December 2019

Midblock Crash Analysis Using Short Segments and Precisely Geocoded Crashes

Adika Iqbal

Clemson University, adika0590@gmail.com

Recommended Citation

Iqbal, Adika, "Midblock Crash Analysis Using Short Segments and Precisely Geocoded Crashes" (2019). *All Dissertations*. 2506.
https://tigerprints.clemson.edu/all_dissertations/2506

MIDBLOCK CRASH ANALYSIS USING SHORT SEGMENTS AND PRECISELY GEOCODED CRASHES

A Dissertation
Presented to
the Graduate School of
Clemson University

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

---

by
Adika Mammadrahimli Iqbal
December 2019

Accepted by:
Dr. Wayne Sarasua, Committee Chair
Dr. Jennifer Ogle, Committee Co-Chair
Dr. William J. Davis, Committee Member
Dr. Christopher Post, Committee Member

ABSTRACT

This research focused on evaluating how crash geocoding has improved over the years and how this enhanced spatial accuracy of crashes can potentially lead to a new paradigm for midblock crash safety analysis. Robust midblock safety analysis exhibits special challenges because methods of locating crashes have historically not been very accurate. One objective of this research was to assess how the accuracy of crashes has improved over time and what the current state of the art is.

The second objective focused on using segment lengths less than the Highway Safety Manual (HSM) recommended minimum of 0.1 miles for statewide screening of midblock crash locations to identify site specific locations with high crash incidence through a peak search methodology. The research clearly indicates that the use of segments of 0.1 miles (or greater) in many instances' "hides" the severity of a single location if the rest of the segment has few or no additional crashes. The research also evaluated a sliding window approach using short segments. Based on the analysis, the short segment peak search method is recommended for use by state agencies as a network screening approach because it is much less complex to implement than the sliding window approach, locations can be easily ranked, and direct comparisons can be made of segment crash incidence over multiple years.

The final objective of this research was to compare the short segment peak search approach to other HSM methods. The results of the comparison revealed similar results at the highest priority level and thus the former can be used as an alternative in case of insufficient data on driveway and roadway characteristics.

This research shows that improvements in crash geocoding makes short-segment peak search network screening viable for segment lengths less than 0.1 miles. By using short segment network screening, segments of high crash incidence can be displayed with overlayed crashes at their actual crash locations which can minimize the need for developing collision diagrams. Secondly, one of the hypotheses is that the current intersection to intersection process aggregates crashes to long segments which can mask the crash severity of point locations.

DEDICATION

This dissertation is dedicated to the memory of my beloved grandmother, Solmaz Rahimova, who passed away earlier this year. I am sure she is watching me and is very proud.

ACKNOWLEDGMENTS

First and foremost I want to thank my advisor Dr. Wayne Sarasua, for allowing me to be a part of the team for this project and giving me the opportunity to complete my PhD. Dr. Sarasua showed a tireless attitude towards guiding me through my program and sharing his knowledge. I am truly grateful for the support and dedication he has constantly provided for me during this research and throughout my stay at Clemson University.

I would also like to thank Dr. William J. Davis for his input in my research and writing of my dissertation.

I also would like to thank Dr. Jennifer Ogle and Dr. Christopher Post, for their technical guidance, encouragement and support during my research.

I would like to acknowledge my team of hard-working students that helped throughout the duration of this project. Completing this study would not have been possible without your hard work and diligence. Thank you Afshin Famili, for all of your help and hard work with data analysis. Additional thanks go out to Devesh Kumar, Saurbah Basnet and Emmanuel Adjei. Lastly, I want to give thanks to Dr. Kweku Brown from Citadel for his help and expertise during the analysis process. Once again, I want to thank you all for your dedication and hard work for this project was truly a team effort.

I would like to thank my husband Waqas Iqbal for his love and encouragement throughout my life's journey up until now. I would also like to thank my parents, my sister, my nephew, my host parents and my husband's parents for always encouraging me.

Lastly, I want to thank everyone at Clemson ITE Student Chapter who made my stay in Clemson very memorable.

TABLE OF CONTENTS

LIST OF TABLES

Paper I: Assessment of Crash Location Accuracy in Electronic Crash Reporting Systems

Table Page

Paper II: Short Segment Statewide Screening of Midblock Crashes in South Carolina

Table Page

Paper III: Assessing the Predictability of Short Segment Crash Analysis in the State of South Carolina

Table Page

LIST OF FIGURES

Paper I: Assessment of Crash Location Accuracy in Electronic Crash Reporting Systems

Figure Page

Paper II: Short Segment Statewide Screening of Midblock Crashes in South Carolina

Figure Page

List of Figures (Continued)

Figure Page

Paper III: Assessing the Predictability of Short Segment Crash Analysis in the State of South Carolina

Figure Page

CHAPTER ONE

INTRODUCTION

Over the last several years, traffic fatality rates in South Carolina have been consistently ranked amongst the highest in the country *(1)*. Throughout the nation, a lot of emphasis has been put on intersection crashes over the years. Intersection crashes are typically geocoded more accurately than other crashes because they are explicitly associated with intersections as part of the crash attributes *(2)*. Midblock crashes are segment oriented and studies have shown that for the most part most analysis is done on an intersection to intersection basis using very long segments. A review of literature indicates that there has not been a great deal of midblock safety analysis research using smaller segments. Midblock safety analysis exhibits special challenges because methods of locating crashes have historically not been very accurate. Recent developments in crash geocoding techniques have improved spatial accuracy. This research focuses on evaluating how crash geocoding has improved over the years and how this enhanced spatial accuracy of crashes can potentially lead to a new paradigm on midblock crash safety analysis. Several years of South Carolina crash data is used in this research. It is anticipated that the findings of this research are transferable to other states because of the sample size of the data used. In actuality, the research was conducted using *all* reported midblock crashes for all roadway classes over a multi-year period. The research has three primary objectives discussed in the following paragraphs.

The first objective of the research is to assess how the accuracy of crashes has improved over time and what the current state of the art is. Case studies are used in the assessment of the accuracy of South Carolina crash data and a survey of state highway agencies conducted as part of this research will be used to assess the current state of the art in crash geocoding across the US.

The AASHTO Highway Safety Manual (HSM) presents a variety of methods for quantitatively estimating crash frequency or severity at a variety of locations *(3)*. The HSM predictive methods require the roadway network to be divided into homogeneous segments and intersections, or sites populated with a series of attributes. It recommends a minimum segment length of 0.1 miles. A review of literature indicates that segments lengths less than 0.1 miles are not advisable because findings are highly variable. These findings are based on crash data with questionable spatial accuracy. The second objective of this research focuses on segment lengths of less than 0.1 miles for statewide screening of midblock crash locations to identify site specific locations with high crash incidence. The hypothesis is that improved spatial accuracy of crashes can result in worthwhile analysis using segments less than 0.1 miles. Different analysis methods will be used to look at short segments.

The final objective of this research is to compare the new network screening identified upon completion of the second objective to other HSM methods.

It is anticipated that the findings of this research will show how improvements in crash geocoding can enhance safety analysis. This research could potentially lead to a changing paradigm of how network screening of midblock crashes is done by state agencies.

This dissertation document consists of three research papers on transportation safety. These papers make use of South Carolina Crash Data over fourteen years (2004 – 2018). Each paper focuses on one objective of this research and accounts for one chapter of the dissertation. The objectives are restated below along with the titles of each paper and the tasks performed towards achieving the research objectives.

PAPER I: Assessment of Crash Location Accuracy in Electronic Crash Reporting Systems

Objective 1

Assess how the accuracy of crashes has improved over time and what the current state of the art is. Case studies are used in the assessment of the accuracy of South Carolina crash data and a survey of state highway agencies conducted as part of this research will be used to assess the current state of the art in crash geocoding across the US.

Tasks

Task 1: Deploy a survey and send to all state transportation agencies to better understand data collection and network screening methods.

Task 2: Acquire 2010 – 2018 South Carolina Crash Data and geocode on ArcMap.

Task 3: Use of different analysis methods to compare the accuracy of the different crash geocoding methods used in South Carolina throughout the years. Mi

PAPER II: Short Segment Statewide Screening of Midblock Crashes in South Carolina

Objective 2

Develop fixed-length segmentation network screening approach to identify the top midblock segments for each roadway type that has the highest crash incidence in the state.

Tasks

Task 1: Create a GIS layer representing the road surface variable buffer using the roadway width attribute in the SCDOT road characteristics database.

Task 2: Test different segment lengths and width and compare results.

Task 3: Segment the buffered layer using the different fixed segment length.

Task 4: Aggregate crash data to segment buffers.

Task 5: Compare peak search method to NKDE (Network Kernel Density Estimation) method.

PAPER III: Assessing the Predictability of Short Segment Crash Analysis in the State of South Carolina

Objective 3

Compare fixed-length segment approach to other Highway Safety Manual (HSM) methods.

Tasks

Predicted SPF (Safety Performance Function) for Intersection to Intersection Midblock

Task 1: Create Midblock segmentation from Intersection to Intersection for the entire state.

Task 2: Obtain AADT data from Database (DOT).

Task 3: Calculate SPF's for each segment and predict number of crashes

Task 4: Find excess and rank the obtained segments and compare with short segments method.

Task 5: Obtain number of Buffers from Short Segment method that are also in the high ranked SPF's segment.

Predicted SPF on Driveway using only AADT

Task 6: Segment Roadway based on Short Segments screening method and obtain buffers for the entire roadway.

Task 7: Calculate predicted SPF driveway value considering

Task 8: Find excess and rank the buffers based on excess

Task 9: Obtain high ranked buffers from Short Segment method that match with the high ranked predicted SPF for driveway.

Predicted SPF for driveways adjusting with CMFs

Task 10: Segment Roadway based on short segment screening method and obtain buffers for the entire roadway.

Task 11: Calculate predicted SPF for each driveway adjusted based on driveway characteristics.

Task 12: Find excess and rank the buffers based on excess.

Task 13: Obtain high ranked buffers from short segment method that match with the high ranked CMF buffers,

****Two papers have been submitted to Transportation Research Board Journal (one being already published) and third paper is in the process of being submitted to scholarly journals.*

The next three chapters (Chapter Two, Chapter Three and Chapter Four) contain the three research papers introduced in this chapter, followed by the dissertation conclusion in Chapter Five and then appendices.

REFERENCES

1. South Carolina Department of Public Safety (SCDPS). *South Carolina Traffic Collision Fact Book 2013-17. http://www.scdps.gov/ohsjp/stat_services.asp* , Accessed March 10, 2017.

2. AASHTO, 2010. *Highway safety manual*, 1st Edition. American Association of State Highway and Transportation Officials, Washington, D.C.

3. AASHTO, 2010. *Highway safety manual*, 1st Edition. American Association of State Highway and Transportation Officials, Washington, D.C.

CHAPTER TWO

Paper I: Assessment of Crash Location Accuracy in Electronic Crash Reporting Systems

Co-authors of the paper: Wayne A. Sarasua, Kweku Brown, Jennifer H. Ogle, Afshin Famili, William J. Davis, Saurabh B. Basnet, and Devesh Kumar.

Paper I got accepted by the Annual Meeting of the Transportation Research Board and publication decision is pending.

ABSTRACT

Over the past several years, traffic fatality rates in South Carolina have been consistently ranked amongst the highest in the country. Furthermore, South Carolina incurs an annual economic loss of over two billion dollars due to roadway traffic crashes. The South Carolina Department of Transportation (SCDOT), in collaboration with the South Carolina Department of Public Safety (SCDPS), has undertaken a series of initiatives to reduce the number of annual vehicle crashes, with a particular emphasis on injury and fatal crashes. One of these initiatives is the deployment of a map-based geocoded crash reporting system that has greatly improved the quality of crash location data. This paper provides an assessment of improvements in crash geocoding accuracy in South Carolina and how improved accuracy is beneficial to systematic statewide safety analysis. A case study approach is used to demonstrate practical applications and

analysis techniques based on spatially accurate crash data. A survey of state highway agencies indicates that there are disparate crash reporting systems used across the country with regard to crash geocoding procedures and accuracies. Survey results indicate that not only does geocoded accuracy of crash locations vary by state, accuracies often vary by jurisdiction within each state. Research results suggest that poorly geocoded crash data can bias certain types of safety analysis procedures and that many state safety initiatives, analysis methods, and outcomes can benefit from improving crash report geocoding procedures and accuracies.

INTRODUCTION

From 2014 through 2018 approximately 4,852 motor vehicle-related deaths occurred in South Carolina resulting in an average of 970 traffic fatalities per year over the five-year period. These rates are considerably higher than the national averages of 1.16 fatalities per 100 million VMT and 11.52 fatalities per 100,000 populations. In 2017 alone, there were 988 traffic fatalities in South Carolina resulting in rates of 1.78 fatalities per 100 million vehicle miles traveled (VMT) and 19.70 fatalities per 100,000 population (1). Further, South Carolina incurs an annual economic loss of over two billion dollars due to road traffic crashes (2).

Recent efforts by the South Carolina Department of Transportation (SCDOT) to reduce vehicle crashes, in particular injury and fatal crashes, within the state led to development of the 2015-2018 South Carolina’s Strategic Highway Safety Plan (SHSP):

Target Zero. Published in 2015, the SHSP was the result of concerted efforts by SCDOT, South Carolina Department of Public Safety (SCDPS), South Carolina Division Office of the Federal Highway Administration (FHWA) and other local, state, and federal road safety advocacy groups and agencies. The primary goal of SHSP is to eventually eliminate traffic fatalities and significantly reduce injuries in South Carolina. SHSP emphasizes data-driven, evidence-based recommendations for appropriate strategies and countermeasures to achieve its safety goals (2).

In the previous SHSP published in 2007, improved crash reporting was deemed as essential for safety analysis (4) which led to the development and deployment of an electronic crash reporting system. The 2015 SHSP has recommended continued enhancement of the system (3). This system, known as the South Carolina Collision and Ticket Tracking System (SCCATTS), has grown substantially in its development and implementation since 2007. Starting in 2010, the electronic collision report form component of SCCATTS was deployed to the South Carolina Highway Patrol (with 100% compliance by January 2012) and has since been adopted by local law enforcement agencies throughout the state. The main reason of using SCCATTS was to improve accuracy and timeliness from date of crash to date of data available in the collision master file (5). Recent estimates indicate more than 75% of collision report forms are being submitted to SCDPS electronically which has decreased the number of days for processing a collision report from 35 or more days in 2010 to 5 days or less currently. One of the biggest benefits envisioned for SCCATS implementation was use of mapping software integrated within the electronic reporting hardware that would allow for more accurate reporting of collision locations. The 2015 SHSP specifically states "…proper identification of where a collision occurred is of utmost importance to SCDOT for planning purposes." (3)

Research indicates that accurate crash location data improves reliability of safety analyses and evaluation of countermeasure effectiveness (6,7,8 and 9). Among multiple attributes in a crash data set, the location of a crash is of utmost importance because, crash records with inaccurate locations cannot be considered in the analysis. Excluding crash records can result in under-reporting crash rates, which creates bias in prediction models. This paper provides an assessment of improvements in crash geocoding accuracy in South Carolina and how this improved accuracy can benefit safety analysis. A case study approach is used to demonstrate practical applications and analysis techniques based on spatially accurate crash data. The emphasis of these case studies is predicated on analysis of midblock crashes because locations of midblock crashes are more prone to error compared to intersection crashes (16). Intersection crashes are point oriented and associated with the intersection of two cross-streets on a map; whereas, locating midblock crashes has historically been based on a police officers' estimate of distance from the nearest intersection. In many cases, officers estimate this distance to the nearest ¼ mile. Surveys of state highway agencies are presented that provide the current state of practice in crash reporting and crash geocoding across the country. It is anticipated that many states can enhance safety analysis by improving their crash report geocoding methods. A discussion is also included on how South Carolina's crash reporting system can potentially be further improved by taking advantage of some of the best practices found in other states.

LITERATURE REVIEW

Many states have replaced old methods of data collection with new technologies such as the use of laptops or other electronic devices to collect crash report data and the use of barcode scanners to record the licenses of drivers involved in crashes (11). From an infrastructure standpoint, systems developed to improve crash location characteristics are inherently important, because without a spatial context for the crash problem, it is much more difficult to identify potential contributing factors. Understanding the crash context is critical to identifying appropriate countermeasures, as well as where the improvements should be implemented to have the greatest potential impact.

For many decades, DOTs have defined crash location using route identifiers along with distances to reference points (e.g. route mile post system, route reference point system and link node system) (10). While these methods may appear appropriate, there are several problems associated with their use for crash locationing. For instance, there is not always a single universal route identifier used by all agencies within the state and often a route has multiple designations (e.g. the section of interstate going through downtown Atlanta, Georgia which is designated as both I-85 and I-75). Furthermore, some secondary routes have multiple names and numbers, and many change names over time. Distance measurements are similarly difficult. For example, some police officers may not have the proper equipment or time to measure the distance. They may estimate distances using a value such as ¼ mile which may result in a spatially inaccurate cluster of crashes exactly 1320 feet from an intersection. Additionally, when measurements are based on reference points or crossing streets, the notation becomes complex and the location may be misconstrued. Lastly, these methods of identifying crash locations may not provide precise locations of crashes in the travelway (or adjacent to the travelway in many instances). Due to drawbacks associated with these methods, many states have added coordinate locations using GPS technology.

By the mid 2000's, states such as Iowa, Illinois, Kentucky and Massachusetts developed and deployed electronic crash data collection systems for widescale use by law enforcement officers (11). Iowa's Traffic and Criminal Software (TraCS) consists of bar code scanners, swipe-card readers, digital cameras, GPS technology, a GIS viewer and touch pads to aid digital data entry (11) As of 2007, TraCS had been adopted in 18 states and 2 Canadian provinces (12). More recently, Alabama combined an electronic citation (E-Citation) application with the state's crash database analysis software into a system called Critical Analysis and Reporting Environment (CARE). The system includes a GIS platform where police officers map vehicle crash and traffic citation locations (11). Other states including Louisiana and Tennessee have also recently adopted similar systems and have achieved improvement in the quality of their crash data (13). Florida uses a web based geospatial crash analytical tool called Signal Four Analytics. It is designed to support the crash mapping and analytical needs of law enforcement, traffic engineering, transportation planning agencies, and research institution in the state of Florida.

The transition to the use of GPS technology in crash data collection in South Carolina began in 2004 when SCDOT purchased hand-held GPS units for law enforcement officers to collect coordinate (latitude, longitude) information for crash reports. The use of these hand-held GPS units was not automated, and officers had to read coordinates from the GPS unit and manually record values on the crash report. Information from the paper report would later be keyed into a digital database. Although use of GPS units was advantageous over traditional location referencing methods, there were many issues associated with operation of GPS units, recording of location data on paper crash reports and processing of data. (14, 15).

The deployment of SCCATTS currently used by South Carolina highway patrol and nearly half of local jurisdictions has resulted in considerable improvement in crash data quality. The system enables law enforcement officers to spatially identify and locate crashes via a GIS-based GPS enabled mapping platform operational within police vehicles. The GPS displays the vehicle's location on the GIS map display and officers can pinpoint the actual location of the crash rather than where an officer's vehicle is situated (e.g. on the side of the road or in a parking lot, etc.). Officers can key in all other information related to the crash, which is later uploaded to the SCDPS database and later transferred to SCDOT. Pinpointing crash locations not only lead to more accurate coordinate data provided on crash reports, but also populates other location information data fields on the crash report automatically (5).

SURVEY OF STATES

The literature review indicated that numerous states have transitioned to computer-based crash reporting and geocoding methods, while many others are still using paper-based systems. In an attempt to better understand data collection and network screening methods, the research team developed and distributed a survey that was sent to all state transportation agencies during the Spring, 2019. The survey included 36 questions and was subdivided into five sections: 1) contact information; 2) crash report collection methods; 3) crash data collection training; 4) crash geocoding; and 5) network screening **(Appendix A).** In total, 29 responses were received representing 24 different states. For states providing multiple responses, their data was combined into a single response to eliminate redundancy.

Of particular interest regarding basis of comparable crash data, was a survey question in Section 1 that asked, "What is the most recent full year of crash data that your state department of transportation is working with?" Of the 24 states responding, 10 (42%) indicated 2018, 12 (50%) indicated 2017, and two states indicated 2016 and 2015, respectively.

Crash Report Collection Method

The survey included 13 questions related to crash report collection methods. It also indicated that all the states surveyed are using electronic crash report data collection methods to some extent with most having transitioned from fully paper-based reporting in the last 10 years. In a "check all that apply" question, the most common reasons for the transition were to minimize coding errors, enable consistency checks, and automate uploading. Improved geo-locating was also mentioned by some respondents, but to a lesser extent. Figure 1 shows the approximate proportion of crash reporting using paper or electronic reporting methods. The figure indicates that all responding states have at least 50% electronic reporting of crash reports with the majority (17 of 24) having 90% or more recorded electronically. Montana has multiple electronic crash reporting systems however the system used by major cities cannot currently be directly uploaded into the state database. These reports are printed and input manually. Some states indicated other methods, such as Oregon which uses self-report for all property damage only crashes.

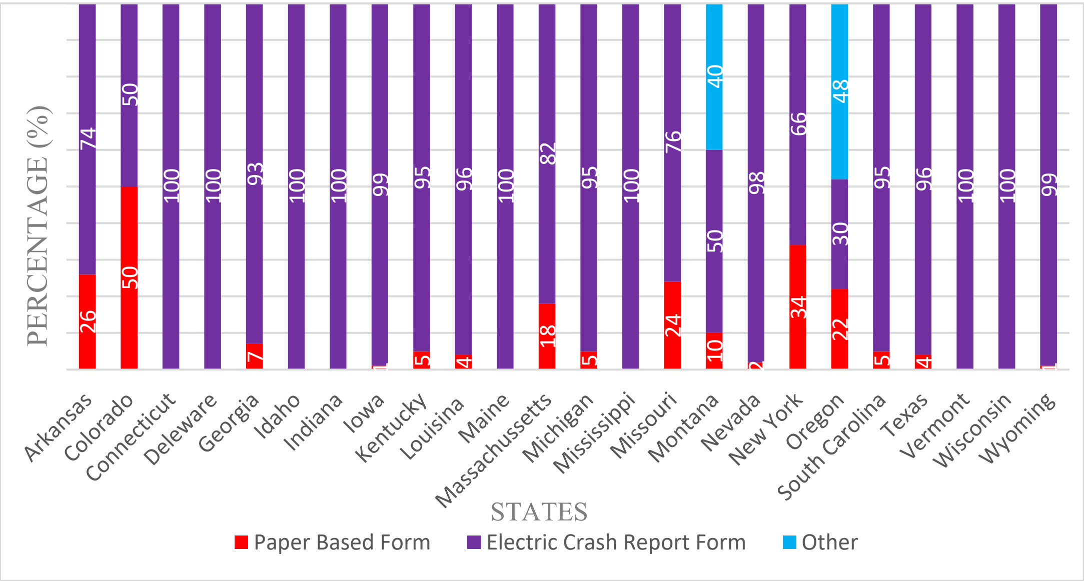


Figure 1: Proportions of paper based, electronic crash, and other report forms by state

In terms of the electronic crash reporting systems used by the state, most of the states use a system developed by a commercial vendor while some states use a system developed in-house. Five different commercial systems were identified: CAPS, CTA SmartCop, LexisNexis, ReportBeam, and TraCS. South Carolina’s SCCATTS uses the ReportBeam platform.

Crash Geocoding

When asked which methods (check all that apply) are used to capture the location of a crash in the field, 12 states use Mile Point Method, 18 use Primary/Secondary Street Name and

Distance from Intersection, 6 use Handheld GPS (coordinates entered or written on crash report manually), 18 said Map-based (with or without integrated GPS), and 6 answered other. Two of the "other" responses (Louisiana and Montana) had integrated GPS but not map-based fine tuning of the position. At least one state (Massachusetts), allows entry of nearest landmarks.

For map-based systems, states were asked what map background was used and 18 responded. The four answers given were Road Centerline Map [6], Street View Map [5], Aerial/Satellite Imagery such as Google Maps [5], and other [5]. Kentucky indicated that the officer can use a centerline map background with or without aerial imagery.

One question asked "For the most recent year, what percentage of crash locations are accurately geolocated on the total road network?" The response indicated that roughly 1/3 of the respondents believe that less than 80% of crashes in their state are correctly geocoded and nearly 1/2 of the respondents believe that greater than 90% of crashes in their state are correctly geocoded. Nearly all of the states indicated that they go through a process to validate the location accuracy of crashes and indicated that if a poorly geocoded crash location was identified a correction would be made.

Survey Discussing

The survey of state highway agencies indicates that there are disparate crash reporting systems used across the country from a crash geocoding standpoint. All of the states have implemented electronic crash reporting to some extent however, geocoding methods and accuracy varies by state and also may vary by jurisdiction within a state.

SOUTH CAROLINA CRASH DATA EVALUATION AND GEOCODING

Over the past decade, the major initiatives by SCDOT and SCDPS have proven to be effective in improving crash data. Crash location data accuracies are compared and contrasted between data recorded using hand-held GPS units from 2004 to 2010 and GIS-based map location system data, beginning in 2011. Comparisons are based on geocoding 15 years (2004 – 2018) of South Carolina crash data including an analysis of over 1.4 million crashes contained in the statewide geocoding database.

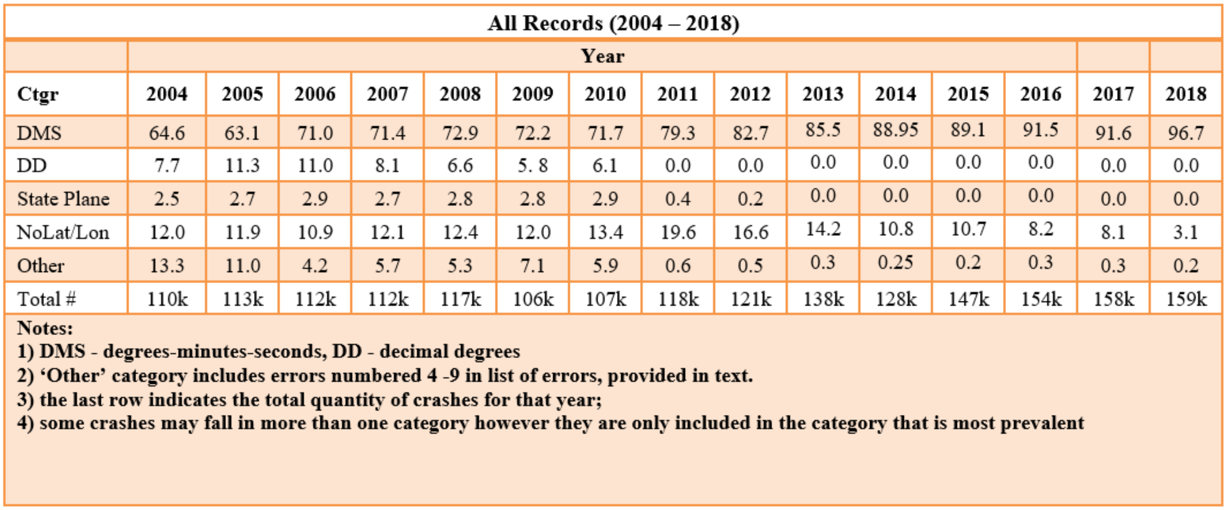
2004 crash location dataset was first geocoded as a baseline to assess the quality of the crash data. Geocoding results indicated 28% of 2004 crash data was geocoded outside of the state boundary, which provide a finding of great concern. Crash location results did not improve considerably until after 2010. By 2018, nearly all crashes geocoded within the state bundary. A review of the data geocoded by handheld GPS for all 15 years resulted in the identification of several systematic errors and erroneous inputs that were consistent with findings from a previous study by Sarasua in 2008 (17). Common and recurring problems in the crash database include:

1. Several crash records were in Decimal Degrees (DD), not Degrees-Minutes-Decimal Seconds (DMS) as referenced in the crash data reporting manual.
2. Some crash records were in state plane coordinates, not latitude and longitude
3. Several crash records were missing either longitude or latitude or both
4. Some crash records had their longitude and latitude values swapped

5. Most of the latitude values did not include a negative sign
6. Several coordinates were recorded with insufficient precision by one or two decimal places
7. Some crash records had spaces and letters as part of the coordinate entry
8. Some coordinates included additional zeroes to make up for the insufficient precision
9. Some crash records had erroneous coordinate values

Many crash records contained a combination of errors. For example, a crash record could have swapped latitude and longitude and at the same time have insufficient precision. The causes of errors include improper settings of the GPS equipment, errors by officers recording the coordinates, and errors by data entry personnel who transcribe information from the handwritten crash report into a digital database (17). A summary of the percentages of the geocoded data in each category by year is provided in **Table 1**. Trends in Table 1 indicate significant improvements in the consistency of geocoded crash data after 2010. The use of decimal degrees and state plane coordinates was virtually eliminated. While much of this improvement can be attributed to the use of SCCATTS, methods for using handheld GPS by local jurisdictions also improved. SCDOT indicated that this was possibly due to enhanced training. More recent improvements in consistency can be attributed in part to the increased proliferation of SCCATTS to local jurisdictions. It should be noted that nearly all highway patrol officers were using SCCATTS by the beginning of 2012.

Table 1: Percent of Crash Data by Geocoded Category and by Year



All Records (2004 – 2018)

	Year														
Ctgr	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
DMS	64.6	63.1	71.0	71.4	72.9	72.2	71.7	79.3	82.7	85.5	88.95	89.1	91.5	91.6	96.7
DD	7.7	11.3	11.0	8.1	6.6	5. 8	6.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
State Plane	2.5	2.7	2.9	2.7	2.8	2.8	2.9	0.4	0.2	0.0	0.0	0.0	0.0	0.0	0.0
NoLat/Lon	12.0	11.9	10.9	12.1	12.4	12.0	13.4	19.6	16.6	14.2	10.8	10.7	8.2	8.1	3.1
Other	13.3	11.0	4.2	5.7	5.3	7.1	5.9	0.6	0.5	0.3	0.25	0.2	0.3	0.3	0.2
Total #	110k	113k	112k	112k	117k	106k	107k	118k	121k	138k	128k	147k	154k	158k	159k

Notes:
1) DMS - degrees-minutes-seconds, DD - decimal degrees
2) 'Other' category includes errors numbered 4 -9 in list of errors, provided in text.
3) the last row indicates the total quantity of crashes for that year;
4) some crashes may fall in more than one category however they are only included in the category that is most prevalent

A separate analysis was conducted for geocoded crash data collected and recorded by highway patrol only. The consistency of highway patrol data is much better, even before the deployment of SCCATTS. Coordinates in DMS format was 88% and improved steadily to reach 96.6% in 2010. This is most likely due to the highway patrol receiving better training in proper use of handheld GPS than local jurisdictions. By 2013, after full deployment of SCCATTS within the highway patrol, virtually 100% of crash data was consistently geocoded in DMS format.

GIS ANALYSIS OF SOUTH CAROLINA CRASH DATA

Additional spatial analysis focused on the accuracy of geocoded crash data was conducted to further evaluate the improved spatial accuracy of geocoded crash data. ESRI's ArcGIS was used in all GIS analysis discussed in the paper. Nine years of crash data (2010-2018), with systematic errors removed or corrected (e.g. swapped longitude and latitude), were used. The highest-ranking corridors from a crash standpoint were the focus of this study.

Highway patrol officers, using hand-held GPS units, collected the majority of 2010 crash data. While much of 2011 and nearly all the 2012 data and beyond were collected using SCCATTS. An indication of the difference in precision of the two methods can be seen in **Figure 2**. The US-25 corridor example in Figure 2 shows that while 2010 crashes are mostly located on the sides of the roadway, or in parking lots, most of the 2012 crashes are shown on the roadway and in the location most likely to be where the crash actually occurred. A probable explanation for why 2010 data were mostly off the roadway is that most police officers would park their vehicles on the side of the roadway, or in parking lots, when filling out parts of the crash report and would read and record GPS coordinates on the GPS unit wherever they were parked.

The 2011 and 2012 data collection using the GPS enabled GIS-based map provided the police officers the tools to identify the approximate crash location using GPS, and then accurately locate (or pin map) the crash at the precise location it occurred on the map, even when parked on the side of the road, or in a parking lot.

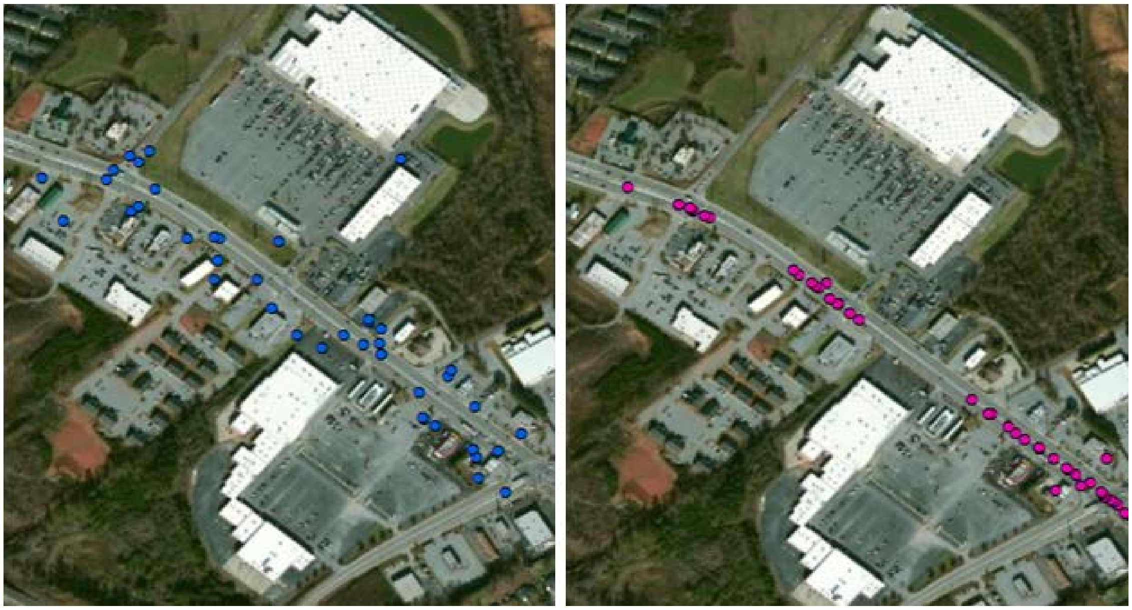


Figure 2: Rear-end and angle crashes on US 25 in Greenville, SC for 2010 (left) and 2012 (right) (images from Bing Maps)

Proximity Analysis

A proximity analysis was conducted to determine if there was a change in crash location relative to a roadway's centerline before and after the implementation of the SCCATTS. The distance of each crash from its corridor centerline was calculated and averaged by corridor using spatial analysis tools in ArcGIS for the 3-years of comparable data. **Table 2** shows the results of the proximity analysis for the top five selected corridors, based on average crash rank.

