AM to 3:59 AM	309	257	310	299	264
4:00 AM to 5:59 AM	345	302	325	335	362
6:00 AM to 7:59 AM	556	611	505	641	600
8:00 AM to 9:59 AM	525	508	436	491	537
10:00 AM to 11:59 AM	400	421	379	473	445
Noon to 1:59 PM	418	476	414	479	439
2:00 PM to 3:59 PM	566	564	567	536	530
4:00 PM to 5:59 PM	603	540	601	540	577
6:00 PM to 7:59 PM	451	512	497	474	437
8:00 PM to 9:59 PM	453	433	452	439	372
10:00 PM to 11:59 PM	412	401	427	390	371

Results indicate that the time bins corresponding to normal morning and evening commute times contain the highest frequencies of crashes, as expected. Generally, the afternoon commute time was the higher time bin with the morning commute generally the second highest. Most crashes occurred between 4:00pm to 5:59pm for all of the years except for 2016 in which most crashes occurred from 2:00pm to 3:59pm. The least number of crashes occurred from 2:00 am to 3:59 am for all of the years as well. Beyond this, the year-to-year variation appears unremarkable and consistent. Also, though daytime hour frequencies are generally higher than nighttime hour frequencies, the reduction does not appear as marked as expected for all crashes. Potentially, this could be related to the rural SVROR nature of the database. Similar to the crash severity discrepancy with other studies, a particular time of the day could be frequent for the current data but not in all cases. Perhaps some aspect (e.g., horizontal curvature, particularly sharp curves) of an unlit roadway might lead to increased SVROR crashes in rural areas. If so, an expectation for clusters at these locations would exist. Figure 3.8 displays these results graphically.

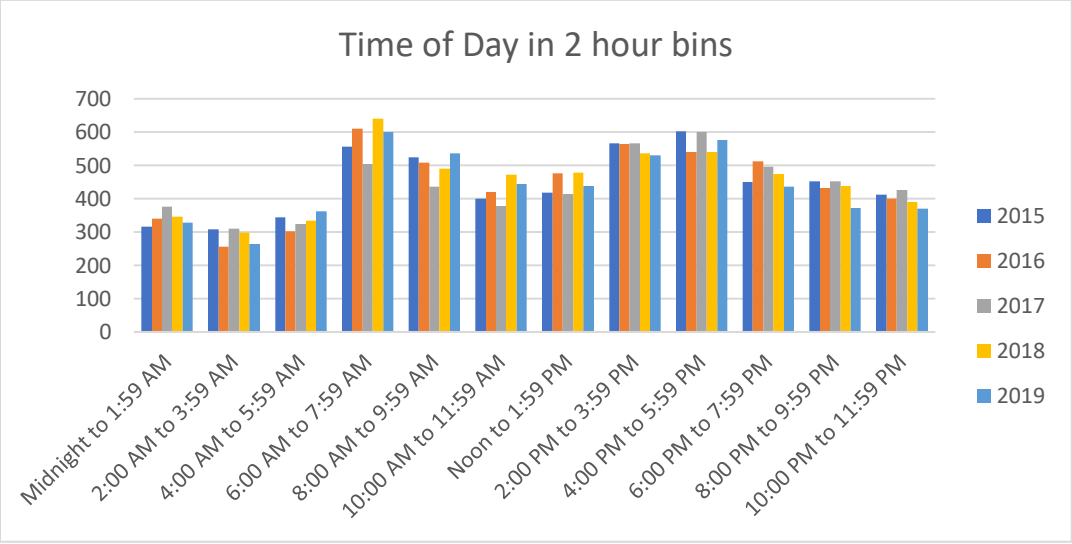

Figure 3.8: Annual SVROR crashes by time of day.

SVROR Lighting

Related to time of day, lighting is similarly interesting. Again, typically crashes during daylight are markedly higher than for darkness situations due to normal commuter traffic and generally increased activity during the day. Table 3.3 displays the summary of the lighting, based on time of day when compared with civil twilight times (NOAA, 2019) for rural, SVROR crashes. Morning and evening twilight durations are based on the civil twilight definition and are thus only 30 minutes each; thus, these frequencies should normally be small.

Table 3.4: Annual SVROR crashes by lighting (daylight/darkness).

	2015	*2016*	*2017*	*2018*	*2019*
Daylight	2,997	3,068	2,913	3,062	2,970
Darkness	2,110	2,065	2,160	2,132	2,054
Morning Twilight (dawn)	130	139	101	140	134
Evening Twilight (dusk)	118	93	116	110	105

Results from Table 3.4, indicate that most rural, SVROR crashes occurred during the daylight consistently across years. Again, this result is not unexpected. However, the relative proportion of darkness crashes to daylight crashes seems rather high and worth further consideration. Again, perhaps some aspect (e.g., horizontal curvature, particularly sharp curves) of an unlit roadway might lead to increased SVROR crashes in rural areas. If so, an expectation for clusters at these locations would exist. Additionally, a potential for seasonal variations related to the length of sunlight compared with darkness might prove interesting. Figure 3.9 displays these results graphically.

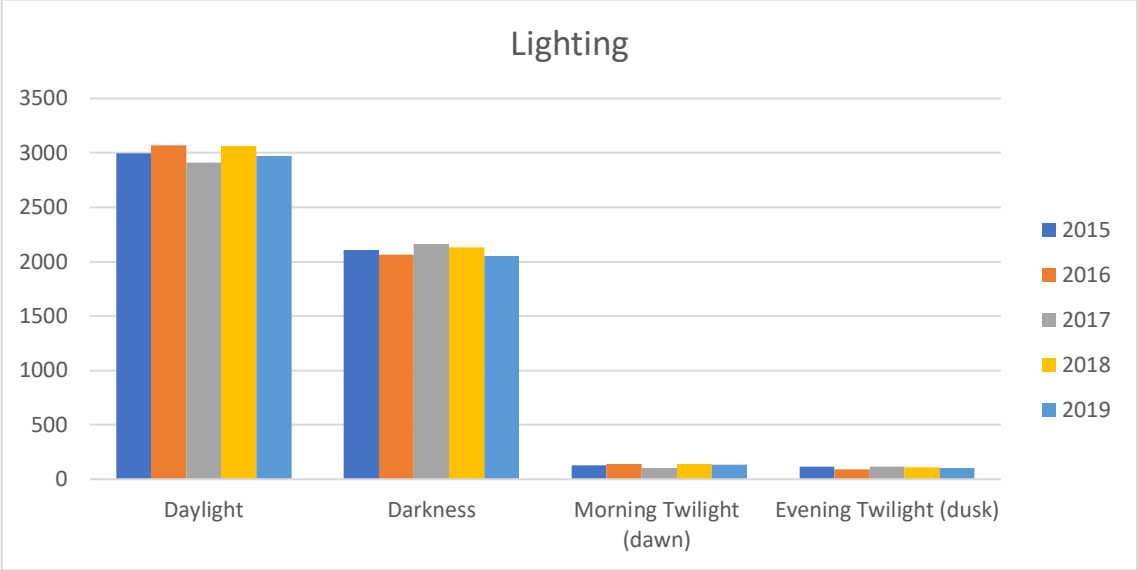

Figure 3.9: Annual SVROR crashes by lighting (daylight/darkness).

SVROR Location of First Harmful Event

The location of the first harmful event field indicates where the initial harmful event for a crash occurred. The first harmful event is not necessarily the initiating event but the first resulting in harm or damage. The options are shown, along with annual frequency of occurrence, in Table 3.5. For SVROR crashes, an expectation would be that harmful events would not occur on the roadway but off the roadway (e.g., shoulder, median, roadside, gore, outside trafficway) where some fixed object would be encountered. However, an animal strike followed by a roadway departure would be included as SVROR. Figure 3.10 displays these results graphically.

For rural, SVROR crashes during the analysis timeframe, the predominant first harmful event location was the roadway with indication of a decreasing trend. The other primary locations were the shoulder, median, roadside, and outside trafficway which might be collectively referred to as “off roadway”. Only primary locations will be shown

in the figure to avoid redundancy. When summed, the frequency of these “off roadway” crashes are similar to the “on roadway” frequencies. This could also be related with design characteristics depending on the location in which these are occurring in the map.

Table 3.5: Annual SVROR crashes by location of first harmful event.

	2015	*2016*	*2017*	*2018*	*2019*
On Roadway	2,620	2,741	2,660	2,482	2,515
Shoulder	1,066	1,045	1,008	1,088	988
Median	279	250	251	353	403
Roadside	723	732	721	805	718
Gore	48	24	37	31	39
Outside trafficway	550	513	534	604	521
In parking lane/zone	13	8	5	2	0
Separator	0	1	1	1	0
Other (explain in narrative)	2	3	3	9	5
Unknown	43	37	58	49	59

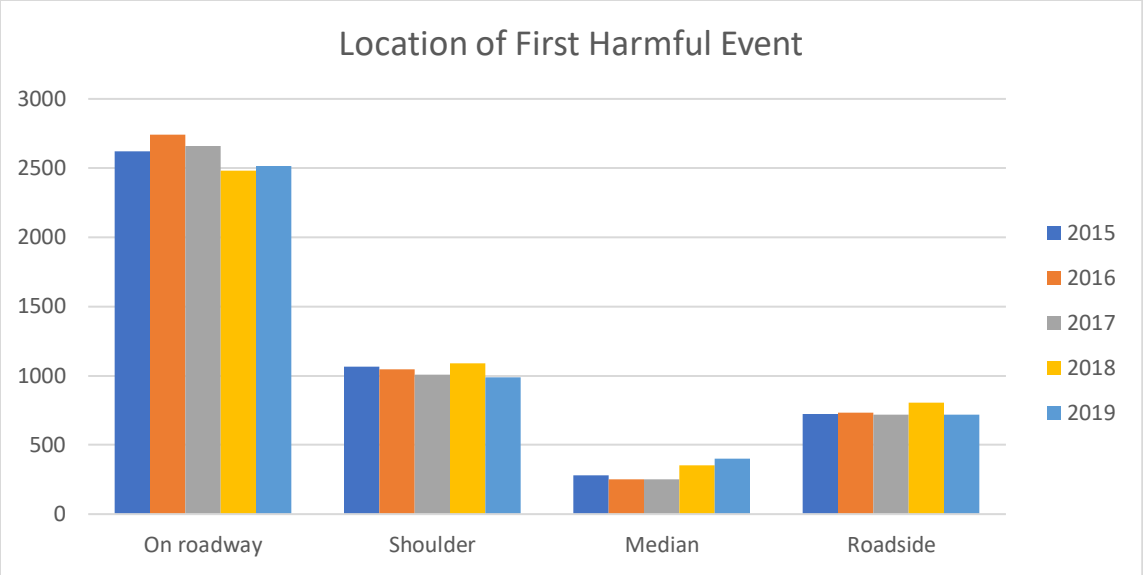

Figure 3.10: Annual SVROR crashes by location of first harmful event.

SVROR Weather Conditions

The weather conditions field indicates the circumstances of the weather while the crash occurred. Table 3.6 shown below includes the weather conditions for each year with the respective crash frequencies. For SVROR crashes, it would be expected that most crashes occur during clear weather conditions since these types of crashes are not dependent on weather conditions. However, the presence of snow or perhaps rain, could increases the chances of a vehicle leaving the road depending on the severity of the weather condition experienced. Figure 3.11 displays these results graphically.

Table 3.6: Annual SVROR crashes by weather conditions.

	2015	*2016*	*2017*	*2018*	*2019*
Clear	2,992	2,940	3,061	2,762	2,609
Cloudy	936	989	968	901	875
Fog, smoke, smog	78	92	127	102	77
Freezing rain/drizzle	186	151	171	321	284
Rain	330	281	361	390	335
Sleet, hail	17	21	10	63	17
Snow	584	593	405	689	709
Blowing snow	99	152	67	118	256
Severe winds	21	39	31	24	35
Blowing sand, soil, dirt	8	4	1	3	2
Unknown	0	0	0	0	0
Not Reported	79	64	64	60	50

For rural, SVROR crashes during the analysis timeframe, the predominant weather condition while the crashes occurred was clear weather conditions as expected with indication of a decreasing trend. The other primary weather conditions were cloudy conditions, snow, rain, and freezing rain/drizzle. Only the primary weather conditions that seemed to have an influence on crashes are shown in the figure. Overall, it can be said that most SVROR crashes occurred while clear weather conditions are observed, as expected. The presence of snow, rain, and freezing rain/drizzle do contribute. Future analyses could examine clusters of this subset of SVROR crashes, especially as these

weather conditions are relatively infrequent, thus perhaps crash frequencies are overrepresented.

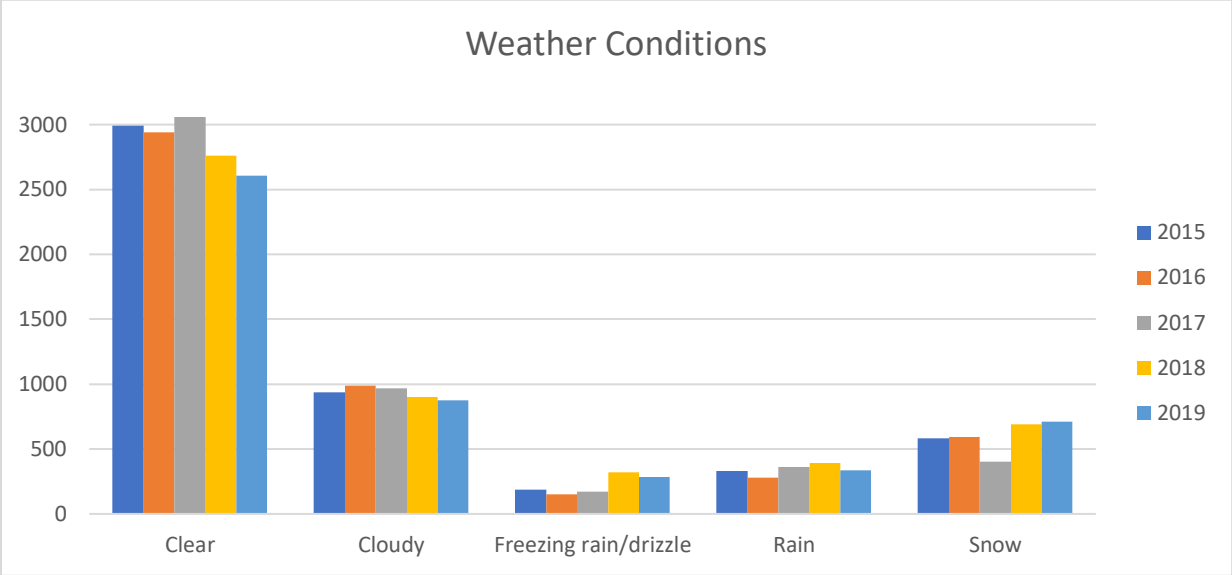

Figure 3.11: Annual SVROR crashes by weather conditions.

SVROR Surface Conditions

The surface conditions field indicates the circumstances of the surface while the crash occurred. Table 3.7 shown below includes the surface conditions description for each year with the respective frequencies. The surface conditions include dry, wet, ice, snow, and other conditions of the surface. For SVROR crashes, it would be expected that most crashes occur during dry surface conditions, similar to the clear weather conditions of the previous field. However, the presence of ice on the road could increase the chances of a vehicle leaving the road. Figure 3.12 displays these results graphically.

Table 3.7: Annual SVROR crashes by surface conditions.

	2015	*2016*	*2017*	*2018*	*2019*
Dry	2,802	2,865	3,108	2,613	2,420
Wet	485	473	534	591	549
Ice/frost	717	695	541	850	1028
Snow	568	527	392	651	701
Slush	74	121	40	181	115
Mud, dirt	31	30	22	37	33
Water (standing or moving)	3	4	3	5	3
Sand	3	10	10	5	2
Oil	1	0	0	1	1
Gravel	604	580	590	455	372
Other (explain in narrative)	18	9	6	13	11
Unknown	34	24	28	37	26
Not Reported	15	27	16	5	2

From Table 3.7, most rural SVROR crashes occurred when the surface presented dry conditions. This result is expected for the evaluated types of crashes and relates to the weather conditions category. Following dry surface conditions, the other predominant surface conditions were wet, ice, snow, and gravel. Similar to the weather conditions, most crashes occurred during normal or dry conditions followed by snow/rain residues on the surface. As mentioned earlier, the presence of snow, water, and especially ice, could increase the chances of a vehicle to leave the road. Supporting that point, for all three

categories, the number of crashes per year is increasing. On the other hand, the number of crashes occurring on gravel roads could be due to the high amount of gravel roads located in rural areas, however for this category, decreasing crashes are observed. Those surface conditions that presented the highest number of crashes will be shown in the figure.