Table 2: Average Distance of Crashes from Route Centerline by Year

Route	Ave Distance (FT)		
	2010	2011	2012
US 1 Richland	14.6	3.7	3.2
US 25 Greenville	17.8	2.4	1.3
SC 146 Greenville	18.6	1.8	1.0
US 176 Richland	15.3	1.7	1.1
US 1 Lexington	14.7	4.4	4.7

The table clearly shows considerable change in crash locations relative to roadway centerlines from 2010 crashes (predominantly recorded with a hand-held GPS unit) to 2011 (predominantly SCCATTS). A paired t-test was conducted to test the null hypothesis that the difference between the 2010 and 2011 means is zero. Results of the t-test indicated a p-value < 0.001 and thus the null hypothesis was rejected with 99% confidence level indicating that the means are different. A similar finding was found for the difference between the 2010 and 2012 means. However, the null hypothesis that the difference between the 2011 and 2012 means is zero had a p-value of 0.08 and thus cannot be rejected for 95% confidence level (indicating that the difference in the means can be explained by random error).

Underlying Roadway Centerline Map Consistency

SCCATTS uses a roadway centerline map as a background reference for officers to use to geocode crashes. The centerline map is the same one used for SCDOT's Roadway Information Management System (RIMS). RIMS is a comprehensive geospatial-based database system that accounts for all data for SCDOT's transportation roadway inventory. A closer look at the RIMS centerlines indicates some problematic issues at some locations. For example, in some instances the RIMS centerline is erroneously offset from its actual location. Police officers using an erroneous centerline as reference to geocode a crash will locate the crash offset from where it actually happened. SCDOT periodically corrects errors it identifies in the RIMS centerline map which are eventually uploaded to officer's laptops however these uploads are currently done manually and very sporadically. **Figure**

3 shows a RIMS centerline map comparison between 2012 and 2018. An aerial map is shown in the background to provide reference. In Figure 3(a), the 2012 RIMS centerline is offset to one of the roadway directions near the major intersection. The 2018 crash data (shown as black squares) in the westbound direction overlaid on this RIMS centerline does not follow the 2012 centerline. The 2018 RIMS map shown in Figure 3(b) shows corrected centerlines and the 2018 crash data for the westbound direction does follow the new centerlines (as they should). Changes to the RIMS centerline map used by officers might be a problem when comparing crash data between different years that were geocoded using different RIMS centerline files. Inconsistency in using the RIMS data between different jurisdictions was also observed. For instance, the 2018 crashes described earlier for the westbound approach are pinpointed for both the directions and follows the 2018 RIMS centerline map. For the same year, the crashes in the northbound seem to be coded referencing the 2012 centerline map. Thus, police officers are not using consistent RIMS centerline data for the same years.

Figure 3: a) 2012 RIMS centerline map; b) Updated RIMS centerline map (image from Google Earth)

Travelway Analysis

While results from the proximity analysis indicate a distinct change occurred in average distance from centerline for crash data collected after 2010, an additional evaluation was conducted to identify the proportion of crashes that fell within the roadway corridor's travelway before and after implementation of SCCATTS. The same five corridors were used in this analysis. Offset lines such as lane lines, edge of pavement, and travelway limits are not included as RIMS GIS data layers, however, travelway width is included as a RIMS attribute. The buffer by attribute capability was used in ArcGIS to synthetically generate edge of travelway polygons for all five corridors. Buffering using buffer by attribute creates a polygon based on an attribute of individual segments, which in this application, buffered the roadway centerline segments using the buffer distance as half of the RIMS travelway width attribute value.

Using GIS point-on-polygon spatial aggregation, the crash data was overlayed with the travelway buffer polygons to identify crashes that are geocoded within the travelway corridors. **Table 3** shows the results of this analysis. It shows that only 35% of the 2010 crashes fall within the travelway even though it is likely that nearly all of the types of crashes used in this analysis occurred in the travelway. It should be noted that fixed object and run-off-the-road crashes were omitted from the analysis because of the likelihood that these crashes could occur outside the travelway. Further analysis of the sections of the routes listed in **Table 3** reveals that 2010 crash percentages do not accurately represent the potential conflict points where crashes are expected to be most prevalent (in the travelway). Conversely, most of the 2011 and nearly all of 2012 crash data do fall within the travelway in locations where potential conflicts are common. The improved performance between 2011 and 2012 is because SCCATTS was not fully deployed until the beginning of 2012.

Table 3: Percent of Highway Patrol Crash Data Identified by Corridor by Year

		2010 Crashes			2011 Crashes			2012 Crashes		
Route	**Miles**	HP	In TW	In	HP	In TW	In TW%	HP	In	In TW%
US1 Richland	22.2	571	218	**38.2**	524	506	**96.6**	639	632	**98.9**
US25 Greenville	66.0	653	176	**27.0**	498	443	**89.0**	718	691	**96.2**
SC146 Greenville	13.9	348	111	**31.9**	301	280	**93.0**	533	527	**98.9**
US176 Richland	24.7	377	131	**34.7**	280	261	**93.2**	493	485	**98.4**
US1 Lexington	38.0	346	169	**48.8**	242	228	**94.2**	406	389	**95.8**
Total	164.8	2295	805	**35.1**	1845	1718	**93.1**	2789	2724	**97.6**

Notes:
1.) HP – SC Highway Patrol
2.) In TW – Number of crashes located by GPS within defined corridor travelway
3.) In TW% – Number of crashes located by GPS within defined corridor travelway as percentage of total known corridor crashes, based on SC HP crash records

Case Study: Analysis of Driveway Related Crash Data

A recent SCDOT research project focused on developing safety performance functions (SPFs) and crash modification factors (CMFs) for commercial driveways in South Carolina *(16)*. Spatial analysis focusing on the accuracy of geocoded driveway crash data was performed as part of this research. Three years (2010-2012) of crash data was used for the geocoded accuracy analysis. Crashes that were potentially driveway related (i.e. coded with junction type –'driveway' or coded with a 'manner of collision' of 'rear-end' or 'angle' or 'side-swipe' or 'head-on') were extracted for use in this study. The improved spatial accuracy of crashes makes it possible to pinpoint the locations where clusters of crashes occur in relation to a driveway. This is evident at the location shown in **Figure 4** on US 17 in Berkeley County, South Carolina. The image shows a number of driveway related crashes (junction type 'driveway' shown with stars) occurring when vehicles attempt to enter or exit from adjacent high turnover driveways across a left-turn bay. The accuracy of crash data prior to 2010 (Blue Color) would not produce evidence of these clusters making it difficult to identify where crashes occur relative to driveways unless the sketches made by officers on the original crash reports are analyzed individually.

Figure 4 Crashes over a three-year period on US highway 17 in Berkeley County, SC *Coded driveway related crashes shown with stars. Note the proximity of the crashes relative to the left-turn bay. (Image from Google Earth)

To model the CMFs based on driveway characteristics it is necessary to associate driveway crashes with driveways. The junction type code included in the crash report includes driveway as one of the options. Unfortunately, a detailed analysis of the driveways along several selected corridors in South Carolina indicated that the driveway code was not used for more than 60% of crashes that were clearly driveway related based on crash type and proximity to driveways. Thus, a spatial analysis approach was developed to associate crashes with driveways. After querying possible crash types that could be associated with driveways (e.g. angle, rear-end, etc), the analysis assumption is that crashes in an influence area of a driveway is a driveway related crash of that driveway. It is crucial that the driveway influence areas are as precise as possible in order to evaluate the driveways effectively. One approach is to use ArcGIS buffer techniques to buffer an area on the travelway adjacent to each driveway to delineate the influence area. Once these buffers are created, they can be overlaid with underlying crashes to do the association. One problem with this approach is that the resulting driveway buffers would be circles around the point that represents the location of the driveway. This would bias crashes that occur closer to the side of the road. Ideally, rectangular buffers would give a better indicator of a driveway's influence area. Thus, the researchers developed a model that could make rectangular buffers that stretched across the roadway *(16).* Two models were created depending on driveway type: a model for full access driveways that creates buffers extending across all travel lanes; and a model for right-in right-out (RIRO) driveways that creates buffers that extend to the roadway centerline. Both models used the driveway width attribute from the driveway database to create the driveway buffer. The driveway buffer width is the driveway width plus thirty feet to accommodate about a car length on each side of the driveway. **Figure 5** shows resulting driveway influence area buffers along with 2012 driveway related crash data that fall within the buffers. Note that none of the 2010 crashes shown in the figure fall within the driveway buffers.

Figure 5: Driveway Influence Buffers Overlayed with 2010 and 2012 Crashes (image from Google Maps)

The analysis revealed an average crash incidence of 0.46 crashes per driveway using 2012 data. The analysis showed a much lower crash incidence (less than 0.1) for the same corridors using 2010 data. The 2010 rates are biased because poor geocoding precision placed most of the driveway related crashes outside of the driveway buffers. The driveway research made it very apparent that accurate crash geocoding was necessary to provide valid statistical results.

Case Study: Analysis of Driveway Related Crash Data Within 150 feet of intersections

A similar illustration of how accurately coded crash data can benefit crash spatial analysis is a case study for identifying problem driveways within 150 feet of intersections in which the corner clearance of the driveway does not comply with published standards in the SCDOT Access Management Guidelines *(4)*. A travelway polygon layer delineating edge of pavements for 5 major corridors were used for this study. Travelway polygons were overlayed with 50 foot buffer polygons of a selection set of driveways that fell within 150 feet of intersections. The resulting polygon layers were then overlayed with the driveway crash layer to determine the number of driveway related crashes within the hatched area shown in **Figure 6**. This analysis used only highway patrol data to ensure the before data (2010 driveway related crashes) was using predominantly GPS coordinates and the after data (2018 driveway related crashes) used SCCATTS. The number of crashes that fell within the driveway buffer and within the street travelway buffer for the 5 corridors totaled 64 crashes in 2010, and 525 crashes in 2018 (see **Table 4**). This represents a 700% increase in the quantity of driveway crashes that occurred in the travelway in close proximity to intersections. While some of this increase may be due to changes in landuse over this period the dramatic increase is undoubtedly due, in large part, to improved crash geocoding rather than a change in actual crash incidence. A closer look at these locations show that many of the 2010 crashes occur outside of the travelway and thus are ignored by the GIS operation.

While, the analysis shows how a GIS combined with precisely located crash data can be used to quickly identify potentially dangerous driveways with inadequate corner clearance, the omission of crashes due to poor geocoding will result in bias.

Table 4: Number of Driveway Crashes Occurring within Close Proximity to Intersections

Corridor	# of driveways	2010 Crashes	2011 Crashes	2018 Crashes
US 1 Richland	219	18	63	97
US 25 Greenville	177	9	36	167
SC 146 Greenville	29	8	18	73
US 176 Richland	102	16	30	88
US 1 Lexington	167	13	29	100
Total	694	64	176	525

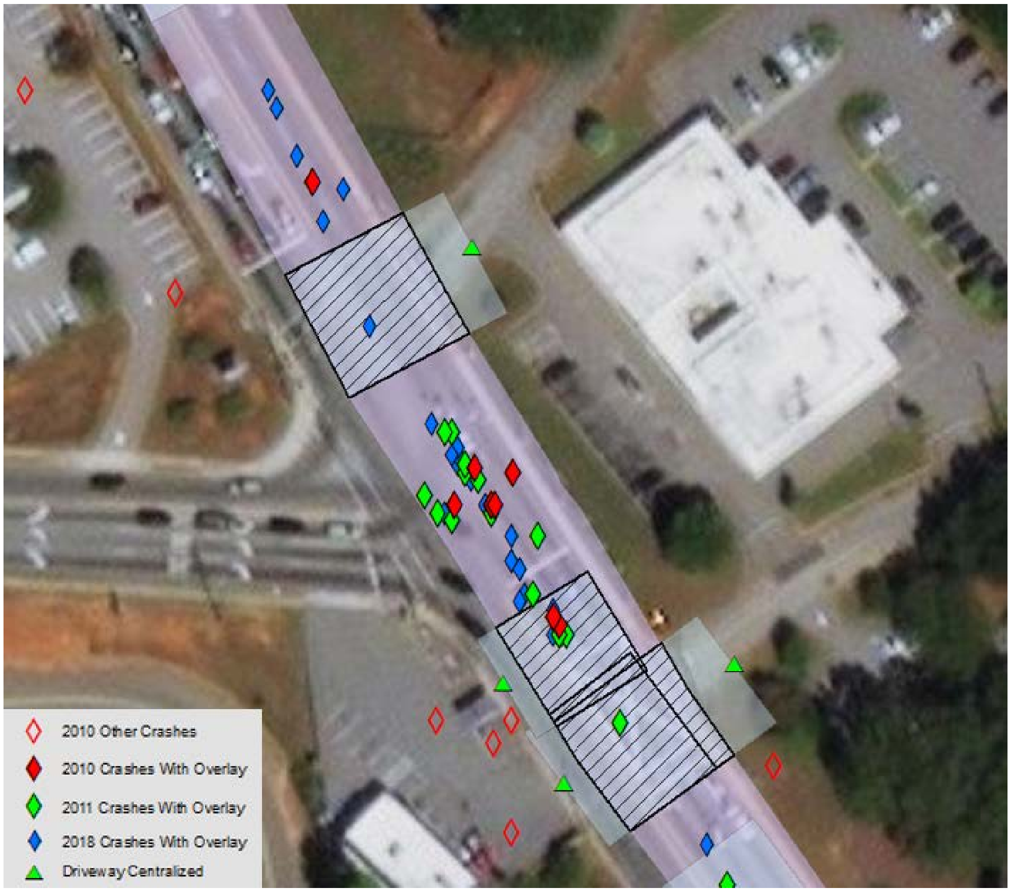


Figure 6: Driveway related crashes on US Highway 176 in Richland County, South Carolina. (image from Bing Maps)

Case Study: Statewide Screening of Mid-block Crash Locations Using Short Segments

The AASHTO Highway Safety Manual (HSM) (18) presents a variety of methods for quantitatively estimating crash frequency or severity at a variety of locations. The HSM recommends a minimum segment length of 0.1 miles for developing predictive models. The research discussed in this case study focuses on segment lengths of less than 0.1 miles for statewide screening of midblock crashes to identify site specific locations with high crash incidence. Famili et al (17) makes an argument that many midblock crashes can be concentrated along a very short segment due to undesirable characteristics of a specific site. The use of longer segments may “hide” the severity of a single location, if the remaining portion of the segment has few, or no, additional crashes. In this case study, South Carolina’s statewide road network is divided into short segments buffers, 50 to 100 feet in length. Intersection crashes were excluded from the analysis through the use of intersection buffers. Midblock crash data were aggregated along pre-designated short roadway network segments using the spatial overlay (spatial join) operation in ArcGIS (17). Figure 5 shows critical 50-foot segments identified along a section of Broad River Road in Columbia, South Carolina, with yellow and red polygon segments representing an occurrence of 4 or more cashes recorded during 2012. The red polygon segment has 7 or more crashes. The 4-crash threshold was chosen based on discussions with SCDOT. 2010 crashes are mostly located on the sides of the roadway, or in parking lots and were not captured by the short segments. Figure 6 shows a location associated with a dangerous horizontal curve where a 50-foot segment has 6 crashes in one year. One critical finding from the research described in this case study is short segment roadway network screening is viable as a safety analysis approach, only if accurately geocoded crash data is available for use as the basis for this methodology (17).

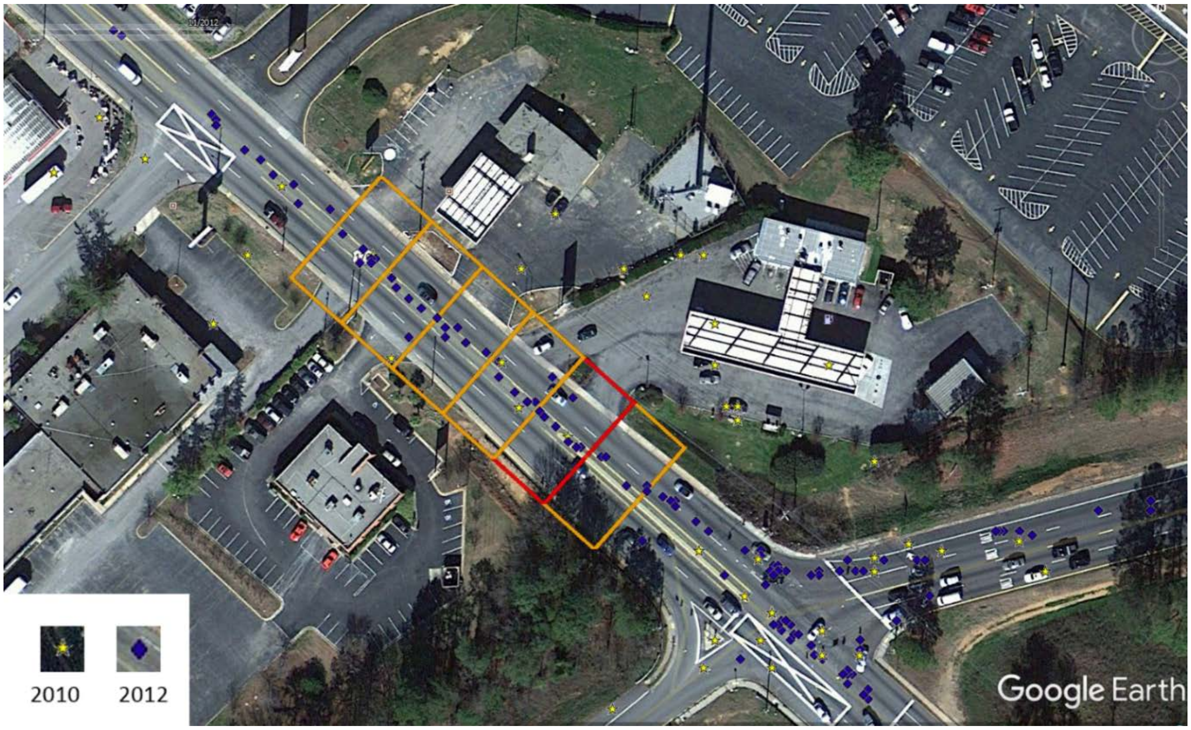


Figure 7: A length of Broad River Road in Columbia, SC showing 50 ft road segments with 4 or more crashes in 2012 . (image from Google Earth)

Figure 8: Dangerous horizontal curve on Reid School Rd, Taylor, SC . (image from Google Earth)

CONCLUSION

South Carolina has taken considerable strides towards improving crash data quality with the implementation of SCCATTS. Accurate crash data is an essential requirement for performing robust safety analysis and developing data-driven programs and policies. GIS spatial analyses methods and case study applications described in this paper would produce misleading and biased results if geocoded crash data used in the procedures contained systematic locational inaccuracies. Safety analysis along five major corridors using crash data geocoded with handheld GPS (2010 data) revealed that only 35% of the crash locations (not including run-off-the-road and fixed object crashes) were geocoded within the travelway, while SCCATTS geocoded crash data indicated that the proportion of crashes occurring within the travelway was virtually 100% for the same corridors. In the case study analysis estimating driveway CMF, findings revealed dramatically different values before and after the implementation of SCCATTS, which upon adopting and using 2012 produced an average crash incidence of 0.46 crashes per driveway. For the case study of driveways within an 150-foot proximity to intersections, procedures to analyze geocoded data failed to identify a substantial number of driveway crash clusters using 2010 crash data, whereas analysis of 2018 data readily revealed identifiable patterns and driveway crash concerns for similar locations. In the case study of short roadway segments, systematic statewide screening of midblock crash locations provided an effective approach for identifying problematic locations experiencing 4 crashes on more in a 50-foot segment for further evaluation. Furthermore, use of this method as a viable safety analysis approach for effective network screening is largely dependent on availability of accurately geocoded systemwide crash data (17).

Availability of accurately geocoded systemwide crash data is emerging as one of the most consequentially important transformational and essential elements for advancing roadway safety analysis of crash data in the future. The case studies presented in this paper are unique in their methods to aggregate crashes to driveways and short roadway segments. These approaches are largely reliant upon availability of accurately geocoded statewide crash data. An important caveat should be noted for the short segment analysis in that the HSM does not recommend use of segments less than 0.1 miles, however, the case study approach demonstrated use of segment lengths less than 0.1 miles (50-feet) as a viable means for effectively screening network midblock crashes, given accurately geocoded crashes. Specific benefits of enhanced systemwide crash geocoding include:

- Supports enhanced midblock crash cluster analysis and network screening;
- Provides an ability to associate midblock crashes with systemwide planimetric roadway features such as driveways, dangerous curves or roadside features;
- Allows efficient network-based analysis for specific types of midblock crashes and associated causation factors that would have previously required painstaking review of individual crash report illustrations;
- Promotes systematic network-based safety countermeasure analysis to pinpoint factors and locations where countermeasures can provide the greatest benefit.

Results from the survey of state highway agencies (n=24) indicates disparate crash reporting systems are used across the country with regard to crash geocoding data and procedures. All of the states responding to the survey have implemented electronic crash reporting to some extent; however, geocoding methods and accuracy varies from state to state, and also varies across jurisdictions within many states. Best practices for crash geocoding centers on providing reporting officers with an aerial image background and integrated GPS. The system should be deployed statewide and extend across all law enforcement jurisdictions. A few states responding to the survey indicated use of integrated GPS without a reference map, however, a major concern with this configuration is that patrol vehicles record their own location via GPS coordinates and fail to accurately geocode the crash location, which is a critical aspect for all GIS analysis methods.

While deployment of SCCATTS has led to substantial improvement in crash geocoding, there is potential for additional improvement. Based on assessment of SCCATTS crash accuracy and results from the survey of states, South Carolina SCCATTS system could be further improved by making an aerial background image available to reporting officers in the field. An aerial image background is an improvement over a centerline only background because the officers can use visible landmarks and lane geometry to more precisely locate crashes. For centerline only backgrounds, positional errors in the centerline will result in positional errors in crash locations and lead to limitations for use of the data in safety analysis.

Any safety analysis can only be as good as the data being used in the procedures. The ability to collect spatially accurate crash data constitutes an essential element in enhancing a state transportation agency's ability to conduct reliable safety analysis, as well as foster other transportation related research, resulting in more effective safety programs and policies for the traveling public.

REFERENCES

1. *Fatality Analysis Reporting System (FARS)* - National Highway Traffic Safety Administration (NHSTA). Traffic Safety Facts South Carolina, 2013-2017. *https://cdan.nhtsa.gov/SASStoredProcess/guest* , Accessed July 9, 2019.

2. South Carolina Department of Public Safety (SCDPS). *South Carolina Traffic Collision Fact Book 2013-17. http://www.scdps.gov/ohsjp/stat_services.asp* , Accessed March 10, 2017.

3. South Carolina Department of Transportation (SCDOT). "Target Zero South Carolina's Strategic Highway Safety Plan, South Carolina, 2015.

4. South Carolina Department of Transportation (SCDOT). *A Strategic Highway Safety Plan: The Road Map to Safety*. South Carolina Department of Transportation (SCDOT), South Carolina, 2007.

5. Scopatz, R. A., & Brown, R. South Carolina Safety Data Improvements through Electronic Crash System Deployment: Roadway Safety Data and Analysis Case Study, 2018 (No. FHWA-SA-16-109).

6. Abdel-Aty, M., Wang, X. "Crash estimation at signalized intersections along corridors: analyzing spatial effect and identifying significant factors", Transportation Research Record, 1953 (2006), pp. 98–111

7. El-Basyouny, K., Sayed, T. "Urban arterial accident prediction models with spatial effects", Transportation Research Record, 2102 (2009), pp. 27–33

8. Kim, K.E., Yamashita, E.Y. “Using K-means clustering algorithm to examine patterns of pedestrian involved crashes in Honolulu, Hawaii” Journal of Advanced Transportation, 41 (1) (2007), pp. 681–689

9. Siddiquia, C., Abdel-Aty, M., Choi, K. “Macroscopic spatial analysis of pedestrian and bicycle crashes Accident Analysis and Prevention”, 45 (2012), pp. 382–391

10. Smith, R., Graettinger, A. J., Keith, K., Hudnall, M., and Parrish, A. “Using GIS for Law Enforcement”. Journal of Safety Research, Vol. 36, No. 5, 2005, pp. 477-479.

11. Cherry, E., R. Floyd, T. Graves, S. Martin, and D. Ward. “Crash Data Collection and Analysis System”. 537, Arizona Department of Transportation, 2006.

12. National Cooperative Highway Research Program. NCHRP Synthesis 367 - Technologies for Improving Safety Data. NCHRP Synthesis 367, Transportation Research Board, Washington, D.C., 2007.

13. Federal Highway Administration (FHWA) Office of Safety. Roadway Safety Professional Capacity Building (RSPCB) - Peer-to-Peer Program. FHWA, 2011.

14. Sarasua, W., J. Ogle, and K. Geoghegan. “Use of Global Positioning System to Identify Crash Locations in South Carolina”. Transportation Research Record:

Journal of the Transportation Research Board, Vol. 2064, No. -1, 2008, pp. 43-50.

15. Stantec, and CDM Smith. South Carolina Multimodal Transportation Plan: Safety and Security. South Carolina Department of Transportation, 2013.

16. Stokes, A., Sarasua, W. A., Huynh, N., Brown, K., Ogle, J. H., Mammadrahimli, A., ... & Chowdhury, M. (2016). Safety Analysis of Driveway Characteristics Along Major Urban Arterial Corridors in South Carolina (No. 16-6766).

17. Famili, A., Sarasua, W., Iqbal, A. M., Kumar, D., & Ogle, J. H. (2019). Short Segment Statewide Screening of Midblock Crashes in South Carolina. Transportation Research Record. https://doi.org/10.1177/0361198119851068

18. AASHTO, 2010. Highway safety manual, 1st Edition. American Association of State Highway and Transportation Officials, Washington, D.C.

19. Lord, D. and Bonneson, J.A. Development of accident modification factors for rural frontage road segments in Texas. Transportation Research Record 2023, Transportation Research Board of the National Academies, Washington, DC (2007); pp. 20–27.

CHAPTER THREE

Paper II: Short Segment Statewide Screening of Midblock Crashes in South Carolina

Co-authors of the paper: Afshin Famili, Wayne A. Sarasua, Devesh Kumar, Jennifer H. Ogle.

Paper II was published at the 2018 Annual Meeting of the Transportation Research Board.

ABSTRACT

The AASHTO Highway Safety Manual (HSM) presents a variety of methods for quantitatively estimating crash frequency or severity at a variety of locations. The HSM predictive methods require the roadway network to be divided into homogeneous segments and intersections, or sites populated with a series of attributes. It recommends a minimum segment length of 0.1 miles. This research focuses on segment lengths of less than 0.1 miles for statewide screening of midblock crash locations to identify site specific locations with high crash incidence. The paper makes an argument that many midblock crashes can be concentrated along a very short segment due to an undesirable characteristic of a specific site. The use of longer segments may “hide” the severity of a single location if the rest of the segment has few or no additional crashes. In actuality, this research does not divide sections of roads into short segments. Instead, a short window approach is used. The underlying road network is used to create a layer of segment polygons using GIS buffering.

Crash data are then overlaid and aggregated to the segment polygons for further analysis. The paper makes a case for the use of short fixed segments to do statewide screening and how accurately geocoded crash data is key to its use. A comparison is made with a sliding window approach (Network Kernel Density). The benefits of using fixed segments is that they are much less complex then using the sliding window approach. Because the segmentation can be the same from year to year, direct comparisons can be made over time while spatial integrity is maintained.

INTRODUCTION

In Part B of the AASHTO Highway Safety Manual (HSM), a variety of network screening methods are presented for identifying and prioritizing sites with potential for safety improvement. These range from simplistic approaches such as calculating crash rates or severity indices to determining the excess predicted average crash frequency using safety performance functions (SPFs). The latter is the approach taken by SafetyAnalyst – an AASHTOWare software tool developed to support the safety management process at a state agency. Using this method, a site's observed average crash frequency is compared to a predicted average crash frequency found using an SPF. If the difference between the observed and predicted is greater than zero, the site experiences more crashes than predicted and might be considered as a candidate for further study. two represents the excess predicted crash frequency.

To correctly apply any of the screening methods, the road network and crash data needs to be divided into road segments and intersections, and then further grouped into reference populations based on select characteristics (e.g., rural two-lane highways, or urban three-legged stop-controlled intersections). Prioritization of sites is made within a reference population. Some common characteristics used to define reference populations are listed in Table 1:

Table 1: Common characteristics used to define reference populations

Intersections:	Segments:
• Traffic control	• Number of lanes
• Number of approaches	• Access density
• Number of lanes	• Traffic volume range
• Functional classification	• Median type
• Area type	• Operating speed
• Traffic volume range	• Terrain
• Terrain	• Functional classification

Within each reference population, sites may be further disaggregated into homogeneous units by factors such as traffic volume, lane width, curve presence, median type, etc. Homogeneous, with respect to a roadway segment, implies that all of the characteristics of that segment are the same.

While all state Departments of Transportation maintain a number of roadway attributes, few, if any, have a statewide database that contains all of the variables found to be of significance in the HSM safety performance functions. Driveway density, for example, is only collected by a few DOTS that have extensive asset management programs (2). Furthermore, state-specific SPFs tend to be limited to only one or two significant parameters either due to the lack of comprehensive data or due to lack of variability in the design parameters. For example, the Safety Performance Function (SPF) for midblock segments used by Caltrans *(3)* only takes average annual daily traffic (AADT) into account as an explanatory variable. The use of just segment AADT will not identify dangerous driveway locations regardless of segment length. Further, if driveway density is incorporated in the analysis, high-turnover driveways are treated the same as low-turnover driveways. Stokes *(4)* showed that the characteristics of driveways and the land use they serve have a significant impact on crash incidence. Studies *(5, 6)* have shown that geometric design has a significant effect on safety, especially on rural highways. They suggest that there is a relationship between geometric consistency and crash frequencies. Although one can find consistency among the effects of segment length and AADT, there is a variation in safety performance of horizontal alignment and access management strategies *(7)*. When possible, it is best to disaggregate segments to identify such design features, and the HSM procedures reflect this strategy.

The HSM recommends a minimum segment length of 0.1 miles for Safety Performance Function (SPF) development. One reason for this is that variability in crash location data can allow for incorrect assignment of crashes to the appropriate road segment *(1)*. A number of researchers have indicated that spatially inaccurate crash data can adversely affect safety analysis *(4)*. In 2008, the South Carolina Department of Public Safety undertook a major initiative to improve crash data quality through implementation of an automated crash data collection system called the South Carolina Collision and Ticket Tracking System (SCCATTS) to be used by law enforcement *(8)*. This system enables officers to spatially see and locate crashes via a GIS-based GPS enabled mapping platform in police vehicles. The GPS displays the vehicle's location on the GIS map display and then the officer can pinpoint the actual location of the crash rather than where the officer's vehicle is (e.g. on the side of the road or in a parking lot, etc.). The deployment of the system began in 2010 and as of April 2013, all SC highway patrol vehicles have been equipped with SCCATTS *(8)*. SCCATTS has resulted in dramatically improved geocoded crash positioning. An indication of this improvement is shown in Figure 1. The US-25 corridor example in Figure 1 shows that while 2010 crashes are mostly located on the sides of the roadway, or in parking lots, most of the 2012 crashes are shown on the roadway and in the location most likely to be where the crash actually occurred, as they were visually verified by officers using a map application.

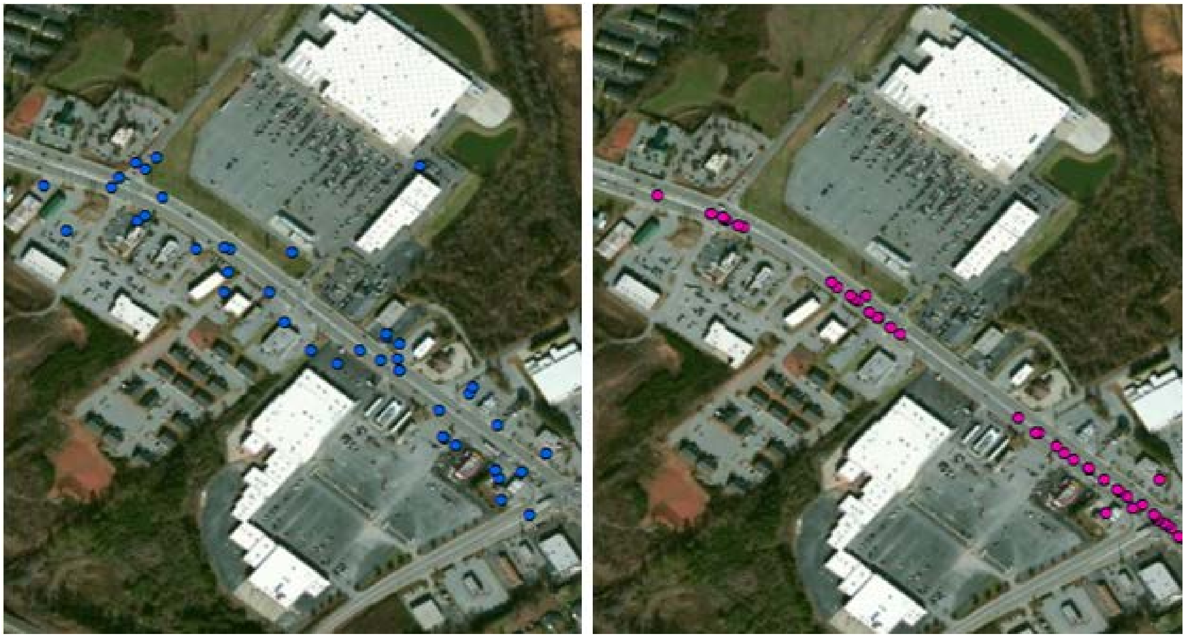


Figure 1: Rear-end and angle crashes on US 25 in Greenville, SC for 2010 (left) and 2012 (right) (images from Bing Maps)

This research focuses on segment lengths of less than 0.1 miles for statewide screening of midblock crash locations to identify site specific locations with potential for safety improvement. The primary motivation for this research is that precise crash location data can overcome previous issues with using short segments. Further, the use of short segments can potentially identify locations with a concentration of midblock crashes that may be related to an undesirable characteristic of a specific site. The use of longer segments may “hide” the severity of a single location if the rest of the segment have few or no additional crashes. In this research, short segment buffers are created from the underlying road network to create a GIS layer of segment polygons. Crash data can then be overlaid and aggregated to the segment polygons for further analysis. The statewide screening methodology presented in this paper has benefits over more complex spatial statistical approaches because of its ease to implement within a DOT. Because the segmentation can be maintained from year to year, direct comparisons can be made over time while also ensuring spatial integrity.

LITERATURE REVIEW

At mid-block segments and on intersection approaches, researchers have evaluated the impact of different methodologies and criteria on the accuracy of high risk locations (hotspots). The literature review summarizes previous research and studies on hotspot identification using segment based analysis methods.

Crash Screening using Segment Based Analysis Methods

Crash screening methods are used widely to quickly characterize observed crash data from a large study area which will lead to identifying a smaller set of locations (hotspots) that can then be analyzed in more detail. A description of crash screening methods is provided in HSM part B *(1)*. The pros and cons of using different types of statistical methodologies in hotspot detection is also provided in the manual.

The expected crash frequency where collision frequency is statistically modeled as a function of relevant features (e.g. road characteristics, traffic exposure, and weather factors) is a method for identifying potential hot spot locations *(9,10,11,12)*. The Negative Binomial approach is one of the popular methods for modeling crash frequency. The approach is data intensive and requires significant effort in processing the related data and calibrating the corresponding models *(13)*. Expected crash frequency can also be calculated through geostatistical techniques such as kernel density estimation (KDE) *(14,15,16),* K-means clustering *(17)*, Getis-Ord Gi statistics *(14,18),* and nearest neighbor clustering *(16, 19)*.

Kwon et al *(20)* evaluated the performances of three different segment analysis methods (Sliding Moving Window (SMW), Peak Searching (PS) and Continuous Risk Profile (CRP)) to analyze freeways. They used the same input requirements for each of the three methods in the evaluation. Each of these methods were used to prioritize the detected sites for safety investigation and the lists were compared with previously confirmed hotspots. The length of segments defined in the approaches varied from 0.04 to 3.64 miles

In 2008, Xie and Yan *(21)* employed a network KDE (NKDE) approach to estimating the density of spatial point events. The results showed that the NKDE is more appropriate than standard planar KDE for density estimation of traffic crashes because KDE covers space beyond the roadway network and is more likely to overestimate the density values. In follow-up research they conducted a study *(22)* integrating NetKDE with local Moran'I for hotspot detection of crashes. Using the combination of approaches with a 328-ft neighbor for Moran's I computation, they found fewer statistically significant "high-high" (HH) segments and hotspot clusters.

Dai et al. *(23)* have applied network-based geospatial techniques to identify crash clusters on the University of Georgia campus. They used network-based Kernel Density Estimation to identify high- density road segments and intersections, then used network-based K-function to examine the clustering of pedestrian crashes. The results suggested that crashes occurred more frequently in road segments with strong street compactness and mixed land use present and were significantly ($p<0.05$) clustered in these high-density zones.

Nie et al. *(24)* applied NKDE with Network-constrained Getis-Ord Gi* to detect spatial cluster patterns and identify hotspots in midblock segments. The methods were applied to one-year crash data in China. The results indicated that both methods performed well in identifying risky segments.

Pande et al. *(25)* presented a classification tree based alternative to crash frequency analysis for analyzing crashes on mid-block segments of multilane arterials. The classification tree models provided a list of significant variables as well as a measure to classify crash cases. They provided the safety analysis community an additional tool to assess safety without having to aggregate the corridor crash data over arbitrary segment lengths.

Segmentation Length (Window size for the Peak Search Method)

In the Peak Search method, the roadway network is divided into equal length segments and then each segment is examined using one of the hotspot identification methods. Previous researches *(26)* show that segment length has significant impact on identifying high crash locations and can affect the consistency of high crash locations *(27)*. The results of a study conducted by Green and Agent *(28)* showed that up to 8 percent of crashes may be incorrectly located by over 500 ft because of the accuracy of recorded crash location. Hence, the accuracy of recorded crash location should be considered in defining the segment length.

Lu et al. *(29)* employed a Negative Binomial model for divided segments using fixed length, variable length, and Fisher's clustering methods. They applied a minimum segment length of 0.05 mile (264 ft) in Fisher's clustering approach. In this approach, the roadway section can be considered as a set of samples while, the crash frequency can be considered as the crash indices. This study found that the relationship between crashes and independent variables is facilitated using Fisher's clustering which improves the precision of SPF calibration over variable length and fixed length methods.

Medury et al. *(30)* proposed a dynamic programming-based hotspot identification approach, which provides efficient hotspot definitions for pedestrian crashes. They compared the proposed approach with the sliding window method and an intersection buffer-based approach. The results suggested that the dynamic programming method generates more hot spots with a higher number of crashes using small hot spot segment lengths. In addition, the sliding window method was shown to suffer from shortcomings due to a first-come-first-serve approach in hot spot identification.

Literature Review Summary

It is very clear from the literature that the research consensus is that short segments should be avoided in segment based safety analysis. Clemson University has done a great deal of research on the accuracy of crash geocoding. Based on this prior research, the authors contend that one of the reasons that short segment crash analysis does not provide meaningful results is in part because of the inadequacy of the locations of crashes. This research explores how spatially accurate crash locations can facilitate short segment crash analysis.

METHODOLOGY

South Carolina has made great strides to improve crash data quality within the state with the implementation of SCCATTS. The methodology for this research involved two approaches including 1) segmentation through polygon buffers of the underlying roadways at different intervals and 2) Network KDE/Sliding Window method. For this analysis multiple years of crash data was used.

Segmentation (Peak Search) Approach using Polygon Buffers

Crash Accuracy

The deployment of a map-based crash geocoding system has greatly improved the quality of crash location data in South Carolina. Improved crash data helps to improve the reliability of crash location identification and evaluation of countermeasure effectiveness. Among multiple attributes in a crash data set, the location of the crash has special value because in many cases, crash records with the wrong location cannot be considered in the analysis. Before 2011, the South Carolina highway patrol used hand-held GPS units to geocode crash locations. This led to many systematic errors when officers transcribed the coordinates to the crash report. The current SCCATTS system has now made virtually all crash data usable for safety analysis from a spatial location standpoint (Figure 2).

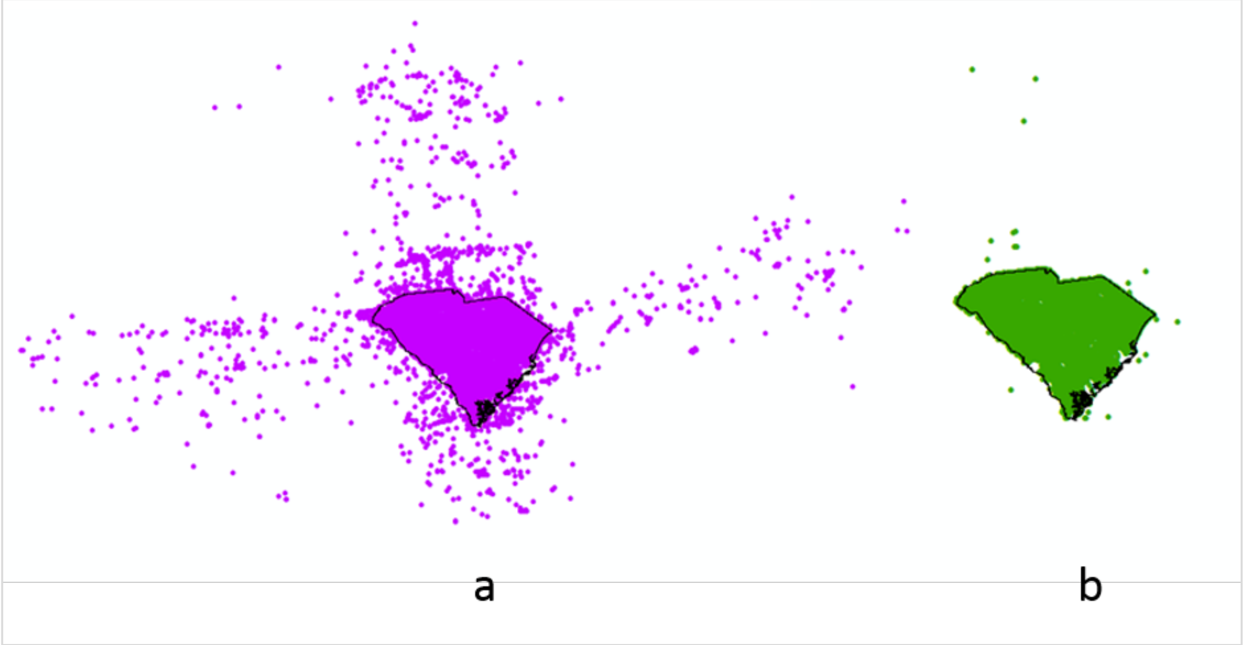


Figure 2: Geocoded Crashes in South Carolina: a) 2004 all; b) 2012 Highway Patrol

Short Segment Buffers

In the first method, the research team initially used a variable buffer using the roadway width attribute in the SCDOT road characteristics database to create a GIS layer representing the road surface. After careful examination, the SCDOT GIS centerline in many cases was found to be displaced from its exact position. In this scenario, creating buffers with surface width might result in displaced buffers which might not be able to cover the whole roadway width. An example of this is presented in Figure 3. The green dots symbolize 2016 crashes while the orange colored strip represents the roadway. After experimenting with different buffer widths a fixed buffer of 50 ft on each side of the centerline was used to help ensure that all crashes along the roadway are accounted for in the analysis.

Figure 3: Roadway Buffer created using Roadway Width

Once the 50 foot wide roadway buffers were created in ArcGIS, the buffered layer was segmented using different fixed segment lengths. The first segment length used was 50 ft. Once created, crash data could be overlaid with the buffers to identify critical segments with unusual number of high crashes. This was done using the spatial join operation in ArcGIS. Because the analysis focused on midblock crashes segments within a

150 ft radius of intersections were removed before the spatial join operation was performed. Figure 4 shows continuous as well as critical crash segments as created using the methodology. On the left side, continuous 50 ft buffer segments are shown while on the right side, critical crash segments with a crash count of 4 or more crashes are presented. The remaining segments were turned off in the right-side image.

Figure 4: Continuous vs Critical Crash Segments (50 ft) in Anderson County, SC

The total network length assessed with this approach is 41,282 miles. The initial 50 ft segment length was due to Stokes *(4)* who found that a driveway's primary influence area is roughly equal to the driveway width plus a car length before and after the driveway. Subsequently, the length was increased with an increment of 50 ft up to the recommended HSM minimum length of 1/10 mile (~500 ft). Hence, a total of six different segment lengths (50 ft, 100 ft, 150 ft, 200 ft, 250 ft and 500 ft) were created with an objective of evaluating the different lengths to determine the most appropriate length of fixed segment buffers for network screening analysis.

Aggregating Crash Data to Segment Buffers

Once the polygon buffers with different segment lengths were created, the midblock crash data was aggregated to the segment buffers. Prior to associating crash data with these segment buffers, the crashes within 150 ft of intersections were removed. The 150 ft length was used after discussions with SCDOT to identify the intersection area of influence. Crashes within the intersection influence area were not considered as midblock. The crashes were aggregated to the segment buffers using the spatial overlay (spatial join) operation in ArcGIS. The resulting segments were stratified based on the number of crashes they contained. Figure 5 shows critical segments identified for 50 ft polygon segments. This is a section of Broad River Road in Columbia, South Carolina and the colored polygon segments represent an occurrence of 4 or more cashes occurring in the year 2012. The 4 crash threshold was chosen based on discussions with SCDOT. They identified that choosing a crash threshold that produced less than 500 segments would be management for network screening purposes. The red polygon segment has 7 or more crashes.

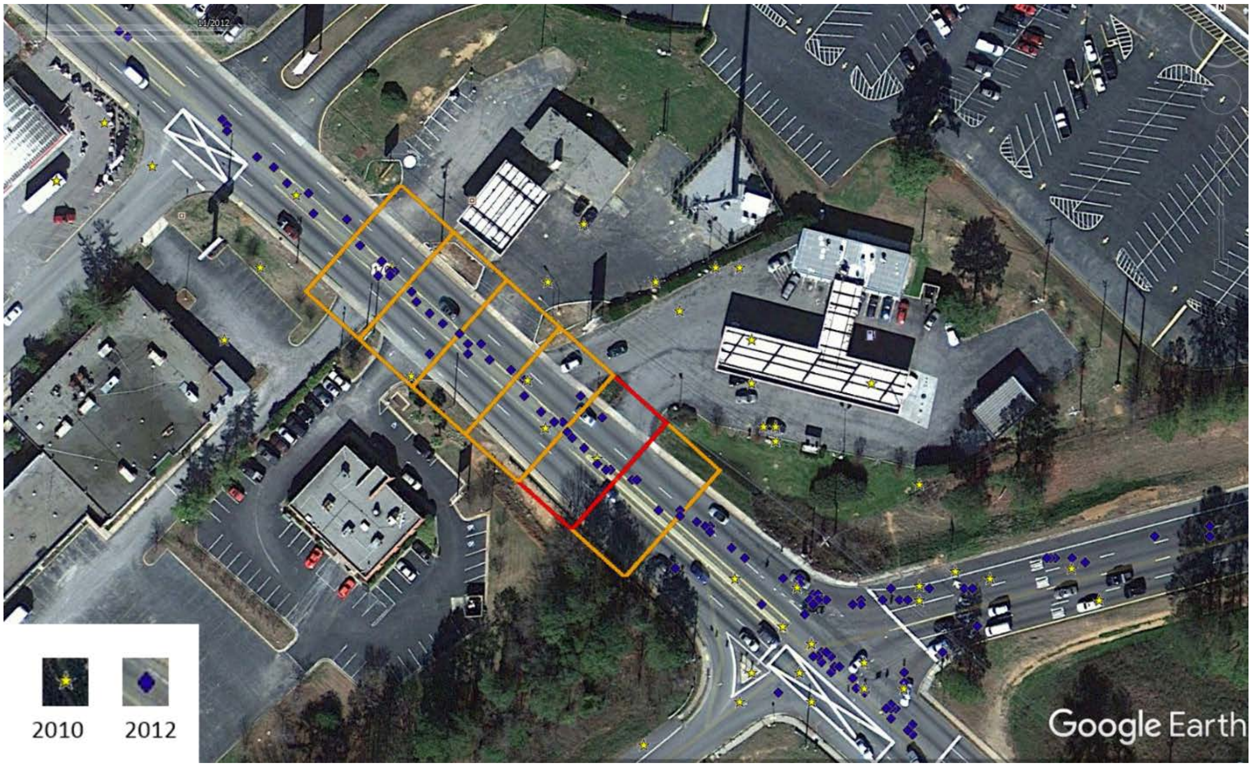


Figure 5: A length of Broad River Road in Columbia, SC showing 50 ft Road Segments with 4 or more Crashes in 2012.

Creating these buffers helped in identification of various type of critical midblock segments. Figure 6, 7, and 8 show how the short segment overlay analysis identified locations with unusually high number of crashes. Figure 6 shows a location associated with a dangerous horizontal curve. Figure 7 shows a partial clover interchange ramp using 2012 data. A look at 2016 data showed a significant drop in the number of crashes. The 2016 imagery shows that the ramp was realigned which served as a successful countermeasure. Figure 8 shows a midblock driveway location with an unusually high incidence of crashes.

Figure 6: Dangerous horizontal curve at Reid school Rd towards Wade Hampton Blvd, Taylor

Figure 7: Partial clover interchange ramp using 2012 and 2016 data at Liberty Highway, Anderson

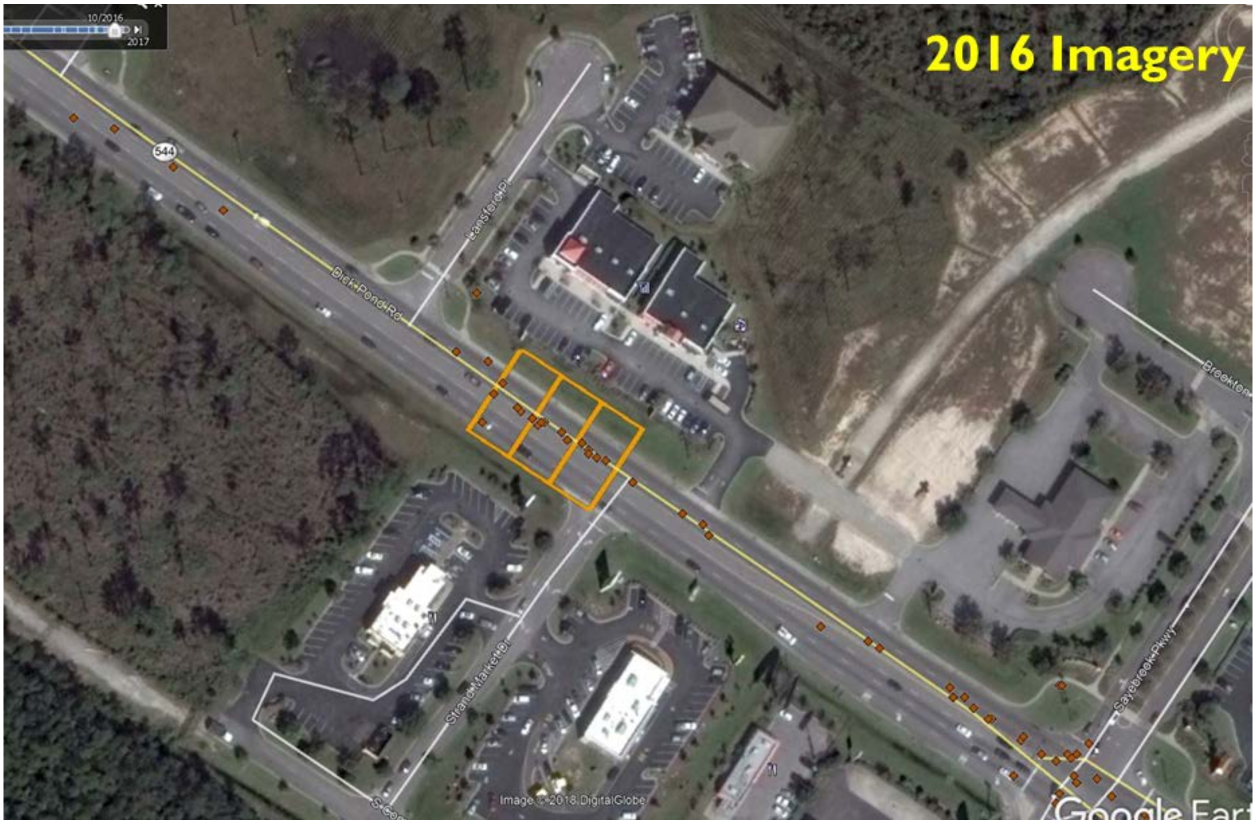


FIGURE 8: Midblock driveway location with an unusually high incidence of crashes at Dick Pond Rd / Strand Market Dr. near Myrtle Beach

Network KDE

Planar KDE considers a planar area of influence for each crash. However, in NKDE each crash has impact on a chosen distance (bandwidth) along the network. In other words, a 0.1 mile bandwidth on a network means 0.1 mile from a crash in shortest path distance. The NKDE estimators are as follows:

$$f_s = \sum_{i=1}^{n} \frac{1}{h} k\left(\frac{d_{i,s}}{h}\right) \qquad (1)$$

Where:

h: defined bandwidth

k=function of network kernel density

d: shortest path distance from the center of the bandwidth along the network

Multiple models can be used to estimate the distance in the NKDE function including Gaussian, Quartic, Conic, negative exponential, and epanichnekov *(31, 32).* The Gaussian function of k is defined as:

$$k(\frac{d_{is}}{h}) = \begin{cases} \frac{1}{\sqrt{2\pi}} \exp(-\frac{{d_{is}}^2}{2h^2}) & 0 < d_{is} \le r \\ 0 & d_{is} > r \end{cases} \qquad (2)$$

It is often that the density values calculated from NKDE are considered to be positively spatially auto-correlated, and nearby density values considered to be similar to each other since neighboring points within the distance of a bandwidth are used in the NKDE process. Therefore, it is assumed that there is no need to carry out additional spatial auto-correlative analysis on density values. One of the main limitations for KDE and NKDE is that no statistical inference can be evaluated in the process and there is no indication of a density threshold which a hot spot can be confidently stated. It might be guessed that locations with high density values could possibly be hot spots, but no mechanism has been available in KDE to assess their statistical significance. By conducting a statistical significance analysis on density values it is possible to evaluate formally the statistical significance of the extensiveness of locations with high density values, and to determine if hot spots of traffic accidents actually exist consistently along certain portions of a roadway network *(33)*.

RESULTS AND DISCUSSION

Peak Search Method Using Small Window Size

In this approach, whole South Carolina's road network (41,282 miles) was assessed. The initial use of 50 ft segment buffers produced 274 segments with 4 or more crashes using 2016 data. This turns out to be 0.000062% of the total midblock 50 ft segments. This was deemed manageable to look at on a case by case basis by SCDOT. A closer look indicated that many of the segments did not include crashes that were probably associated with the driveway (see Figure 9 for example). This finding justified trying longer segment lengths with the intent of identifying roadway midblock locations with an unusual number of crashes. Thus, the researchers tried a series of segment lengths in 50 ft increments up to the recommended HSM minimum of 1/10 mile. For each increment, the researchers tabulated the number of segments with a range of crashes (see Table 2). A segment length of 100 ft with 4 or more crashes increased the number of segments to 554 out which is 0.00025% of the total 100 ft midblock segments. This was close to the threshold identified by SCDOT as manageable for network screening purposes. A closer look at these segments showed that in nearly all cases, the crashes seem to be concentrated at a single driveway or a geometric event such as a curve (see Figure 10).

The increase to 150 ft identified 1,017 segments, 0.00068% of the total 150 ft midblock segments. A closer look at these showed many segments with more than 1 driveway. Further, many of the 50 ft segments and even the 100 ft segments with the highest number of crashes were not as highly ranked when 150 ft was used. This indicated that a combination of locations was beginning to play a role in the crash incidence. Increasing the segment length to 1/10 mile (500 ft) and ranking the top segments from a crash incidence standpoint resulted in only 23% of the 100 ft segments and less than 14% of the 50 ft segments were included in the highest ranked 1/10 mile segments. This indicates that the longer segment length diluted the effect of a dangerous driveway or geometric situation. Our segment length analysis indicated that 100 ft would be an ideal segment length for short fixed segment analysis.

Figure 9: Woodruff Road, Greenville

Table 2: Critical Crash Segments for various lengths of segments

Buffer Length (ft)	50	100	150	200	250	500
Crash Segments (>=4 crashes)	274	554	1017	1,243	1,437	1,911
Total Segments	4,415,467	2,221,710	1,49,7088	1,132,157	913,038	475,049

Figure 10: Crashes concentrated at a single driveway (Woodruff Road, Greenville)

NKDE Method

In this approach, the road network in Anderson County (1,302 miles) was selected on a sample basis for the assessment. NKDE has been adopted as a practical method for decision making in this research. Combination of SANET 4th Edition and ArcGIS pro has been used to implement NKDE. To make it comparable with peak search method, this study used the same window size (50 ft, and 100 ft bandwidth) and 5 ft *lixel* size and a Gaussian kernel function. The method calculates the density at the center of each lixel and the whole lixel will have that value. Figures 11a and b illustrates the spatial pattern of crashes for Anderson County with these two chosen bandwidth.

The red circle on the figure represents the identified high crash locations. As it can be noticed, the bandwidth plays significant role in structuring the network density pattern. The density pattern showing longer segments as high density with increasing search bandwidth (50 ft and 100 ft, respectively), when the kernel function is the same (Gaussian) and at the same lixel size (5 ft). It appears that the narrow bandwidths (50 ft) may produce patterns suitable for presenting local effects or "hot spots" at smaller scales. As the search bandwidth increases from 50 ft to 100 ft, the high crash locations are gradually combined with their neighbors, and larger clusters appear. A high bandwidth might affect a safety professional's decision as it suggest longer segments for improvement.

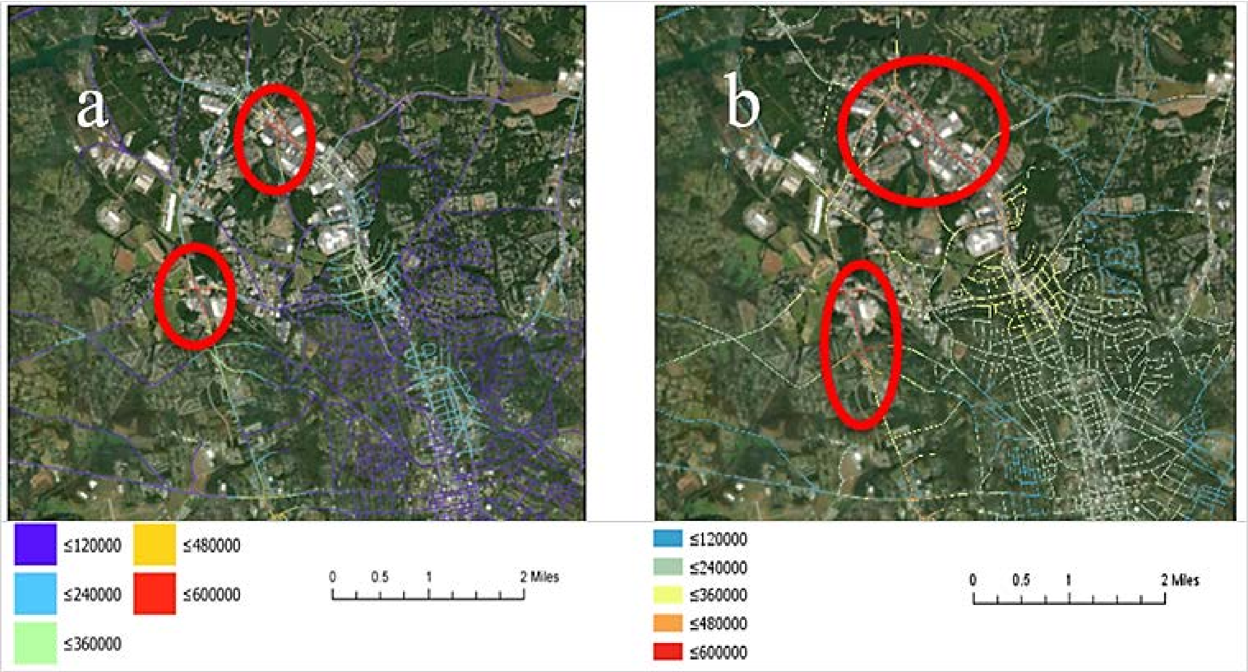


Figure 11: NKDE method with bandwidth of a) 50 ft b) 100 ft

Comparison of two methods

Figure 12a and b represents a hotspot locations identified using the peak search method (100 ft window size) and NKDE method (100 ft bandwidth) respectively. As can be seen, the NKDE method will identify continuous local hotspots. However, the short segmentation approach will result in discrete identification. As visible in Figure 4, the NKDE method generated continuous corridors while peak search method generated discrete locations which can be attributed to single driveways or other geometric feature which is not possible in former method. Also, due to discrete locations being short in length (100 ft segments) it may be more economical to make improvements as compared to improving a whole corridor. However, proper analysis of prioritized corridors should also enable discovery of these point issues.

One of the advantages of using fixed segments is the same window slots with potential minor updates can be used for subsequent years facilitating temporal analysis. Different window sizes can be created once and can be used to find an optimal size for the study area. This can help state DOTs in terms of implementation because of the simplicity of the short segment method. Further, the result of the segmentation is easier for safety professionals to interpret and implement countermeasures.

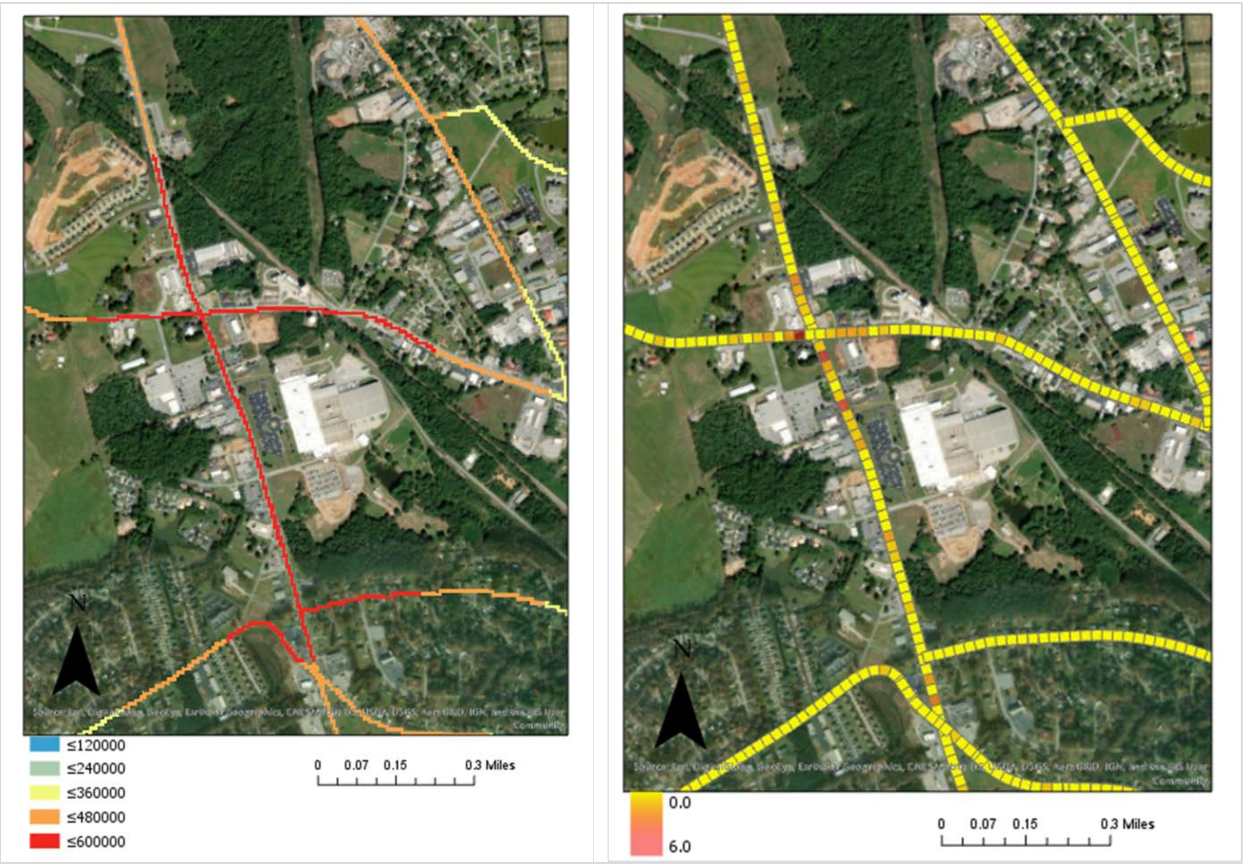


Figure 12: (a) NKDE using 100 ft bandwidth (b) Peak Search using 100 ft window size

NETWORK SCREENING USING 100 FOOT SEGMENTS STRATIFIED BY ROADWAY TYPE

To evaluate the accuracy of short-window peak search method, three specific road configuration has been stratified (rural two lane, rural multilane and urban and suburban arterial). The urban arterial includes undivided two-lane and a three-lane and five-lane section with a center of TWLTL (U2U, U3T and U5T respectively). Rural sites include two way two lane undivided and rural 4 lane divided (R2U and R4D respectively).

Comparison of Highest Ranked Segments Based on Excess Predicted Frequency and Absolute Crash Frequency

SPFs and CMFs for driveways have been calculated based on data obtained from *(4)*. The "Excess predicted average crash frequency using SPFs" method is compared against absolute crash frequencies obtained at each site. Using the excess prediction method, a site's observed average crash frequency is compared to a predicted average crash frequency found using an SPF. If the difference between the observed and predicted is greater than zero, the site experiences more crashes than predicted and might be considered as a candidate for further study. The reasons of choosing this method are more accurately calculating the potential for safety improvement and acknowledging the complex, non-linear relationship between crash frequency and volume *(1)*. Figure 13 represents the comparison of highest ranked 100' segments based on excess predicted and absolute crash frequencies stratified by roadway type. Since most DOTs do not collect detailed data regarding driveway and geometric design features throughout the state, finding an alternative to use of SPFs can be cost effective and can save time in data processing. The results show that short-segment absolute frequency method can fairly accurately obtain the top 20% of sites with accuracies ranging from 76.9%-100% when compared to the robust approach of using excess predicted average crash frequency. The exception to this was in the rural four-lane divided reference population where sample sizes were small. However, as you move down the priority list, the short window approach becomes less effective.

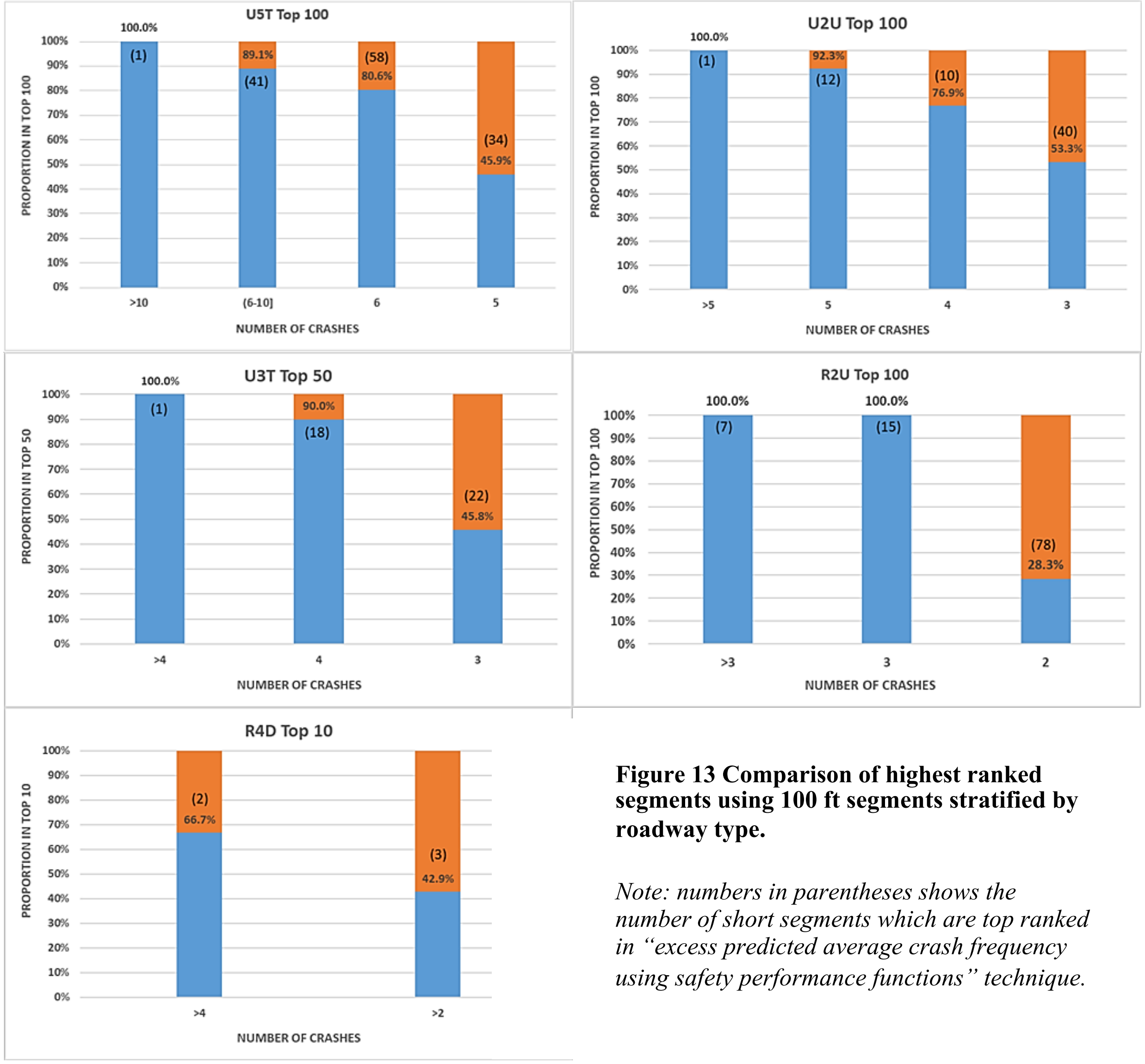


Figure 13 Comparison of highest ranked segments using 100 ft segments stratified by roadway type.