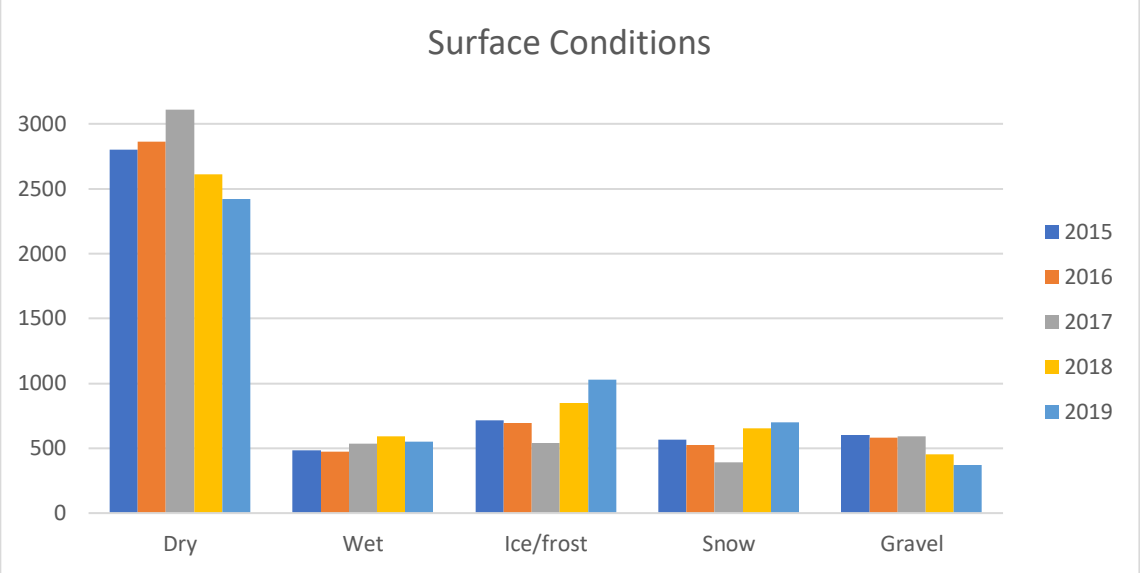

Figure 3.12: Annual SVROR crashes by surface conditions.

SVROR First Harmful Event

The location of the first harmful event field indicates where the initial harmful event for a crash occurred and those options are shown, along with annual frequency of occurrence, in Table 3.8. For SVROR crashes, an expectation would be that harmful events would not occur on the roadway but off the roadway (e.g., shoulder, median, roadside, gore, outside trafficway) where some fixed object would be encountered. However, an animal strike followed by a roadway departure would be included as SVROR. Figure 3.13 displays these results graphically.

The first harmful event field indicates what the initial harmful event for a crash to occurred was and those options are shown, along with annual frequency of occurrence, in Table 3.8. For SVROR crashes, an expectation would be that the first harmful event would include a rollover, or collision with the ditch. Figure 3.13 displays the results graphically.

Table 3.8: Annual SVROR crashes by first harmful event.

	2015	*2016*	*2017*	*2018*	*2019*
Non-collision events: Overturn/rollover	1478	1521	1440	1276	1376
Collision with fixed object: Ditch	1750	1842	1923	1915	1767
Collision with fixed object: Others	1479	467	441	578	528
Collision with: Other	191	162	196	206	204
Non-collision events	266	251	225	325	341
Miscellaneous events	96	97	118	154	163

From Table 3.8, it can be observed that most rural SVROR crashes occurred when the first harmful event was collision with ditch. This result is expected for the evaluated types of crashes since most likely, that is the first object that the vehicle will hit after a SVROR crash. Following that event, the other relevant first harmful events were non-collision events - overturn/rollover, collision with cable barrier, collision with animal, and collision with embankment. The values through the years appear to be constant since no major variation is observed.

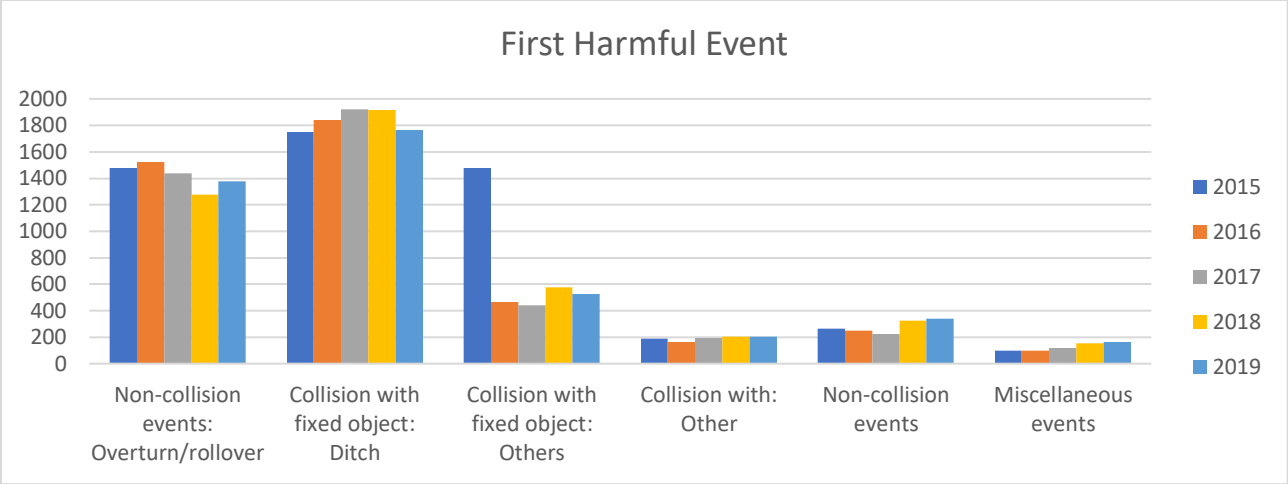

Figure 3.13: Annual SVROR crashes by first harmful event.

SVROR Road System

Safety analysis is also often performed related to the road system. The road system field somewhat indicates the ownership of the road on which the crash took place. For this analysis, interstate, US routes, Iowa routes, farm to market routes, local roads, and construction roads are included for each evaluated year. For SVROR crashes, the type of road in which the crash occurs should not have such a great impact since SVROR crashes are common in all rural roads. Figure 3.14 displays these results graphically.

Most rural SVROR crashes took place on farm to market routes. It is expected for that road system to have the highest number of crashes due to the mileage in rural areas. Typically, however, these roads have lower volumes. The reasoning could also be due to the lower maintenance that this type of secondary roads receives when compared to primary roads. Local

Table 3.9: Annual SVROR crashes by road system.

	2015	*2016*	*2017*	*2018*	*2019*
Interstate	951	870	853	1,068	1,071
US Route	767	782	705	738	777
Iowa Route	535	635	611	602	588
Farm to Market Route	2,107	2,096	2,076	2,145	1,967
Local Road	993	980	1,045	890	860
Construction	2	2	0	1	0

roads, interstates, and US routes were the other predominant roads in which SVROR crashes took place. Farm to market and local roads present a highest total of crashes when compared to interstates, US routes, and Iowa routes when combined together. Perhaps this indicates that major attention should be offered to these roads. Overall, all road systems do not present any significant changes throughout the evaluated years and as expected, roads with construction were not significant in this case.

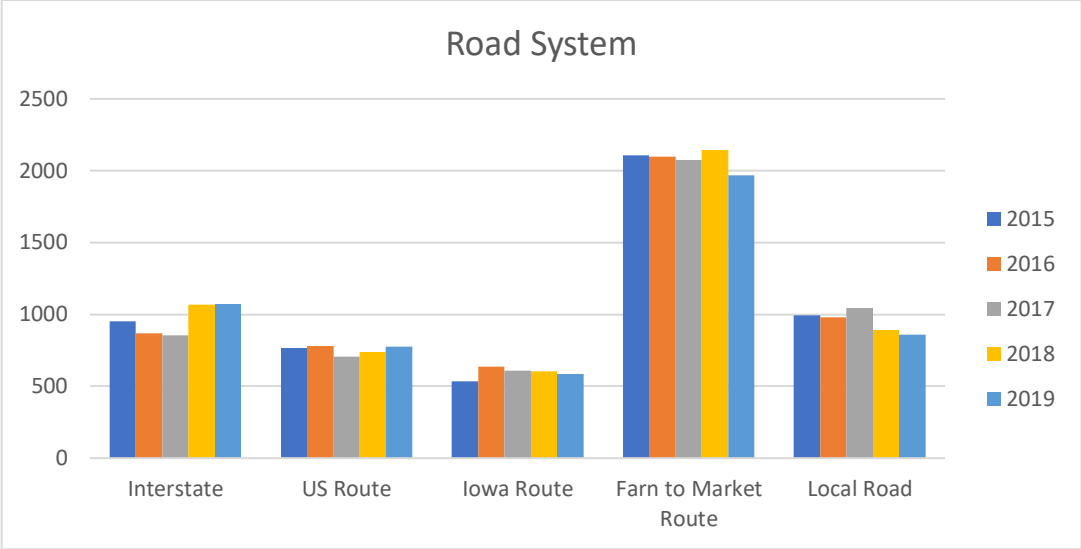

Figure 3.14: Annual SVROR crashes by road system.

SVROR Roadway Contributing Circumstances

Roadway contributing circumstances field indicates any roadway related circumstances that might have an effect on the crash. Those contributing circumstances are shown, along with annual frequency of occurrence, in Table 3.10. For SVROR crashes, and from what has been observed in previous fields, an expectation would be that there will not be any roadway contributing circumstances for this type of crashes. However, it was previously noted that ice, snow and wet surface conditions could have an impact on the increased number for the current type of crashes. Figure 3.15 displays these results graphically, only including those relevant categories.

Table 3.10: Annual SVROR crashes by contributing circumstances - roadway.

	2015	*2016*	*2017*	*2018*	*2019*
None apparent	3704	870	3907	3422	3105
Surface condition (e.g., wet, icy)	1341	782	1057	1725	1826
Debris	12	635	11	16	14

Ruts/holes/bumps	37	3062	25	19	31
Other	261	16	290	262	287

Most SVROR crashes presented none contributing roadway conditions, as it was expected. As mentioned earlier, the previous fields indicated that was going to be the case for roadway contributing circumstances. Also, surface conditions on the roadway did present an effect on SVROR crashes. That was also indicated earlier on the surface conditions and weather conditions fields. It was also observed that 2016 did not show consistency with the rest of the years, which can be due to an error in the data. In a general note, SVROR crashes are not greatly affected by roadway contributing circumstances but could become an influencing factor when surface conditions are present.

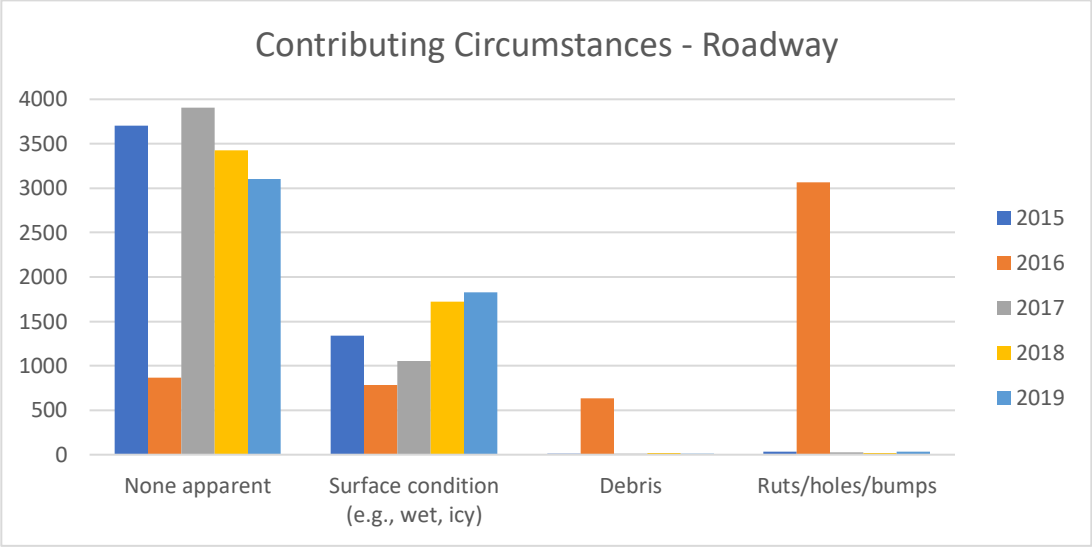

Figure 3.15: Annual SVROR crashes by contributing circumstances - roadway.

SVROR Horizontal Alignment

The horizontal alignment field was used in order to further evaluate the roadway horizontal alignment design in which crashes took place. Those include whether the alignment was straight or had curves, and if other characteristics were involved. Those horizontal alignment features are shown, along with annual frequency of occurrence, in Table 3.11. For SVROR crashes, it is expected that most crashes occur on straight horizontal alignment. The presence of curves along the alignment in which these crashes occurred might have an impact on the number of crashes or severity, but that is not as expected. These results are displayed graphically in Figure 3.16, and only relevant categories are included.

Table 3.11: Annual SVROR crashes by horizontal alignment.

	2015	*2016*	*2017*	*2018*	*2019*
Straight	3907	3900	3921	4059	3955
Traversing curve to left	730	742	738	741	689
Traversing curve to right	495	545	494	536	537
Other (explain in narrative)	24	26	32	22	32
Unknown	76	37	29	19	15
Not Reported	123	115	76	67	35

As expected for SVROR crashes, straight horizontal alignments presented the most crashes with a very steady number of crashes on the evaluated years. Straight horizontal alignments take approximately 70% of SVROR crashes while approximately

20% of SVROR crashes occur along horizontal alignment curves. Comparing those crashes that occur along curves with those that do not, the horizontal alignment design factors do not significantly affect SVROR crashes. The previous statement indicates that most likely, there are no improvement that could be applied to the horizontal alignment in order to decrease the number of crashes that occurred in that location.

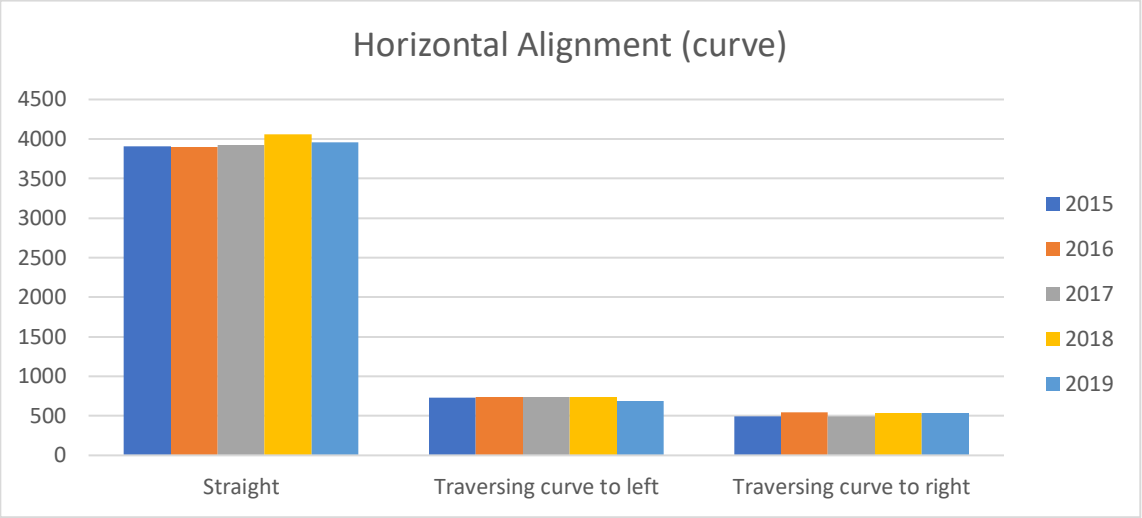

Figure 3.16: Annual SVROR crashes by horizontal alignment.

SVROR Vertical Alignment

The vertical alignment field was used in order to further evaluate the roadway vertical alignment design in which crashes took place. Those include whether the alignment was level, at crest, traversing uphill or downhill, and if other characteristics were involved. Those vertical alignment features are shown, along with annual frequency of occurrence, in Table 3.12. For SVROR crashes, it is expected that most crashes occur along level alignment. Traversing uphill or downhill along the alignment in which these crashes occurred might have an impact on the number of crashes or severity, but that is

not as expected. These results are displayed graphically in Figure 3.17, and only relevant categories are included.

Table 3.12: Annual SVROR crashes by vertical alignment.

	2015	*2016*	*2017*	*2018*	*2019*
Level	3,711	3,651	3,736	3,762	3,649
At crest	182	171	168	153	169
Traversing uphill	410	486	447	519	538
Traversing downhill	740	787	760	807	755
At sag (bottom of hill)	104	98	68	99	78
Other (explain in narrative)	17	15	6	15	21
Unknown	64	35	25	19	15
Not Reported	127	122	80	70	38

As expected for SVROR crashes, level vertical alignments presented the most crashes with a very steady number of crashes on the evaluated years. Similarly, to the horizontal alignments, level vertical alignments take approximately 70% of SVROR crashes while approximately a little over 20% of SVROR crashes occur along traversing vertical alignments. With those results, it can be implied that vertical alignment design factors do not significantly affect SVROR crashes. The previous statement indicates that most likely, there are no improvement that could be applied to the vertical alignment design to decrease crashes that occurred in that location.