Note: numbers in parentheses shows the number of short segments which are top ranked in "excess predicted average crash frequency using safety performance functions" technique.

CONCLUSIONS

This paper focused on using short fixed length segments for statewide screening of midblock crash locations. In this analysis, segmentation is only accomplished through polygon buffers of the underlying roadways at different intervals. Thus, varying link based attributes are not necessary. Crash data is aggregated to the buffers through a spatial overlay (spatial join) operation in ArcGIS and buffer segments with the highest number of crashes can be identified. This research indicated that fixed length segments are a viable alternative to sliding window approaches. There are two benefits of using short fixed segments: 1) the GIS polygon layer consisting of the segments can be used from year to year to allow for temporal comparisons; and 2) the use of fixed segments combined with a spatial join is a much simpler screening approach than using the sliding window approach. Further, the results of the fixed segment approach are easy to interpret. A prioritized ranking of the most hazardous segment locations can be easily tabulated and can be displayed thematically on a GIS map.

The ideal segment length for identifying candidate locations for counter measures was found to be 100 ft. At this resolution, crashes can usually be associated with a single location specific characteristic such as the presence of a hazardous driveway or geometric characteristic. Longer segment lengths were found to dilute the impact of point source crash location. It is noteworthy that application of location specific countermeasures might be more cost effective than trying to do corridor length improvements.

Network screening using stratified 100 ft segments of different roadway types showed that SPF method (using excess predicted average crash frequency) and the peak search method (using short window size) reveals the similar results at the highest priority level and the later can be used as an alternative in case of insufficient data on detailed driveway and roadway characteristics.

One significant finding of this research is that short segment screening is only viable if accurately geocoded crash data is available. This is likely only possible with GIS based crash management approaches combined with high accuracy GPS data locations.

REFERENCES

1. *AASHTO, 2010. Highway safety manual, 1st Edition. American Association of State Highway and Transportation Officials, Washington, D.C.*

2. *Scopatz, R. A., & Brown, R. (2016). South Carolina Safety Data Improvements through Electronic Crash System Deployment: Roadway Safety Data and Analysis Case Study (No. FHWA-SA-16-109).*

3. *Caltrans, 2002. Table C Task Force: Summary Report of Task Force's Findings and Recommendations, California Dept. of Transportation, Sacramento.*

4. *Stokes, A., Sarasua, W. A., Huynh, N., Brown, K., Ogle, J. H., Mammadrahimli, A., ... & Chowdhury, M. (2016). Safety Analysis of Driveway Characteristics Along Major Urban Arterial Corridors in South Carolina (No. 16-6766).*

5. *Cafiso, S., Di Graziano, A., Di Silvestro, G., La Cava, G., & Persaud, B. (2010). Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables. Accident Analysis & Prevention, 42(4), 1072-1079.*

6. *Camacho-Torregrosa, F. J., Pérez-Zuriaga, A. M., Campoy-Ungría, J. M., & García-García, A.(2013). New geometric design consistency model based on operating speed profiles for road safety evaluation. Accident Analysis & Prevention, 61, 33-42.*

7. *Mitra, S., Haque, M., & King, M. J. (2017). Effects of access, geometric design, and heterogeneous traffic on safety performance of divided multilane highways in India. Journal of Transportation Safety & Security, 9(sup1), 216-235.*

8. *Stantec, and CDM Smith. South Carolina Multimodal Transportation Plan: Safety and Security. South Carolina Department of Transportation, 2013.*

9. *Garber, N. J., Haas, P. R., & Gosse, C. (2010). Development of safety performance functions for two-lane roads maintained by the Virginia Department of Transportation (No. FHWA/VTRC 10-R25).*

10. *Tegge, R. A., Jo, J. H., & Ouyang, Y. (2010). Development and application of safety performance functions for Illinois.*

11. Rajabi, M., Ogle, J. H., & Gerard, P. (2018). Highway Safety Manual Calibration: Assessing Alternate Definitions of the Calibration Factor. Transportation Research Record, 0361198118788205.

12. Cheng, W., & Washington, S. (2008). New criteria for evaluating methods of identifying hot spots. Transportation Research Record, 2083(1), 76-85.

13. Kuo, P. F., Lord, D., & Walden, T. D. (2012). Using geographical information systems to effectively organize police patrol routes by grouping hot spots of crash and crime data. Zachry Department of Civil Engineering, Texas A&M University.

14. Manepalli, U. R., Bham, G. H., & Kandada, S. (2011, September). Evaluation of hotspots identification using kernel density estimation (K) and Getis-Ord (Gi*) on I-630. In Proceedings of the 3rd international conference on road safety and simulation. Transportation Research Board of the National Academies, Washington. http:// onlinepubs. trb. org/onlinepubs/conferences/2011/RSS/2/Manepalli, UR. pdf.

15. 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), 106.

16. Shafabakhsh, G. A., Famili, A., & Bahadori, M. S. (2017). GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran. Journal of traffic and transportation engineering (English edition), 4(3), 290-299.

17. Kim, K., & Yamashita, E. Y. (2007). Using a k-means clustering algorithm to examine patterns of pedestrian involved crashes in Honolulu, Hawaii. Journal of advanced transportation, 41(1), 69-89.

18. Prasannakumar, V., Vijith, H., Charutha, R., & Geetha, N. (2011). Spatio-temporal clustering of road accidents: GIS based analysis and assessment. Procedia-Social and Behavioral Sciences, 21, 317-325.

19. Kundakci, E., & Tuydes-Yaman, H. (2014). Understanding the Distribution of Traffic Accident Hot Spots in Urban Regions (No. 14-3811).

20. Kwon, O. H., Park, M. J., Yeo, H., & Chung, K. (2013). Evaluating the performance of network screening methods for detecting high collision concentration locations on highways. Accident Analysis & Prevention, 51, 141-149.

21. Xie, Z., & Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. Computers, environment and urban systems, 32(5), 396-406.

22. *Xie, Z., & Yan, J. (2013). Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. Journal of transport geography, 31, 64-71.*

23. *Dai, D., Taquechel, E., Steward, J., & Strasser, S. (2010). The impact of built environment on pedestrian crashes and the identification of crash clusters on an urban university campus. Western Journal of Emergency Medicine, 11(3), 294.*

24. *Nie, K., Wang, Z., Du, Q., Ren, F., & Tian, Q. (2015). A network-constrained integrated method for detecting spatial cluster and risk location of traffic crash: A case study from Wuhan, China. Sustainability, 7(3), 2662-2677.*

25. *Pande, A., Abdel-Aty, M., & Das, A. (2010). A classification tree based modeling approach for segment related crashes on multilane highways. Journal of safety research, 41(5), 391-397.*

26. *Cook, Dan, Reginald Souleyrette and Justin Jackson. "Effect of Road Segmentation on Highway Safety Analysis." Transportation Research Board 90th Annual Meeting, no. 111995, 2011.*

27. *Geyer, J., Lankina, E., Chan, C., Ragland, D., Pham, T., & Sharafsaleh, A. (2008). Methods for identifying high collision concentration locations for potential safety improvements*

(California PATH Research Report UCB-ITS-PRR-2008-35). California PATH Program, Institute of Transport Studies, University of California, Berkeley, CA, USA.

28. *Green, Eric R., and Kenneth R Agent. "Evaluation of the Locations of Kentucky's Traffic Crash Data." Proc. of 3rd International Conference on Road Safety and Simulation, Indianapolis, IN. 2011.*

29. *Lu, Jinyan, Albert Gan, Kirolos Haleem and Wanyang Wu. "Clustering-based roadway segment division for the identification of high-crash locations." Journal of Transportation Safety & Security, vol. 5, no. 3, 2013, pp. 224-239.*

30. *Medury, A., & Grembek, O. (2016). Dynamic programming-based hot spot identification approach for pedestrian crashes. Accident Analysis & Prevention, 93, 198-206.*

31. *Levine, N. (2004). Crime Stat III: a spatial statistics program for the analysis of crime incident locations (version 3.0). Houston (TX): Ned Levine & Associates/Washington, DC: National Institute of Justice.*

32. *Gibin, M., Longley, P., & Atkinson, P. (2007, April). Kernel density estimation and percent volume contours in general practice catchment area analysis in urban areas. In the Proceedings of GISRUK (pp. 11-13).*

33. *Xie, Z., Yan, J. (2013, July). Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach, pp. 64-71).*

CHAPTER FOUR

Paper III: Assessing the Predictability of Short Segment Crash Analysis in the State of South Carolina

Co-authors of the paper: Wayne A. Sarasua, Afshin Famili, Jennifer H. Ogle, Devesh Kumar, Saurabh Bikram Basnet, and Emmanuel Adjei.

Paper III is in the process of being submitted to scholarly journal.

ABSTRACT

The primary objective of this research is to evaluate the predictability of a short segment peak search method with lengths of less than 0.1 miles for the statewide screening of midblock crash locations. Three different approaches, based on Highway Safety Manual Manual (HSM) Safety Performance Functions (SPFs), are used to evaluate reliability of the short segment method to identify problematic network crash sites. These approaches include, 1.) state specific SPFs, 2.) driveway SPFs (using only AADT) and 3.) driveway SPFs with adjusted crash modification factors (CMFs). Frequency based identification of short segments stratified by six different roadway types (R2U, R4D, U2U, U4D, U3T, and U5T) was compared with three SPF based screening methods to identify short segments with the highest excess predicted average crash frequency. For short segment sites with highest crash frequencies (3 for U3T, U4D and U2U; 4 for U5T, and 2 for R4D and R2U) comparisons indicated similar results (Top 90% agreement). Thus, in the event sufficient data in not available to apply SPFs, a frequency-based short segment approach provides an

effective means to identify problematic top sites exhibiting highest number of crashes. As we move down the list, the reliability of this method wanes.

INTRODUCTION

South Carolina consistently ranks among the highest rates for fatalities per VMT and fatalities per 100,000 population in the United States. Furthermore, South Carolina incurs over two billion dollars in economic loss annually due to roadway traffic crashes. In 2018, 158,448 motor vehicle crashes were recorded in South Carolina *(1)*, resulting in 977 fatalities and 38,393 injuries. There is considerable body of research and guidance on network screening methods for identifying and prioritizing sites with potential for safety improvement, however much of the focus is placed on reducing intersection-related crashes. While the largest proportion of all crashes in 2018 occur at intersections (51%), more than half of the fatal crashes (62%) occur along midblock highway sections *(1)*.

In 2008, the South Carolina Department of Public Safety (SCDPS) embarked upon a major initiative to improve crash data quality through implementation of an automated crash data collection system entitled the South Carolina Collision and Ticket Tracking System (SCCATTS) to be used by law enforcement agencies *(2)*. One of the principal objectives of SCCATTS is to obtain more timely, accurate, and complete crash and citation data, resulting in improved quality in law enforcement crash data. Deployment of the system began in 2010, and as of April 2013, all SC highway patrol vehicles were equipped with SCCATTS *(2)*. SCCATTS has resulted in dramatically improvement in geocoded crash positioning data. Currently, South Carolina collects 80 percent of its crash reports electronically, and this effort will continue to reduce submission of paper crash reports on an incremental basis with the ultimate goal of eventual elimination *(3)*. With the benefit of obtaining precisely geocoded locations of midblock crashes, a new opportunity emerged in network safety analysis to identify short road segments with a high incidence of crashes.

Network screening methods for midblock crashes range from simplistic approaches such as calculating crash rates to advanced methods using Empirical Bayes (EB) to determine the excess predicted average crash frequency using safety performance functions (SPFs). The latter is the approach used by SafetyAnalyst – an AASHTO software tool developed to support the safety management process within state agencies. A SPF consists of two primary elements: (1) Estimates of the mean of the expected number of crashes (μ) of each unit (road segments, intersections, grade crossings, etc.) in each population and the standard deviation of this estimate; (2) the standard deviation of the μ's in each population, which considers the diversity of within the population unit *(4)*. For SPF application, one has to include important population defining traits (variables including traffic, geometry, operation, etc.) for practical applications *(4)*.

Although traditional methods using crash frequencies needs minimal requirements for input data, values may be biased to regression-to-the-mean (RTM). This bias may result in overestimating the need for a countermeasure at a location with a high amount of crashes experiencing extreme random events within a single year due to random fluctuations in crash occurrences *(5)*. While Empirical Bayes method focuses on use of expected average crash frequencies for both the before and after periods to address RTM bias, use of longer segments may "hide" the severity of a single location, if the remaining portion of the segment has few or no additional crashes. The use of short segments in roadway corridor safety analysis can potentially resolve this issue by identifying locations with a concentration of midblock crashes that may be related to undesirable characteristics of a specific site. Furthermore, application of fixed length short segments may have additional benefits over more complex spatial statistical approaches (e.g. moving windows) within state DOT agencies, due to relative ease of implementation *(6)*. However, adoption of segmentation definition parameters for this method may need to be revised to reflect common roadway geometry characteristics.

This research focuses on evaluating the predictability of a short segment peak search method with lengths of less than 0.1 miles for a statewide network screening of midblock crash locations. Short segment buffers are created based on the underlying road network to create a GIS layer of short segment polygons. Specifically, geocoded crash data is overlaid based on superimposing crash locations onto the short segment polygon projections and aggregated for further analysis through the use of roadway classifications and application of network screening criteria *(6)*. The primary objective of this research is to perform test of applications investigating the effectiveness of the short segment approach for network safety analysis screening, using existing HSM SPF methods to demonstrate methodology feasibility and to document research results.

LITERATURE REVIEW

Numerous hotspot identification procedures or network screening techniques are used by agencies responsible for highway safety to identify and prioritize problematic crash locations for potential safety improvement. Some of the most common methods include “sliding moving window”, “Peak Searching”, “Continuous Risk Profile”, and “Latent Class Clustering”. The literature review summarizes previous research and studies conducted on evaluation of peak search methods using safety performance functions, the choice of segmentation length, and examples of segmentation studies conducted in South Carolina.

Evaluation of Network Segmentation with Other Safety Methods

The use of segmentation of roadways has been applied successfully to effectively address safety analysis criteria with regard to producing meaningful results. Casifo et al. *(7)* compared the peak search segmentation method based on five different safety traits (traffic volume, radius of curvature, vertical gradient, type of section, roadside attributes) and used goodness of fit of the SPF to evaluate each approach. Their evaluation determined that a fixed length segment with two tangents and two curves resulted in the best fitting SPF. They also used a fixed length segmentation technique by dividing the roads sample into segments in which all the highway characteristics (exposure, geometric, consistency and context-related variables) were constant and used to establish a minimum length of significant for accident expectation.

Kwon et al. *(8)* evaluated the performances of three different segment analysis methods (Sliding Moving Window (SMW), Peak Searching (PS) and Continuous Risk Profile (CRP)) to analyze freeways. They used similar input requirements for each of the three methods in conducting the evaluation. Each of these methods was used to prioritize the detected sites for safety investigation, and the lists were compared with previously confirmed hotspots. The length of segments defined using these approaches varied from 0.04 to 3.64 miles. The study concluded that the Continuous Risk Profile (CRP) screening method out-performed the Sliding Moving Window and 13 Peak Searching methods. Qin and Wellner *(9)* conducted research concluding that a sliding window (variable) method provides more reliable predictive results than use of fixed length. They explained the impact of segmentation technique on traffic safety with the prevalence of Empirical Bayes (EB) methods.

The Impact of Segment Length in Peak Search Method

For application of the peak search method, the roadway network is divided into equal segment lengths, after which each segment is examined using one of the safety analysis methods. Several studies *(10, 11, 12)* have shown that the choice of segment length significantly affects the consistency in identifying high-crash locations, which adversely influences reliably estimating safety analysis outcomes. Results from studies such as this have led to wide-ranging professional discussions and debates within highway safety analysis fields on criteria and outcome thresholds for selecting and applying optimal roadway segmentation lengths for network-based midblock crash analysis.

Conversely, findings from similar safety analysis studies *(9, 13, 14)* have shown that using short segment lengths exhibit tendencies of producing undesirable results of high crash variation leading to uncertainties in SPF performance. However, it should be noted for the purpose of this research, only a small portion of midblock segments may require improvement, while choosing long segmentation intervals may be economically impractical. Additionally, a number of common safety countermeasures are likely to only be feasible if applied over relatively short distances, such as an increased turn radii for driveways/intersections, shielding for protection from roadside hazards, or a host of other safety countermeasures due to associated absorbent cost of improving longer segments.

Examples of Segmentation Studies in South Carolina

Beginning in 2010, South Carolina has made great strides in improving crash data quality within the state through the systematic implementation and roll-out of South Carolina Collision and Ticket Tracking System (SCCATTS). Deployment of the system began in 2010 and as of April 2013, all SC highway patrol vehicles have been equipped with SCCATTS *(1)*. The primary goal of this research is to examine how precisely geocoded crash locations in SCCATS can potentially be used to screen, investigate and evaluate safety analysis outcomes based on use of short segments.

Famili et al. *(6)* conducted an evaluation of evaluation midblock safety based on fixed-short segment length (less than 0.1 mile) using 2016, 2017 and 2018 crash data in South Carolina. A premise of the research was focused on address the occurrence of midblock crashes clustered along a very short segment length due to undesirable characteristics of a specific site within a predetermined longer segment length. Researchers used the underling road network to create a layer of segment polygons using GIS buffering. Findings supported use of 100-foot length polygons as an ideal length to best identify individual candidate locations for application of safety countermeasures, which can be associated with specific physical problematic roadway characteristics such as the presence of hazardous driveways or undesirable geometric characteristics.

Rajabi et al. *(15)* collected and compiled a comprehensive data set needed to calibrate each of the 18 SPFs, identified in the HSM, for the state of South Carolina. The study developed a robust database and calibration factors for all roadway segment and intersection combinations commonly occurring throughout the state. State geography is best described by three distinct subregions, coastal, midlands, and upstate, with each exhibiting unique terrain, weather conditions, and traffic patterns; specific SPF calibration factors were developed for each of the three subregions.

Another study conducted for SCDOT *(16)* created an estimation of CMFs determined directly from coefficients developed from a negative binomial model, based on a sample size of 3,774 driveways. The method used for developing CMFs was based on multiple studies referenced in a US DOT/FHWA study reviewing methods for developing CMFs including overview of procedures, sample size considerations and strengths and weaknesses of each approach (*17)*. Variables considered for use in developing CMFs include driveway spacing, driveway class, roadway AADT, driveway access, number of entry lanes, driveway width, and corridor speed limit.

Literature Review Summary

As identified in the literature, use of short-length segments need to address safety analysis concerns that results from the approach may be adversely affected by inconsistent results for high crash locations. Conversely, use longer segments may also be economically impractical to use due to a host of practical considerations pertaining to implementation of cost effective and targeted site-specific safety countermeasures. With an overarching objective of finding an optimal medium between these two methodological limitations, this research provides a framework to evaluate and investigate the predictability of using a short segment approach based on HSM SPF prediction method as the basis for identifying network-level midblock crash locations exceeding specified threshold criteria.

METHODOLOGY

This section provides a comparative overview of the frequency-based identification approaches using short segments with three SPF-based screening procedures for evaluating application of a short segment methodology for conducting network-level safety analysis. Approaches include, state specific SPFs, Driveway SPF (using AADT data only) and driveway SPFs with adjusted CMFs (for specific driveway characteristics), all methodologies of which are adopted for this use in conducting research and described individually. **Figure 1** provides a visual conceptual representation steps in the procedures identified in a flowchart of the study.

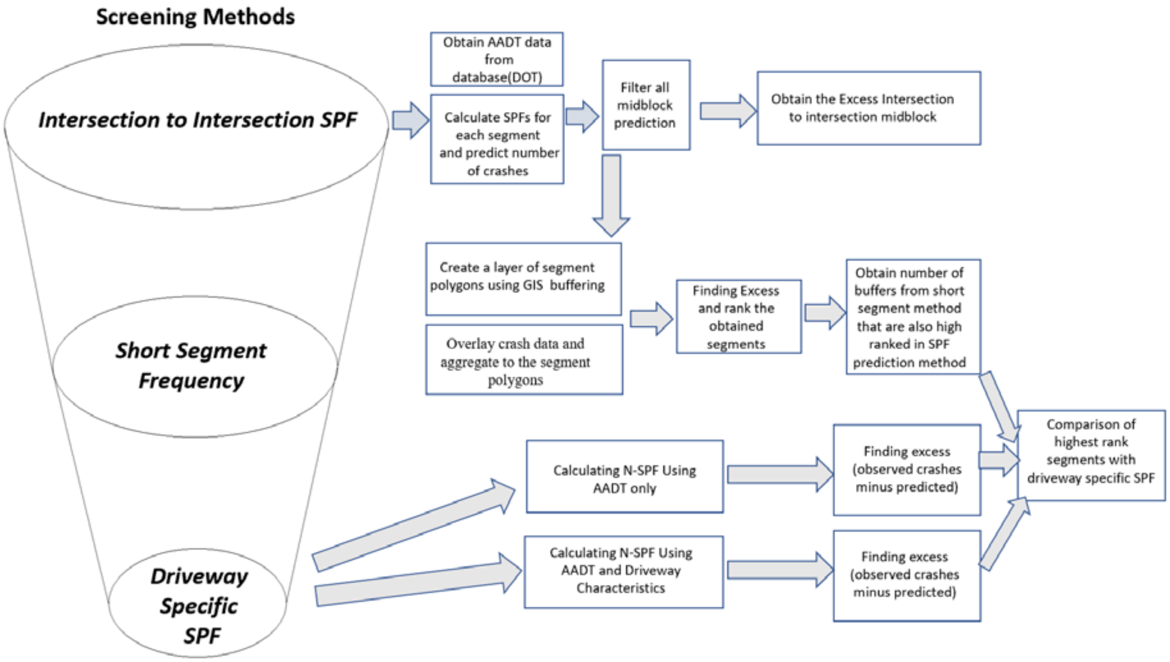


Figure 1: Flowchart of the study

The AASHTO Highway Safety Manual (HSM) recommends segmenting roadways/highways based on their geometric characteristics to generate homogenous segment lengths. HSM Part C includes the division of roadway segment models based on geometric characteristics such as the number of lanes, lane width, median etc. Segments are classified and identified by road type, number of lanes and median type. The first character represents the type of land use, U for urban areas and R for rural areas. A second character is a number that represent the total number of lanes in both directions and the last character is the median type for the specific roadway ‘D' for divided medians, ‘U' for undivided median and ‘T' for two way left turn lane. For this research, the roadways were also divided as suggested in the HSM. Based on the state DOT’s classification system in South Carolina, roadways are divided into six distinct categories summarized in **Table 1**.

Table 1: Roadway Types by Definitions

Roadway type	***Definition***	***No. of lanes***	***Median***	***Mileage***
R2U	Rural	2	Undivided	16,055
R4D	Rural	4	Divided	829
U2U	Urban	2	Undivided	8,761
U4D	Urban	4	Divided	357
U3T	Urban	3	TWLTL	278
U5T	Urban	5	TWLTL	795

Three years of crash data (2016, 2017 and 2018) was used as the basis for conducting the analysis. Prior to applying any of these network-screening methods, it is essential to create buffers and identify driveway locations. All midblock crashes within a 150- foot radius of intersections were eliminated for this research in GIS platform (See **Figure 2**). Fixed polygon buffers start at the beginning point of network and continue along the entire route until the ending point of network roadway. Researchers collected and entered data describing driveway characteristics such as width, access type, class, etc. that was used as the basis for screening methods included analysis.

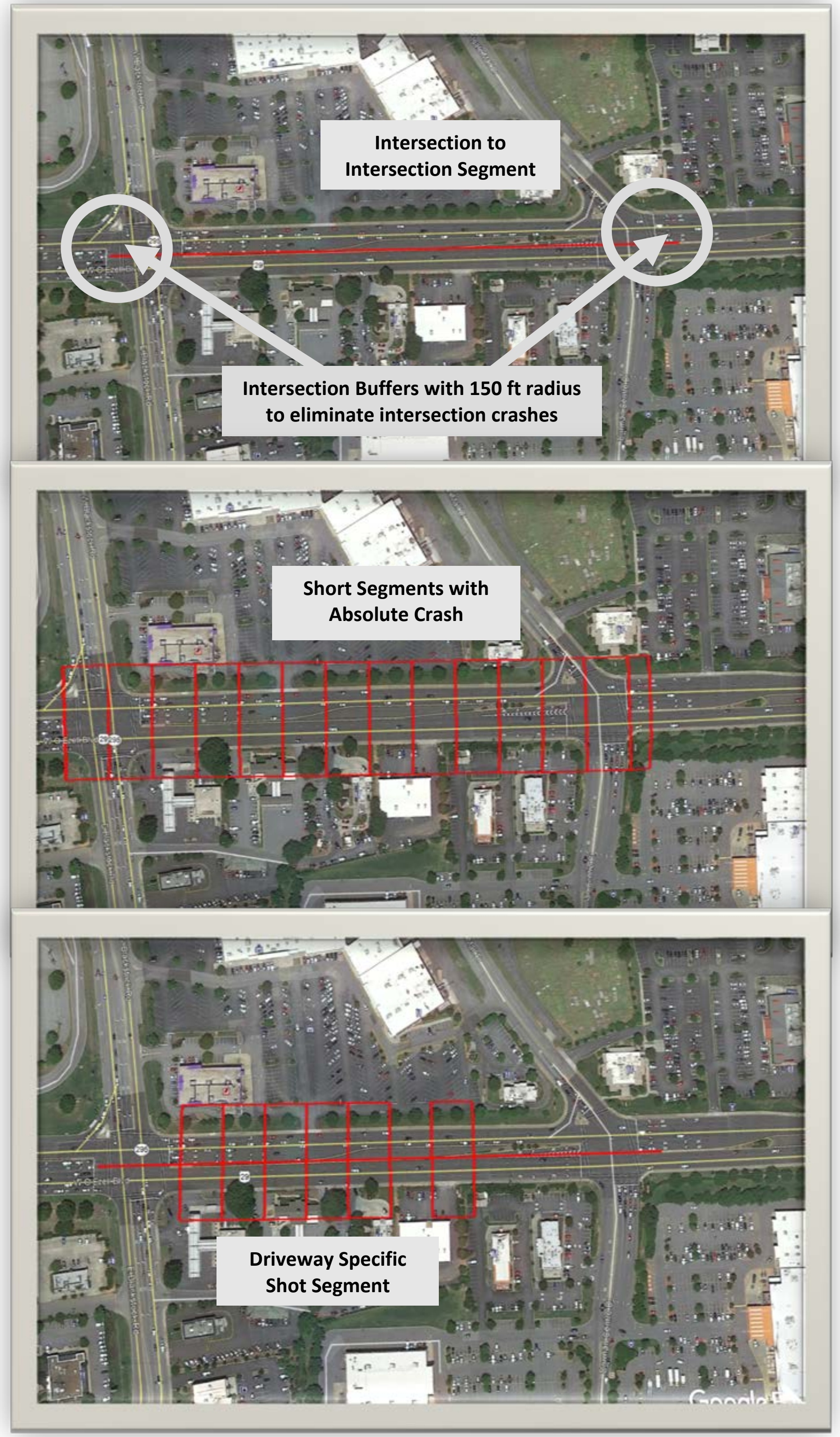


Figure 2: Comparison of 3 methods

Short Segment Network screening method

Uniform short segments of 100-feet in length were established using specialized buffer routines in ArcGIS. Depending on roadway type, buffer widths of either 75-feet or 125-feet were used in the transverse direction to encompass the entire road right-of-way (ROW) and roadside limits. Researchers initially used a variable buffer based on roadway width obtained from the SCDOT's Roadway Information Management System (RIMS), however, systematic positional issues of the roadway centerline resulted in portions of the actual roadway not being included in the buffers. Thus, larger buffer widths were tested. After several trial and error investigations, a 75-feet and 125-feet width buffered on either side of the centerline were determined to provide the best fit for roadway group types of R2U, U2U, U3T and U5T, U4D, R4D. These adopted buffer widths would capture all the crashes occurring along any particular segment in question. These roadway buffers were established using GIS using the basis of fixed 100-foot length. The 100-foot segment length was selected based on reflecting delineation results from Rajabi et al. *(15)* asserting that the influence area of a driveway is most nearly equal to the driveway width, plus a car length before and after the driveway. After the buffer creation process, crashes were overlaid for each year (2016, 2017, 2018) individually to identify high-risk locations. To ensure that the crashes influenced by the intersections were not included in the analysis, crashes occurring within 150-foot radius of an intersection were eliminated from the database. The 150-ft segment length was used after discussions with SCDOT to best reflect representative intersection influence areas *(17)*.

The total network length included in the database and assessed in this short segment network screening approach was 41,282 miles of roadways. In overlaying crashes with short segment buffers, each segment was ranked based on the absolute number of crashes occuring within each preestablished 100-foot long segment buffer. Through application of this method researchers were able to effectively identify the most critical midblock crash segment locations extracted for an expansive statewide network, in a relatively short period of time.

South Carolina Calibrated SPFs

Statewide calibrated SPFs developed by Rajabi et al. *(15)* were used as the basis for predicting average crash frequency. It is important to note that SPF's are state-specific and site-specific for locations sharing similar characteristics. A site's observed average crash frequency is compared to a predicted average crash frequency using an SPF. If the difference between the observed and predicted was greater than zero, then it was concluded that a site experiences more crashes than predicted, indicating the need for consideration as a candidate location for further study. Candidate sites identified using this screening method represent locations experiencing observed crash frequency exceeding predicted crash estimated values. To appropriately apply any of these identified screening methods, the road network and accompanying crash database needs to be subdivided into predetermined midblock segments, and then further grouped into descriptive reference populations based on selected roadway geometric characteristics (e.g., rural two-lane highways, or urban 4-lane with TWLTL). **Table 2** represents site-specific coefficient values for each specific segment facility types that were obtained from *(15)* and used to calculate the predicted number of crashes.

Table 2: Shows the Different Coefficient Values Used for Each Roadway Types

Facility Type	*Variable*	*Estimate*	*p-value*
R2U	AADT	0.6441	<0.00
R4U	AADT	1.3841	<0.00
U2U	AADT	0.5612	<0.00
U3T	AADT	2.7995	<0.00
U4U	AADT	1.2514	<0.00
U4D	AADT	0.979	<0.00
U5T	AADT	0.8943	<0.00

Average predicted crash frequencies calculated for each segment were determined based on equation 1.

$$N_{spf} = e^{\widehat{\beta}_0+\widehat{\beta}_1\times\ln(AADT)+\ln(L)+\sum_{i=1}^{n}\widehat{\beta}_i(X_i-X_{b_i})} \qquad (1)$$

Where:

L= Segment Length

AADT = Average annual daily traffic

$\hat{\beta}_0, \hat{\beta}_1$: Coefficients of regression

Predicted Crash for each driveway Based on segment AADT

A separate ranking of buffers was performed based on predicted crashes for each driveway using segment Average Annual Daily Traffic (AADT) data (see equation 2). The predicted SPF was calculated based on findings identified in the study conducted by Rajabi et al. *(15)*. The purpose of this analysis was to create a comparison for predicted driveway crashes within the context of the short segment method. Predicted SPFs for each driveway were determined based on roadway segment AADT values. Within each reference population, driveways were further disaggregated into homogeneous subdivisions based on traffic volume factors. The results from this safety analysis model approach can also be directly compared with the State Specific SPFs as both are based upon the same parameter (roadway segment AADT). Furthermore, results of this analysis would also be useful in determining the effectiveness of traffic volume for estimating predicted crash frequency. This method allows researchers to assess if predicted crashes for the driveway, using only AADT, function as a deterministic factor for predicted crashes.

$$N_{spf} = e^{(-16.52+1.668*\mathrm{Log}\ (\ [AADT]))} \quad (2)$$

Predicted Crash for each Driveway Based on Driveway Characteristics

Another analysis approach used to predict crash frequency was determined based on CMFs that include driveway characteristics. Multiple studies support the use of this method as an effective means for conducting network-based safety analysis *(14, 15)*. The CMF for this approach accounts for AADT, driveway class, and driveway access. These safety-related characteristics were collected for driveways exhibiting high crash frequencies. CMFs for continuous variables are estimated using equation 2 *(16)*.

$$CMF_j = e^{\left(\beta_j \times \left[x_j - y_j\right]\right)} \quad (3)$$

Where:

x_j = range of values or a specific value investigated (e.g., lane width, shoulder width, etc.) for CMF_j;

y_j = baseline conditions or average conditions for the variable x_j (when needed or available); and

β_j = regression coefficient associated with the variable j.

The CMF model is based on the assumption that each variable is independent and therefore, not influenced by other values. The method also assumes the relationship between the change in the variable value and change in crash frequency is exponential (as a negative binomial model). **Table 3** summarizes derived crash modification functions (CMF) for driveway class and access within indicated confidence interval limits *(16)*.

Driveway access and class were collected for high ranked crash frequency buffer locations as identified based on application of the short segment method. Predicted crash frequency for each driveway, based on segment AADT, was adjusted based on CMFs. A segment's observed average crash frequency was compared to a predicted average crash frequency using an SPF. If the difference between observed and predicted crashes was greater than zero, then it was concluded that a driveway experiences more crashes than predicted, indicating the need for consideration as a candidate location for further study.

Table 3: Crash Modification Factors *(14)*

Variables	***CMF***	***95% Confidence Bounds***	
Median (1 for raised, 0 for all others)	0.49	0.00	1.12
D_Class4	2.12	1.94	2.30
D_Class5	2.31	2.01	2.60
D_Control	4.08	3.73	4.44
RIRO	1.26	0.37	1.72

RESULTS AND DISCUSSION

Validation of short segment method using South Carolina Calibrated SPFs

Midblock segmentation from intersection to intersection have been created throughout the state for six different roadway type. Segments were established using statewide roadway network information from the SCDOT RIMS dataset that includes AADT and SPFs using AADT for segments calculated based on data obtained from *(15)*. Excess predicted average crash frequency using SPFs was compared against absolute crash frequencies observed for each segment. Using the excess prediction method, a segment's observed average crash frequency was compared to a predicted average crash frequency using an SPF. If the difference between observed and predicted was greater than zero, a segment was determined to experience more crashes than predicted and indicating the need for consideration as a candidate location for further study. Reasons supporting use of approach as an effective method include: more accurately calculating the potential for safety improvement and acknowledging the complex, non-linear relationship between crash frequency and volume *(6)*. **Table 4** provides a comparison of highest ranked 100-foot short segments based on excess predicted and absolute crash frequencies stratified with regard to roadway type. A segment length of 100-feet with selected 2, 3 and 4 or more crashes for different roadway types were determined based on threshold criteria identified by SCDOT as practical values for network screening purposes. **Figure 3** presents U5T top short segments for crash data from 2017. Using a frequency of four observed crashes as the screening criteria for evaluation of U5T short segments, 358 of the locations of concern were identified through use of this method as candidate locations for further study.

Additional top crash segments for other roadway types, similarly identified through application of this screening method are compiled in **Appendix B**.