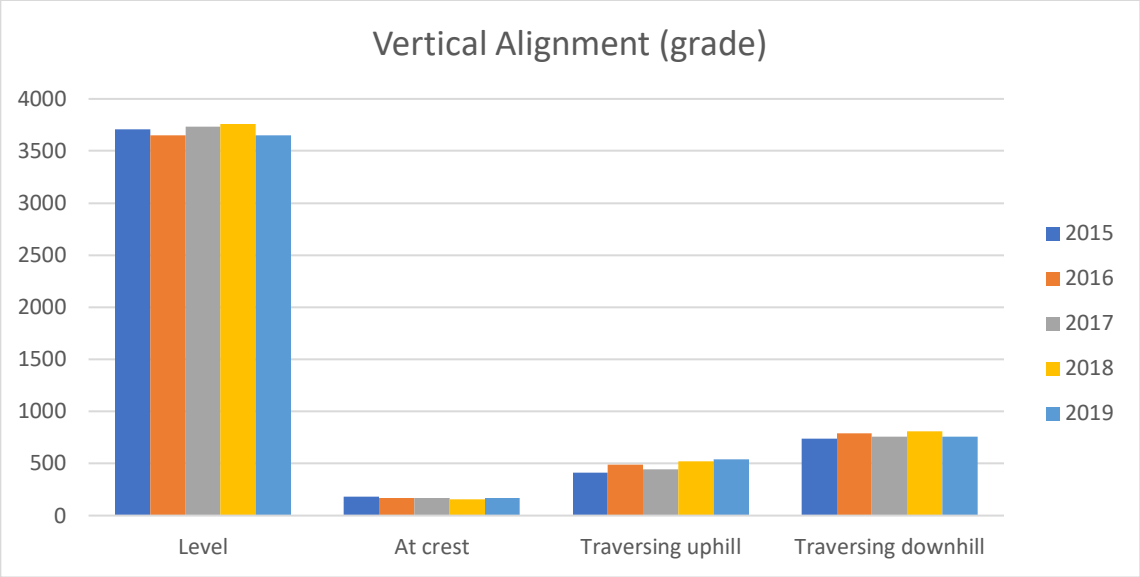

Figure 3.17: Annual SVROR crashes by vertical alignment.

Methodology

As was mentioned earlier, ArcGIS was used to visualize and manipulate the data as the software enables creation and viewing of maps, compilation of spatial and non-spatial data, and analysis of spatial data. Geographic Information System (GIS) software (such as ArcGIS) provides multiple functions to analyze crashes, having been used for simple linear analyses, map development, identification of risk areas, and spatial analysis (*13, 18, 29, 31, 32, 84*). The software has been used in recent studies to generate maps, create models, and for risk estimation. Spatial characteristics have been previously evaluated using the tools of proximity analysis, density analysis, hotspot analysis, directional distribution analysis, group analysis, overlay analysis, spatial-temporal analysis, and network analysis (*18*). Additionally, GIS has been a useful tool when understanding and analyzing data to identify hotspots. Multiple studies have used GIS as a tool to identify hotspots but in most cases, clusters are not created after that (*7*). The

Highway Safety Information System (HSIS) uses GIS safety tools and screening techniques to analyze data, but it mainly determines crash counts (*29*).

In previous studies, the spatial join operation in GIS has been adopted to aggregate the segment buffers which have allowed the identification of various types of midblock segments (*15*). Other studies have used some toolsets by GIS to detect hotspots (*21*). Particularly, a study had the objective to identify hotspots by location while using a spatial statistical visualization and modeling techniques. The used methodology involved data analysis that included million vehicle kilometers travelled (MVKT) measures and the emerging hotspot analysis tool by ArcGIS Pro. From ArcGIS, several tools were used to summarize spatial distribution, identify clusters, and explore patterns over time. From the study, it was found that hotspots were identified and a better development for roads could be implemented thanks to the methodology that was used (*18*). In previous studies and overall, ArcGIS has been found to be a useful tool to identify potential hazardous locations and implement safety measures (*31*).

For my research, ArcGIS was used to extract data for the descriptive statistics previously presented and to develop data subsets (e.g., SVROR) for the subsequent processes. ArcGIS Pro was used with these data subsets to develop the clustering. To facilitate the ArcGIS Pro processes, Python code was developed to automate selection of each crash then, for each crash, traveling along the roadway network to identify frequency of crashes within bins of distance based on tenths of stopping sight distance (SSD). Sight distance is one of the primary elements of roadway design. Sight distance is the available visibility provided to a driver in order to be able to observe stationary or moving objects along the road surface. Design criteria and guidance for sight distance is

relative to its application depending on the highway and street classification (*88*). As different roads were evaluated, the sight distance to be used was determined depending on the type of road and corresponding speed limit. The data were manipulated within Excel and mapped with ArcGIS to visualize the data in an alternate way, both graphs and within maps, to verify and make sense of the KDE output from R.

R is widely used to evaluate data and to develop software and data analysis. R was used on the ArcGIS Pro/Python output to generate the KDE scatterplots (using the smoothScatter function).

Applying the KDE method to the obtained data was a crucial step in the analysis. It was important to determine whether the proposed and created clustering method was significant or not. Also, evaluating the proposed method and observing the resulting clustering distribution was important and R allowed the simplification of that process. To apply KDE to the obtained data, the smoothScatter function was used. The application of this function allows the creation of a scatterplot with a smoothed color density representation using a 2D kernel density estimate. The obtained output will not only illustrate the points where the crashes occurred, but also the color of the plot will indicate whether a cluster is found or not, and if the cluster is significant.

CHAPTER 4. RESULTS

To evaluate the effectiveness of the initial methodology, three different case study roadways were developed: I-35 from Ankeny to Ames (north of Des Moines), I-380 from Iowa City to Cedar Rapids, and US20 from I-35 east to Dike. Both I-35 and I-380 are heavily travelled, high-speed routes in Iowa with a high percentage of daily commuters but with relatively few curves. US 20 is a less heavily travelled, slightly lower speed highway with some notable curvature. All three routes travel through terrain that undulates with alternating farm field (e.g., corn, soybean) and tree cover, as is typical of much of Iowa.

Summary descriptive data regarding, and the case study sections are shown in Table 4.1. General information regarding each case study location is provided, including annual average daily traffic (AADT) volumes, length (in miles), and vehicle miles travelled (VMT), the latter used for crash rate calculations. For each section, two subsets and two stopping sight distance (SSD) lengths were developed. The "rural" subset includes all crashes along the sections that are outside corporate limits per the Iowa road data. The "SVROR" subset includes rural crashes that were also coded as single-vehicle, ROR crashes. For all of the rural crashes, I-380 presented the highest crash rate which is expected since I-380 has high volumes. The lowest crash rate was found when single vehicle run off road crashes were evaluated with a total sight distance along I-380. All these types of crashes will be evaluated in the following section.

Table 4.1: Case study descriptive data and 5-year crash data.

						Crash	
	AADT	Length	VMT	Subset	SSD Length	Frequency	Rate
I-35	46,358	15	686,426	Rural	*Full*	6,173	24.5
					Half	3,703	14.7
				SVROR	*Full*	666	2.6
					Half	406	1.6
US 20	9,176	48	441,210	Rural	*Full*	2,277	14.1
					Half	1,048	6.5
				SVROR	*Full*	412	2.6
					Half	288	1.8
I-380	53,025	9	479,560	Rural	*Full*	8,085	46.2
					Half	4,801	27.4
				SVROR	*Full*	445	2.5
					Half	286	1.6

The output from ArcGIS Pro/Python contained individual crash point coordinates and annual crash frequencies within SSD bins, each iteratively 1/10th (i.e., $\frac{1}{10}, \frac{2}{10}, \ldots, half\ SSD, \ldots, full$ SSD). These data included the crash coordinates, each year from 2015 through 2019, and crash counts for each individual bin. Within Excel, these data were summarized to 5-year totals and subsets for half SSD and full SSD were extracted. From the subsets, Excel graphs were developed using the bubble graph option. Additionally, the subsets were imported into ArcGIS to develop thematic maps using the

coordinates and the crash counts. These maps were exported as KMLs for importation into GoogleEarth to produce the images showing crash "clusters" coincident with aerial imagery. Furthermore, the data subsets, after minor additional preparation, were imported into R for KDE evaluation. Using R, the smoothScatter function which involved kernel density estimation was applied to the data. The set of different plots created via Excel, ArcGIS/GoogleEarth, and R allowed comparative visualization of results, mainly to aid understanding of the KDE output. Individual case study discussion occurs in the following sections.

Interstate 35 – Iowa

I-35 is classified as an interstate and stretches north to south from Texas to Minnesota. I-35 traverses the State of Iowa from south to north, from Missouri to Minnesota, passing through 9 counties: Worth, Cerro Gordo, Franklin, Wright, Hamilton, Story, Polk, Warren, Clarke, and Decatur. I-35 is a primary travel route through the center of the country, serving commercial truck traffic but also local commuter traffic (90). I-35 passes through the Des Moines metropolitan area which is the largest metropolitan area in Iowa. Traffic volumes within and in the vicinity of Des Moines are notably higher. The section from Ankeny north to Ames is a heavily travelled commuter route.

The speed limit on rural interstates in Iowa is 70 mph, which applies to this section of I-35. For that speed limit, the corresponding design stopping sight distance is 730 ft, using the standard SSD equation assuming a flat grade (i.e., 0%) and rounding for design purposes (88). However, to add some accommodation for the undulating terrain, we used a 3% grade and, thus, adjusted the SSD to roughly 770 ft. Thus, our tenths were

additive iterations of roughly 77 ft, with half SSD at roughly 386 ft. For this section of I-35, both all rural crashes and rural, SVROR crashes were evaluated.

Interstate 35 Full Stopping Sight Distance – Iowa Rural Crashes

Figure 4.1 displays the I-35 case study area with the "clusters" displayed within both GoogleEarth (KMZ) and Excel. The Excel graph appears to have more curvature due to a distortive effect of the horizontal axis scale. The split between the north and the south portion is due to that section of I-35 being interpreted as within the corporate limits of Huxley, Iowa.

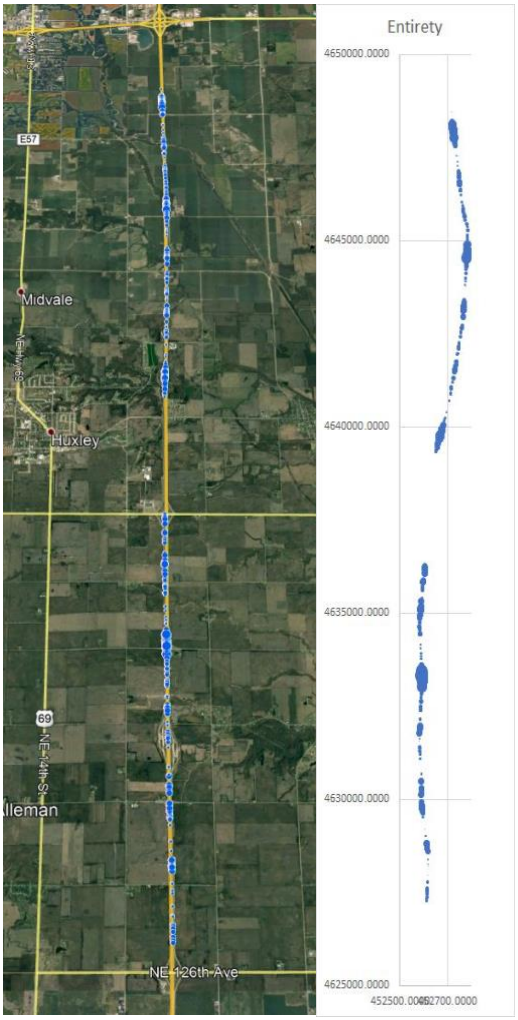

Figure 4.1: I-35 rural clusters, full SSD – KMZ and Excel.

Comparing the GoogleEarth and Excel images, several similarities are apparent. Many of the “clusters” appear to be similarly located, as expected. Additionally, the spread of these “clusters” and relative “intensity” seem to match. Furthermore, based on this, when using the GoogleEarth imagery, the “clusters” can be visually associated to the surrounding landscape. Additionally, as shown in Figure 4.2, the Excel graphs can be divided to provide additional detail to the visualization. Clearly, the same can be done with the GoogleEarth imagery. However, again, the primary purpose of both the GoogleEarth and Excel images is to provide context to the R-based KDE output.

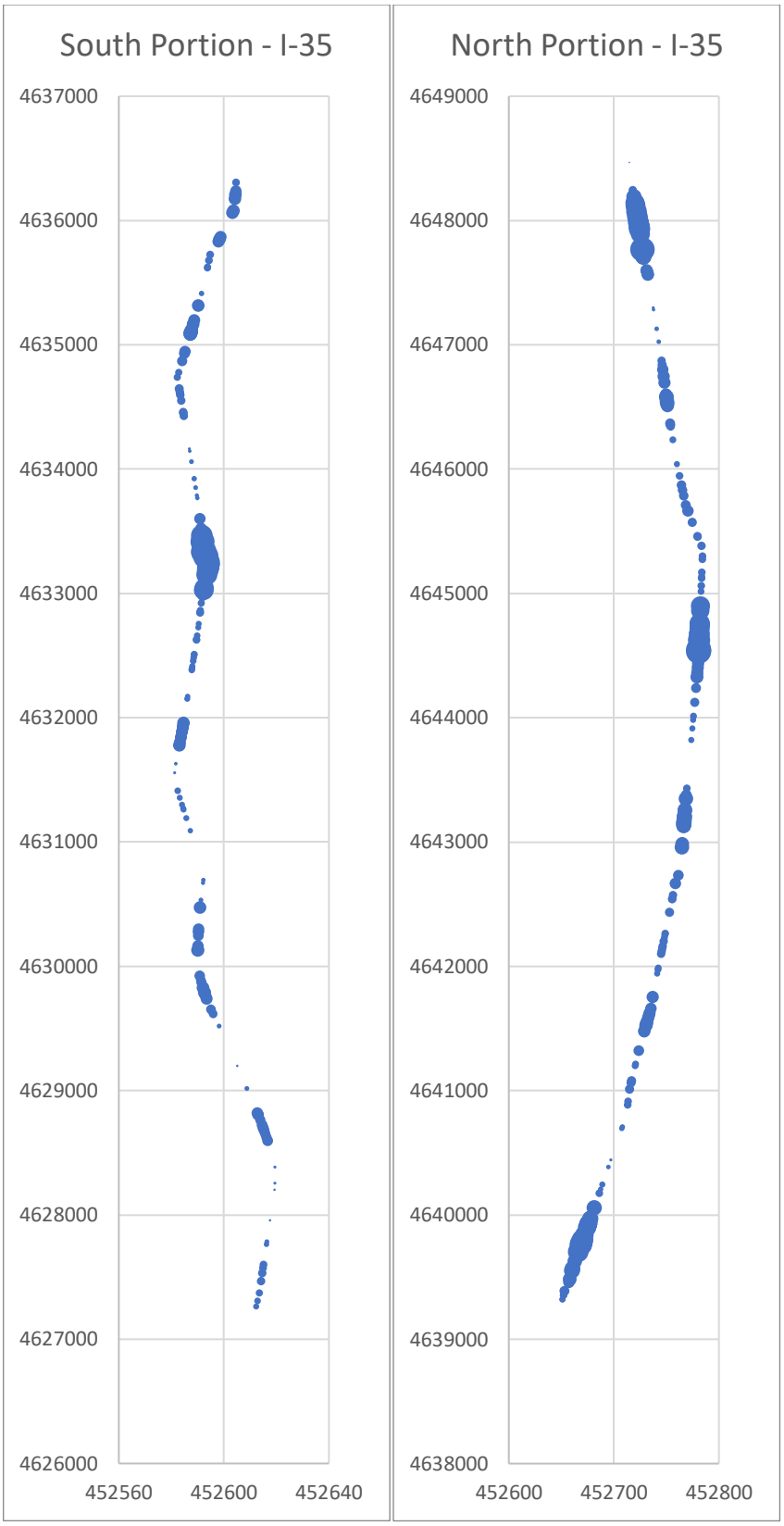

Figure 4.2: I-35 north and south section rural clusters, full SSD – Excel.

With the comparison and verification from the previous GoogleEarth and Excel figures available, the KDE output generated from the smoothScatter function in R can be evaluated with context. Figure 4.3 provides the 2D kernel density estimate for rural crashes along the I-35 case study section including the total count.

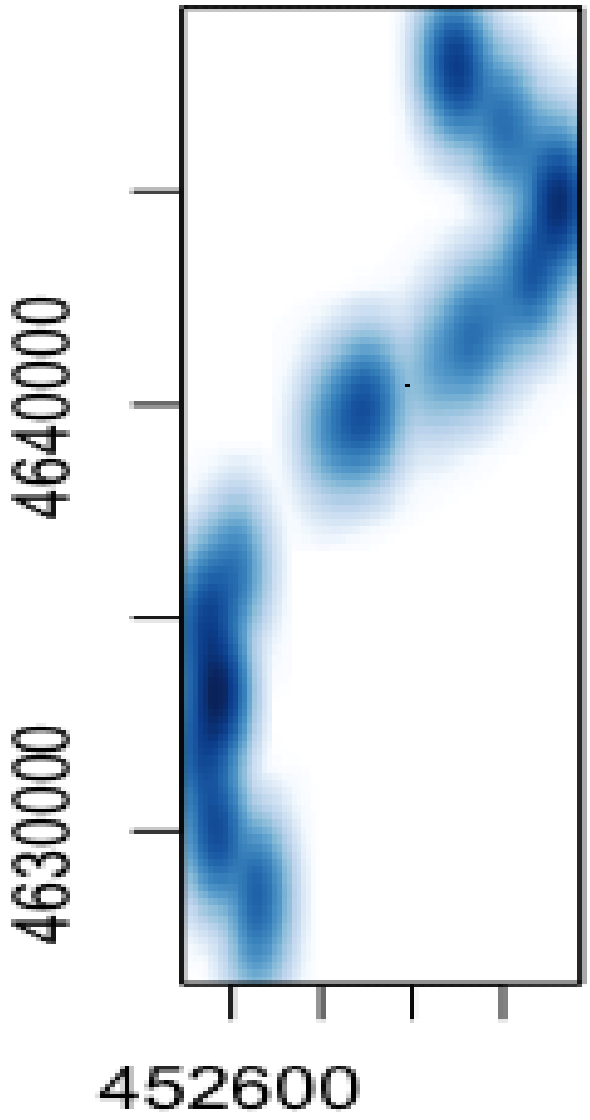

Figure 4.3: I-35 rural clusters, full SSD – KDE.