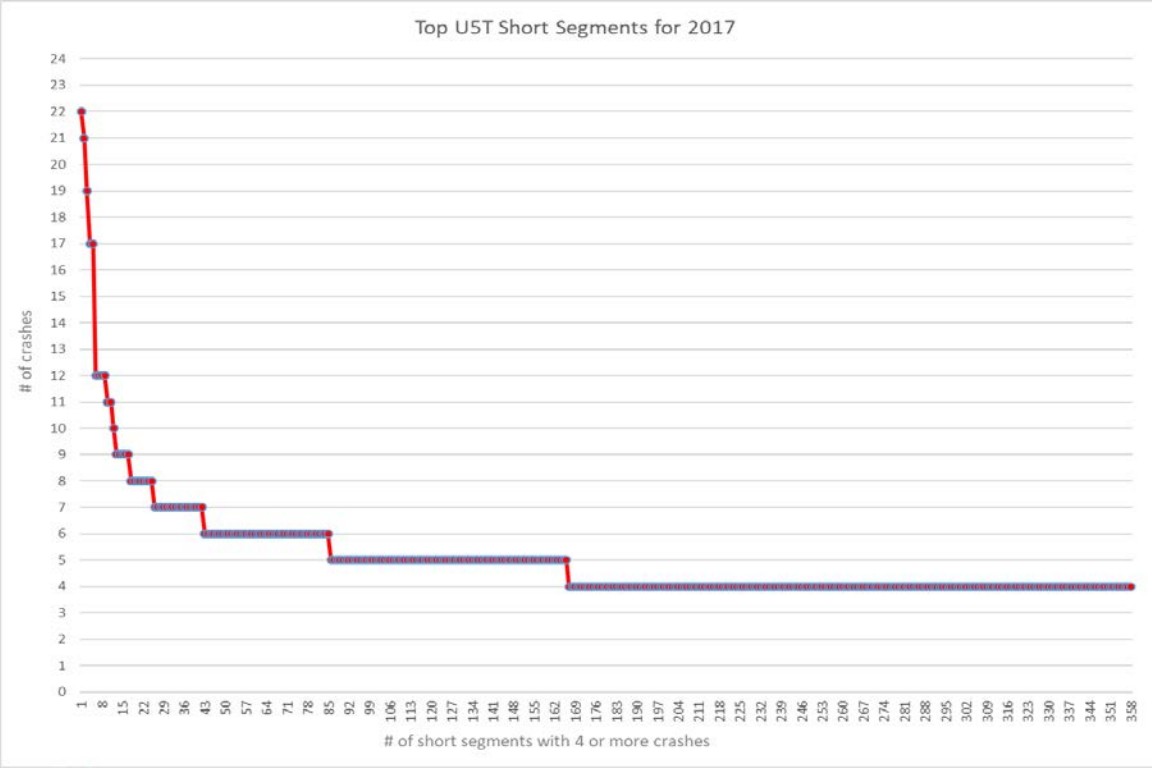


Figure 3: Top U5T Short Segments for 2017

Furthermore, results indicate that use of the short-segment absolute frequency method obtained 98.5%, 99% and 99.3% of site screening for 2018, 2017 and 2016 crash data, respectively. One limitation of this method exists for U3T for which screening methods captured a lower percentage of short buffers that were occurring as indicated from crash data within the excess intersection to intersection segments.

Table 4: Validation of Short Segment Method Using Roadway Predicted SPF

Roadway Type	Total Int to Int Segments	Year	Actual > Predicted SPF	Excess Percentage	100' Short Segments	Segments within excess	# Crashes considered in Short Segments	Segments within excess (%)
U5T	9430	2018	2896	31%	392	392	≥4	100%
		2017	2572	27%	358	358		100%
		2016	2501	27%	329	329		100%
R2U	62376	2018	7613	12%	316	316	≥2	100%
		2017	7443	12%	337	337		100%
		2016	7634	12%	312	312		100%
U3T	3832	2018	711	19%	62	57	≥3	92%
		2017	690	18%	79	74		94%
		2016	680	18%	79	76		96%
U4D	3587	2018	1002	28%	125	124	≥3	99%
		2017	945	26%	191	191		100%
		2016	903	25%	182	182		100%
R4D	3847	2018	902	23%	93	93	≥2	100%
		2017	896	23%	144	144		100%
		2016	886	23%	147	147		100%
U2U	82214	2018	6055	7 %	208	208	≥3	100%
		2017	5901	7%	204	204		100%
		2016	5715	7%	191	191		100%

Validation of short segment method using Driveway SPF (considering only AADT)

SPFs for driveways calculated based on AADT only AADT data, obtained from *(16)*, driveway characteristics were omitted from the screening, so results could be directly compared with results from the previously described method using short segments calibrated from SPFs. Furthermore, results reflect the effect of driveway characteristics through inclusion of "Driveway class" and "Driveway access" on SPF. Similar to the previous described short segment SPF method, "Excess predicted average crash frequency using driveway SPFs" method was compared against absolute crash frequencies observed for each segment. **Table 5** represents the comparison of highest ranked 100-foot segments based on excess driveway SPF predicted and absolute crash frequencies stratified by roadway type. Results indicate that use of the short-segment absolute frequency method can obtain 99%, 98.5% and 98.5% of screening sites for 2018, 2017 and 2016 crash data, respectively. Considering top ranked location identified from application of both screening methods, driveway SPF produces higher values, than short segment SPF for the same locations.

Table 5: Validation of Short Segment Method Using Driveway SPF (Including Only AADT)

Roadway Type	Year	Total # of Short Segment Buffers	# of Short Segment in Top 100 SPF Excess	% of Short Segments in Top 100 SPF Excess	# of Short Segment in Excess DW AADT SPF	% of Short Segment in Excess DW AADT SPF	# of Top SPF AADT Compared	# of Short Segment in Top SPF AADT	% of Short Segments in Top SPF AADT
R2U	2018	316	97	31%	312	99%	100	88	28%
	2017	337	90	27%	336	99%		84	25%
	2016	312	78	25%	309	99%		78	25%
U2U	2018	208	112	54%	204	98%	100	107	51%
	2017	204	96	47%	202	99%		92	45%
	2016	191	106	56%	188	98%		103	54%
U5T	2018	392	198	51%	391	99%	100	193	49%
	2017	358	186	52%	353	99%		176	49%
	2016	329	170	52%	327	99%		163	50%
U3T	2018	62	45	73%	61	99%	100	52	84%
	2017	79	62	79%	78	99%		65	82%
	2016	79	65	82%	79	100%		71	90%
U4D	2018	141	89	63%	140	99%	100	81	58%
	2017	135	99	73%	130	99%		81	60%
	2016	182	128	70%	173	95%		110	60%
R4D	2018	93	67	72%	93	100%	30	61	66%
	2017	74	53	72%	73	99%		49	66%
	2016	85	62	73%	85	100%		59	69%

Validation of short segment method using Driveway SPF (considering driveway characteristic)

Since most DOTs do not collect detailed data regarding driveway characteristics on a statewide basis, identifying an effective alternative to use of SPFs can be cost effective and can save time in data processing. Driveway class and access were collected manually for the high ranked buffer locations as defined form the short segment approach (**Table 4** summarizes the number of crashes considered in short segment method). SPFs and CMFs for driveways were calculated considering segment AADT data only as obtained from *(16)*. Similar to previously described methods, a site's observed average crash frequency is compared to a predicted average crash frequency using an SPF. The short segment buffers with the “observed value more than predicted” were used as the basis for comparison with short segment method. **Figure 4** summarizes a comparison of highest ranked 100-foot segments based on excess predicted and absolute crash frequencies, stratified by roadway type. Results indicate that short-segment absolute frequency method can fairly accurately identify more than 60% of sites compared to the robust approach of using excess predicted average crash frequency. Lowest matches occur for lower ranges of crash frequencies.

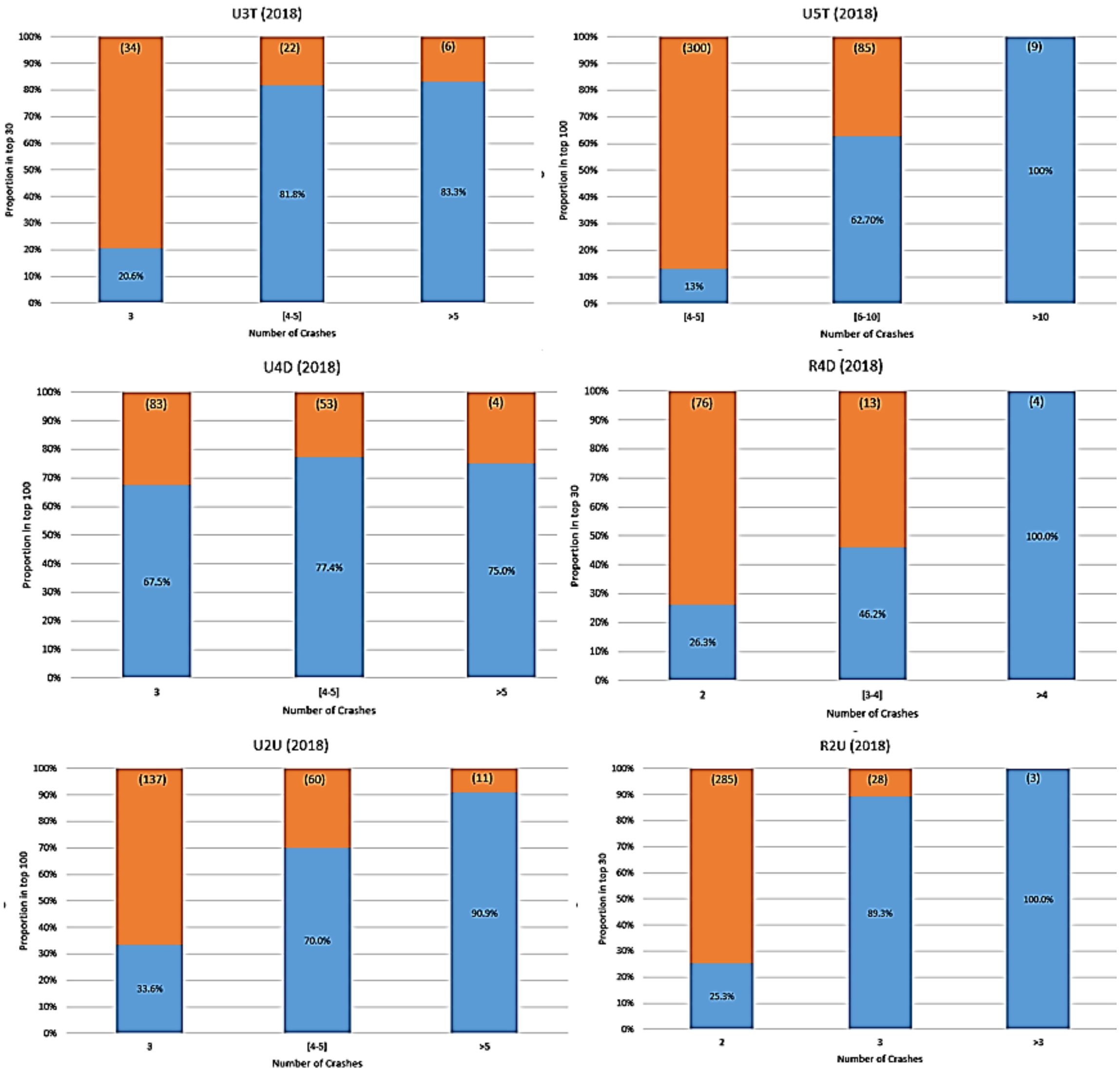


Figure 4: Comparison of year 2018 highest ranked segments using 100’ segments stratified by roadway type with driveway SPF (*Blue: percentages matched with driveway SPF)*

CONCLUSIONS

This research focused on validation of fixed short length segments for statewide network screening of midblock crash locations using three different HSM recommended approaches. Methods include roadway SPF, driveway SPF considering only AADT, and driveway SPF using driveway characteristics.

Network screening was based on preestablished stratified 100-foot segment lengths used along different roadway types. Research findings indicated that in comparing three SPF methods (using excess predicted average crash frequency) and the peak search method (using short window size) representative and comparable results are achievable at the highest priority level, and the later can be used as the basis for effective safety alternative analysis, in the event of insufficient data on detailed driveway and roadway characteristics. Since most state DOTs do not collect detailed driveway characteristic, identifying an alternate means to serve as the basis of SPFs application provides safety focused agencies with an approach offering cost effective results and time savings in data processing.

REFERENCES

1. *Book, S. C. T. C. F. (2018). South Carolina Department of Public Safety. Blythewood, South Carolina.*

2. *Stantec, and CDM Smith. South Carolina Multimodal Transportation Plan: Safety and Security. South Carolina Department of Transportation, 2013.*

3. *Scopatz, R. A., & Brown, R. (2016). South Carolina Safety Data Improvements through Electronic Crash System Deployment: Roadway Safety Data and Analysis Case Study(No. FHWA-SA-16-109).*

4. *Hauer, E. (2015). The art of regression modeling in road safety (Vol. 38). New York: Springer.*
5. *Loudon, W., & Schulte, R. J. (2016). Evaluating Performance of Highway Safety Projects (No. FHWA-ID-16-257). Idaho. Transportation Department.*

6. *Famili, A., Sarasua, W., Iqbal, A. M., Kumar, D., & Ogle, J. H. (2019). Short Segment Statewide Screening of Midblock Crashes in South Carolina. Transportation Research Record, 0361198119851068.*

7. *Cafiso, Salvatore, Alessandro Di Graziano, Giacomo Di Silvestro, and Grazia La Cava. "Safety Performance Indicators for Local Rural Roads: Comprehensive Procedure from Low-Cost Data Survey to Accident Prediction Model." Transportation Research Board 87th Annual Meeting, No. 08-2542, 2008.*

8. *Kwon, O. H., Park, M. J., Yeo, H., & Chung, K. (2013). Evaluating the performance of network screening methods for detecting high collision concentration locations on highways. Accident Analysis & Prevention, 51, 141-149.*

9. *Qin, Xiao and Adam Wellner. "Segment length impact on highway safety screening analysis." Transportation Research Record: Journal of the Transportation Research Board. 2011, No. 12-0644. 2012.*

10. *Geyer, Judy, Elena Lankina, Ching Yao Chan, David Ragland, Trinh Pham and Ashkan Sharafsale. "Methods for identifying high collision concentration locations for potential safety improvements." California PATH Program, Institute of Transportation Studies, University of California at Berkeley, 2008.*

11. *Cook, Dan, Reginald Souleyrette and Justin Jackson. "Effect of Road Segmentation on Highway Safety Analysis." Transportation Research Board 90th Annual Meeting, no. 11- 1995, 2011.*

12. *Lu, Jinyan, Albert Gan, Kirolos Haleem and Wanyang Wu. "Clustering-based roadway segment division for the identification of high-crash locations." Journal of Transportation Safety & Security, vol. 5, no. 3, 2013, pp. 224-239.*

13. *Srinivasan, Bhagwant, Craig Lyon, Bhagwant Persaud, Carol Martell and Jongdae Baek. “Methods for Identifying High Collision Concentration Locations (HCCL) for Potential Safety Improvements: Phase II, Evaluation of Alternative Methods for Identifying HCCL.” California Department of Transportation, 2011.*
14. *Ogle, J H, P Alluri, and W A Sarasua. 2011. "Model Minimum Uniform Crash Criteria and Minimum Inventory Roadway Elements: Role of Segmentation in Safety Analysis" Transportation Research Board 90th Annual Meeting.*
15. *Rajabi, M., Ogle, J. H., & Gerard, P. (2018). Highway Safety Manual Calibration: Assessing Alternate Definitions of the Calibration Factor. Transportation Research Record, 2672(30), 99-108.*

16. *Sarasua, W. A., Davis, W. J., Ogle, J., Huynh, N. N., & Chowdhury, M. (2015). Support for the development and implementation of an access management program through research and analysis of collision data (No. FHWA-SC-15-02). South Carolina. Dept. of Transportation.*

17. *Gross, F., Persaud, B., and C. Lyon. A Guide to Developing Quality Crash Modification Factor. Report No. FHWA-SA-10-032. USDOT, FHWA, 2010.*

CHAPTER FIVE

CONCLUSION

As alluded in Chapter One, this research focuses on evaluating how crash geocoding has improved over the years and how this enhanced spatial accuracy of crashes can potentially lead to a new paradigm on midblock crash safety analysis.

There were three main objectives established and achieved over three research papers in this dissertation that helped to reach the research objectives. The three research papers presented in this dissertation covered several geospatial analysis methods and HSM methods that could be used in various stages in the analysis process.

A survey of state highway agencies conducted as part of this research and discussed in the first paper indicates that there are disparate crash reporting systems used across the country from a crash geocoding standpoint. The survey indicates that not only does the geocoded accuracy of crash locations vary by state, the accuracy can even vary by jurisdiction within each state. Accurate crash data is essential for robust safety analysis. The GIS spatial analyses and case studies described in the first paper gave biased results if the geocoded crash data is poorly located.

The analysis along five major corridors using crash data geocoded with handheld GPS (2010 data) showed that only 35% of the crash locations (not including run-off-the-road and fixed object crashes) geocoded inside the travel way while the SCCATTS crash data indicated that the proportion of crashes occurring within the travel way is nearly 100% for the same corridors. The first paper concluded that any safety analysis can only be as good as the data being used. The ability to collect spatially accurate crash data will enhance a state transportation agency's ability to conduct reliable safety analysis as well as foster other transportation related research resulting in more effective safety programs and policies.

The second paper focused on using short fixed length segments for statewide screening of midblock crash locations. In this analysis, segmentation is accomplished through polygon buffers of the underlying roadways at different intervals. Crash data is aggregated to the buffers through a spatial overlay (spatial join) operation in ArcGIS and buffer segments with the highest number of crashes can be identified. In this approach, whole South Carolina's road network (41,282 miles) was assessed. The initial use of 50 ft segment buffers produced 274 segments with 4 or more crashes using 2016 data. This turns out to be 0.000062% of the total midblock 50 ft segments. Our segment length analysis indicated that 100 ft would be an ideal segment length for short fixed segment analysis. This was deemed manageable to look at on a case by case basis by SCDOT. The research clearly indicates that the use of segments of 0.1 miles (or greater) in many instances "hides" the severity of a single location if the rest of the segment has few or no additional crashes. Different analysis methods were used to look at short segments including a peak search method with fixed segments and a sliding window approach (Network Kernel Density). Based on the analysis, the short segment peak search method is recommended for use by state agencies as a network screening approach because it is much less complex to implement than the sliding window approach and locations can be easily ranked. Because the segmentation can be the same from year to year, direct comparisons can be made over time while spatial integrity is maintained. One significant finding of this research is that short segment screening is only viable if accurately geocoded crash data is available. This is likely only possible with GIS based crash management approaches combined with high accuracy GPS data locations.

The third paper focused on validation of short fixed length segments for statewide screening of midblock crash locations using three different HSM recommended approaches. These methods include roadway SPF, driveway SPF considering only AADT and driveway SPF using its characteristics. Network screening using stratified 100' segments of different roadway types showed that comparing three SPF methods (using excess predicted average crash frequency) and the peak search method (using short window size) reveals similar results at the highest priority level and the later can be used as an alternative in case of insufficient data on detailed driveway and roadway characteristics. Since most state DOTs do not collect detailed driveway characteristics, finding an alternative to use of SPFs can be cost effective and can save time in data processing.

Overall, the research in this dissertation document has added to the body of knowledge in the field of transportation safety and geospatial science. The research also applied innovative GIS analysis methods that have not been used before in safety analysis based on the literature review. These include the use of variable buffers to generate roadways from centerlines and pavement width attributes; creation of driveway buffers to determine the influence area of driveways; creation of short segments buffers for use in midblock network screening; the use of spatial join and overlay operations to aggregate crashes to the various buffers; and applying NKDE using a shorter window size than what has been used in previous research.

This research has shown that improvements in crash geocoding can enhance safety analysis. By using short segment network screening, segments of high crash incidence can be identified and displayed with ovelayed crashes at their actual crash locations which can minimize the need for developing collision diagrams. Further, one of the findings of the research is that the current intersection to intersection process aggregates crashes to long segments which can mask the crash severity of point locations.

Specifically, this research could help guide state officials to make decisions with regard to selecting and implementing transportation safety programs and strategies for the safety emphasis areas in South Carolina's current strategic highway safety plan, 'Target Zero' at midblocks. The findings of this research showed how improvements in crash geocoding can enhance safety analysis which could potentially lead to a changing paradigm of how network screening of midblock crashes is done by state agencies. Future research could possibly focus on a more detailed analysis of high crash segments by stratifying data based on manner of collision, crash severity, and environmental conditions.

APPENDICES

Appendix A

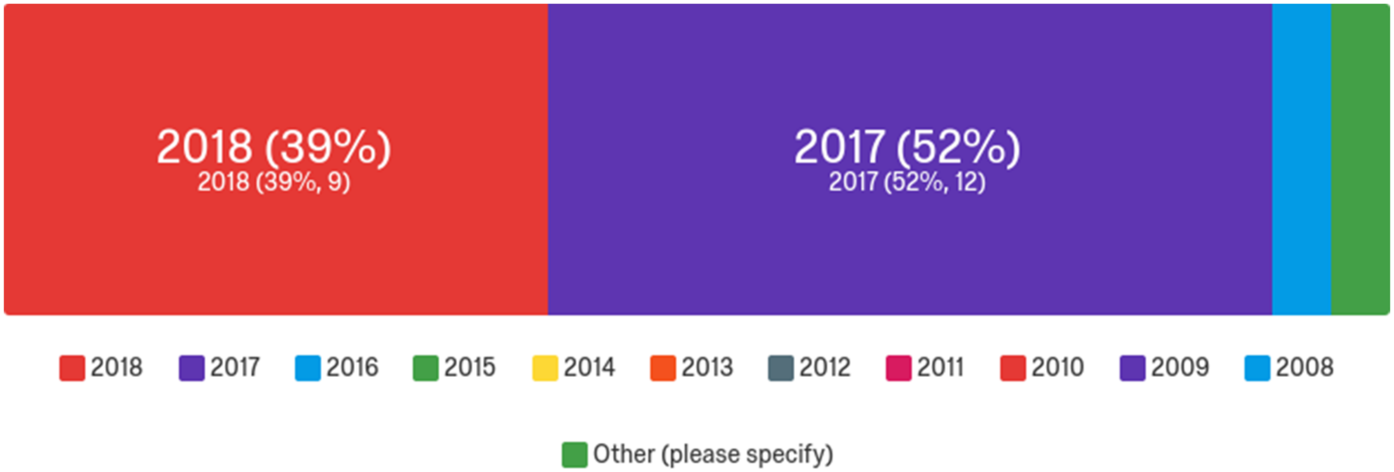
Survey of States

Survey of States

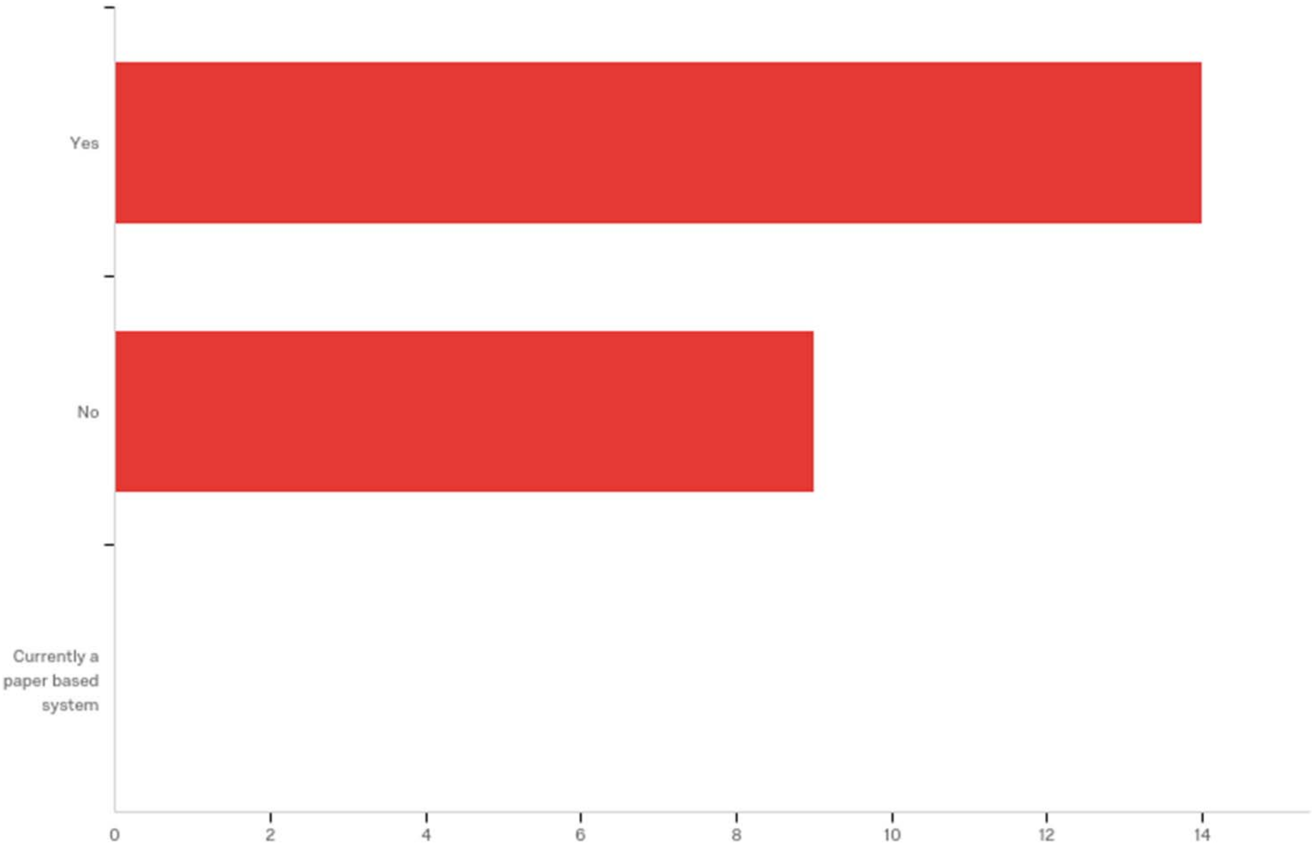
States

Arkansas State (ARDOT)
Colorado State (DOT)
Connecticut State (DOT)
Delaware State (DelDOT)
Georgia State (DOT)
Idaho State (Idaho Transportation Department)
Indiana State (DOT)
Iowa State (DOT)
Kentucky State (University of Kentucky)
Louisiana State (Department of Transportation and Development)
Maine State (Maine DOT)
Massachusetts State (MassDOT)
Michigan State (DOT)
Mississippi State (DOT)
Missouri State (MoDOT)
Montana State (DOT)
Nevada State (NDOT)
New York State (NYSDOT)
Oregon State (DOT)
Texas State
Vermont State (Vtrans)
Wisconsin State (Bureau of Transportation Safety)
Wyoming State (DOT)

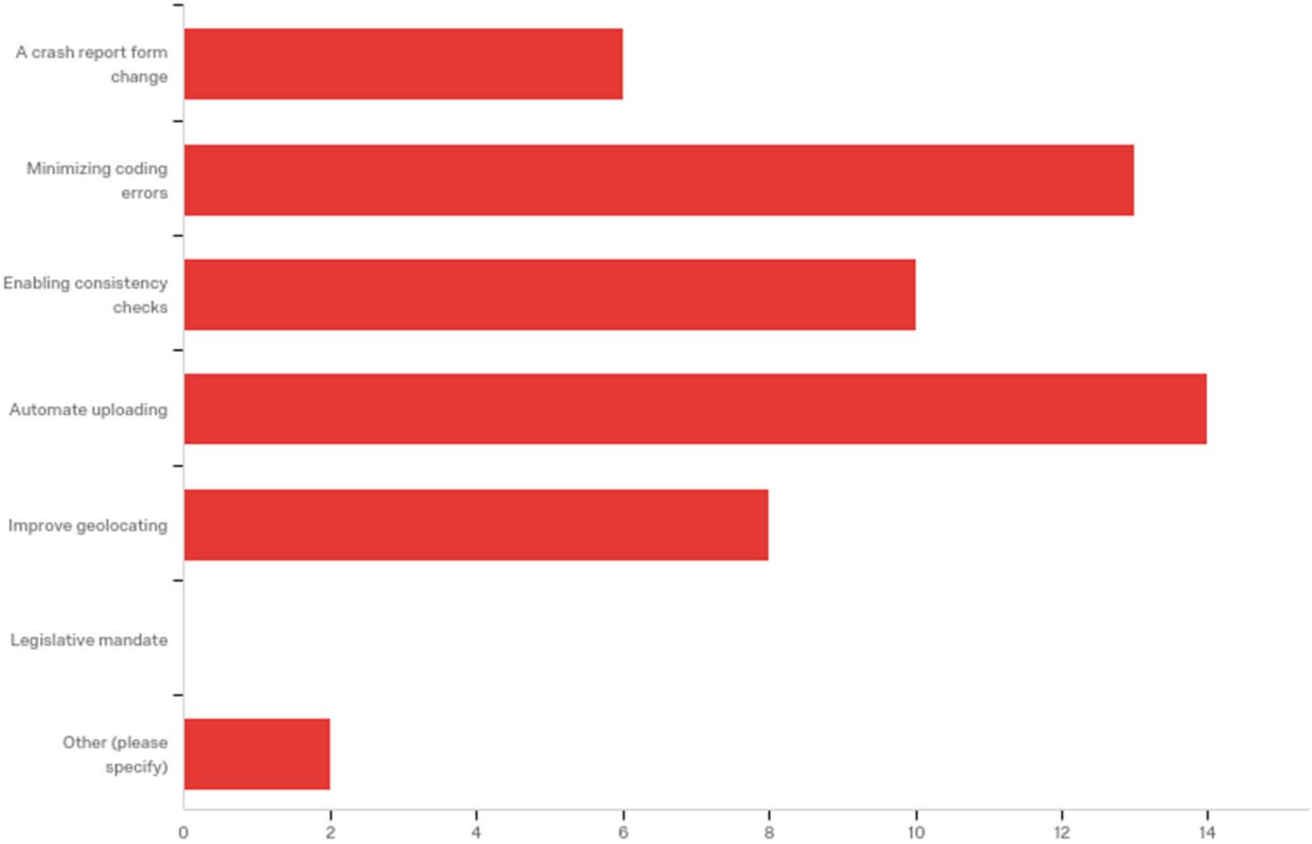
2. We will be asking you a series of questions about crash data collection and network screening analysis for safety programs in your state. In regard to these types of questions, what is the most recent full year of crash data that your state department of transportation is working with?



3. Did your agency transition from a paper-based data collection instrument to a digital/electronic method in the last ten years?



5. Was the transition based on? Please select all that apply.



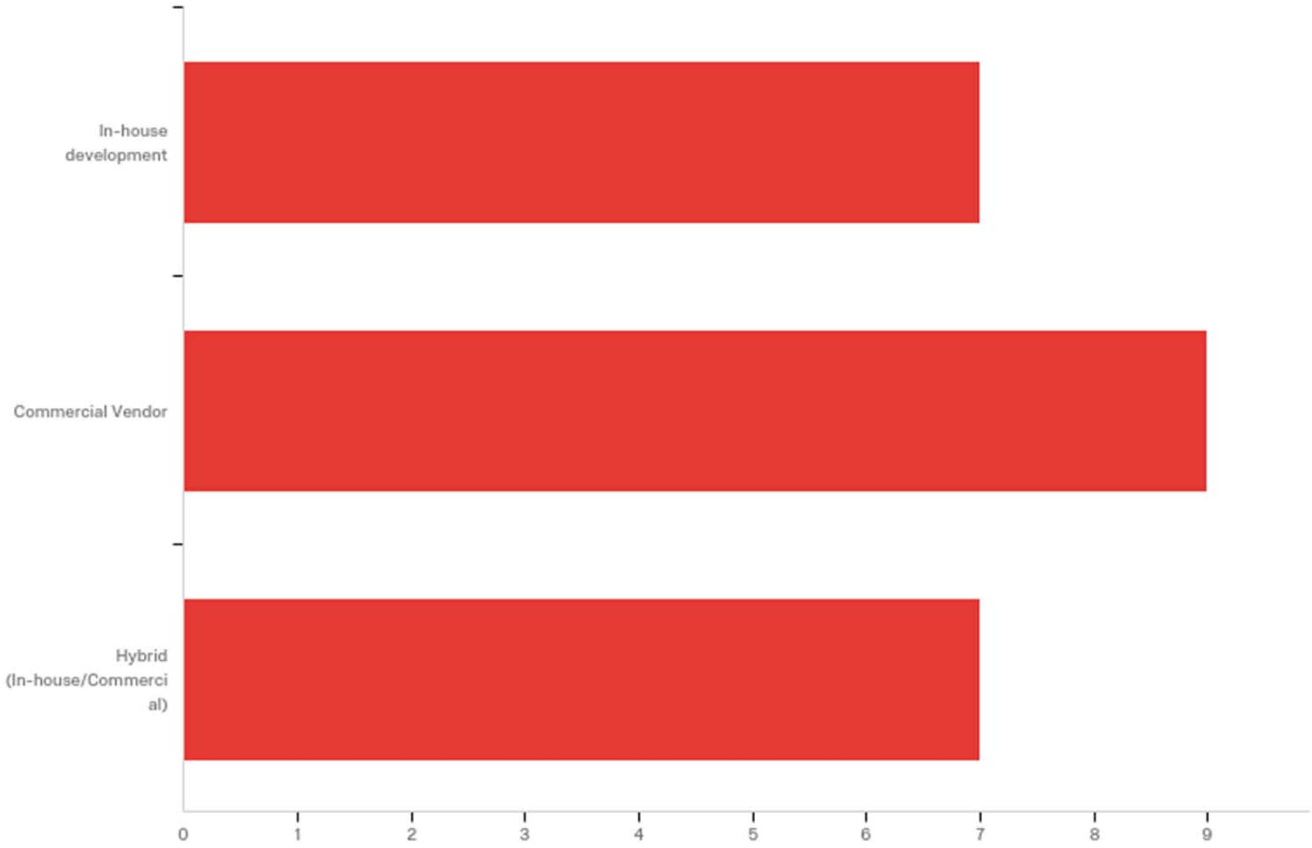
5. Was the transition based on? Please select all that apply.

Answer	Count
A crash report form change	6
Minimizing coding errors	13
Enabling consistency checks	10
Automate uploading	14
Improve geolocating	8
Legislative mandate	0
Other (please specify)	2
Total	53

5. Was the transition based on? Please select all that apply.

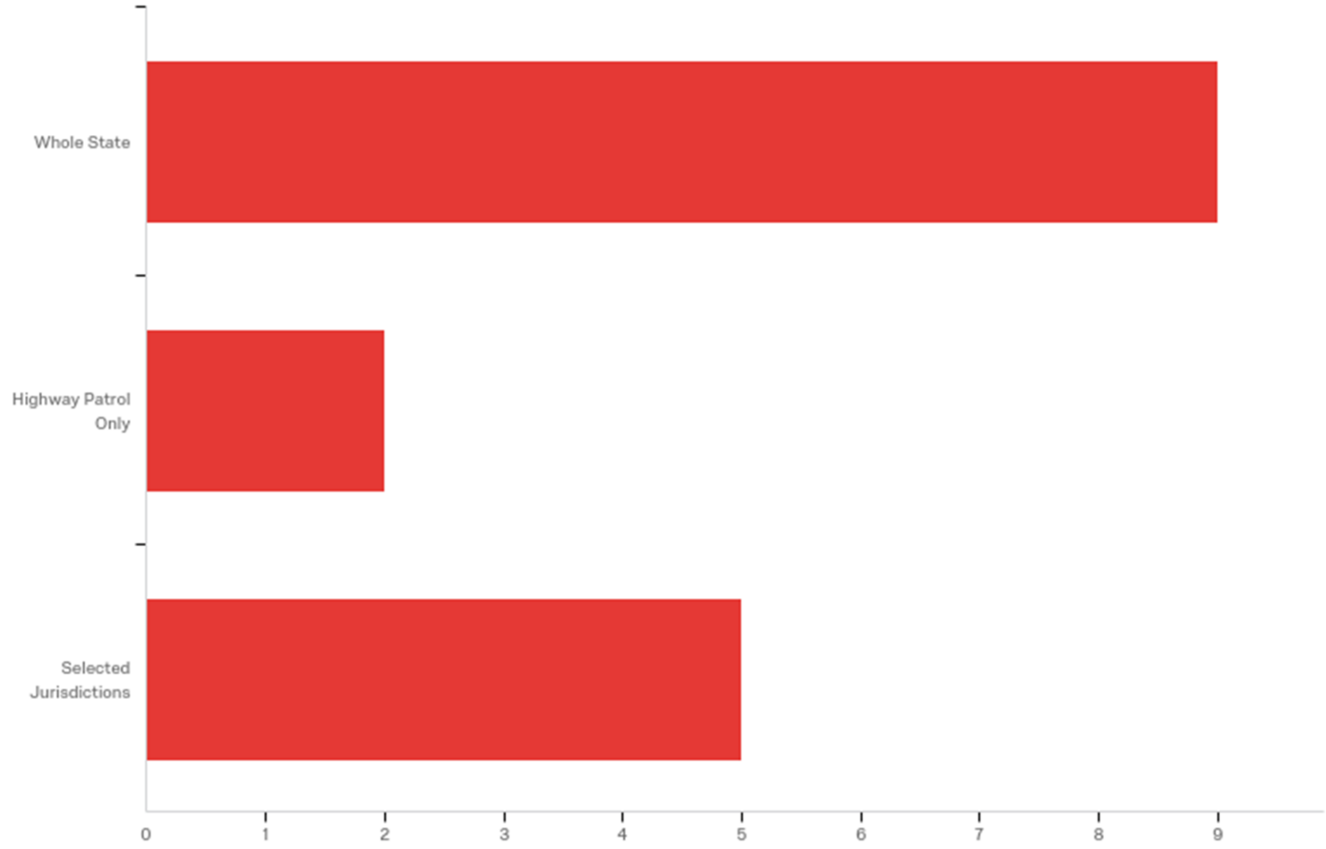
Other (please specify) - Text
We are in the process of converting to electronic reporting. This process has been going on for at least 6 years. (NYSDOT)
Providing timely data to our county Traffic Safety Commissions. (Wisconsin DOT)

8. For your electronic crash report form, was the form developed in-state or adopted from a commercial vendor?

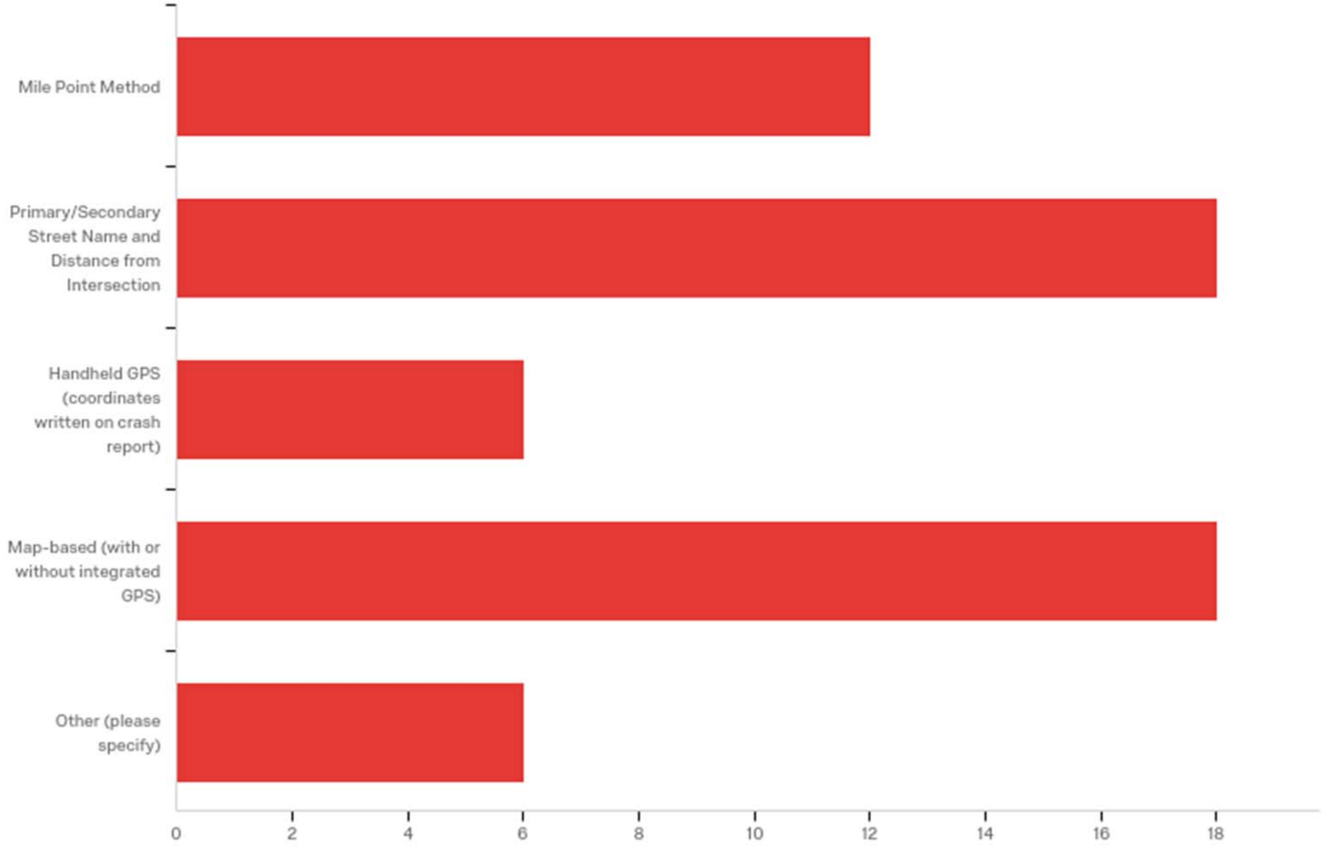


9. For either, commercial vendor or hybrid vendor please indicate vendor or system name.
Statewide Traffic Accident Records System (STARS) – MoDOT
Lexis Nexis – Vtrans, MaineDOT, University of Kentucky
The final form is a product of the state, but the crash software interface is unique to the software vendor. We work with the vendors to ensure the data elements are in alignment with the state’s requirements. –GDOT
CAPS University of Alabama eCrash – ARDOT
TraCS – NYSDOT, Iowa DOT
Don't have information on hand – Colorado DOT
Brazos, Tyler Technology – Nevada DOT
ReportBeam – Mississippi DOT
CTA Smart Cop – Montana DOT
CRIS – Texas
No Answer. Michigan State police developed this – Michigan DOT

10. For electronic data collection methods, is the electronic form being used throughout the entire state, by highway patrol only, or just voluntarily by jurisdiction?

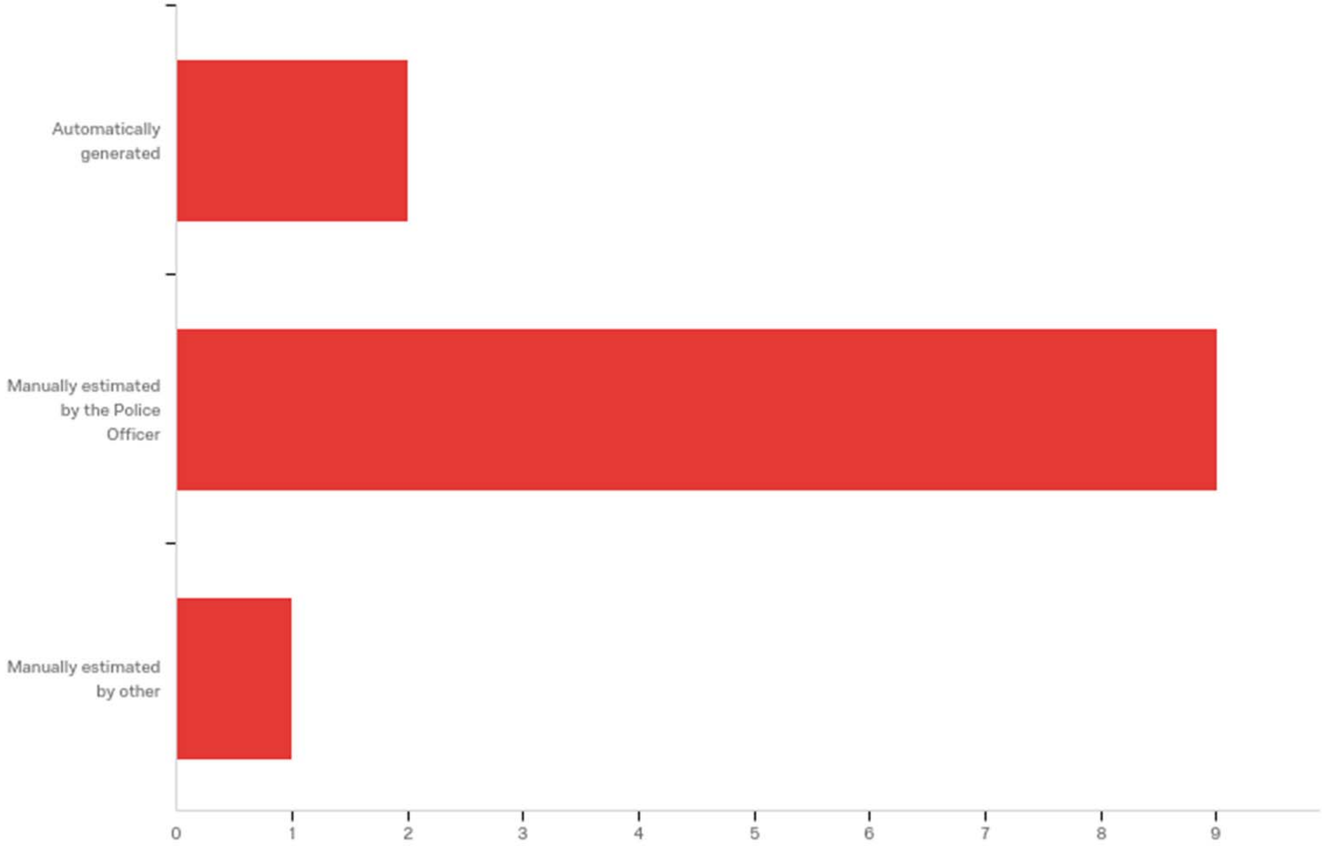


12. Of the total crash reports received for the most recent year in your state, which methods are used to capture data in the field? Please select all that apply.

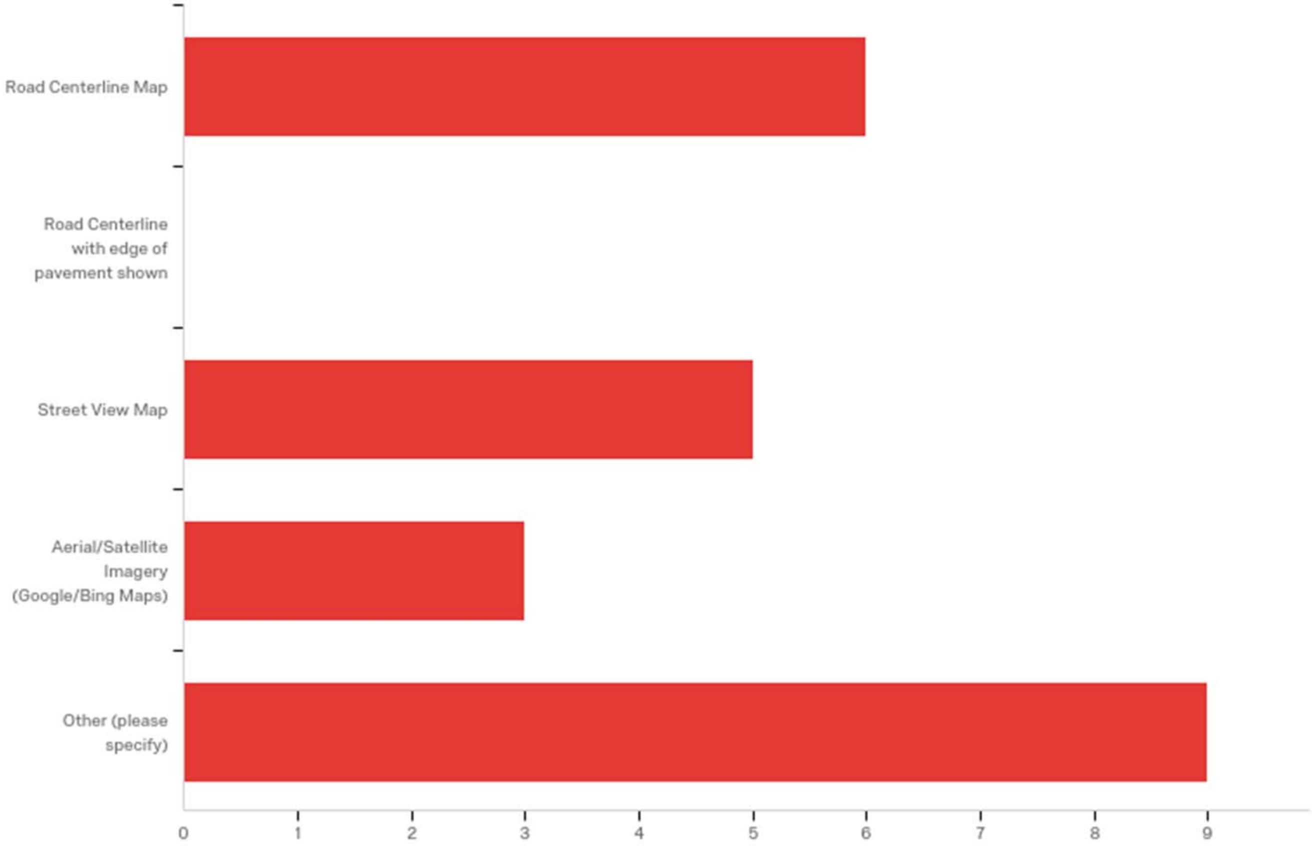


Other (please specify) - Text
Road segments with offsets from nearest intersection. – Maine DOT
Integrated GPS (not map-based) – Louisiana DOT
Auto Capture from PD vehicle – GDOT
MDOT relies on Michigan Sate Police for this – Michigan DOT
Address, Exit number (distance and direction from point of gore), landmarks (we can add them in based on frequency of use). we just built a new system with an API for vendors and a map-based "form“ – MassDOT
ODOT assigns a lat-long themselves, using a custom GIS interface with all the info above and aerial imagery to confirm crash site location. – ODOT

13. For Mile Point method, is the distance generated automatically or manually estimated by the police officer/other?



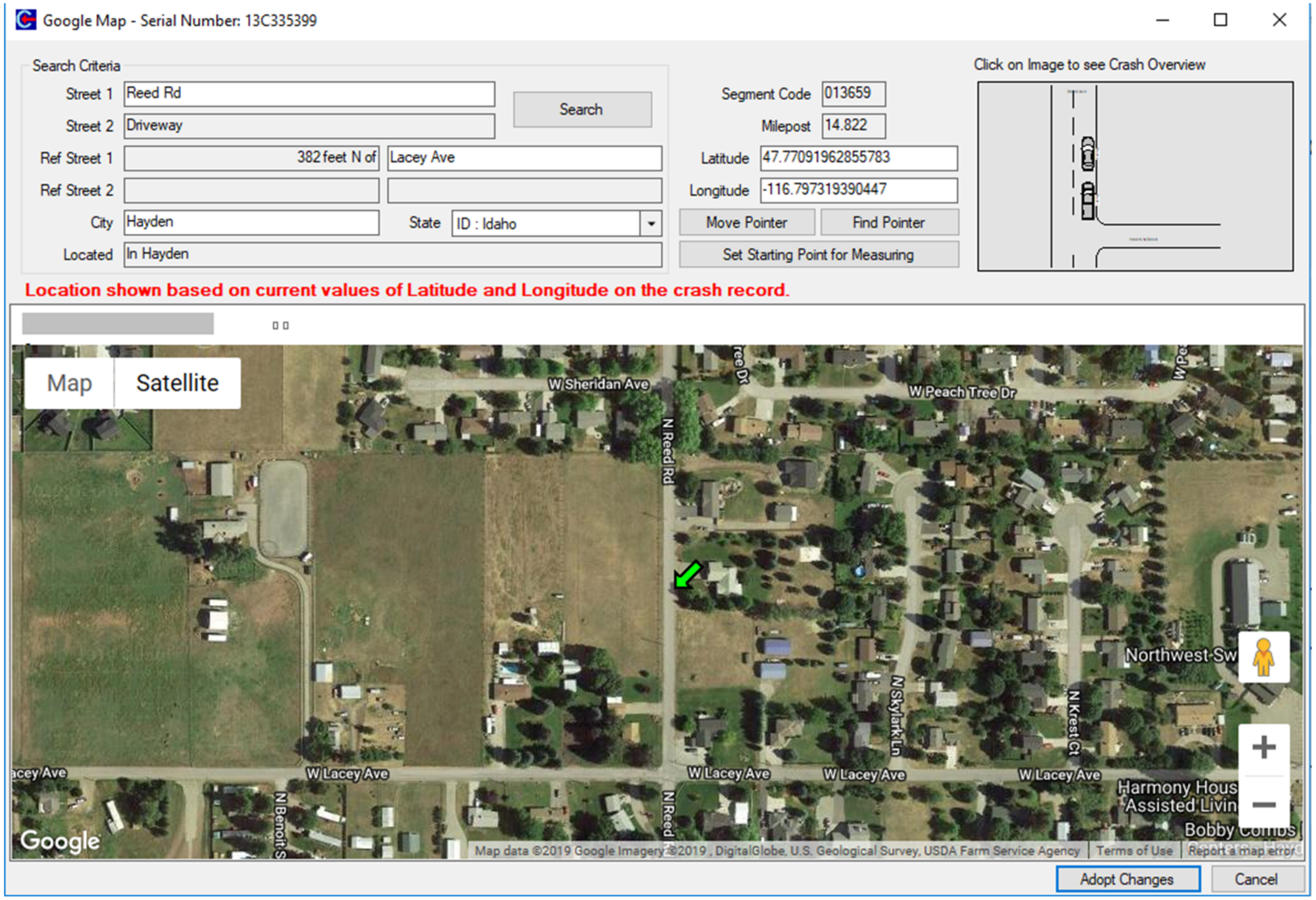
14. For your "Map-Based" system what map background do you use?



Other (please specify) - Text
Road centerline map with imagery behind. Imagery can be toggled off – MoDOT
Road centerline and aerial in some cases – Kentucky Transportation Center
N/A – Colorado DOT
In development, not yet in use – Louisiana DOT
Aerial/Satellite Imagery – Vermont Agency of Transportation, Idaho DOT, Vtrans, ODOT
Michigan State Police developed the system – Michigan DOT
Location Reference System (LRS) based map – Wisconsin DOT
Laptop computer has a gps unit that locates the crash location – Montana DOT
Linework and Aerial and other as layers can be selected – MassDOT

15. If possible, please upload a sample screenshot of the map.

Idaho Transportation Department, Google Maps Road Centerline with Aerial Imagery for the GPS coordinates



Google Map - Serial Number: 13C335399

Search Criteria

Street 1: Reed Rd

Street 2: Driveway

Ref Street 1: 382 feet N of | Lacey Ave

Ref Street 2:

City: Hayden

State: ID : Idaho

Located: In Hayden

Segment Code: 013659

Milepost: 14.822

Latitude: 47.77091962855783

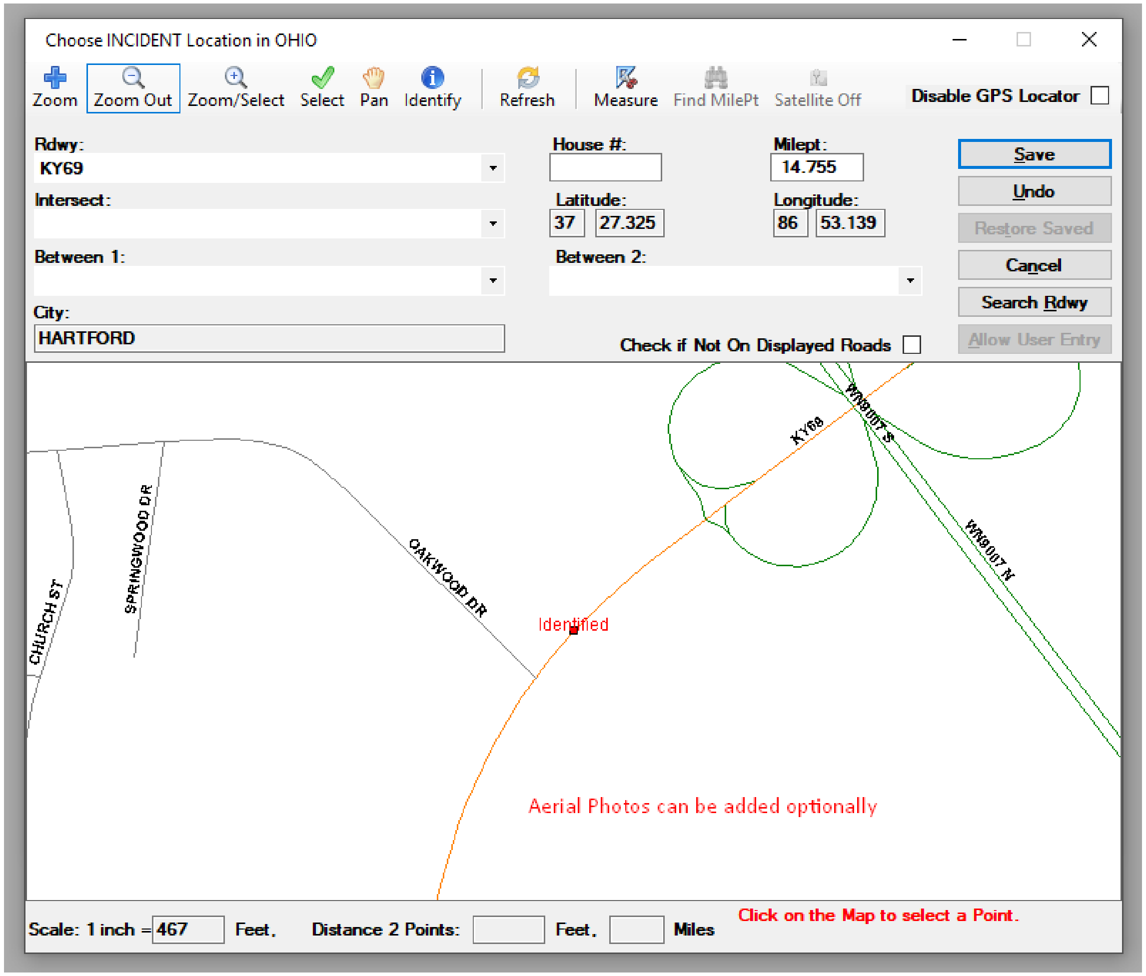
Longitude: -116.797319390447

Click on Image to see Crash Overview

Location shown based on current values of Latitude and Longitude on the crash record.

Map data ©2019 Google Imagery ©2019 , DigitalGlobe, U.S. Geological Survey, USDA Farm Service Agency

Kentucky Transportation Center, Road Centerline (Aerial optional)



Choose INCIDENT Location in OHIO

Disable GPS Locator ☐

Rdwy: KY69

Intersect:

Between 1:

City: HARTFORD

House #:

Latitude: 37 27.325

Between 2:

Milept: 14.755

Longitude: 86 53.139

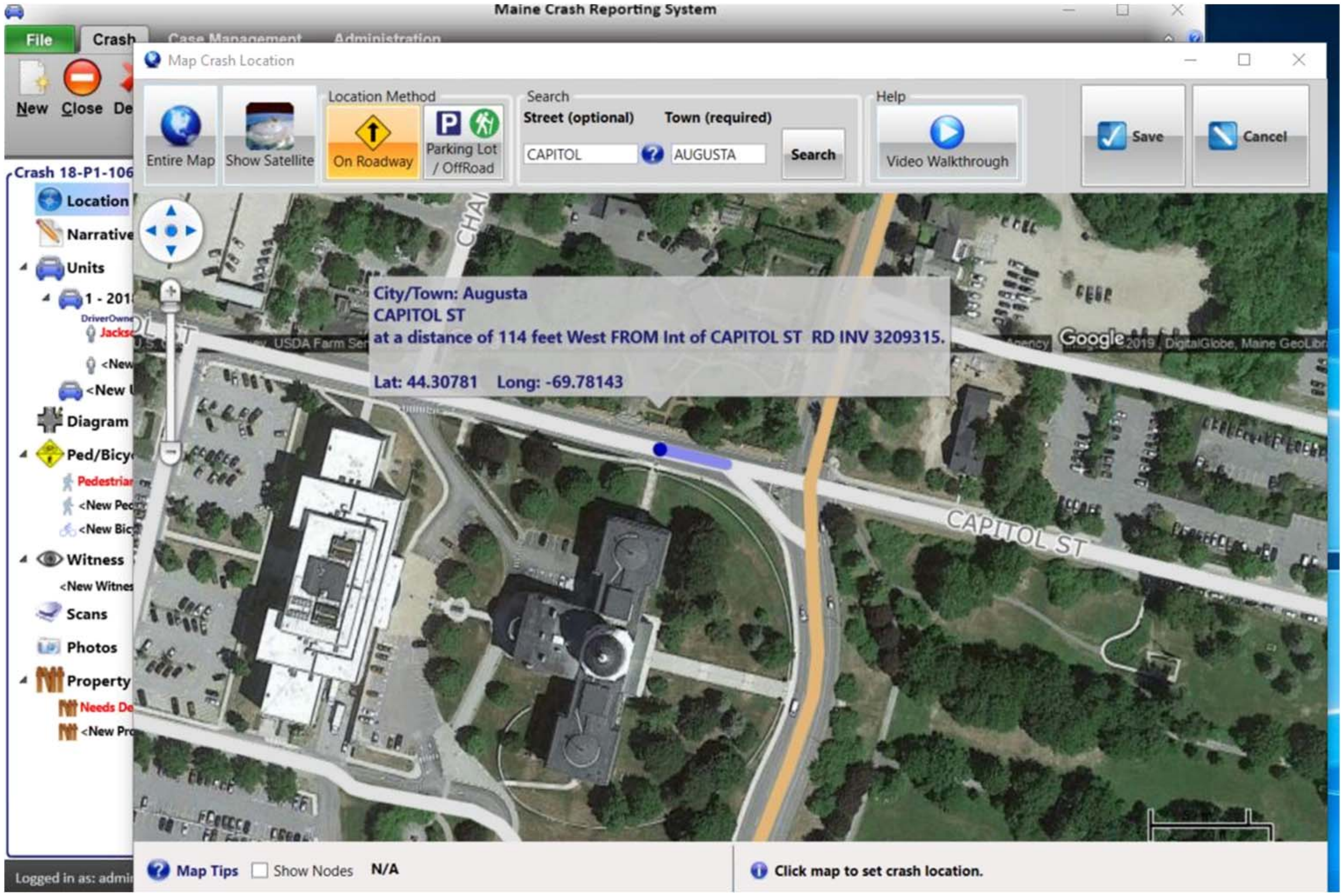
Check if Not On Displayed Roads ☐

Aerial Photos can be added optionally

Scale: 1 inch = 467 Feet. Distance 2 Points: Feet. Miles

Click on the Map to select a Point.

Maine DOT, Road Centerline Map



Street (optional): CAPITOL

Town (required): AUGUSTA

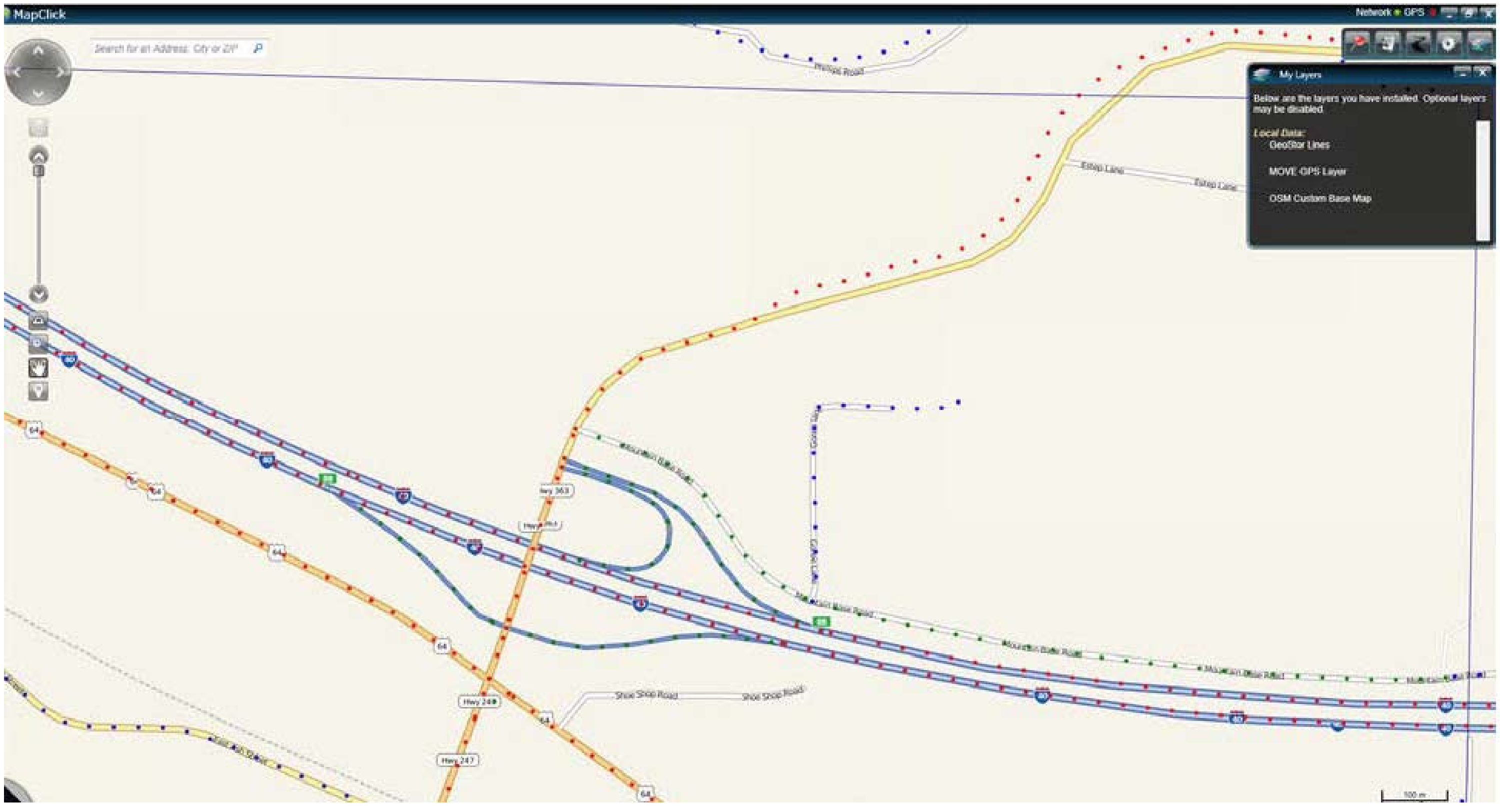
Search

City/Town: Augusta
CAPITOL ST
at a distance of 114 feet West FROM Int of CAPITOL ST RD INV 3209315.

Lat: 44.30781 Long: -69.78143

Click map to set crash location.

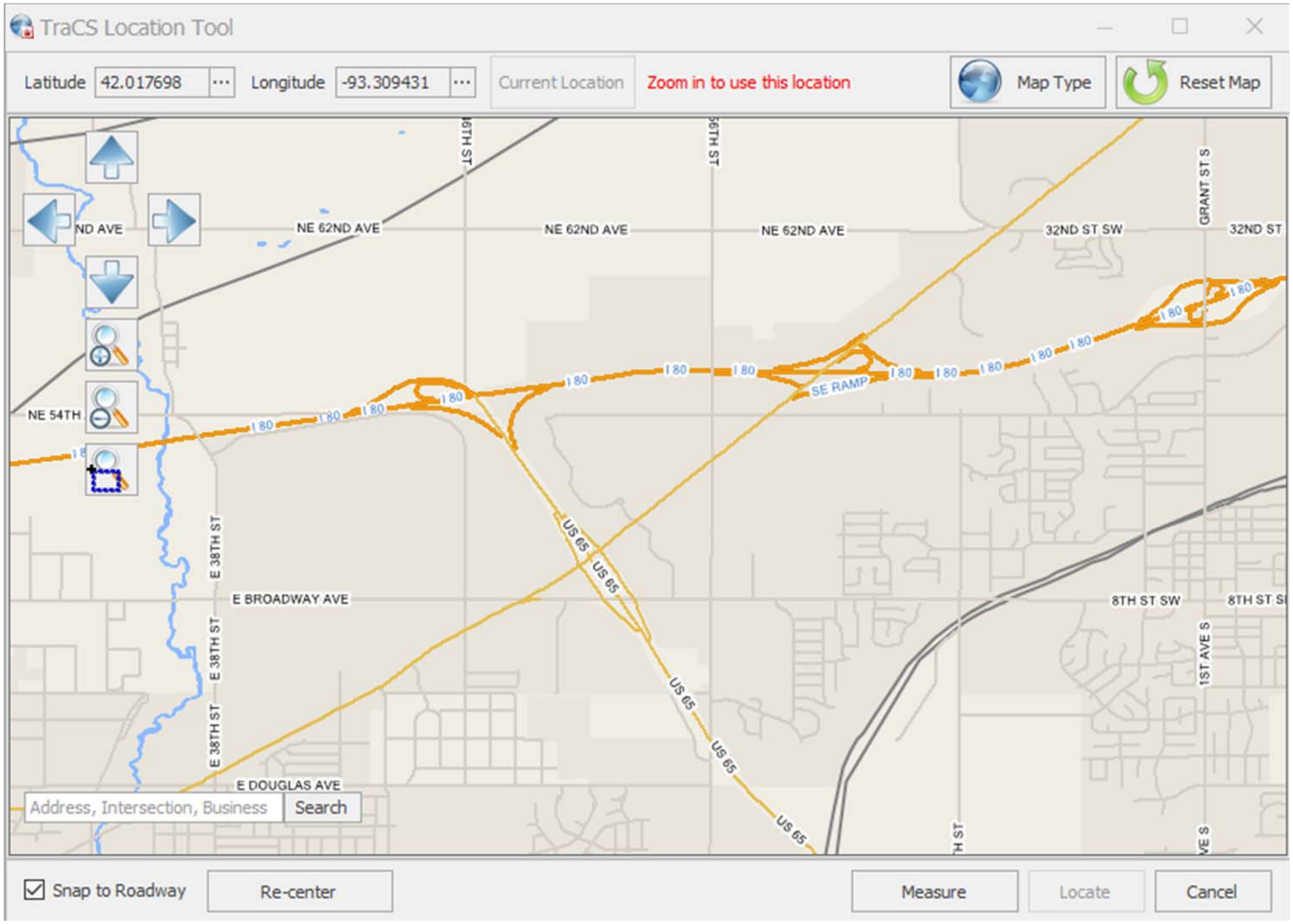
ARDOT Street View Map, CAPS University of Alabama eCrash



Below are the layers you have installed. Optional layers may be disabled.