In the figure, crash occurrence is represented with the color blue. Those locations with darker blue represent higher crash concentrations, while those with lighter blue, display locations with fewer crashes. The KDE output is similar to the Excel and GoogleEarth output, with similar locations and intensities. Four primary "clusters" were evident within the Excel and GoogleEarth output; however, the KDE output seems to indicate three. However, as the KDE output is statistically based, the output results are more reliable.

Interstate 35 Half Stopping Sight Distance – Iowa Rural Crashes

The crashes count involved in half of the sight distance will be evaluated. Figure 4.4 shown below, includes the "clusters" for those crashes in both GoogleEarth and Excel. When the aerial imagery is observed in more detail, it can be observed that similar clusters to those found with the full stopping sight distance is observed, with the difference that less total crashes are present.

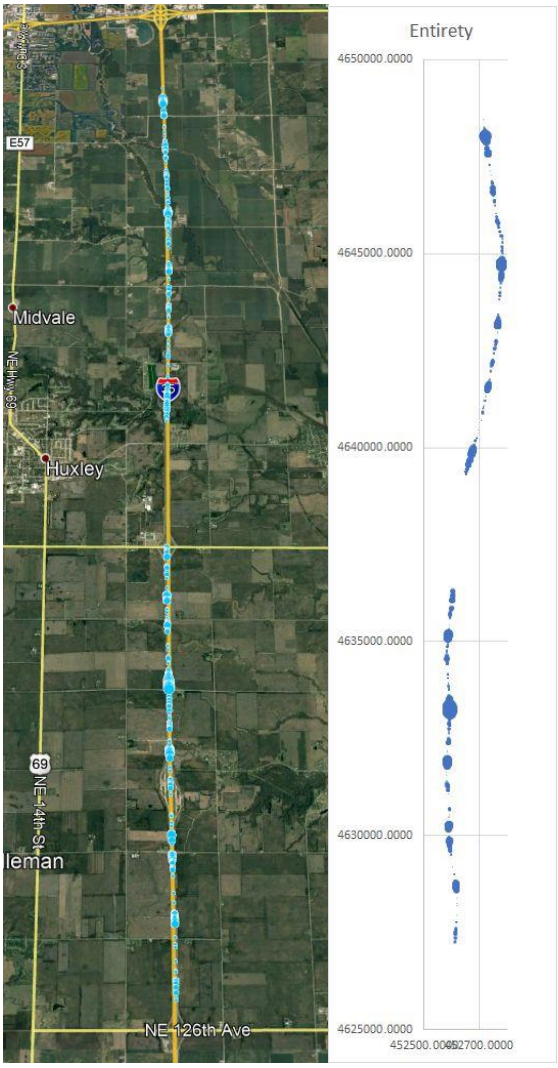

Figure 4.4: I-35 rural clusters, half SSD – KMZ and Excel.

Similar to the total sight distance, from the GoogleEarth and Excel images, several similarities are apparent. Many of the "clusters" appear to be similarly located, as expected, but with less volume when compared to the full sight distance count. In order to compare and verify those results obtained with GoogleEarth and Excel, the smoothScatter function in R will be applied to the data set. Figure 4.5 displays the 2D kernel density estimate for rural crashes along the I-35 case study section including the half count of the sight distance.

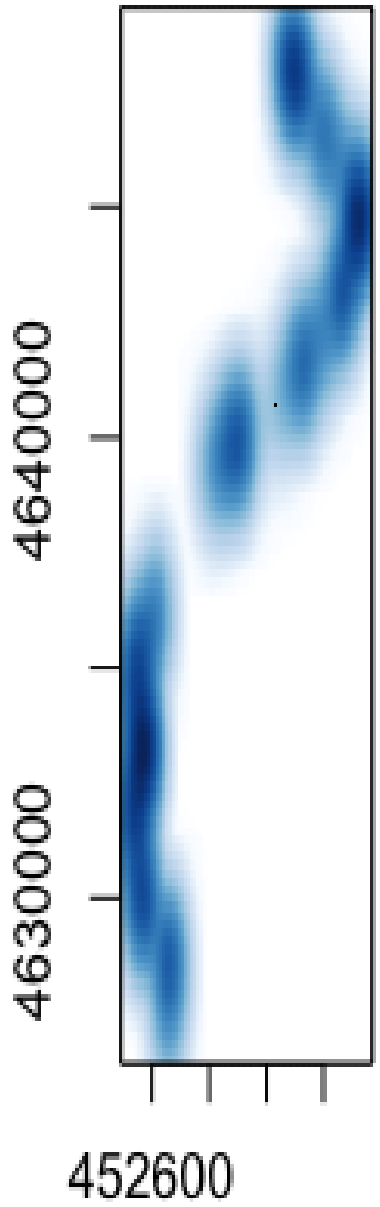

Figure 4.5: I-35 rural clusters, half SSD – KDE.

As in the previous section, the KDE output is similar to the Excel and GoogleEarth output. The KDE output presented four “clusters” as well as the Excel and GoogleEarth images. The location of the found “clusters” is the same location as those observed on the full sight distance situation. However, the density of those “clusters”

when compared to the full sight distance is different for the reason that as expected, there are fewer total crashes when half of the distance is observed. With both methods and both sight distances, most crashes along the selected segment occurred on traversing curves to the right and, along ramps and bridges.

Interstate 35 Full Stopping Sight Distance – Iowa SVROR Crashes

Figure 4.6 shown below, includes an illustration of the I-35 case study area with the "clusters" obtained by both GoogleEarth and Excel for single vehicle run-of-road crashes. Similar to the rural crashes in the same segment, the Excel graph includes more curves due to the characteristics of the horizontal scale.

Figure 4.6: I-35 SVROR clusters, full SSD – KMZ and Excel.

Single vehicle run off road crashes along the selected area of I-35, are similar to those found in the total rural crashes. Many of the "clusters" conserved their location and fewer crashes on those locations are observed. From the figures, it was observed that most crashes along the selected segment occurred along curves. To evaluate and verify the found "clusters", the KDE method was applied to the selected section using the smoothScatter function in R. Figure 4.7 provides the 2D kernel density estimate for single vehicle run off road crashes along the selection section in I-35 including the total count.

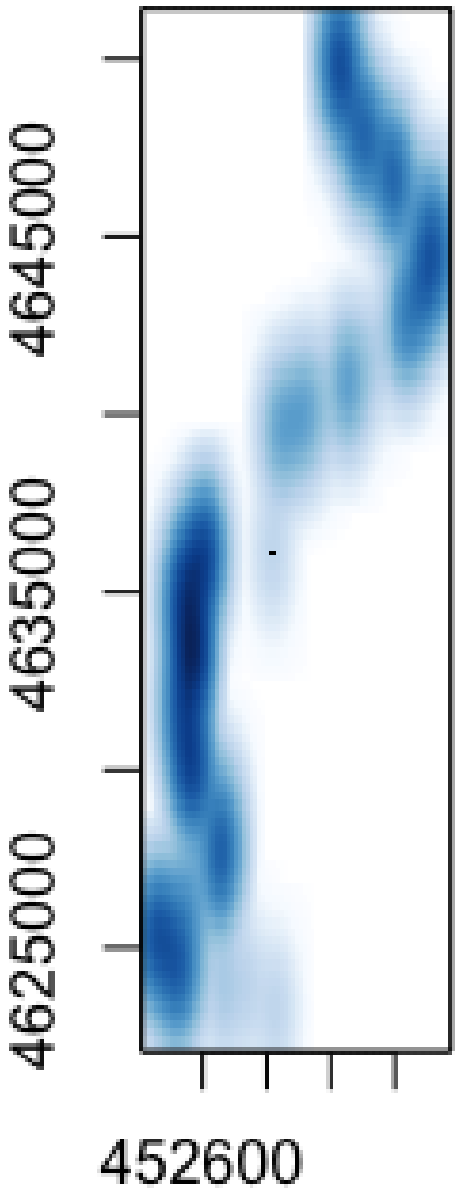

Figure 4.7: I-35 SVROR clusters, full SSD – KDE.

In the figure, locations where crashes occurred are represented with the color blue as in the similar sections. With the KDE, a similar and clearer output can be observed. Properly representation of the “clusters” is represented as expected. Those areas that should be prioritized can also be identified. With both methods, it can be observed that

most crashes along the selected segment occurred on a traversing curve to the right. A similar output as the one obtained with full stopping sight distance for rural crashes is observed.

Interstate 35 Half Stopping Sight Distance – Iowa SVROR Crashes

The single vehicle run-off-road crashes count involved in half of the sight distance will be evaluated. Figure 4.8 shown below, includes the "clusters" for those crashes in both GoogleEarth and Excel. When the aerial imagery is observed in more detail, it can be observed that similar clusters to those found with the full stopping sight distance is observed, with the difference that less total crashes are present.

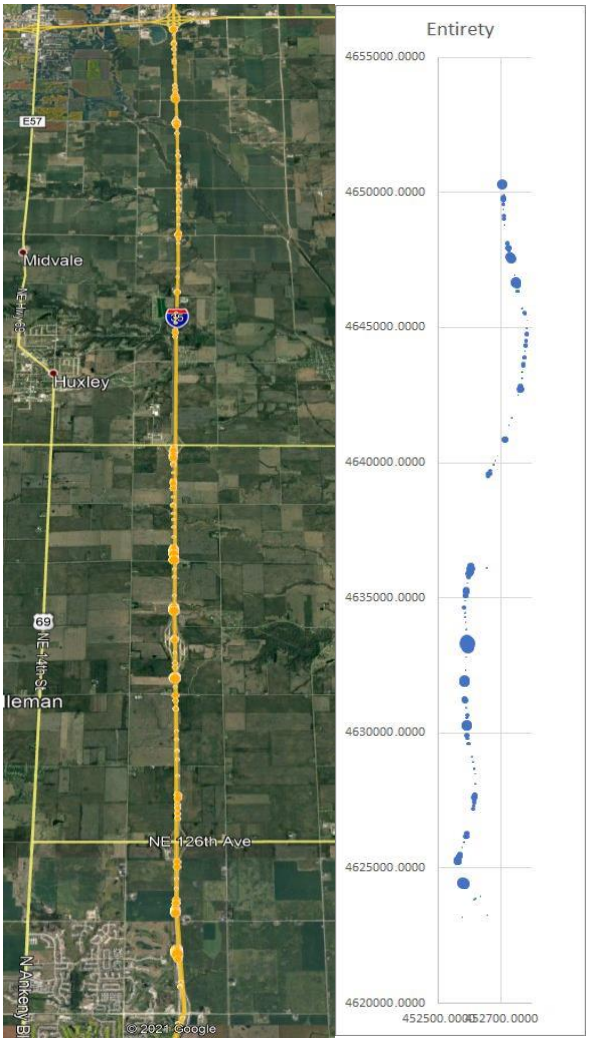

Figure 4.8: I-35 SVROR clusters, half SSD – KMZ and Excel.

Similar to the total sight distance, from the GoogleEarth and Excel images, several similarities are apparent. Many of the "clusters" appear to be similarly located, as expected, with the different that less crashes are observed. In order to compare and verify those results obtained with GoogleEarth and Excel, the smoothScatter function in R will be applied to the data set. Figure 4.9 displays the 2D kernel density estimate for single vehicle run-off-road crashes along the I-35 case study section including the half count of the sight distance.

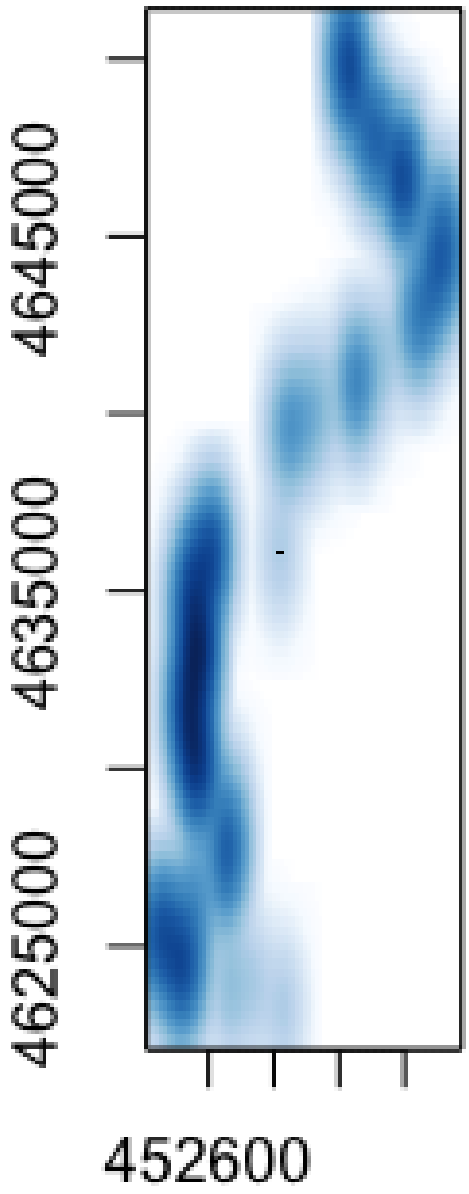

Figure 4.9: I-35 SVROR clusters, half SSD – KDE.

As in the previous section, the KDE output is similar to the Excel and GoogleEarth output. The KDE output presented one “cluster” as well as the Excel and GoogleEarth images. The location of the found “clusters” is the same location as those observed on the full sight distance situation. However, the density of those “clusters” when compared to the full sight distance is different as expected. With both methods and

both sight distances, the areas of interest were identified. Similar outputs were observed when both rural and single vehicle run-off-road crashes were compared. Using the full stopping sight distance, a total of four "clusters" were identified for rural crashes, and for single vehicle run-off-road crashes, only one of those four "clusters" were outstanding. When comparing those results with the half stopping sight distance, a very similar output was observed for both categories with the different that fewer total crashes were observed. For all of the evaluated categories in the case study, most crashes occurred along curves, ramps, and bridges.

US Route 20 – Iowa

US 20 is classified as a non-interstate expressway and stretches east to west from Massachusetts to Oregon. US 20 traverses the State of Iowa from east to west, from Illinois to Nebraska, passing through 12 counties: Woodbury, Ida, Sac, Calhoun, Webster, Hamilton, Hardin, Grundy, Black Hawk, Buchanan, Delaware, and Dubuque. US 20 is a primary travel route for the northern portion of the State, serving commercial truck traffic but also local commuter traffic (90). US 20 passes through the Dubuque, Waterloo/Cedar Falls, and Sioux City metropolitan areas which are larger communities in Iowa. Traffic volumes within and in the vicinity of these metropolitan areas are notably higher as these sections are a well-travelled commuter route. However, the section used for this evaluation, from I-35 eastward to Dike is much more rural in nature and not as heavily travelled.

The speed limit on rural expressways in Iowa is 65 mph, which applies to this section of US 20. For that speed limit, the corresponding design stopping sight distance is 645 ft, using the standard SSD equation assuming a flat grade (i.e., 0%) and rounding for

design purposes (88). However, to add some accommodation for the undulating terrain, we used a 3% grade and, thus, adjusted the SSD to roughly 680 ft. Thus, our tenths were additive iterations of roughly 68 ft, with half SSD at roughly 341 ft. For this section of US 20, both all rural crashes and rural, SVROR crashes were evaluated.

US Route 20 Full Stopping Sight Distance – Iowa Rural Crashes

Figure 4.10 displays the US 20 case study area the "clusters" displayed within both GoogleEarth (KMZ) and Excel. Comparing the GoogleEarth and Excel images, several similarities are apparent. Many of the "clusters" appear to be similarly located, as expected. Additionally, the spread of these "clusters" and relative "intensity" seem to match. Furthermore, based on this, when using the GoogleEarth imagery, the "clusters" can be visually associated to the surrounding landscape. Again, the primary purpose of both the GoogleEarth and Excel images is to provide context to the R-based KDE output.

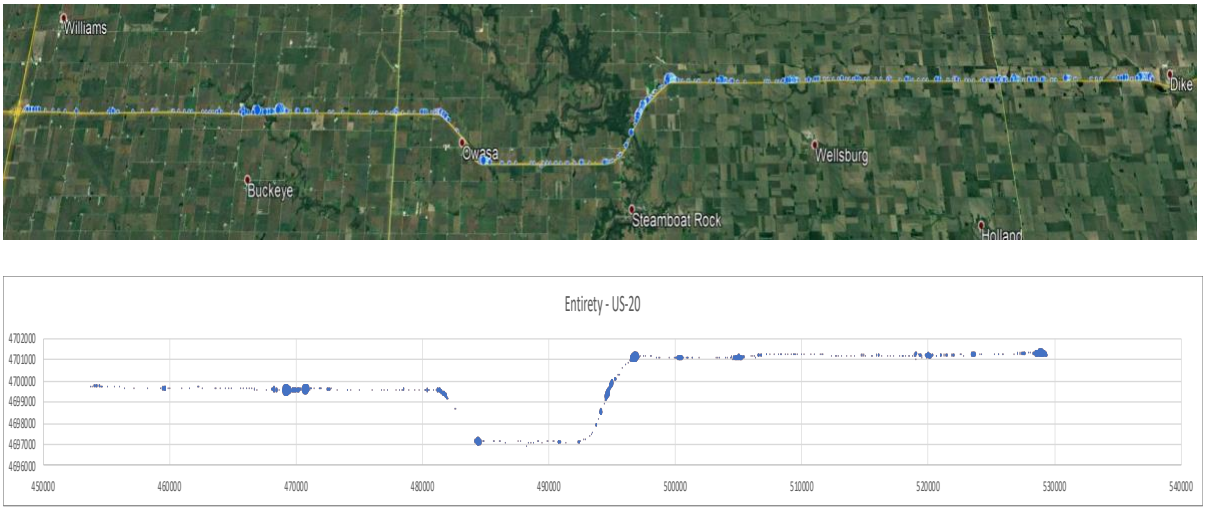

Figure 4.10: US 20 rural clusters, full SSD – KMZ and Excel.

With the comparison and verification from the previous GoogleEarth and Excel figures available, the KDE output generated from the smoothScatter function in R can be

evaluated with context. Figure 4.11 provides the 2D kernel density estimate for rural crashes along the US 20 case study section including the total count.

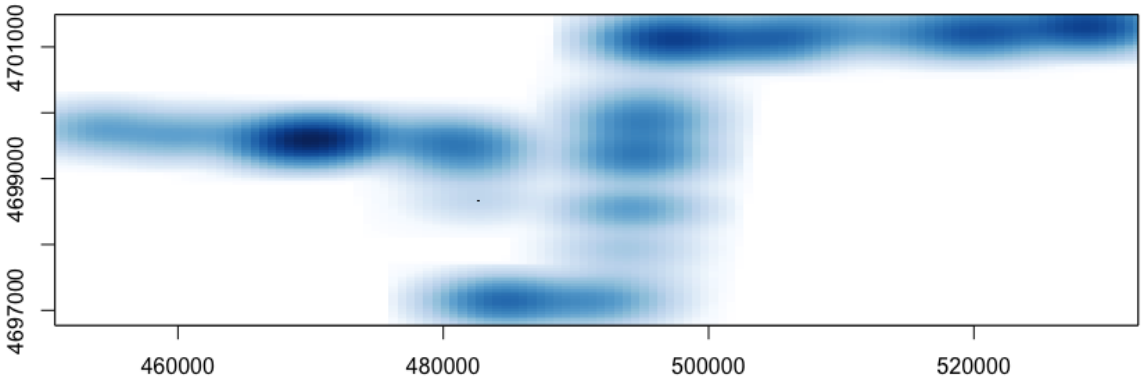

Figure 4.11: US 20 rural clusters, full SSD – KDE.

In the figure, crash occurrence is represented with the color blue. Those locations with darker blue represent higher crash concentrations, while those with lighter blue, display locations with fewer crashes. The KDE output is similar to the Excel and GoogleEarth output, with similar locations and intensities. Primary "clusters" were evident within the Excel and GoogleEarth output; however, the KDE output seems to indicate three. However, as the KDE output is statistically based, the output results are more reliable.

US Route 20 Half Stopping Sight Distance – Iowa Rural Crashes

Figure 4.12 shown below, includes the "clusters" for those crashes in both GoogleEarth and Excel. When the aerial imagery is observed in more detail, it can be observed that similar clusters to those found with the full stopping sight distance is observed, with the difference that less total crashes are present.

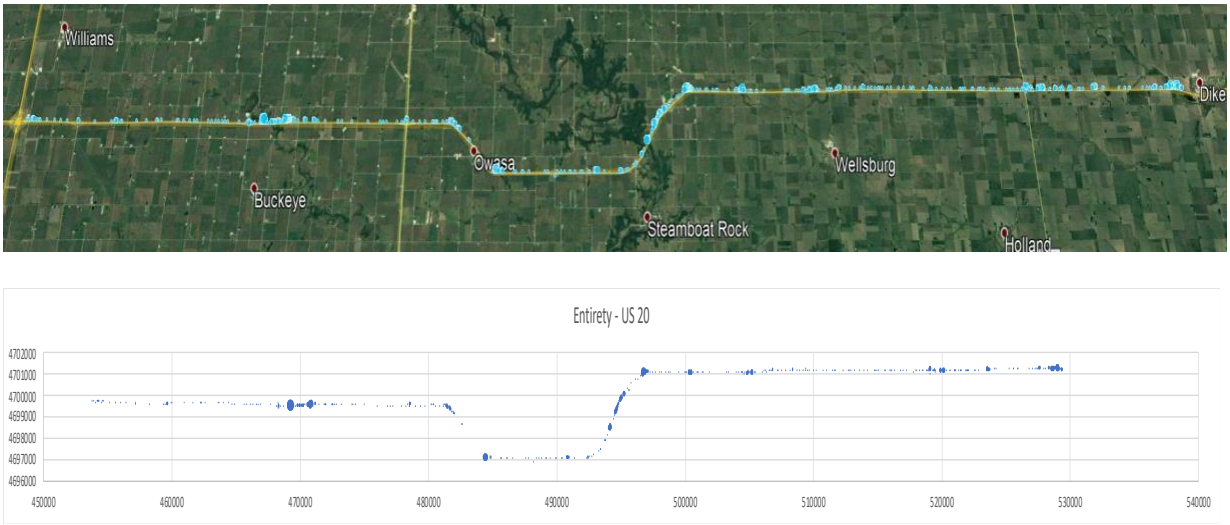

Figure 4.12: US 20 rural clusters, half SSD – KMZ and Excel.

Similar to the total sight distance, from the GoogleEarth and Excel images, several similarities are apparent. Many of the "clusters" appear to be similarly located, as expected, but with less volume when compared to the full sight distance count. In order to compare and verify those results obtained with GoogleEarth and Excel, the smoothScatter function in R will be applied to the data set. Figure 4.13 displays the 2D kernel density estimate for rural crashes along the US 20 case study section including the half count of the sight distance.

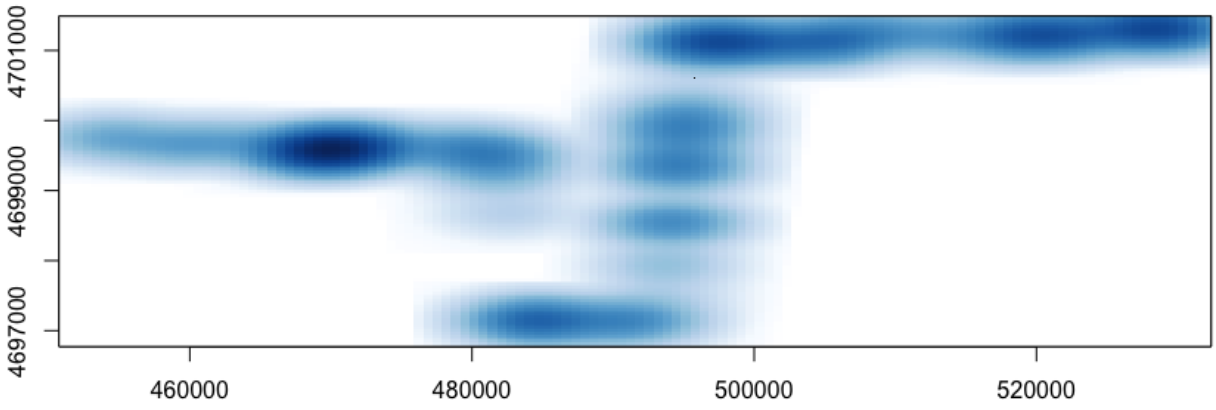

Figure 4.13: US 20 rural clusters, half SSD – KDE.

As in the previous section, the KDE output is similar to the Excel and GoogleEarth output. The KDE output presented three “clusters” as well as the Excel and GoogleEarth images with minor differences. The location of the found “clusters” is the same location as those observed on the full sight distance situation. However, the density of those “clusters” when compared to the full sight distance is different for the reason that as expected, there are fewer total crashes when half of the distance is observed. With both methods and both sight distances, most crashes along the selected segment occurred on traversing curves to the right and, along ramps and bridges.

US Route 20 Full Stopping Sight Distance – Iowa SVROR Crashes

Figure 4.14 shown below, includes an illustration of US 20 case study area with the “clusters” obtained by both GoogleEarth and Excel for single vehicle run-of-road crashes. Similar to the rural crashes in the same segment, the Excel graph includes more curves due to the characteristics of the horizontal scale.

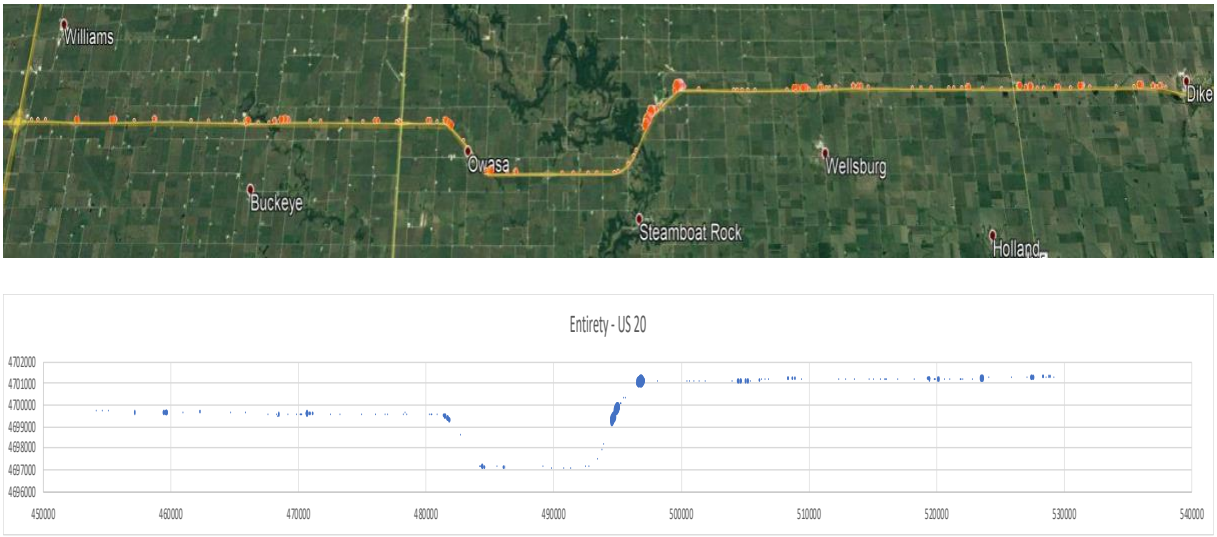

Figure 4.14: US 20 SVROR clusters, full SSD – KMZ and Excel.

Single vehicle run off road crashes along the selected area of US 20, are similar to those found in the total rural crashes. Many of the "clusters" conserved their location and fewer crashes on those locations are observed. From the figures, it was observed that most crashes along the selected segment occurred along curves. To evaluate and verify the found "clusters", the KDE method was applied to the selected section using the smoothScatter function in R. Figure 4.15 provides the 2D kernel density estimate for single vehicle run off road crashes along the selection section in US 20 including the total count.

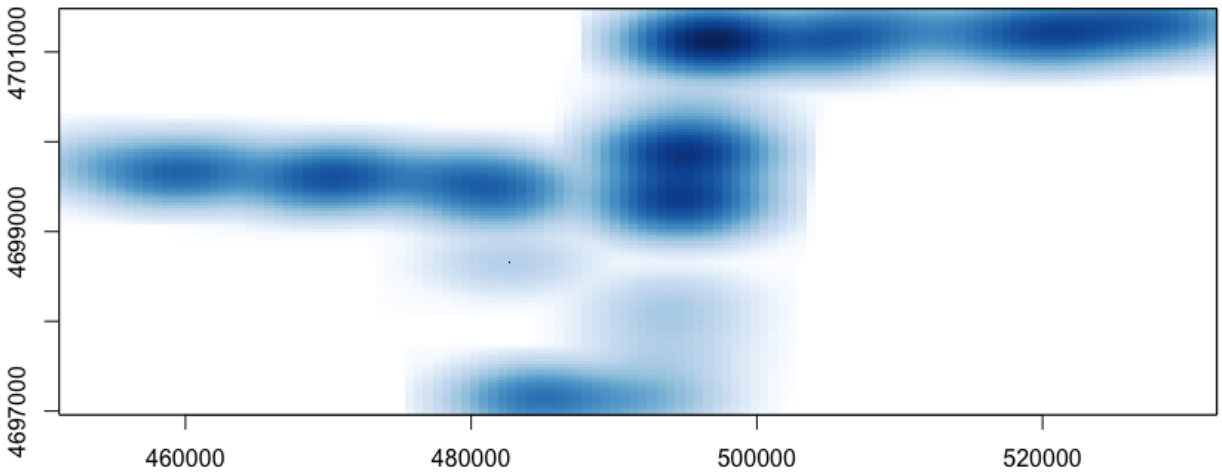

Figure 4.15: US 20 SVROR clusters, full SSD – KDE.

In the figure, locations where crashes occurred are represented with the color blue as in the similar sections. With the KDE, a similar and clearer output can be observed. Properly representation of the "clusters" is represented as expected. Those areas that should be prioritized can also be identified. With both methods, it can be observed that most crashes along the selected segment occurred along curves. A similar output as the one obtained with full stopping sight distance for rural crashes is observed but with less crashes, as expected.

US Route 20 Half Stopping Sight Distance – Iowa SVROR Crashes

The single vehicle run-off-road crashes count involved in half of the sight distance will be evaluated. Figure 4.16 shown below, includes the "clusters" for those crashes in both GoogleEarth and Excel. When the aerial imagery is observed in more detail, it can be observed that similar clusters to those found with the full stopping sight distance is observed, with the difference that less total crashes are present.

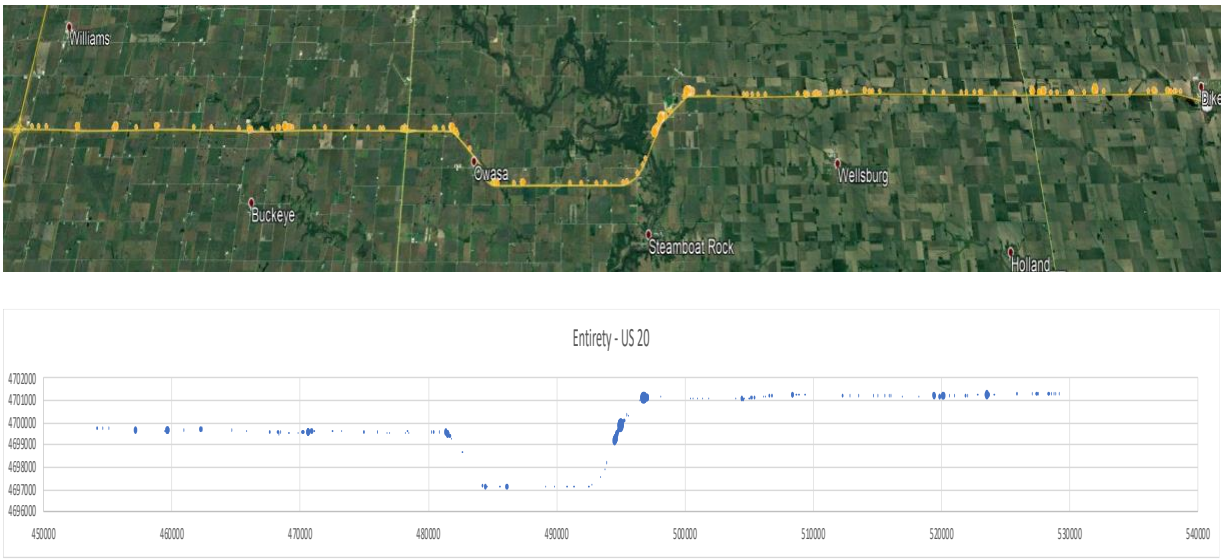

Figure 4.16: US 20 SVROR clusters, half SSD – KMZ and Excel.

Similar to the total sight distance, from the GoogleEarth and Excel images, several similarities are apparent. Many of the "clusters" appear to be similarly located, as expected, with the different that less crashes are observed. In order to compare and verify those results obtained with GoogleEarth and Excel, the smoothScatter function in R will be applied to the data set. Figure 4.17 displays the 2D kernel density estimate for single vehicle run-off-road crashes along the US 20 case study section including the half count of the sight distance.

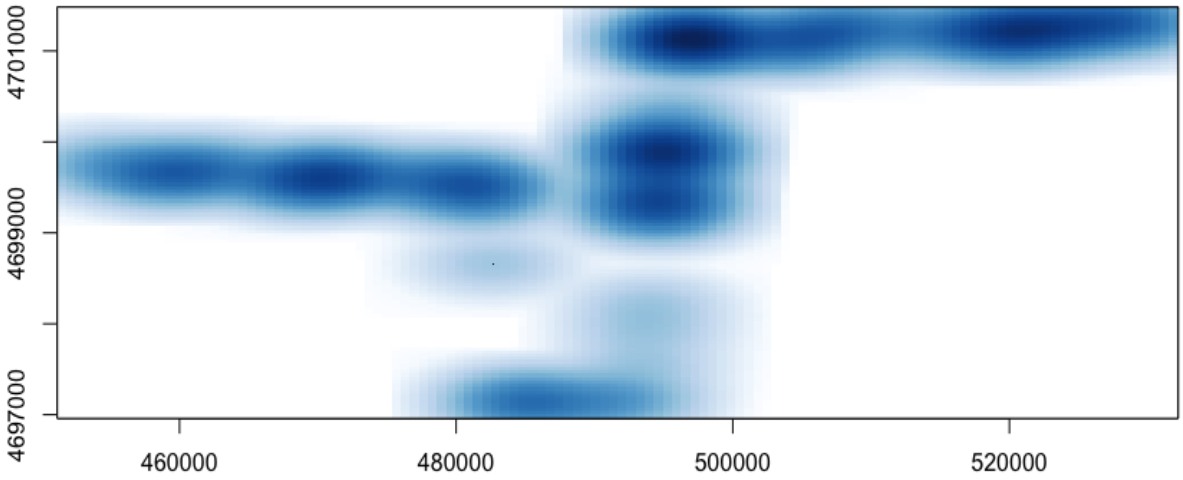

Figure 4.17: US 20 SVROR clusters, half SSD – KDE.