Local Data:
GeoStor Lines
MOVE GPS Layer
OSM Custom Base Map

Iowa DOT, Road Centerline Map



TraCS Location Tool

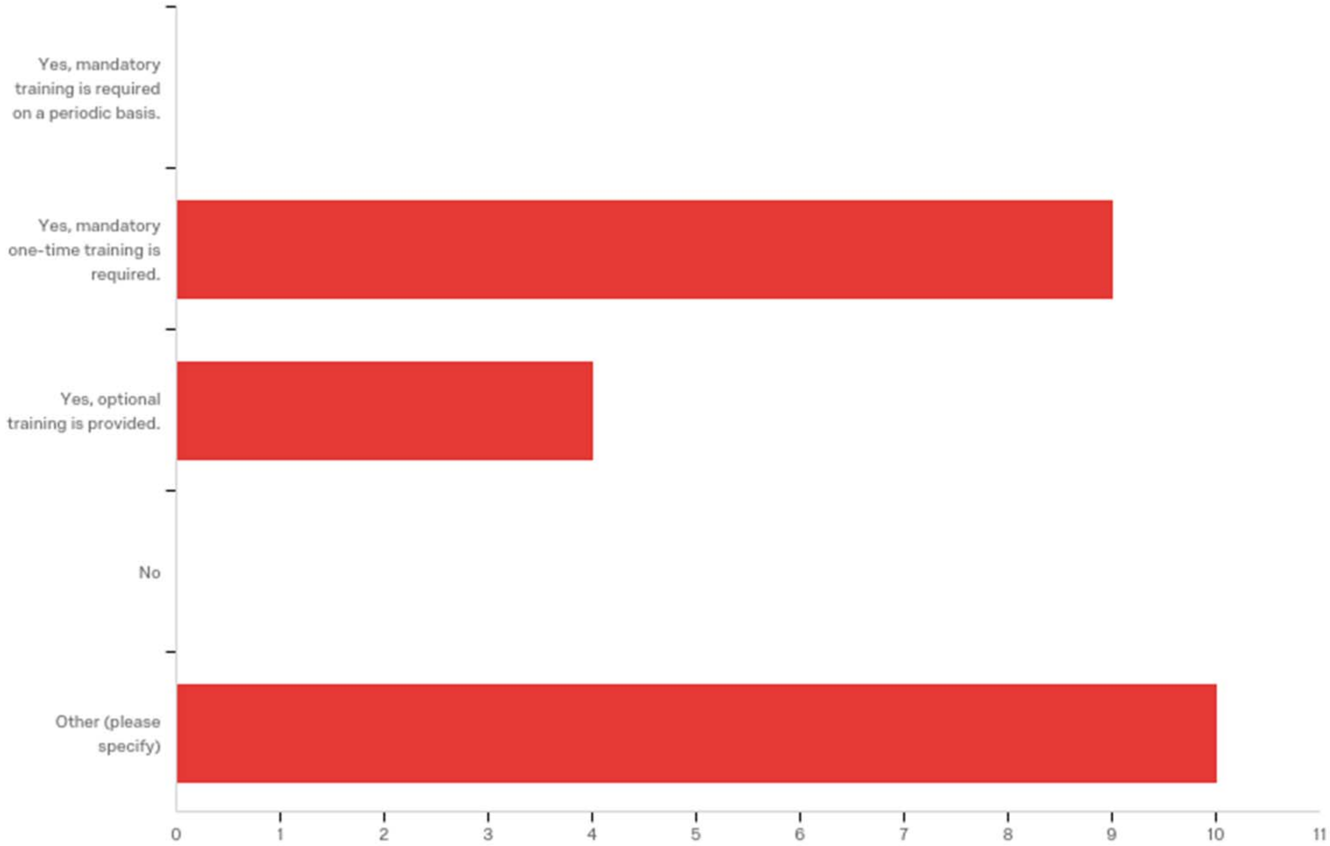
Latitude 42.017698

Longitude -93.309431

Zoom in to use this location

☑ Snap to Roadway

16. Is there a training program in place for police officers on crash data collection?



16. Is there a training program in place for police officers on crash data collection?

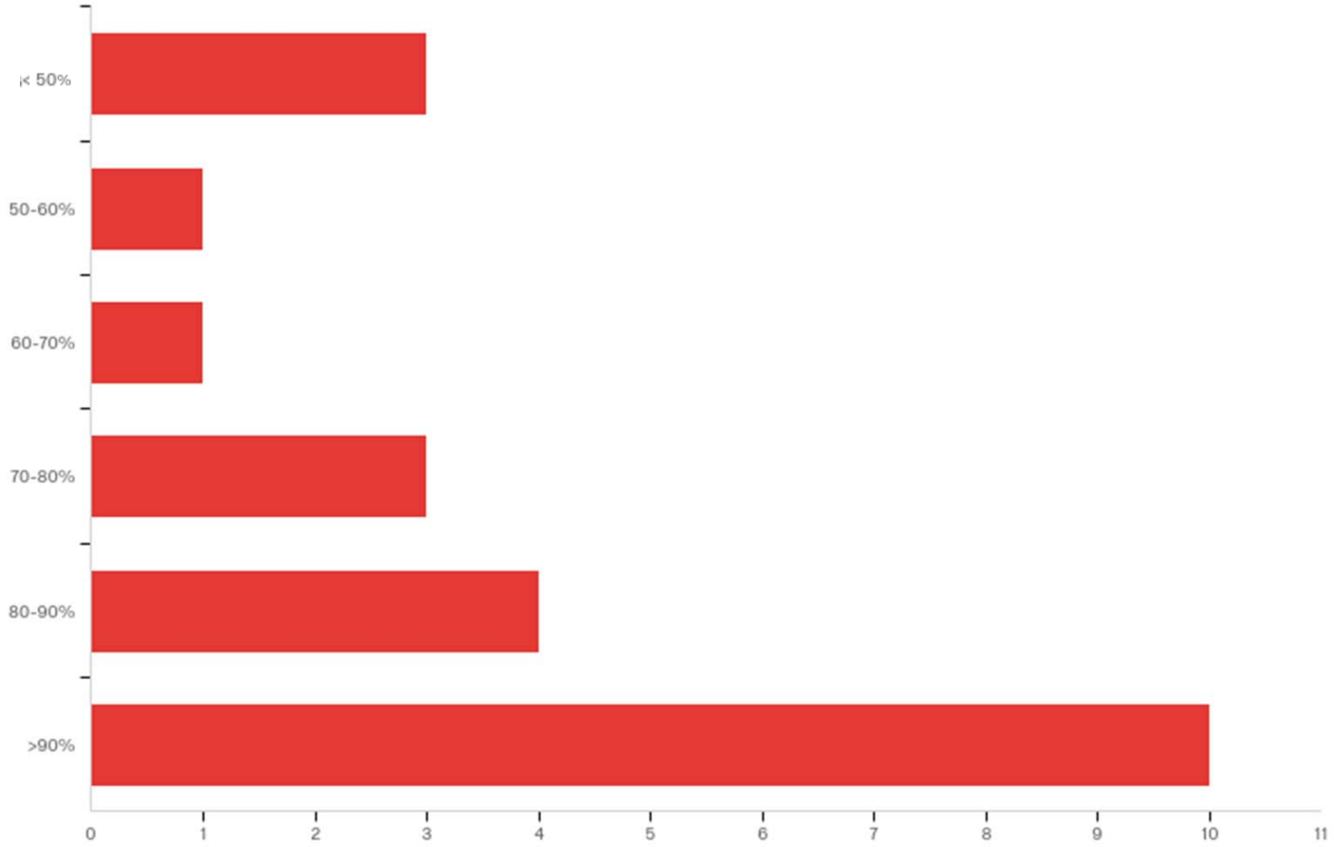
Other (please specify) - Text
Developing training for new crash form – Colorado DOT
There is a guidance document/handbook. – Iowa DOT
Training is provided. Not sure if it is mandatory or optional. Training has also been provided at safety conferences and via a podcast that is available online -NYSDOT
I'm not positive. I know there is training, but I'm unsure how frequent/required – Mississippi DOT
Don't know the answer to this one. – Montana DOT
Mandatory training at police academy. Additional training is available when requested – Wyoming DOT
Again MSP does training – Michigan DOT
Yes, however frequency and requirement is unknown, as it is not administered through DelDOT – DelDOT
Not really but our Law Enforcement Liaison offers as needed in AdHoc way – MassDOT
Recently the basic crash report training has been reduced from 16 hours to a total of 6 hours at the police academy and that may result in more errors or omissions. Basic traffic crash report training is brief, there is optional certification and reconstruction investigation training available at a cost. That course will teach skills for conducting a thorough crash investigation and properly documenting findings for court room presentations. Measuring, photographing and preparing physical evidence learn fundamentals of mathematics, engineering and physics to accurately analysis crash factors and evidence. – ODOT

17. Please describe the training program.
Basic Training Course provided by department/training school – CTDOT
In development – Colorado DOT
In - person training at police departments is offered for updates and new employees. We also are part of the crash reporting training at the police academy. – Vermont Agency of Transportation
new cadet training for most officers, while some larger agencies have their own training. State Police have their own training. – Kentucky Transportation Center
All officers receive training at the Maine Criminal Justice Academy as part of their initial training. – Maine DOT
The eCrash system is taught at the Police Academies as part of their curriculum. – ARDOT
I am not sure of mandatory training, but there is a handbook located on the internet for them to use when using the software. – Iowa DOT
The training primarily consists of covering information within the Missouri Uniform Crash Report Preparation Manual. That information includes definitions and crash classification standards from the ANSI.D16 Manual on Classification of Motor Vehicle Accidents publication, instructions for completing each field on the Missouri Uniform Crash Report form and specifics for locating motor vehicle crashes. - MoDOT
All officers attend the training academy that includes crash reporting. The PDs then provide training using the Field Training Officers and most software vendors provide training. - GDOT

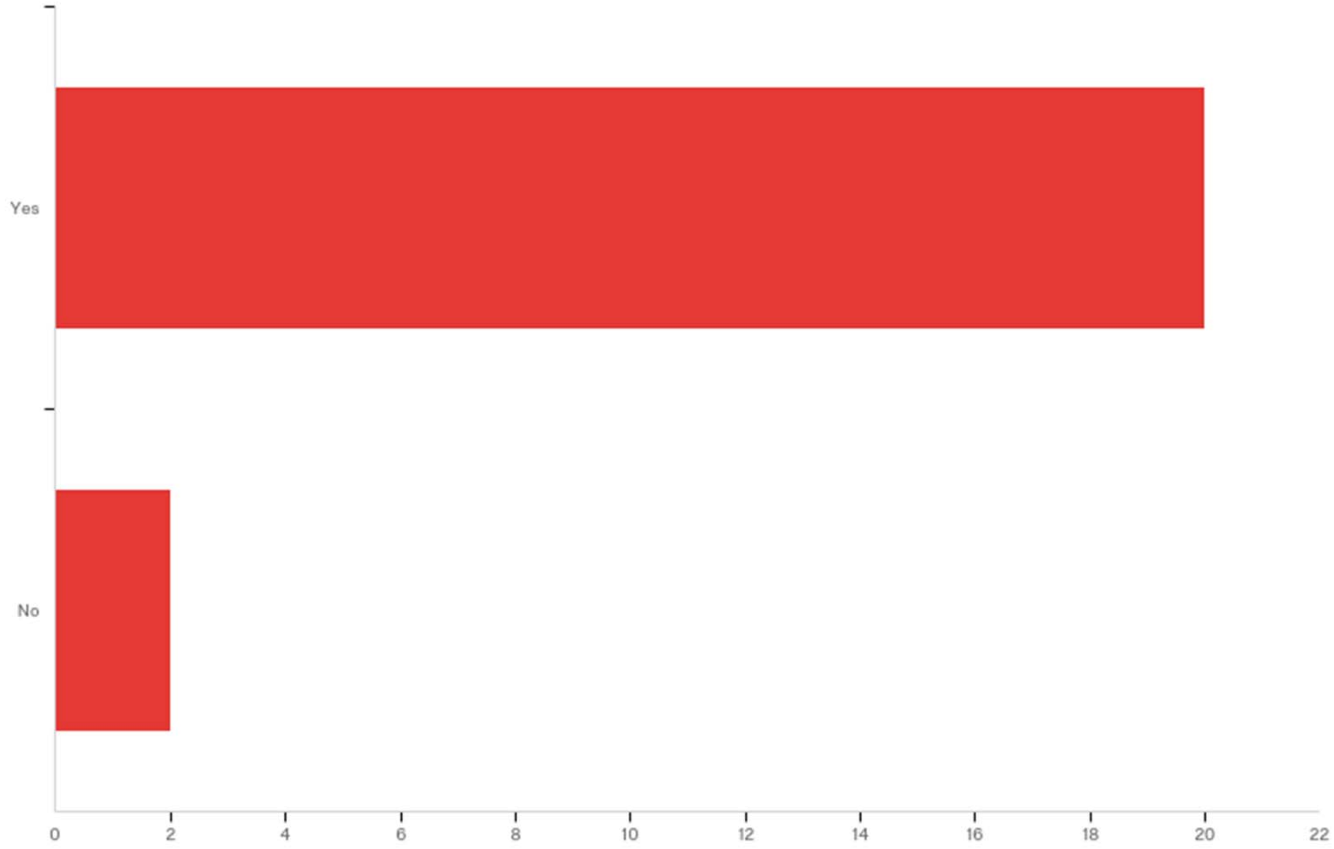
17. Please describe the training program.
Training is provided. Not sure if it is mandatory or optional. Training has also been provided at safety conferences and via a podcast that is available online. The Governors Traffic Safety Committee could provide more detailed information on the specific training provided to law enforcement – NYSDOT
Depends on agency, but usually about an hour at multiple week training academy. DOTD has a Law-Enforcement Expert (LEE) who provides additional training to agencies based on issues. – Louisiana DOT
Training is in person and as needed at this point. When we went electronic it was an in-person training also at each department or regionally. VTrans
Agencies request training or trainer contacts agencies with common problems. Training covers a step by step description of fields and values of fields, including why they should be filled in certain ways. – Idaho Transportation Department
I am unsure of the training program – Mississippi DOT
All officers receive training at the MCJA as part of their basic training. – Bureau of Highway Safety Maine
Beginning in late 2016 we did statewide training – Wisconsin DOT
MSO does training – Michigan DOT

17. Please describe the training program.
Unknown – Indiana DOT, Colorado DOT, Montana DOT
Short term refresher course is 8 hr course on training for police officers when requested. All law enforcement trains at the academy when hired. – Wyoming DOT
TCOLE approved training for the CRASH application is offered and consists of a 2 hours Configure (User Management) training and a 4 hours CRASH (User submission) training; it is not required as an agency can opt out and receive the Quick Training session in which they only receive a quick overview of the application and does not receive the TCOLE credit, or they can opt out of the CRASH training and only receive the Configure portion to get their agency kicked off on the CRASH application. This too they do not receive TCOLE credit. - Texas
Unknown, as it is not administered through DelDOT – DelDOT
It is more AdHoc as needed. We do have some powerpoint presentations that we have provided on common issues – MassDOT
Training is included at the Peace Officer standard training (POST) academy – Nevada DOT

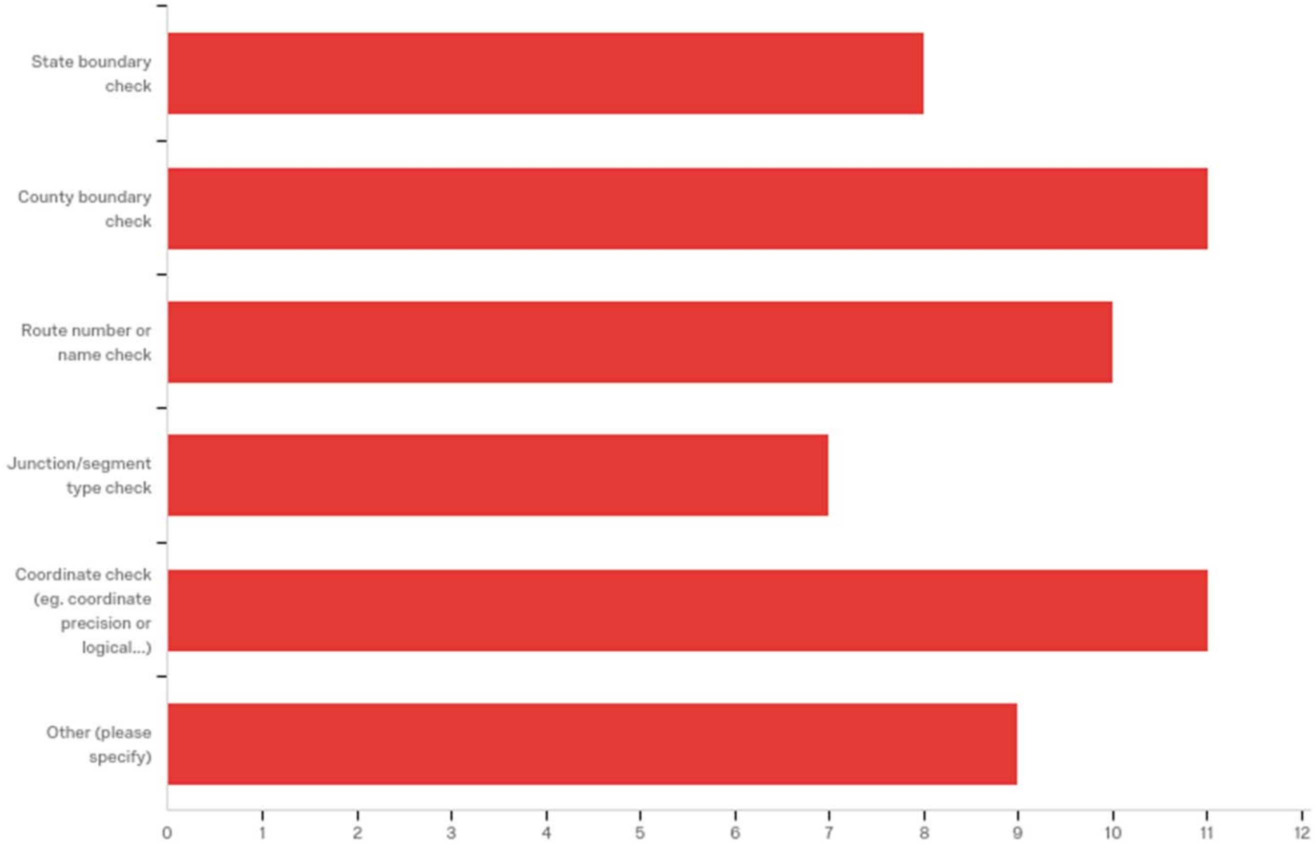
18. For the most recent year, what percentage of crash locations are accurately geolocated on the total road network? (23 Responses)



5.3 - 20. Do you validate the accuracy of geolocation processes in any way? (23 Responses)



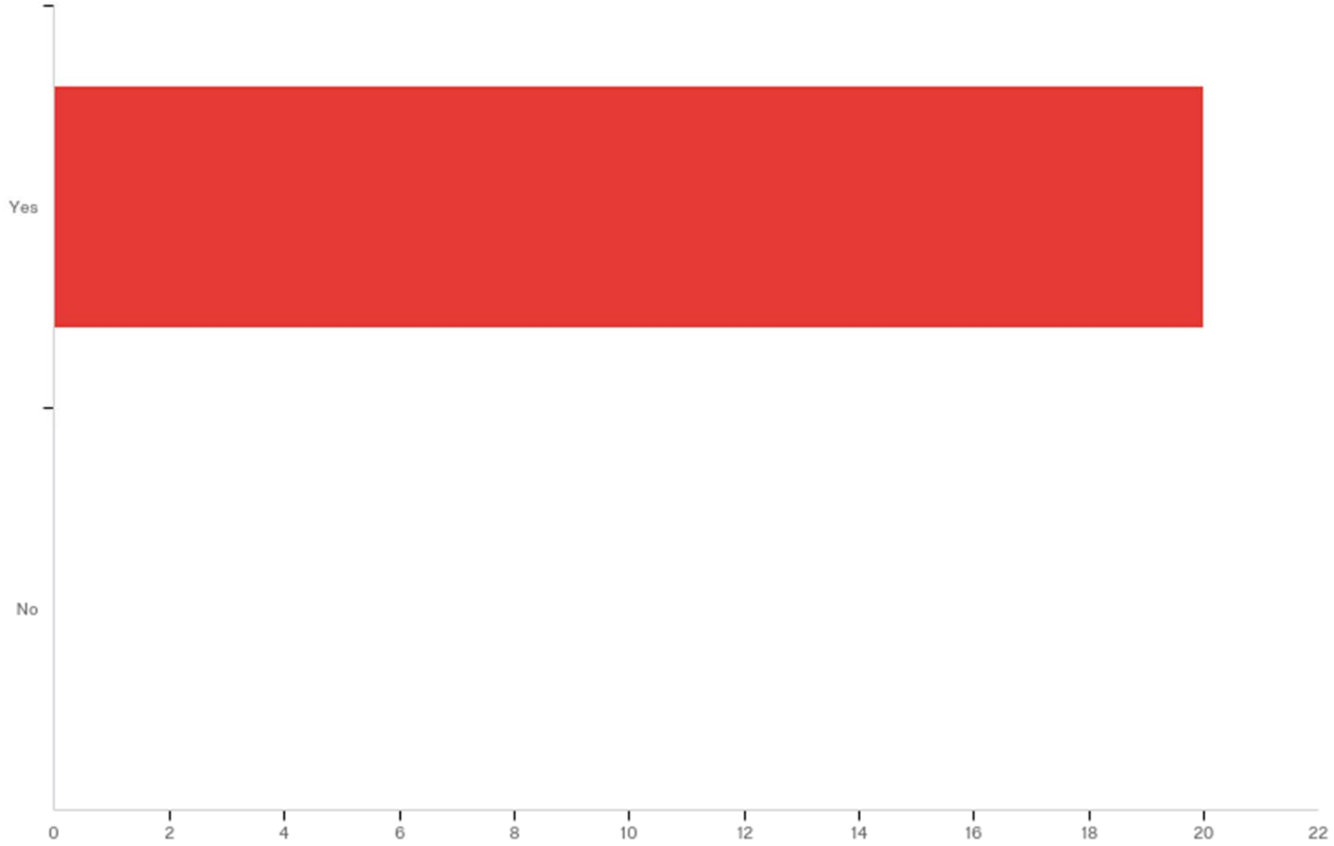
5.4 - 21. What processes or metrics do you use? Please select all that apply.



21. What processes or metrics do you use? Please select all that apply.

Other (please specify) - Text
Manual review is conducted of each crash location. – Maine DOT
This is only done on KA crashes. - ARDOT
Manual review of crash locations of state roads. – Wisconsin Department of Transportation
Manually as reports are pulled/read. – Montana DOT
Spot check locations on map vs crash report data – Wyoming DOT
We have a whole post process geocoding system (the new system actually runs a compare betwen police provided GPS and derived coordinates – MassDOT
location contained in the crash report – Nevada DOT
Aerial imagery, snapped to linework, city limits, etc. - ODOT

22. Is there any attempt to fix poorly geolocated crash data? (20 Responses)

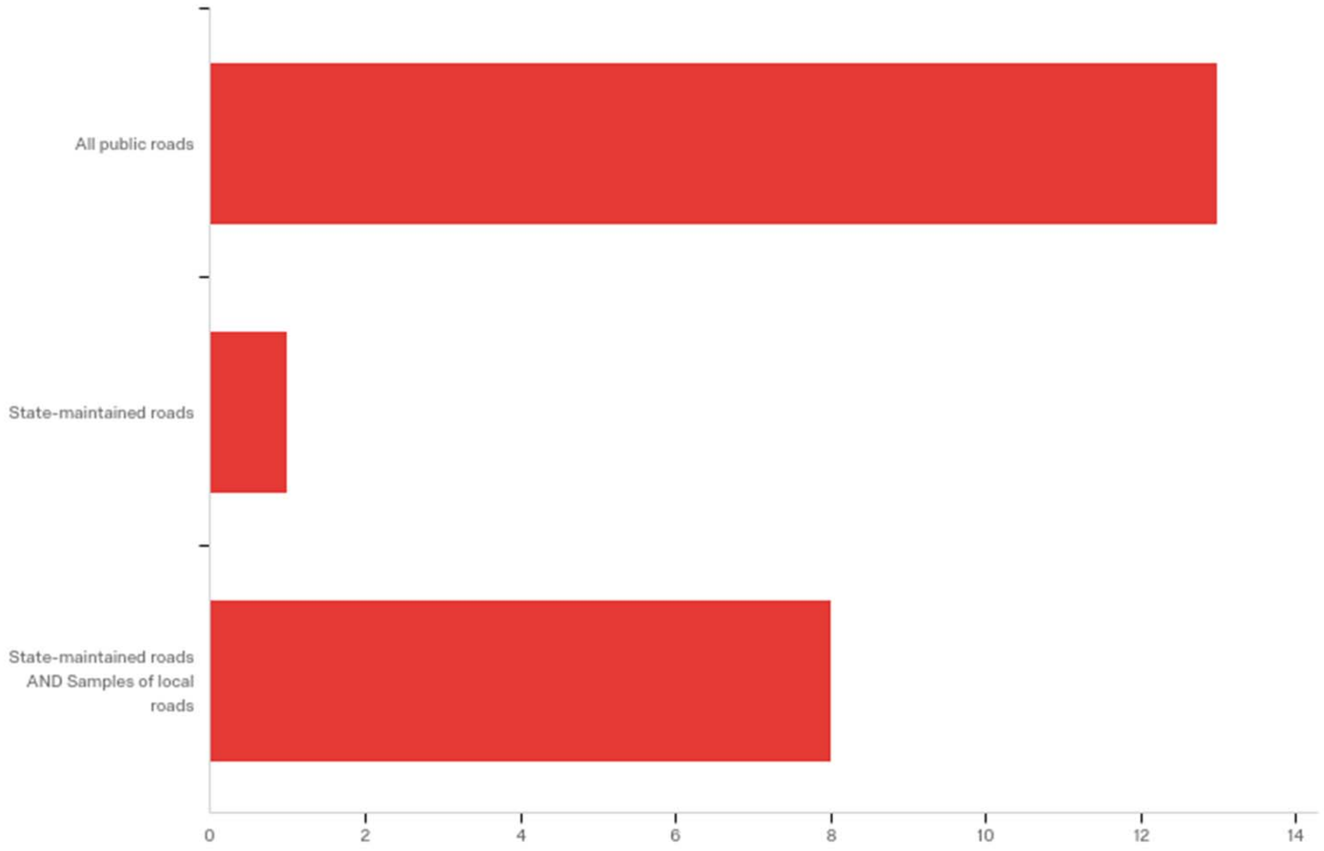


23. Please describe processes/methods used to correct poorly geolocated crash data.
We have queries and spatial maps that we use to help identify crashes that are incorrectly located. - MoDOT
We have a contract that includes a team of mappers that correct/validate crash locations. We sample their work and grade. – GDOT
Locate – Texas
Crash data users will submit issues to the data's owner (DelJIS), and corrections will be made, usually within 24-48 hours. – DelDOT
About 86% of crashes are automatically geocoded. MassDOT staff manually review the remainder. Even if State police crashes are automatically geocoded, before the crash file closes, some crashes meeting specific criteria (flagged as conflicting linked road attributes with information provided like speed limit below 45 but linked onto an interstate)are manually reviewed – MassDOT
manually located so low error rate – NDOT
GIS FME workbench. We do not use the police collected coordinates as they are not often collected properly at the POI and not precise enough to aid accurate engineering analysis and development of productive and cost efficient safety countermeasures – ODOT
As we are make aware of poorly located crashes (by end users), we correct that data if is not more than 3-years old. – LADOT

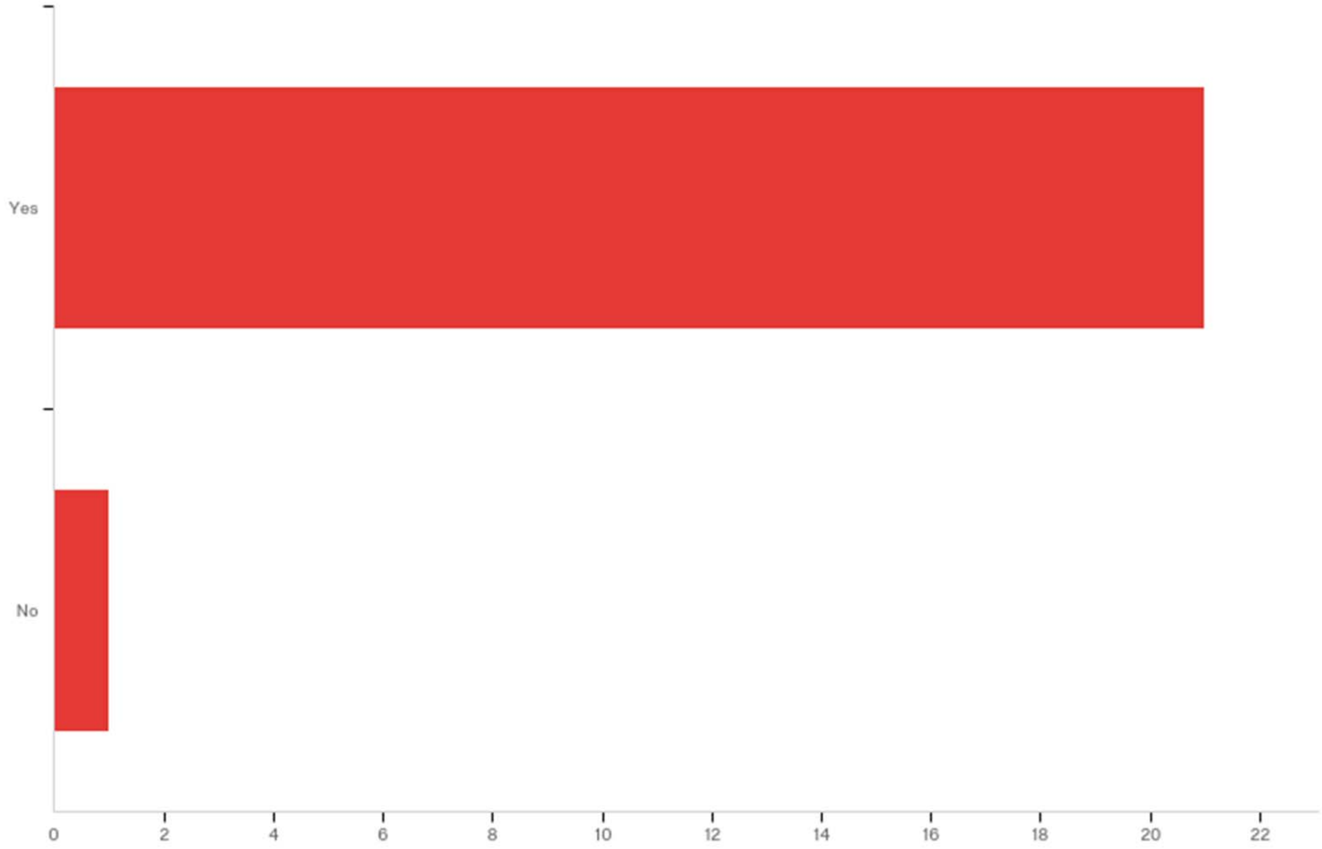
23. Please describe processes/methods used to correct poorly geolocated crash data.
Use of linear reference system to check or populate geolocations on state highways only – Colorado DOT
We have a team of personnel who review and correct where necessary the location of each and every crash that is reported. - Maine DOT
We have built an in house system to manually correctly geo locate crashes – Mississippi DOT
manual review and correction – Wisconsin DOT
Human review – CTDOT
We are in the process of changing the base map of the road network for the software. The current linework is outdated and we are currently updating it to our version of ESRI's Roads and Highways. – Iowa DOT
Crash data is updated when errors are found – Wyoming DOT
Crashes found to be inaccurate are corrected. – Idaho DOT
It is checked and if the location is wrong, it is then updated in the system and the database. - ARDOT
Crash report is read, state plane XY coordinates are found and then updated in the database. – Montana DOT

24. Please give a reason why there is no attempt for validation of accuracy.
It has been recommended but not implemented. – Kentucky Transportation Center
I don't know. – VTrans

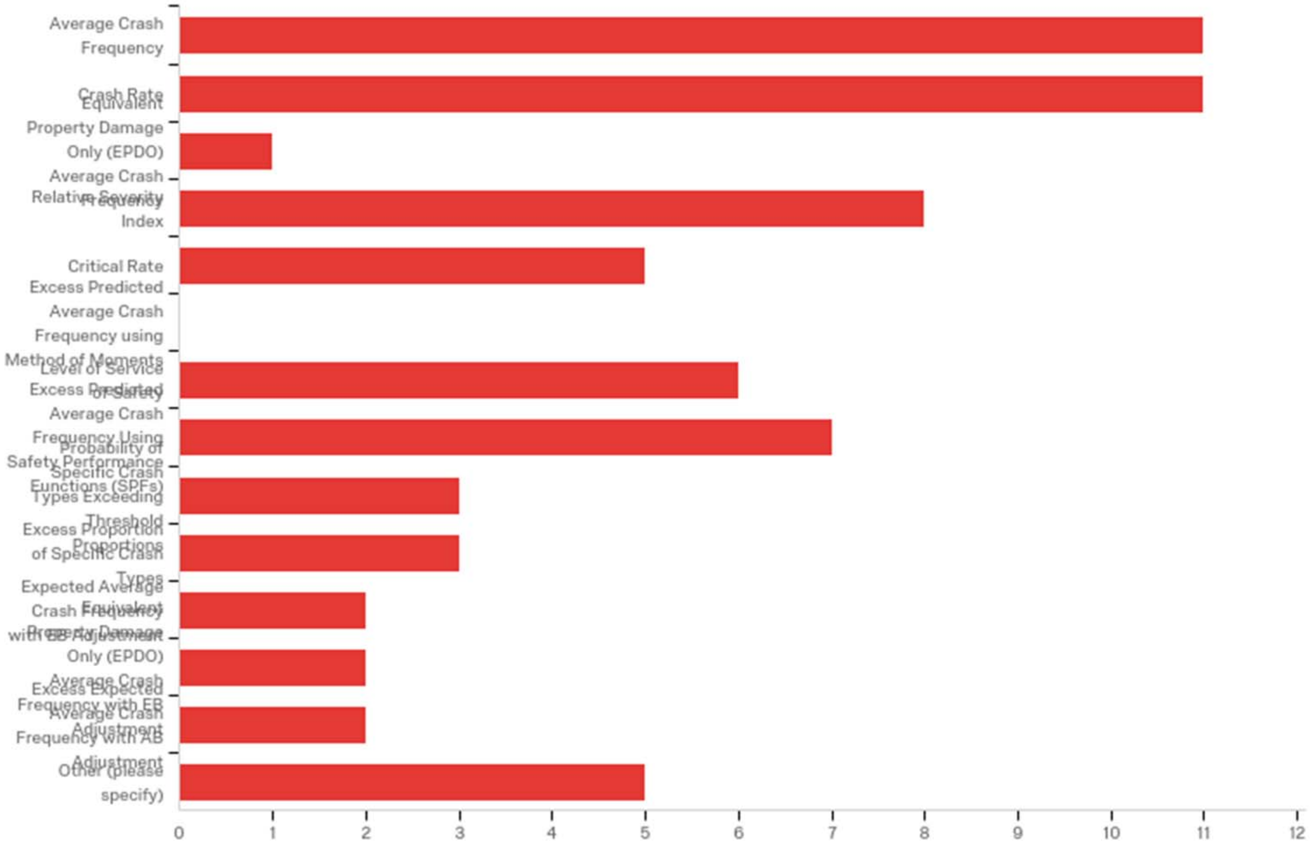
26. In your state, do you maintain data (e.g. roadway, crash, AADT) for safety network screening on? (22 Responses)



27. Do you do safety network screening? (22 Responses)



6.3 - 28. What measures do you typically use for safety network screening? Please select all that apply.