As in the previous section, the KDE output is similar to the Excel and GoogleEarth output. The KDE output presented one "cluster" as well as the Excel and GoogleEarth images. The location of the found "clusters" is the same location as those observed on the full sight distance situation. However, the density of those "clusters" when compared to the full sight distance is different as expected. With both methods and both sight distances, the areas of interest were identified. Similar outputs were observed when both rural and single vehicle run-off-road crashes were compared. Using the full stopping sight distance, multiple "clusters" were identified for rural crashes, and for single vehicle run-off-road crashes, only two of those "clusters" were outstanding. When comparing those results with the half stopping sight distance, a very similar output was observed for both categories with the different that fewer total crashes were observed. For all of the evaluated categories in the case study, most crashes occurred along curves, ramps, and bridges.

Interstate 380 – Iowa

I-380 is classified as an interstate but as an auxiliary interstate, is contained wholly within the State of Iowa. I-380 stretches north to south from Coralville/Iowa City to Waterloo and passes through 5 counties: Black Hawk, Buchanan, Benton, Linn, and Johnson. I-380 serves as a primary travel route between Coralville/Iowa City and Waterloo, serving commercial truck traffic but also local commuter traffic (90). I-380 passes through the Cedar Rapids/Marion metropolitan area which is the second largest metropolitan area in Iowa. Traffic volumes within and in the vicinity of these larger cities are notably higher, with the section between Coralville/Iowa City and Cedar Rapids/Marion being a heavily travelled route.

The speed limit on rural interstates in Iowa is 70 mph, which applies to this section of I-35. For that speed limit, the corresponding design stopping sight distance is 730 ft, using the standard SSD equation assuming a flat grade (i.e., 0%) and rounding for design purposes (88). However, to add some accommodation for the undulating terrain, we used a 3% grade and, thus, adjusted the SSD to roughly 770 ft. Thus, our tenths were additive iterations of roughly 77 ft, with half SSD at roughly 386 ft. For this section of I-380, both all rural crashes and rural, SVROR crashes were be evaluated.

Interstate 380 Full Stopping Sight Distance – Iowa Rural Crashes

Figure 4.18 displays the I-380 case study with the “clusters” displayed within both GoogleEarth (KMZ) and Excel. Comparing the GoogleEarth and Excel images, several similarities are apparent. Many of the “clusters” appear to be similarly located, as expected. Additionally, the spread of these “clusters” and relative “intensity” seem to

match. Furthermore, based on this, when using the GoogleEarth imagery, the “clusters” can be visually associated to the surrounding landscape.

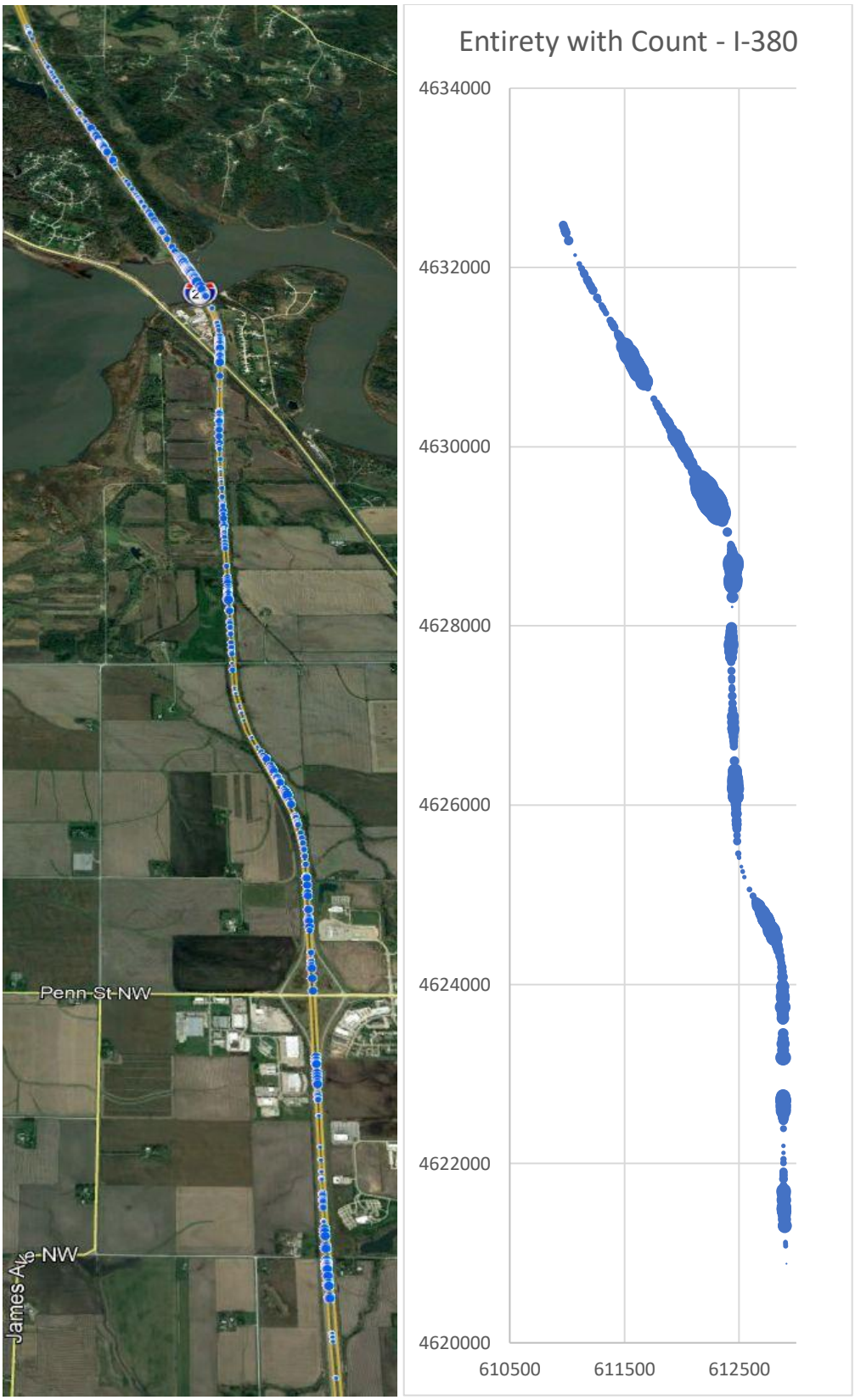

Figure 4.18: I-35 rural clusters, full SSD – KMZ and Excel.

With the comparison and verification from the previous GoogleEarth and Excel figures available, the KDE output generated from the smoothScatter function in R can be evaluated with context. Figure 4.19 provides the 2D kernel density estimate for rural crashes along the I-380 case study section including the total count.

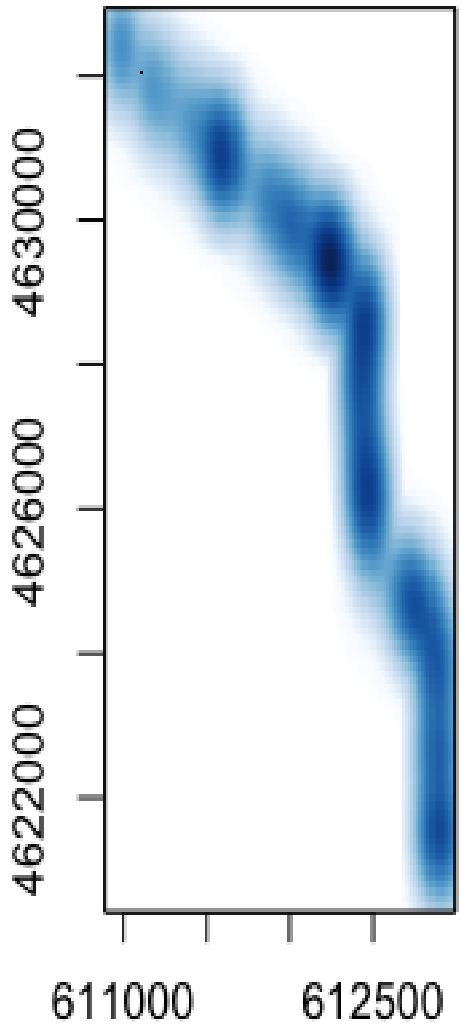

Figure 4.19: I-380 rural clusters, full SSD – KDE.

In the figure, crash occurrence is represented with the color blue. Those locations with darker blue represent higher crash concentrations, while those with lighter blue, display locations with fewer crashes. The KDE output is similar to the Excel and GoogleEarth output, with similar locations and intensities. Three primary "clusters" were evident within all of the outputs from the different methods.

Interstate 380 Half Stopping Sight Distance – Iowa Rural Crashes

The crashes count involved in half of the sight distance will be evaluated. Figure 4.20 shown below, includes the "clusters" for those crashes in both GoogleEarth and Excel. When the aerial imagery is observed in more detail, it can be observed that similar clusters to those found with the full stopping sight distance is observed, with the difference that less total crashes are present.

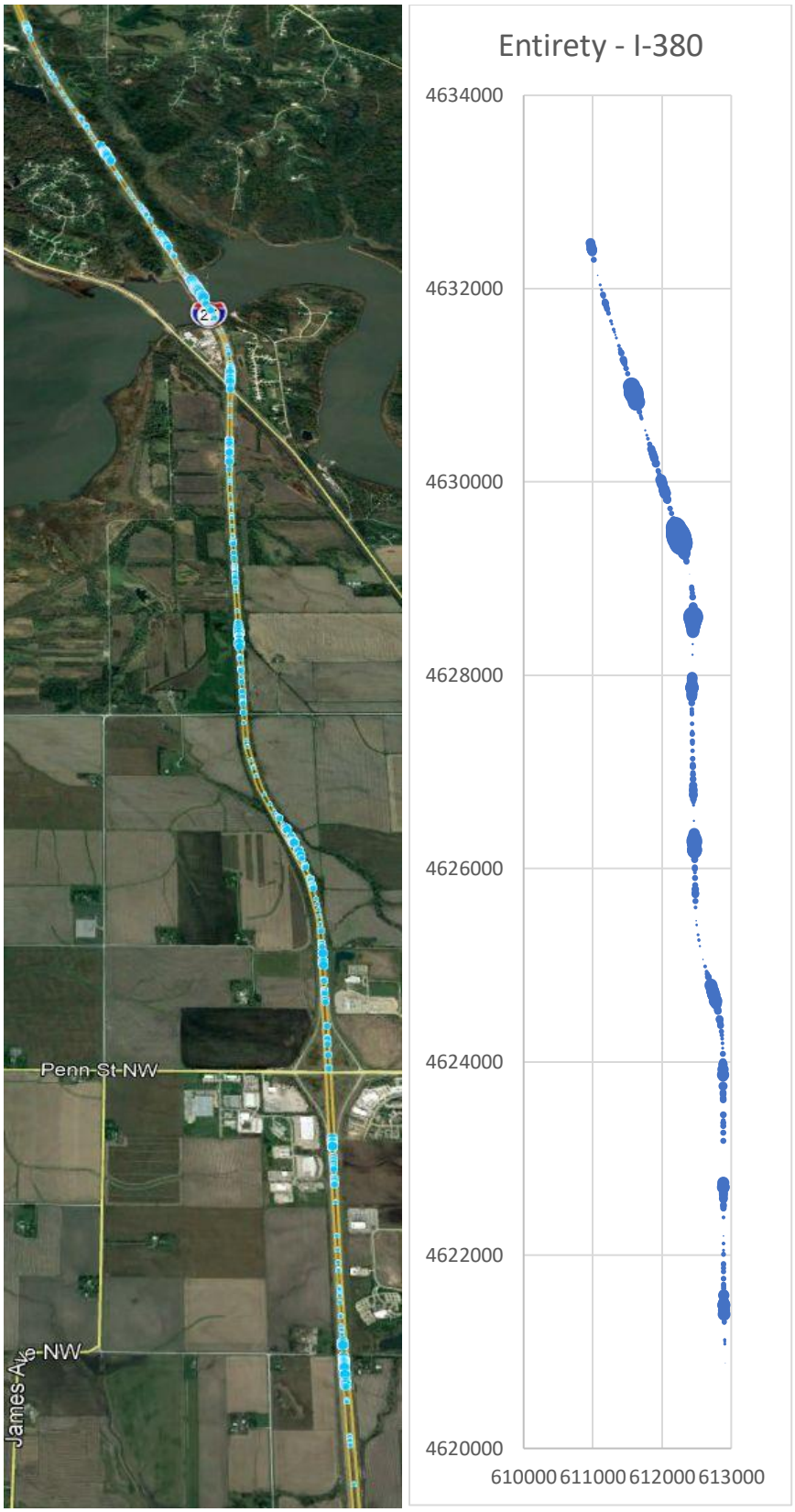

Figure 4.20: I-35 rural clusters, half SSD – KMZ and Excel.

Similar to the total sight distance, from the GoogleEarth and Excel images, several similarities are apparent. Many of the "clusters" appear to be similarly located, as expected, but with less volume when compared to the full sight distance count. In order to compare and verify those results obtained with GoogleEarth and Excel, the smoothScatter function in R will be applied to the data set. Figure 4.21 displays the 2D kernel density estimate for rural crashes along the I-380 case study section including the half count of the sight distance.

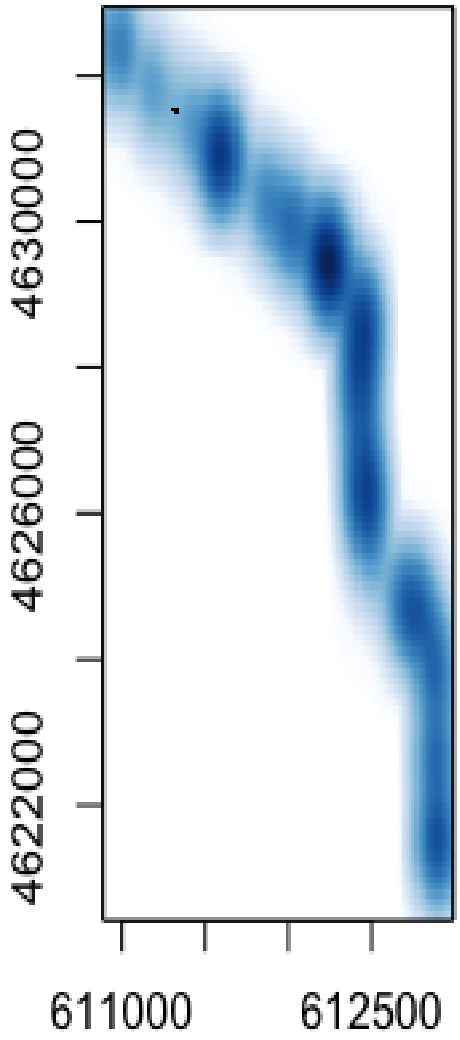

Figure 4.21: I-380 rural clusters, half SSD – KDE.

As in the previous section, the KDE output is similar to the Excel and GoogleEarth output. The KDE output presented three “clusters” as well as the Excel and GoogleEarth images. The location of the found “clusters” is the same location as those observed on the full sight distance situation. However, the density of those “clusters” when compared to the full sight distance is different for the reason that as expected, there are fewer total crashes when half of the distance is observed. With both methods and both sight distances, most crashes along the selected segment occurred along curves and bridges.

Interstate 380 Full Stopping Sight Distance – Iowa SVROR Crashes

Figure 4.22 shown below, includes an illustration of the I-380 case study area with the “clusters” obtained by both GoogleEarth and Excel for single vehicle run-of-road crashes. Similar to the rural crashes in the same segment, the Excel graph includes more curves due to the characteristics of the horizontal scale.

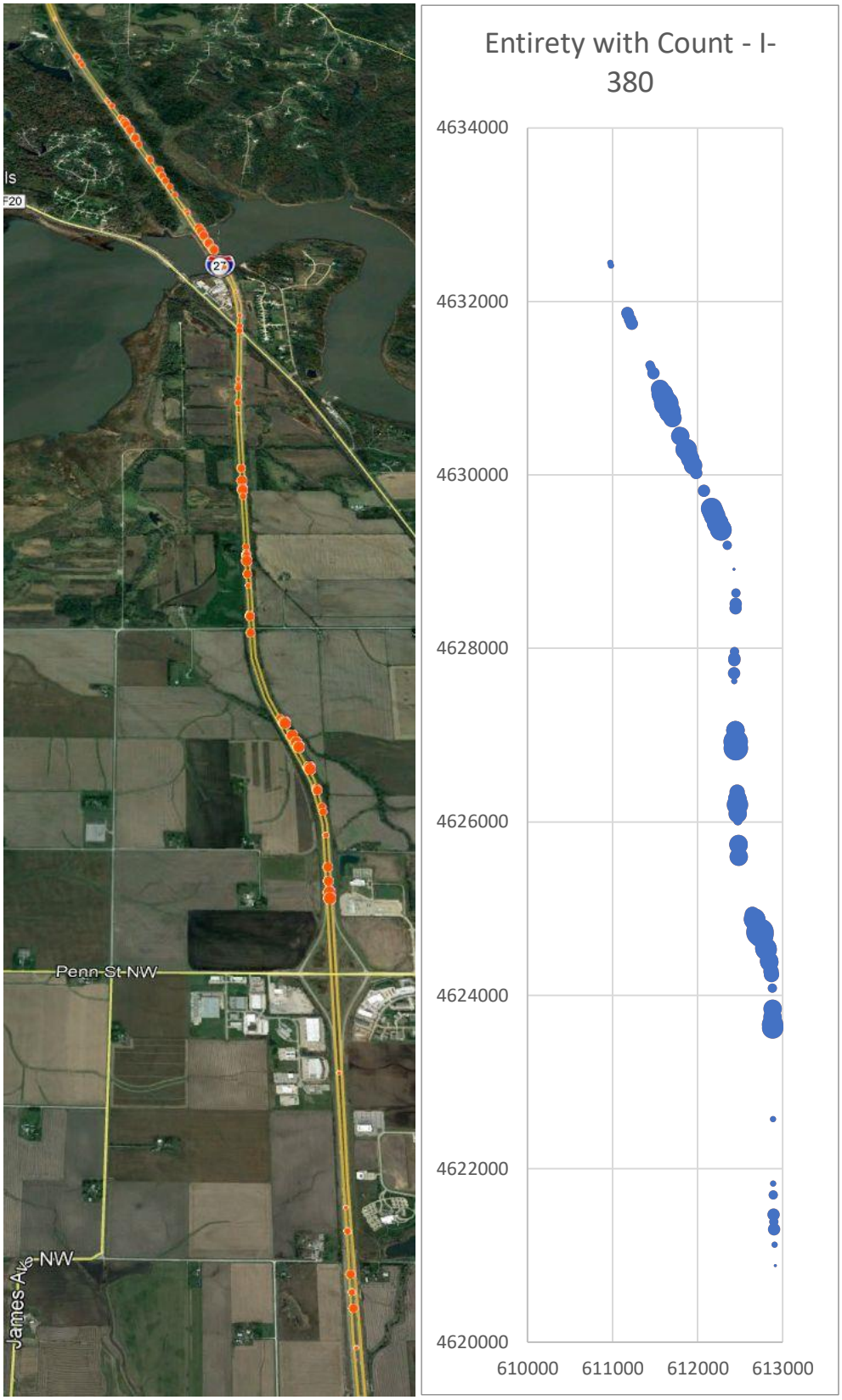

Figure 4.22: I-380 SVROR clusters, full SSD – KMZ and Excel.

Single vehicle run off road crashes along the selected area of I-380, are similar to those found in the total rural crashes. Many of the “clusters” conserved their location and fewer crashes on those locations are observed. From the figures, it was observed that most crashes along the selected segment occurred along curves. To evaluate and verify the found “clusters”, the KDE method was applied to the selected section using the smoothScatter function in R. Figure 4.23 provides the 2D kernel density estimate for single vehicle run off road crashes along the selection section in I-380 including the total count.

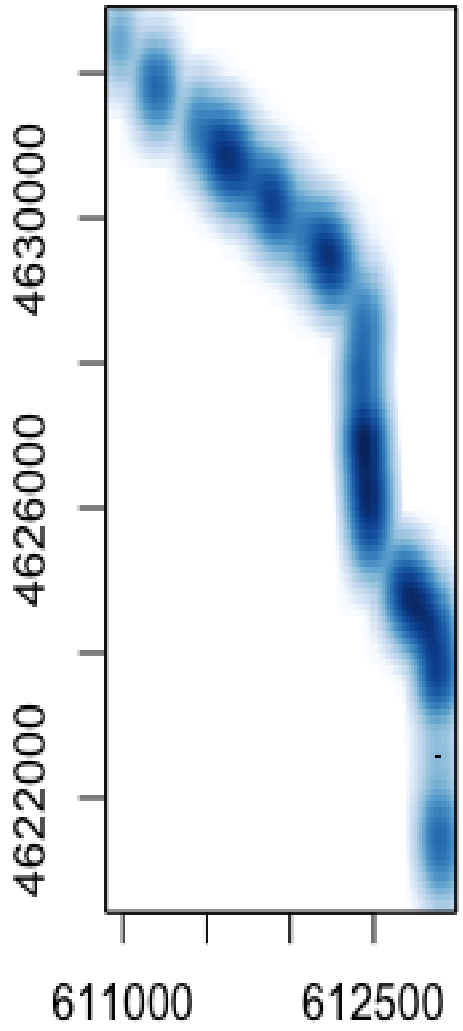

Figure 4.23: I-380 SVROR clusters, full SSD – KDE.

In the figure, locations where crashes occurred are represented with the color blue as in the similar sections. With the KDE, a similar and clearer output can be observed. Properly representation of the “clusters” is represented as expected. Those areas that should be prioritized can also be identified. With both methods, it can be observed that most crashes along the selected segment occurred along curves. A similar output as the one obtained with full stopping sight distance for rural crashes is observed.

Interstate 380 Half Stopping Sight Distance – Iowa SVROR Crashes

The single vehicle run-off-road crashes count involved in half of the sight distance will be evaluated. Figure 4.24 shown below, includes the “clusters” for those crashes in both GoogleEarth and Excel. When the aerial imagery is observed in more detail, it can be observed that similar clusters to those found with the full stopping sight distance is observed, with the difference that less total crashes are present.

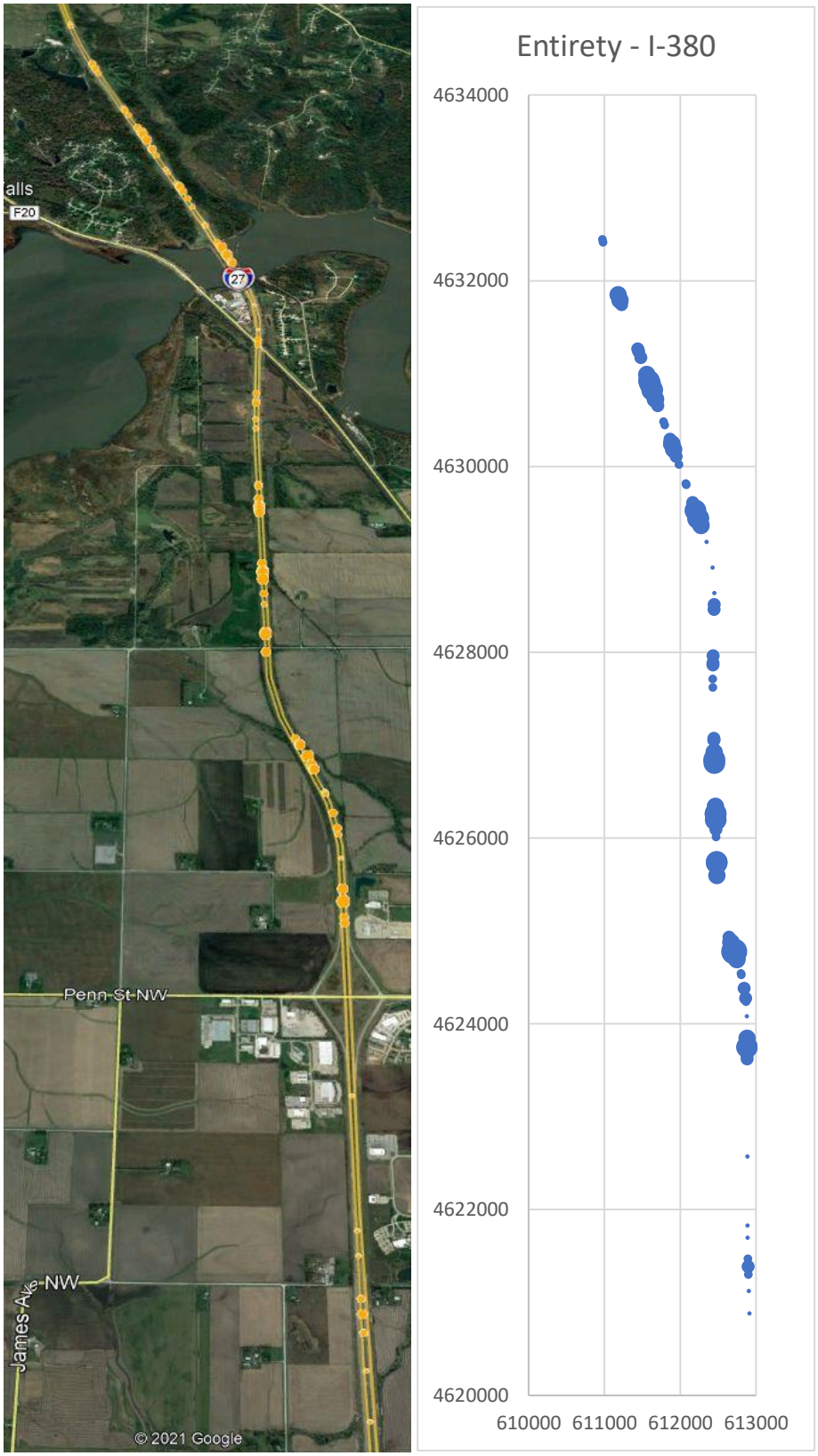

Figure 4.24: I-35 SVROR clusters, half SSD – KMZ and Excel.

Similar to the total sight distance, from the GoogleEarth and Excel images, several similarities are apparent. Many of the “clusters” appear to be similarly located, as expected, with the different that less crashes are observed. In order to compare and verify those results obtained with GoogleEarth and Excel, the smoothScatter function in R will be applied to the data set. Figure 4.25 displays the 2D kernel density estimate for single vehicle run-off-road crashes along the I-380 case study section including the half count of the sight distance.

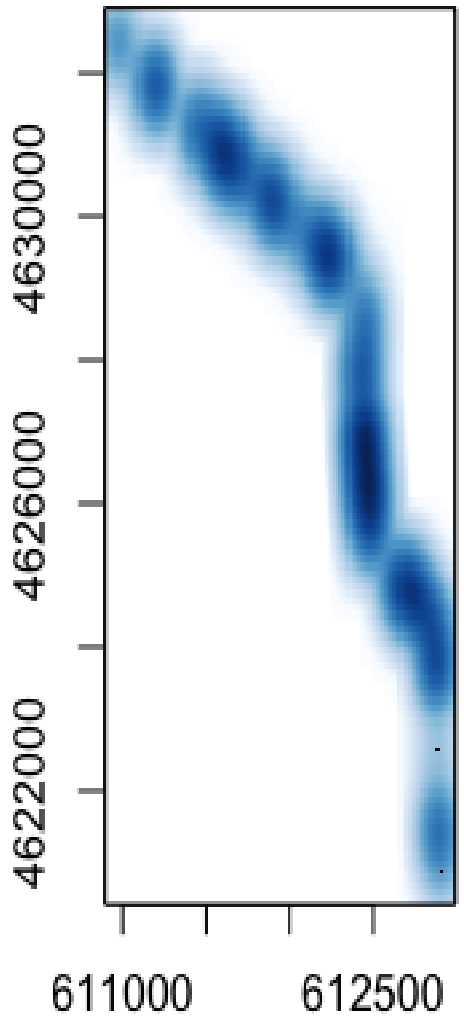

Figure 4.25: Kernel Density SVROR Crashes Half Count I-380.

As in the previous section, the KDE output is similar to the Excel and GoogleEarth output. The KDE output presented one “cluster” as well as the Excel and GoogleEarth images. The location of the found “clusters” is the same location as those observed on the full sight distance situation. However, the density of those “clusters” when compared to the full sight distance is different as expected. With both methods and both sight distances, the areas of interest were identified. Similar outputs were observed when both rural and single vehicle run-off-road crashes were compared. Using the full stopping sight distance, a total of three “clusters” were identified for rural crashes, and for single vehicle run-off-road crashes, only one of those four “clusters” were outstanding. When comparing those results with the half stopping sight distance, a very similar output was observed for both categories with the different that fewer total crashes were observed. For all of the evaluated categories in the case study, most crashes occurred along curves and bridges.

CHAPTER 5. CONCLUSIONS

For the current research, a different approach was followed in order to evaluate single vehicle run off road crashes. Observed crashes were evaluated along rural roads in the State of Iowa for a five-year period using case studies. One of the main objectives was to perform an analysis to determine whether or not clusters were occurring among the observed crashes. Following is a summarization of results and recommendations for future research.

Summary

Single Vehicle Run off Road Crashes

From the summary of the data, it was observed that rural crashes as well as single vehicle run off road crashes have been increasing through the years, which corresponds to what has been observed in the United States. Most single vehicle run off road crashes were comprised of property damage only crashes and occurred during normal peak travel hours and daylight, neither of these results surprising as this is consistent with prior research. For most single vehicle run off road crashes, weather conditions were not found to be a contributing factor; however, weather conditions overall showed a decreasing trend on the effect of these crashes, while weather conditions such as snow and ice, presented an increasing trend. Similar results were found for surface conditions, again not surprising as the two are linked. Additionally, most crashes involved rollovers or collisions with the ditch, which was expected for single vehicle run off road crashes. Farm to market routes, representing the most road mileage but typically lower volumes, had most single vehicle run off road crashes when compared to the other

roads. It was also found that single vehicle run off road crashes were not greatly affected by roadway circumstances in most cases. For the horizontal and vertical alignment contributions, approximately twenty percent of single vehicle run off road crashes were affected by those alignment contributions along the roadway. Lastly, most drivers were not distracted when the crash occurred.

Interstate 35

When I-35 was compared to the other roadways, it was found that all rural crashes presented the second to highest frequency along the second to longest evaluated roadway. From the section along I-35, both rural and single vehicle run off crashes were evaluated.

Interstate 35 Total Sight Distance – Iowa Rural Crashes

When rural crashes were evaluated along this roadway, there were a few locations along the segment that did not present any crashes. As significant clusters were analyzed, four of those were observed along the roadway. These clusters were identified using Microsoft Excel and the kernel density method. These clusters were located along horizontal curves and, on along areas in which transitions from an urban area to a rural area or vice versa were taking place. Including aerial imagery to the segment, allowed a closer observation and a more detailed inspection of the area in which those clusters were occurring was made. From the aerial, it was observed that those clusters and most crashes taking place along I-35 were along ramps, along exits and entrances to the interstate, along bridges, and on regular segments with curves.

Interstate 35 Total Sight Distance – Iowa SVROR Crashes

For single vehicle run off road crashes, there were more locations along the segment that did not present any crashes when compared to all rural crashes, which was expected. Significant clusters were also analyzed for this section and, the same four clusters as those found along the rural evaluation were observed along the roadway. One of those clusters made a more significant impact when compared to the other clusters. That cluster was located along a horizontal curve. When the aerial imagery was observed, a similar result was found as those with the rural crashes. With both methods, it was found that most crashes along the selected segment occurred on a horizontal curve for both, all rural crashes, and single vehicle run off road crashes.

US Highway 20

US-20 was the longest segment evaluated in the current study. Even thought it had the most miles, it presented the least frequency of crashes when compared to the rest of the roads, most likely attributed to the markedly less traffic. Both types of crashes, all rural and single vehicle run off crashes were evaluated for US 20 as well.

US Route 20 Total Sight Distance – Iowa Rural Crashes

Rural crashes were also evaluated along US Route 20. When the analysis was performed, the clusters were also identified as in the previous sections. From this section, a total of five clusters were identified and there were also some locations in which crashes did not occur. In this case, all of those five clusters were located along a horizontal curve. Taking a closer look at the crash cluster locations using aerial imagery, the clusters were found along curves, followed by ramps, and along an existing bridge. Overall, this segment did not include as many crashes when compared to the others.

Perhaps including more crash data from the past, would help the analysis and the effect of curves on rural crashes could be evaluated and verified.

US Route 20 Total Sight Distance – Iowa SVROR Crashes

Single vehicle run of road crashes were ever waited on this segment as well. When this was performed, a similar output like the one obtained from rural crashes was observed. The main difference between these two, was that out of the five clusters observed along the rural crashes, only two of those provided an impact with single vehicle run off road crashes. The two most significant crashes that were found in this case, were also located along horizontal curves. US 20 seems to have crash clusters along curves, perhaps indicating opportunity for some improvements. Also, horizontal curves appear to have an impact on single vehicle run off road crashes.

Interstate 380

When I-380 was compared to the other roadways, the highest number of crashes among the study were found and this roadway represented the shortest alignment of the study, probably explained by the greater traffic volumes. When evaluating the segment of I-380, all rural crashes and single vehicle run off road crashes were also summarized.

Interstate 380 Total Sight Distance – Iowa Rural Crashes

When crash clusters were investigated, most clusters were along horizontal curves, which matches the findings for the other roadways. In this case, a total of three significant clusters were found. Two of those were along horizontal curves and one of those was located on the transition from a curve to the straight segment. When the aerial imagery was observed, crashes taking place along horizontal curves were confirmed, as well as a cluster at a bridge.

Interstate 380 Total Sight Distance – Iowa SVROR Crashes

Significant clusters were also analyzed for single vehicle run off road crashes along this section and, the same clusters as those found along the rural evaluation were observed along the roadway. The identified clusters were on the same locations as with rural crashes, with the difference that less total crashes were presented on the cluster. When the aerial imagery was included, a similar result was found as those with the rural crashes. Most single vehicle run off road crashes were located along horizonal curves, similar to the prior case study sections for I-35 and US 20.