6.3 - 28. What measures do you typically use for safety network screening? Please select all that apply.

Answer	Count
Average Crash Frequency	11
Crash Rate	11
Equivalent Property Damage Only (EPDO) Average Crash Frequency	1
Relative Severity Index	8
Critical Rate	5
Excess Predicted Average Crash Frequency using Method of Moments	0
Level of Service of Safety	6

6.3 - 28. What measures do you typically use for safety network screening? Please select all that apply.

Answer	Count
Excess Predicted Average Crash Frequency Using Safety Performance Functions (SPFs)	7
Probability of Specific Crash Types Exceeding Threshold Proportions	3
Excess Proportion of Specific Crash Types	3
Expected Average Crash Frequency with EB Adjustment	2
Equivalent Property Damage Only (EPDO) Average Crash Frequency with EB Adjustment	2

6.3 - 28. What measures do you typically use for safety network screening? Please select all that apply.

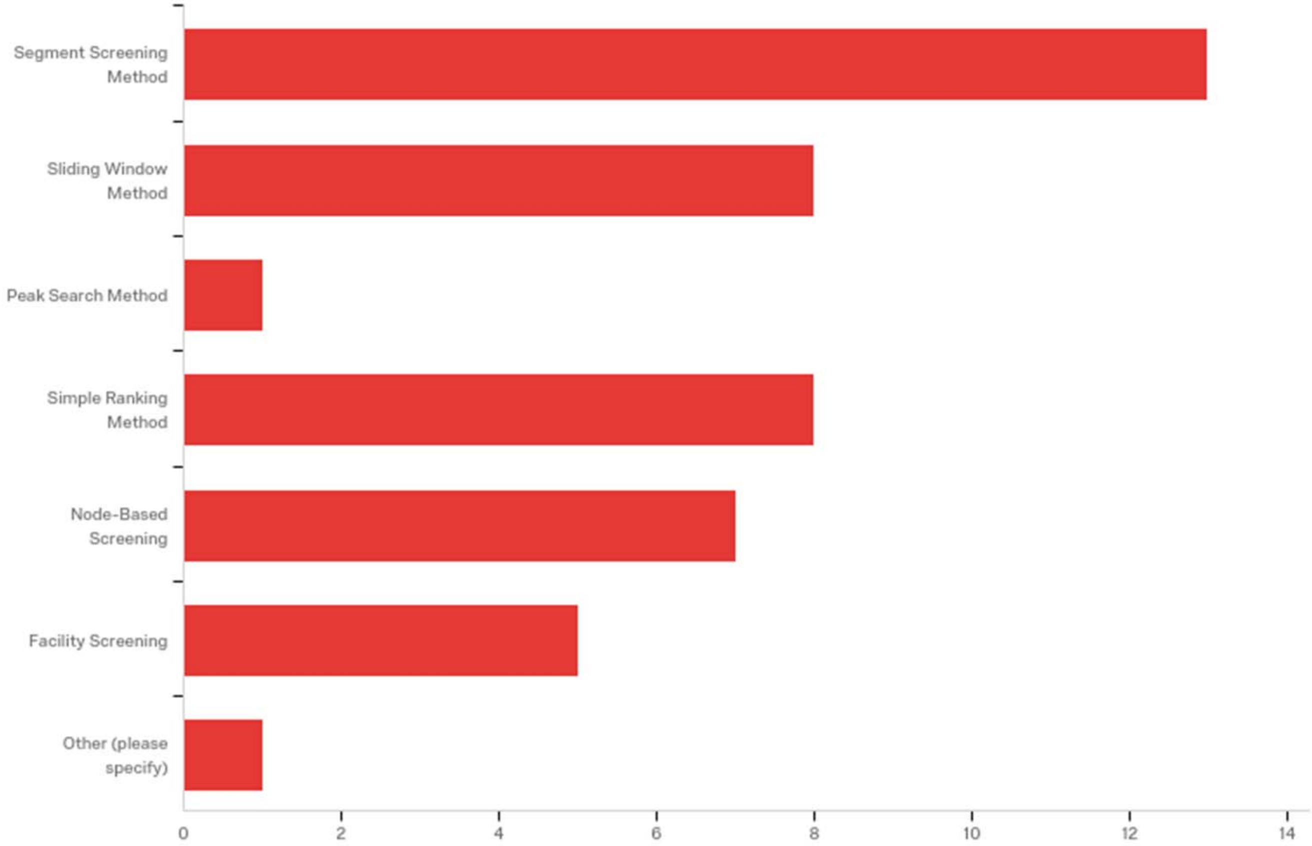
Answer	Count
Excess Expected Average Crash Frequency with AB Adjustment	2
Other (please specify)	5
Total	66

6.3 - 28. What measures do you typically use for safety network screening? Please select all that apply.

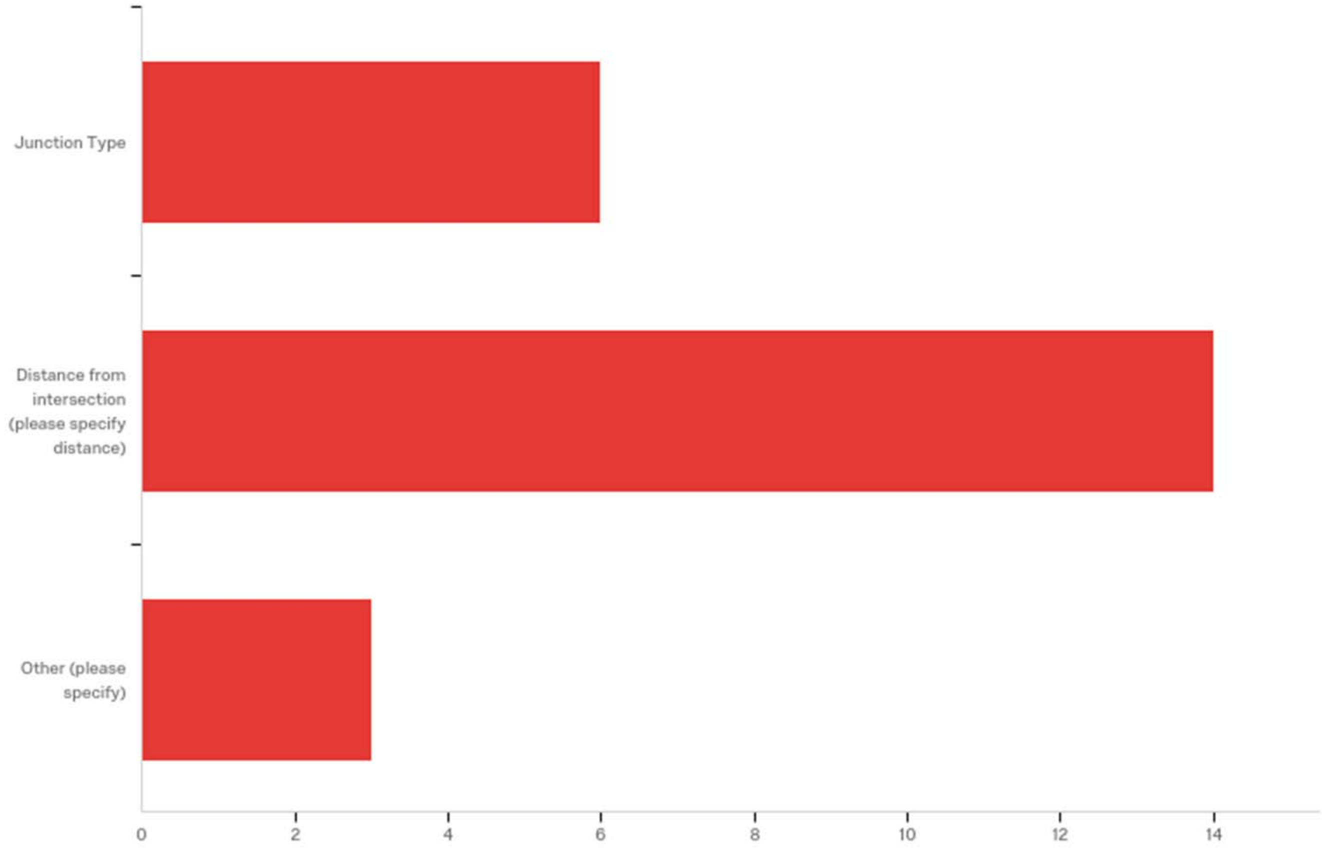
6.3_14_TEXT - Other (please specify)

Other (please specify) - Text
We are updating our network screeening process to make this integrated into our whole system – MassDOT
predictive method out of the HSM – NDOT
KA crash rate – ARDOT
of % of crashes within a half mile segment. – Montana DOT
Not sure as engineers do the screening but assume most of the the above – ODOT

29. Which methods do you use for network screening? Please select all that apply. (43 Responses)

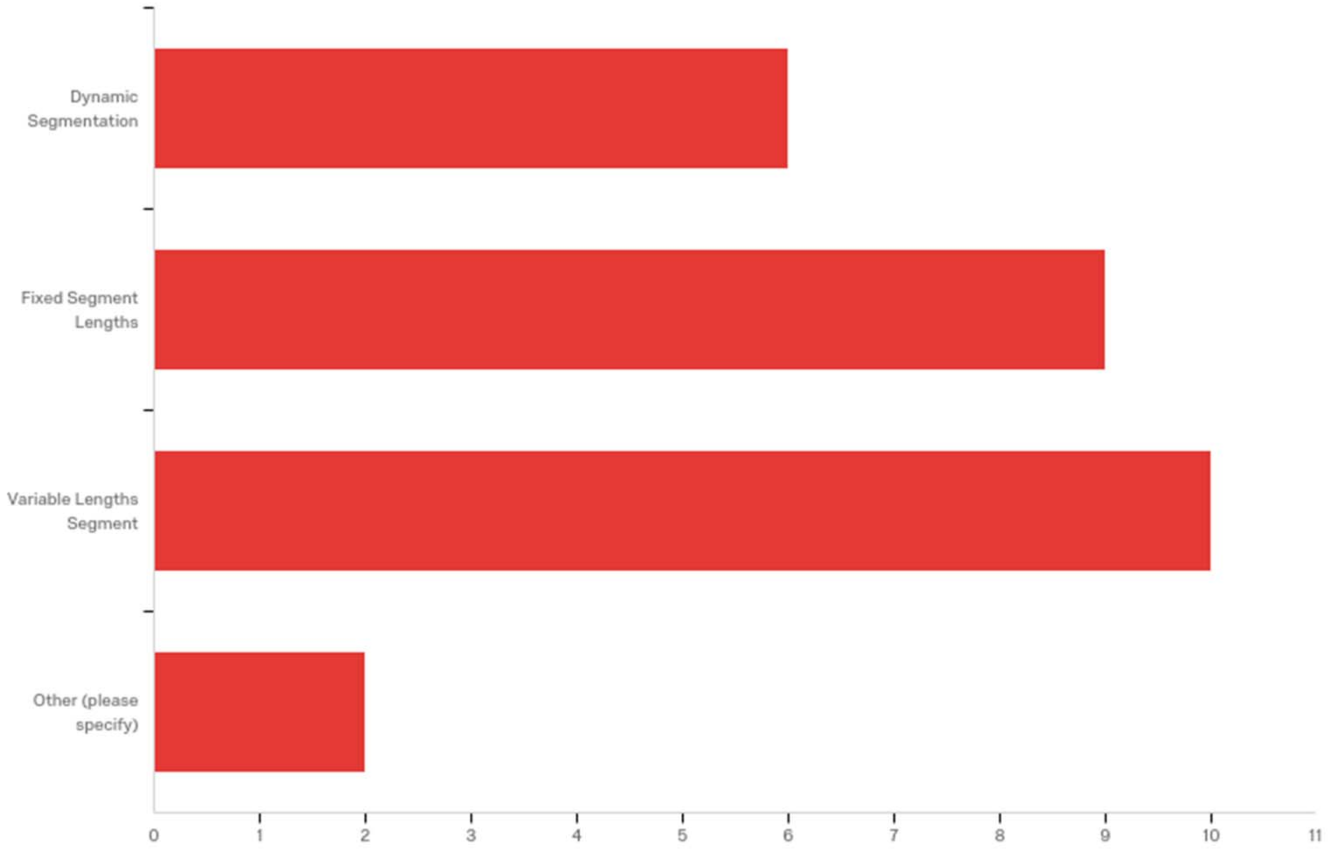


30. Which methods do you use to separate segment crashes from intersection crashes? (22 Responses)



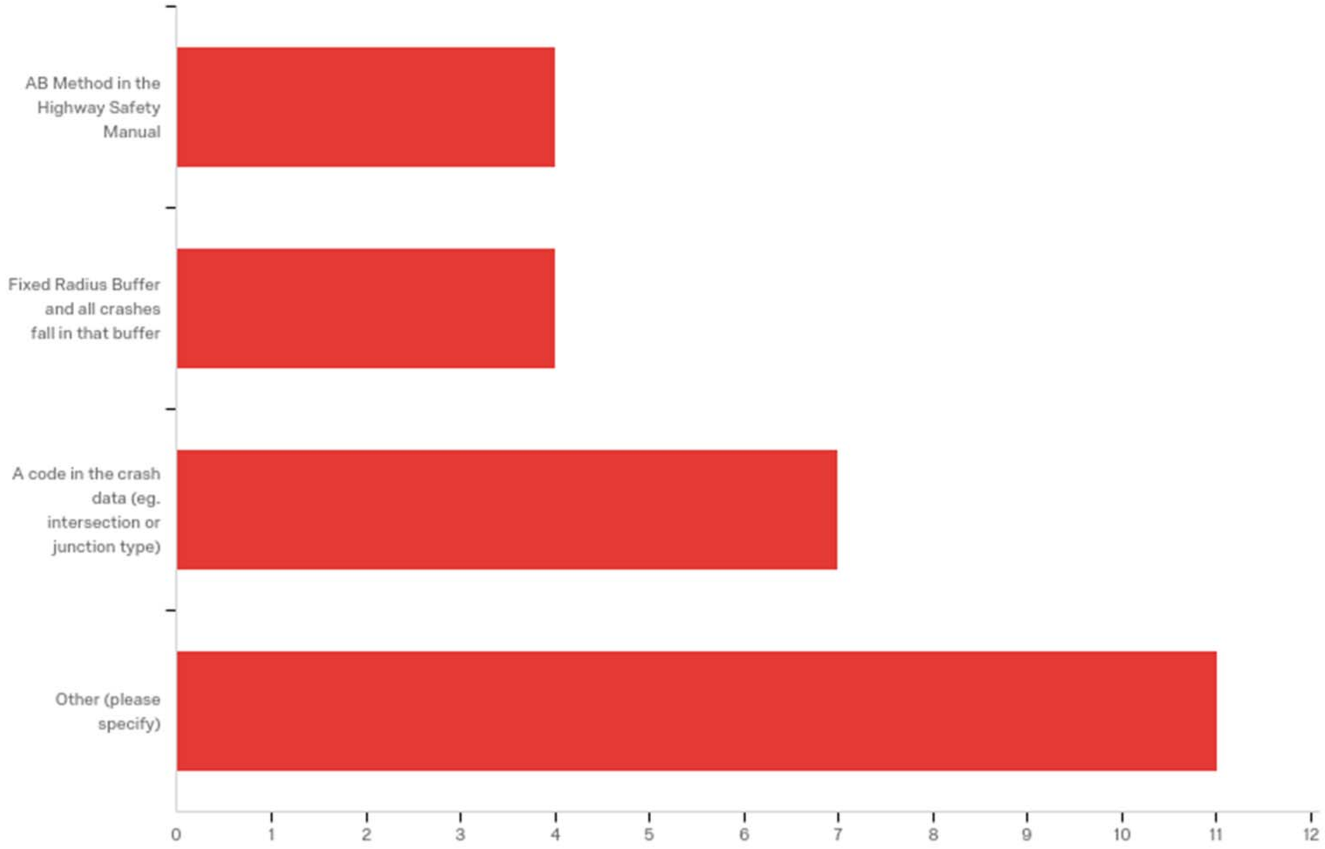
Distance from intersection (please specify distance) - Text
100 ft – Colorado DOT, Maine DOT
varies based on facility type – Kentucky Transportation Center
Safety Priority Index System (SPIS) – ODOT
Within 132 ft of the intersection or officer designated intersection
1/10th of a mile on each leg is counted as part of the intersection unless it encompasses another intersection. – ARDOT
Depends on the intersection type (75ft, 150ft, 300ft). – Iowa DOT
250 ft
40m
Varies based on speed
There is an Intersection-Related field used to separate the crashes – Idaho Transportation Department
Intersection crash: Intersection is True & Intersection ID is not Null. Segment crash: Intersection is False – Louisiana DOT
Sliding 0.3 miles. – VTrans

31. For network analysis, which segmentation methods do you typically use? Please select all that apply.



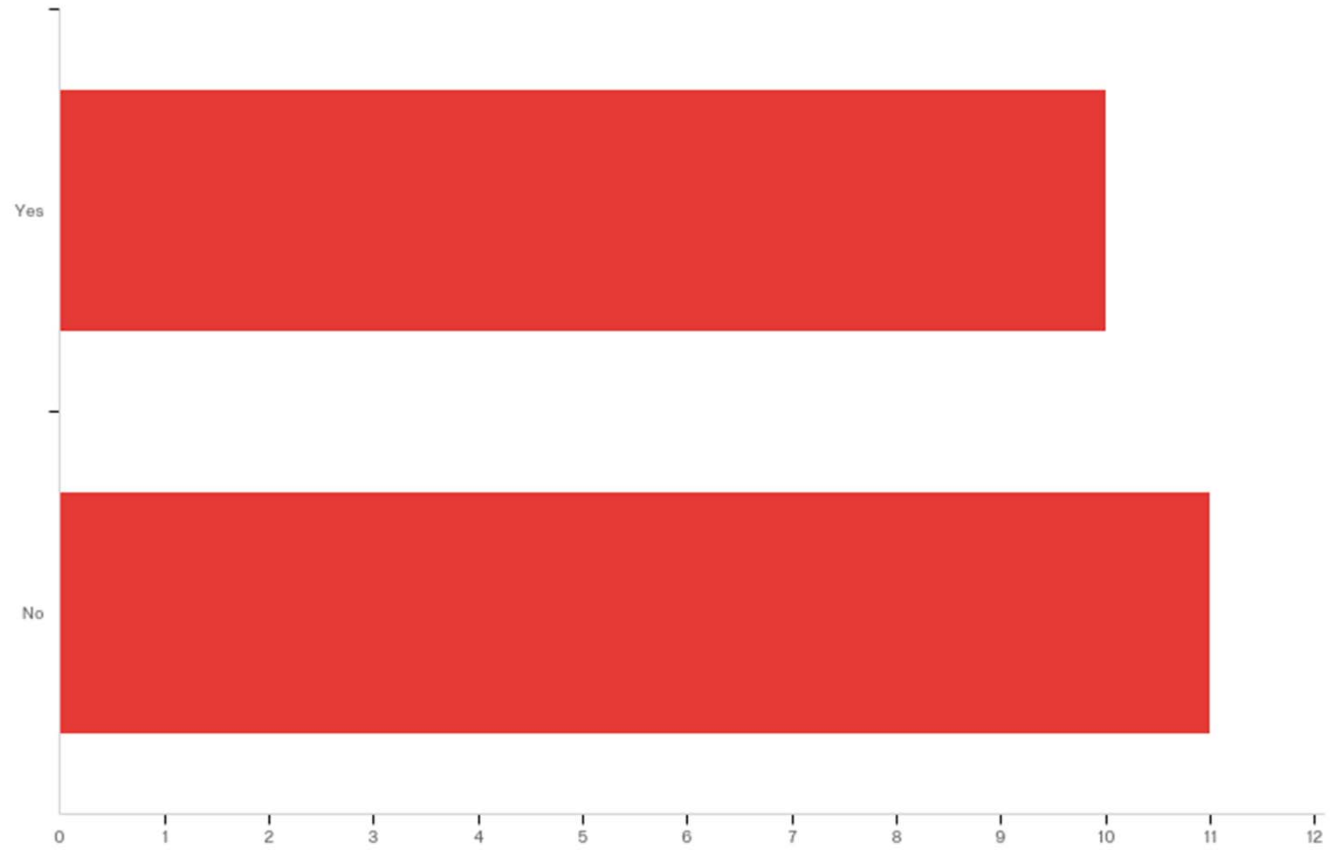
Other (please specify) - Text
Not sure as engineers do the screening but assume Fixed – ODOT
based on changes in attributes – MassDOT

33. What methods do you use for the network segmentation? Please select all that apply.



Other (please specify) - Text
unsure what this question is asking – Kentucky Transportation Center
Variably length roadway segments and intersections. – Maine DOT
Sliding window on defined segment lengths. – ARDOT
segment based upon similar highway class and similar (less than 200% difference) AADT – Louisiana DOT
AADT, Speed are used for defining separate segments – Idaho Transportation Department
Homogeneous segment – Wisconsin DOT
Don't under the question in relation to our process. – Montana DOT
This is becoming automated – MassDOT
Downstream of intersection A to upstream of intersection B – Indiana DOT
Not sure I believe fixed radius buffer and all crashes that fall into that, crash severity, frequency – ODOT

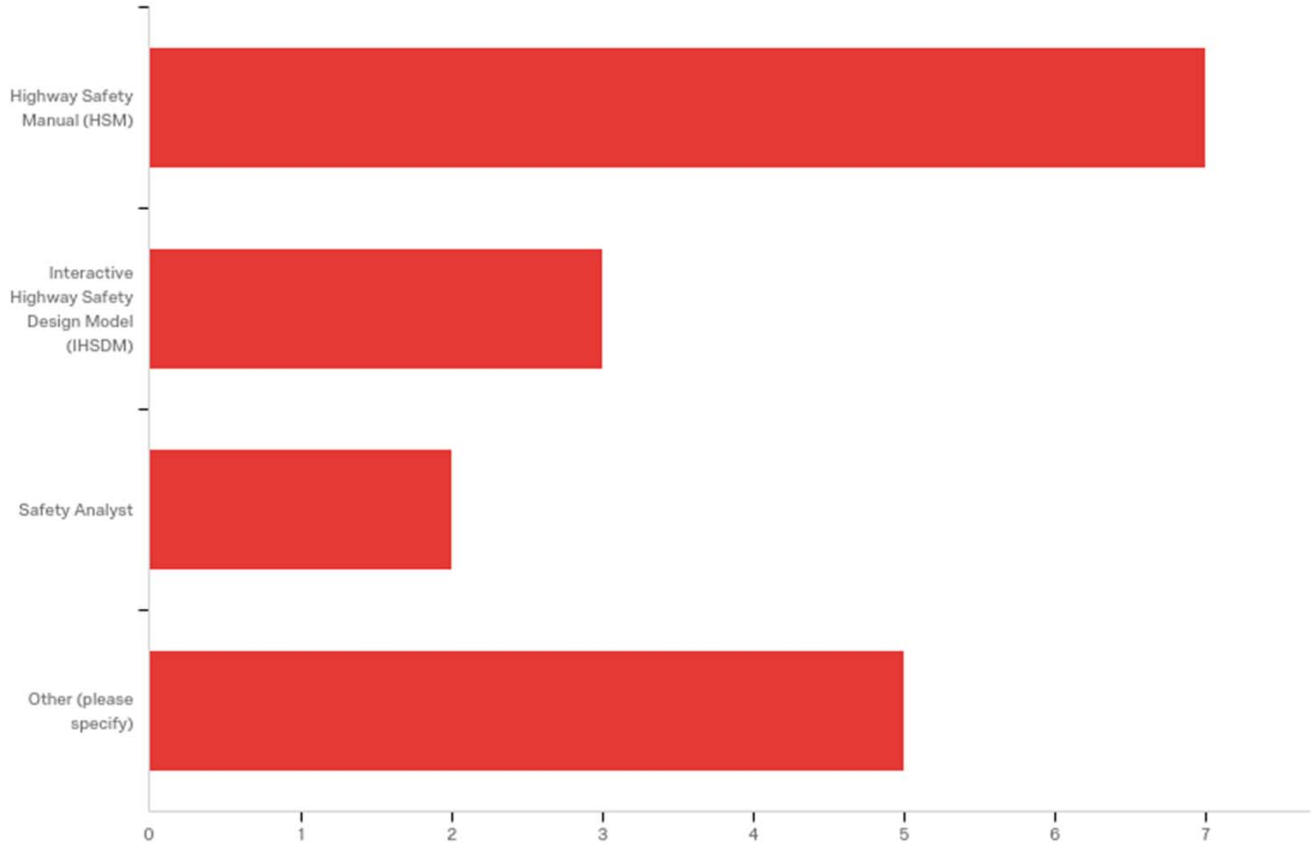
34. Have you adopted new safety predictive methods for any road or intersection types? (21 Responses)



34. Have you adopted new safety predictive methods for any road or intersection types?

Answer	Count
Yes	10
No	11
Total	21

35. Which safety predictive method do you use? Please select all that apply. (17 Responses)



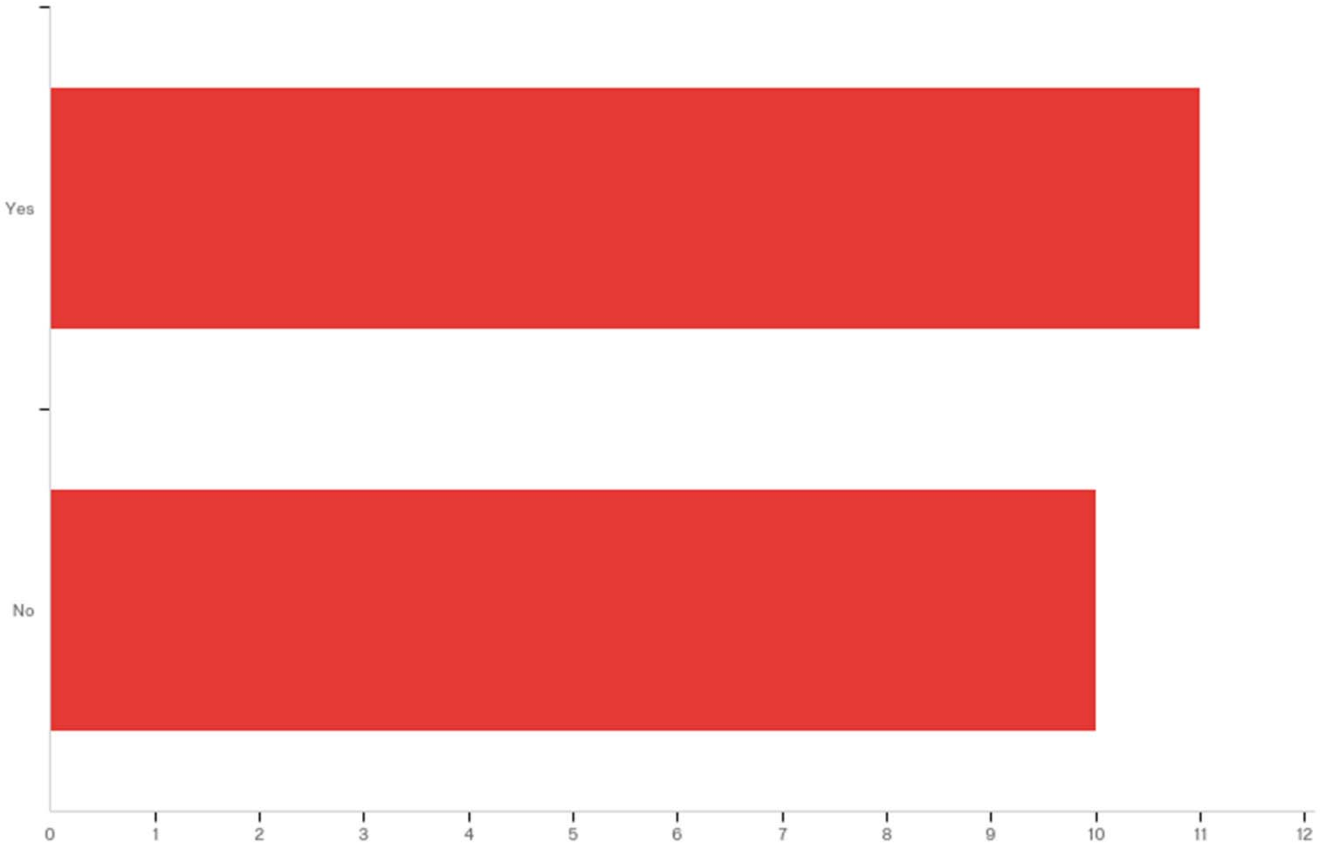
6.10 - 35. Which safety predictive method do you use? Please select all that apply.

Answer	Count
Highway Safety Manual (HSM)	7
Interactive Highway Safety Design Model (IHSDM)	3
Safety Analyst	2
Other (please specify)	5
Total	17

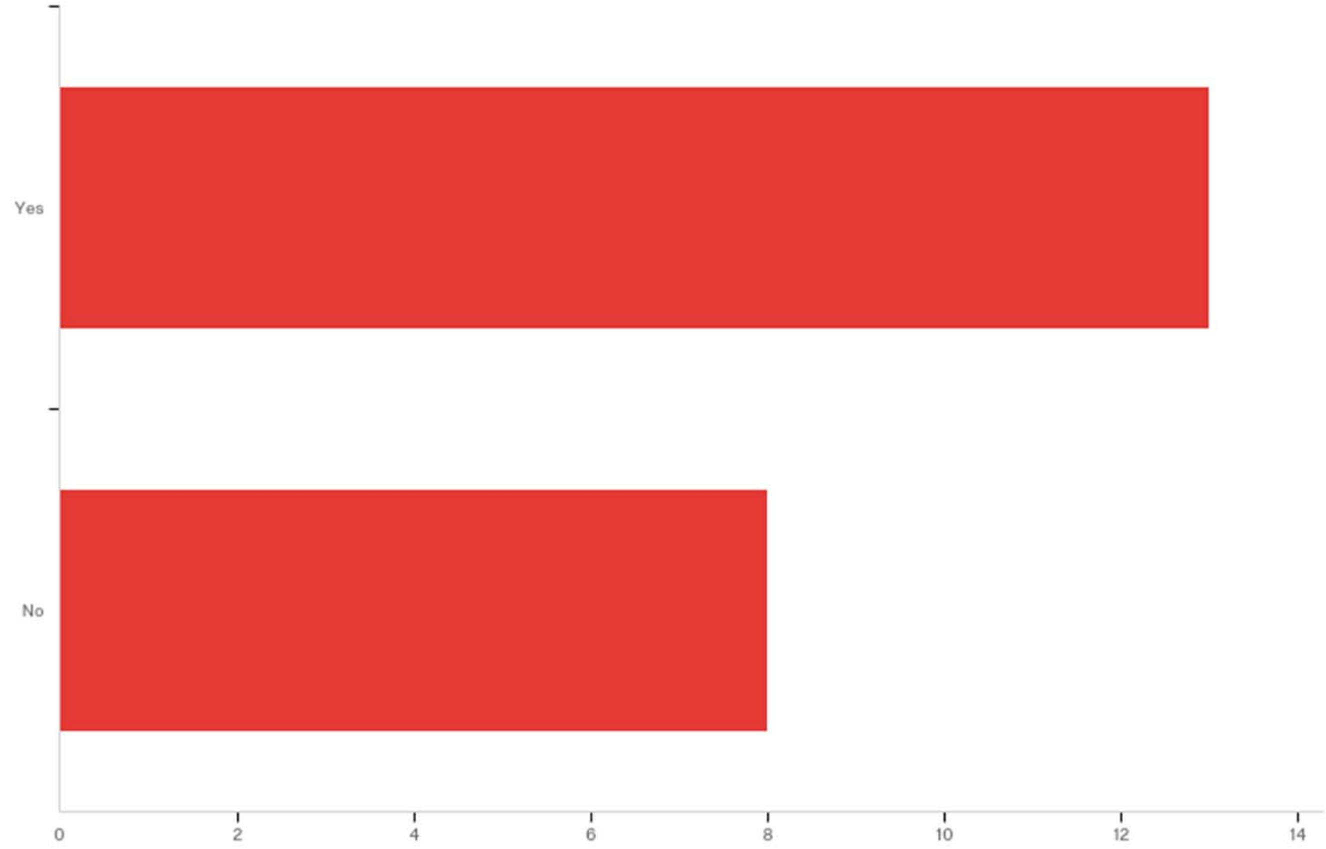
35. Which safety predictive method do you use? Please select all that apply.

Other (please specify) - Text
Vision Zero Suite. Traffic Engineering Software by DiExSys – Colorado DOT
MoDOT continues to develop it's own tool similar to safety analyst.
But we developed our own sreening level SPFs for segmeents and design level SPFs for intersections
Standard deviations from norm factoring in exposure (traffic volume), facility type (e.g. freeway), and setting (i.e. urban or rural), among others.
I think HSM - ODOT

36. Have you developed calibration factors for any road or intersection types? (21 Responses)



37. Do you have state-specific safety performance functions for any road or intersection types? (21 Responses)



Appendix B

HSM Methods Comparison Tables

Please see supplemental file