Recommendation for Future Research

To improve and extend the results and methods thus far, the following are suggestions for future research towards the analysis of crash clusters.

- Implementing different ways of the kernel density estimation and applications to the data. These could involve the inclusion of different software and comparison among them for possible improvement. The application of the current or future methodologies could also be attempted on different road classifications in order to compare observation among those. The current method could also be evaluated for different types of crash data, and even involve different states. Also, using the current methodology, different type of crashes could be evaluated, perhaps including more crash data. This could also be applied to urban areas instead of rural areas.
- Explore different causes for single vehicle run off road crashes by taking a closer evaluation of the identified clusters. The severity of the crashes in the found

clusters could also expand the analysis and determine the reasoning behind less fatal crashes present.

- Evaluate relationships between the roadway design and crashes in a further analysis.
- Examine the analysis per year in order to find any relationships of the crashes or perhaps to evaluate the possibility of the locations with clusters improving or worsen throughout the years. The clusters found in the current exploration could also be evaluated per time of the day or season of the year.

REFERENCES

1. Koepke, F. J., & Levinson, H. S. (1992). Access management guidelines for activity centres.

2. Gluck, J., & Lorenz, M. (2010). NCHRP SyntheSiS 404: State of the practice in highway access management. *Transportation Research Board, Washington, DC*.

3. Albrecht, C. P., & Plazak, D. (1998). Bridging the Gap Between Access Management Ideals and Land Use Planning Practice: Suggested Policies and Potential Benefits. In *Crossroads 2000Iowa State University and Iowa Department of Transportation*.

4. Demosthenes, P. (2016). The Marriage of Roundabouts and Access Management. *Presented as a TRB Webinar for Professional Development*.

5. Hauer, E., Kononov, J., Allery, B., & Griffith, M. S. (2002). Screening the road network for sites with promise. Transportation Research Record, 1784(1), 27-32.

6. Adanu, E. K., Smith, R., Powell, L., & Jones, S. (2017). Multilevel analysis of the role of human factors in regional disparities in crash outcomes. Accident Analysis & Prevention, 109, 10-17.

7. Huang, H., Abdel-Aty, M. A., & Darwiche, A. L. (2010). County-level crash risk analysis in Florida: Bayesian spatial modeling. Transportation Research Record, 2148(1), 27-37.

8. Dadashi, A., & Mirbaha, B. (2019). Prioritizing highway safety improvement projects: A Monte-Carlo based Data Envelopment Analysis approach. Accident Analysis & Prevention, 123, 387-395.

9. Lyon, C., Gotts, B., Wong, W. K. F., & Persaud, B. (2007). Comparison of alternative methods for identifying sites with high proportion of specific accident types. Transportation Research Record, 2019(1), 212-218.

10. Coll, B., Moutari, S., & Marshall, A. H. (2013). Hotspots identification and ranking for road safety improvement: An alternative approach. *Accident Analysis & Prevention*, *59*, 604-617.

11. Gross, F. B., Harmon, T., Albee, M., Himes, S., Srinivasan, R., Carter, D., & Dugas, M. (2016). Evaluation of Four Network Screening Performance Measures (No. FHWA-SA-16-103). United States. Federal Highway Administration. Office of Safety.

12. Tarko, A. P., & Azam, S. (2009). Safety Screening of Road Networks with Limited Exposure Data. Transportation research record, 2102(1), 18-26.

13. Tarko, A. P., & Azam, S. (2009). Safety Screening of Road Networks with Limited Exposure Data. Transportation research record, 2102(1), 18-26.

14. Wang, X., Wu, X., Abdel-Aty, M., & Tremont, P. J. (2013). Investigation of road network features and safety performance. Accident Analysis & Prevention, 56, 22-31.

15. Cafiso, S., Di Graziano, A., & Pappalardo, G. (2015). Safety inspection and management tools for low-volume road network. Transportation Research Record, 2472(1), 134-141.

16. Ghadi, M., & Török, Á. (2019). A comparative analysis of black spot identification methods and road accident segmentation methods. Accident Analysis & Prevention, 128, 1-7.

17. Harirforoush, H., & Bellalite, L. (2019). A new integrated GIS-based analysis to detect hotspots: a case study of the city of Sherbrooke. Accident Analysis & Prevention, 130, 62-74.

18. Pawlovich, M. D. (2003). Evaluating traffic safety network screening: An initial framework utilizing the hierarchical Bayesian philosophy. Iowa State University.

19. Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. Accident Analysis & Prevention, 41(3), 359-364.

20. Shad, R., & Rahimi, S. (2017). Identification of road crash black sites using geographical information system. International Journal for Traffic and Transport Engineering, 7(3), 368-380.

21. Wellner, A., & Qin, X. (2011). Highway Safety Metrics Implementation and Evaluation Using a Geographic Information System–Based Screening Tool. Transportation research record, 2241(1), 1-9.

22. MacNab, Y. C. (2003). A Bayesian hierarchical model for accident and injury surveillance. *Accident Analysis & Prevention*, *35*(1), 91-102.

23. 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*, *2136*(1), 67-76.

24. Nilsson, D., Lindman, M., Victor, T., & Dozza, M. (2018). Definition of run-off-road crash clusters—For safety benefit estimation and driver assistance development. *Accident Analysis & Prevention*, *113*, 97-105.

25. Satterfield, C., Albin, R. B., Noizet, J., Schaftlein, S., Vadakpat, G., Smith, S. F., ... & Kok, R. (2020). *Public Roads Magazine, Winter 2021* (No. FHWA-HRT-21-002). United States. Federal Highway Administration.

26. U.S. Department of Transportation. Federal Highway Administration. *Roadway Departure Safety*. https://safety.fhwa.dot.gov/roadway_dept/. Accessed Aug. 27, 2020

27. U.S. Department of Transportation. Federal Highway Administration. *Travel Monitoring*. https://www.fhwa.dot.gov/policyinformation/travel_monitoring/tvt.cfm. Accessed June 24, 2021.

28. Iowa Department of Transportation. *Iowa in Motion – Planning Ahead*. https://www.iowadot.gov/iowainmotion/files/2.%20Understanding%20Iowa.pdf. Accessed June 24, 2021.

29. Guadamuz-Flores, R., & Aguero-Valverde, J. (2017). Bayesian spatial models of crash frequency at highway–railway crossings. Transportation Research Record, 2608(1), 27-35.

30. Lee, J., Abdel-Aty, M., & Cai, Q. (2017). Intersection crash prediction modeling with macro-level data from various geographic units. Accident Analysis & Prevention, 102, 213-226.

31. Li, Z., Chen, C., Ci, Y., Zhang, G., Wu, Q., Liu, C., & Qian, Z. S. (2018). Examining driver injury severity in intersection-related crashes using cluster analysis and hierarchical Bayesian models. *Accident Analysis & Prevention*, *120*, 139-151.

32. Cheng, W., Gill, G. S., Dasu, R., Xie, M., Jia, X., & Zhou, J. (2017). Comparison of Multivariate Poisson lognormal spatial and temporal crash models to identify hot spots of intersections based on crash types. *Accident Analysis & Prevention*, *99*, 330-341.

33. Aguero-Valverde, J., & Jovanis, P. P. (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis & Prevention*, *38*(3), 618-625.

34. Huang, H., Chin, H. C., & Haque, M. M. (2009). Empirical evaluation of alternative approaches in identifying crash hot spots: Naive ranking, empirical bayes, full bayes methods. *Transportation Research Record*, *2103*(1), 32-41.

35. Aguero-Valverde, J., & Jovanis, P. P. (2008). Analysis of road crash frequency with spatial models. *Transportation Research Record*, *2061*(1), 55-63.

36. Gudes, O., Varhol, R., Sun, Q. C., & Meuleners, L. (2017). Investigating articulated heavy-vehicle crashes in western Australia using a spatial approach. *Accident Analysis & Prevention*, *106*, 243-253.

37. Fishbain, B., & Grembeck, O. (2013). A Multi-Dimensional Clustering Algorithm for Studying Fatal Road Crashes. *Journal of transportation and statistics*, *10*(1), 79-100.

38. Das, S., & Sun, X. (2016). Association knowledge for fatal run-off-road crashes by multiple correspondence analysis. *IATSS Research*, *39*(2), 146-155.

39. Shawky, M., Hassan, H. M., Garib, A. M., & Al-Harthei, H. A. (2016). Examining the factors affecting the severity of run-off-road crashes in Abu Dhabi. *Canadian Journal of Civil Engineering*, *43*(2), 132-138.

40. Wegman, F. (2014). Analyzing road design risk factors for run-off-road crashes in the Netherlands with crash prediction models. *Journal of safety research*, *49*, 121-e1.

41. Gong, L., Fan, W., & Washing, E. M. (2016). Modeling severity of single vehicle run-off-road crashes in rural areas: model comparison and selection. *Canadian journal of civil engineering*, *43*(6), 493-503.

42. Benedek, J., Ciobanu, S. M., & Man, T. C. (2016). Hotspots and social background of urban traffic crashes: A case study in Cluj-Napoca (Romania). *Accident Analysis & Prevention*, *87*, 117-126.

43. Jia, R., Khadka, A., & Kim, I. (2018). Traffic crash analysis with point-of-interest spatial clustering. *Accident Analysis & Prevention*, *121*, 223-230.

44. Shad, R., & Rahimi, S. (2017). Identification of road crash black sites using geographical information system. *International Journal for Traffic and Transport Engineering*, *7*(3), 368-380.

45. Iowa Department of Transportation. *Fatal crashes in Iowa – 1988 to 2019.* https://iowadot.gov/mvd/stats/fatalcrashes.pdf. Accessed June 24, 2021.

46. Stipancic, J., Miranda-Moreno, L., & Saunier, N. (2017). Impact of congestion and traffic flow on crash frequency and severity: application of smartphone-collected GPS travel data. Transportation Research Record, 2659(1), 43-54.

47. Silverman, B. W. (2018). Density estimation for statistics and data analysis. Routledge.

48. Jia, R., Khadka, A., & Kim, I. (2018). Traffic crash analysis with point-of-interest spatial clustering. *Accident Analysis & Prevention*, *121*, 223-230.

49. Thakali, L., Kwon, T. J., & Fu, L. (2015). Identification of crash hotspots using kernel density estimation and kriging methods: a comparison. *Journal of Modern Transportation*, *23*(2), 93-106.

50. Yasmin, S., & Eluru, N. (2013). Evaluating alternate discrete outcome frameworks for modeling crash injury severity. Accident Analysis & Prevention, 59, 506-521.

51. Yasmin, S., Eluru, N., Bhat, C. R., & Tay, R. (2014). A latent segmentation based generalized ordered logit model to examine factors influencing driver injury severity. Analytic methods in accident research, 1, 23-38.

52. Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., & Abdel-Aty, M. (2016). Macro and micro models for zonal crash prediction with application in hot zones identification. Journal of transport geography, 54, 248-256.

53. Davis, G. A., & Yang, S. (2001). Bayesian identification of high-risk intersections for older drivers via Gibbs sampling. Transportation Research Record, 1746(1), 84-89.

54. Gong, L., & Fan, W. D. (2017). Modeling single-vehicle run-off-road crash severity in rural areas: Accounting for unobserved heterogeneity and age difference. *Accident Analysis & Prevention*, *101*, 124-134.

55. Yu, R., Abdel-Aty, M., & Ahmed, M. (2013). Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. Accident Analysis & Prevention, 50, 371-376.

56. Zou, W., Wang, X., & Zhang, D. (2017). Truck crash severity in New York city: an investigation of the spatial and the time-of-day effects. Accident Analysis & Prevention, 99, 249-261.

57. Kim, D. G., Lee, Y., Washington, S., & Choi, K. (2007). Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. Accident Analysis & Prevention, 39(1), 125-134.

58. El-Basyouny, K., Barua, S., & Islam, M. T. (2014). Investigation of time and weather effects on crash types using full Bayesian multivariate Poisson lognormal models. Accident Analysis & Prevention, 73, 91-99.

59. Alarifi, S. A., Abdel-Aty, M., & Lee, J. (2018). A Bayesian multivariate hierarchical spatial joint model for predicting crash counts by crash type at intersections and segments along corridors. Accident Analysis & Prevention, 119, 263-273.

60. Kim, D. G., Washington, S., & Oh, J. (2006). Modeling crash types: new insights into the effects of covariates on crashes at rural intersections. Journal of Transportation Engineering, 132(4), 282-292.

61. Wang, K., Ivan, J. N., Ravishanker, N., & Jackson, E. (2017). Multivariate Poisson lognormal modeling of crashes by type and severity on rural two-lane highways. Accident Analysis & Prevention, 99, 6-19.

62. Mothafer, G. I., Yamamoto, T., & Shankar, V. N. (2016). Evaluating crash type covariances and roadway geometric marginal effects using the multivariate Poisson gamma mixture model. Analytic methods in accident research, 9, 16-26.

63. Findley, D. J., Hummer, J. E., Rasdorf, W., Zegeer, C. V., & Fowler, T. J. (2012). Modeling the impact of spatial relationships on horizontal curve safety. Accident Analysis & Prevention, 45, 296-304.

64. Montella, A., Mauriello, F., Pernetti, M., & Riccardi, M. R. (2021). Rule discovery to identify patterns contributing to overrepresentation and severity of run-off-the-road crashes. Accident Analysis & Prevention, 155, 106119.

65. Jo, J. H., Lee, J. S., Ouyang, Y., & Peng, F. (2011). Integrated decision support for roadway safety analysis. Journal of computing in civil engineering, 25(1), 50-56.

66. Al-Kaisy, A., Ewan, L., & Hossain, F. (2019). Identifying candidate locations for safety improvements on low-volume rural roads: the Oregon experience. Transportation research record, 2673(12), 690-698.

67. Montella, A. (2010). A comparative analysis of hotspot identification methods. *Accident Analysis & Prevention*, *42*(2), 571-581.

68. Ferreira, S., & Couto, A. (2013). Hot-Spot Identification: Categorical Binary Model Approach. Transportation research record, 2386(1), 1-6.

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

70. Guo, X., Wu, L., Zou, Y., & Fawcett, L. (2019). Comparative analysis of empirical bayes and Bayesian hierarchical models in hotspot identification. *Transportation research record*, *2673*(7), 111-121.

71. Cheng, W., & Washington, S. P. (2005). Experimental evaluation of hotspot identification methods. *Accident Analysis & Prevention*, *37*(5), 870-881.

72. Gross, F., & Harmon, T. (2019). Allocating spending between hotspot and systemic approaches to safety management. *Transportation research record*, *2673*(2), 220-231.

73. Dong, N., Huang, H., Lee, J., Gao, M., & Abdel-Aty, M. (2016). Macroscopic hotspots identification: a Bayesian spatial-temporal interaction approach. *Accident Analysis & Prevention*, *92*, 256-264.

74. Bortoloti, F. D., de Oliveira, E., & Ciarelli, P. M. (2021). Supervised kernel density estimation K-means. *Expert Systems with Applications*, *168*, 114350.

75. Young, J., & Park, P. Y. (2014). Hot zone identification with GIS-based post-network screening analysis. *Journal of Transport Geography*, *34*, 106-120.

76. Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention*, *41*(3), 359-364.

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

78. U.S. Department of Transportation. Federal Highway Administration. *Addressing Rural Roadway Departures on All Public Roads.* https://safety.fhwa.dot.gov/FoRRRwD/allpublicroads.cfm. Accessed Aug. 27, 2020

79. Cheng, W., Gill, G. S., Dasu, M., & Jia, X. (2018). An empirical evaluation of multivariate spatial crash frequency models. *Accident Analysis & Prevention*, *119*, 290-306.

80. Zhang, J., Lin, B., & Zhou, Y. (2021). Kernel density estimation for partial linear multivariate responses models. *Journal of Multivariate Analysis*, 104768.

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

82. Khattak, A. J., Wang, X., & Zhang, H. (2010). Spatial analysis and modeling of traffic incidents for proactive incident management and strategic planning. *Transportation research record*, *2178*(1), 128-137.

83. Colak, H. E., Memisoglu, T., Erbas, Y. S., & Bediroglu, S. (2018). Hot spot analysis based on network spatial weights to determine spatial statistics of traffic accidents in Rize, Turkey. *Arabian Journal of Geosciences*, *11*(7), 1-11.

84. Jiang, X., Abdel-Aty, M., & Alamili, S. (2014). Application of Poisson random effect models for highway network screening. *Accident Analysis & Prevention*, *63*, 74-82.

85. Hauer, E., Harwood, D. W., Council, F. M., & Griffith, M. S. (2002). Estimating safety by the empirical Bayes method: a tutorial. *Transportation Research Record*, *1784*(1), 126-131.

86. Khattak, A. J., Wang, X., & Zhang, H. (2010). Spatial analysis and modeling of traffic incidents for proactive incident management and strategic planning. *Transportation research record*, *2178*(1), 128-137.

87. U.S. Department of Transportation. Federal Highway Administration. *Systemic Approach to Rural Roadway Departures*. https://safety.fhwa.dot.gov/FoRRRwD/systemicapproach.cfm. Accessed Aug. 27, 2020

88. Transportation Officials. (2011). A Policy on Geometric Design of Highways and Streets, 2011. AASHTO.

89. Pawlovich, M. D. (2002). Safety Improvement Candidate Location (SICL) Methods. Iowa Department of Transportation, Highway Division, Engineering Bureau, Office of Traffic Safety.

90. Hancock J. (2020). The Iowa Highways Page. https://iowahighways.org/index.html. Accessed Jul. 16, 2021.