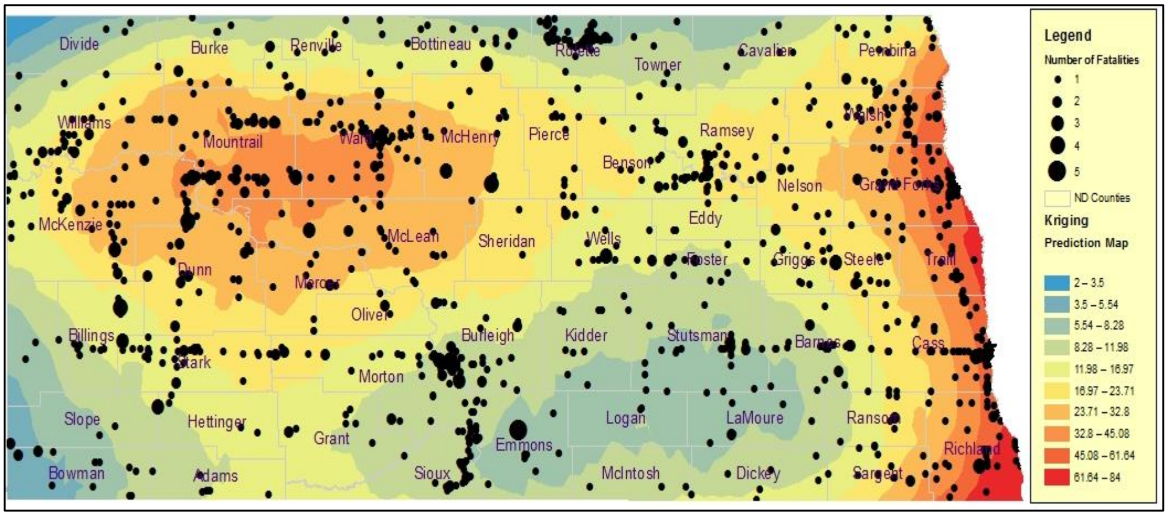
GEOSTATISTICAL APPROACH TO DETECT TRAFFIC ACCIDENT HOT SPOTS AND CLUSTERS IN NORTH DAKOTA



Prepared by:

Mohammad M. Molla, Ph.D. Student
Matthew L. Stone, Ph.D.
Department of Civil and Environmental Engineering

EunSu Lee, Ph.D. (corresponding author)
Upper Great Plains Transportation Institute

North Dakota State University, Fargo

UGPTI Department Publication No. 276
November 2014

Disclaimer

The contents presented in this report are the sole responsibility of the Upper Great Plains Transportation Institute and the authors.

ABSTRACT

This study investigates geostatistical approaches using ordinary Kriging and clustering analysis for traffic crash data in order to identify the traffic accident hot spots in North Dakota counties. The research analyzes 37 years (1975–2011) of fatality crash data available from a U.S. National Highway Traffic Safety Administration official. The resulting geostatistical and tracking analyst model, based on single linkage method clustering, discovered significant facts and features and yielded critical threshold zones for higher accident prone counties/areas in North Dakota counties. Therefore, there is an avenue for improving traffic safety; consequentially, the outcomes of this research would facilitate the efforts or relevant agencies/industry in order to make better decisions for traffic safety planning, and administration.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

The U.S. Census Bureau (2012) stated, "North Dakota is Nation's Fasted-Growing State since 2011. The total populations of North Dakota climbed by 2.17 percent between July 1, 2011, and July 1, 2012. This is the fastest growth of any state, and nearly three times faster than the nation as a whole." (U.S. Census Bureau, 2012). Forbes (2012) stated, "North Dakota leads list of America's Fastest-Growing States" (Forbes, 2012). There is no doubt of the economic prosperity and growth of this state. However, it might be necessary to be more cautious about accident prevention in specific county/regional areas in this state, as this may lead to working strategically toward a goal of zero fatalities on North Dakota roadways.

Thirty-seven years of data revealed a startling increase in fatalities in spite of transportation agencies working for the maximum betterment of the community. Fatality trends in North Dakota from 1975-2011, shown in Figure 1.1, revealed that with the increase of vehicle miles traveled, the number of fatalities in recent years has also increased. It was also discerned that traffic crash fatalities is starting to increase more rapidly than the increase in vehicle miles traveled (VMT) in North Dakota. This exponential increase of traffic fatalities may reach the all-time high set in the year 1976. Larsen (2010) in the Philadelphia Traffic Accident Cluster analysis using GIS and SANET said that road traffic crashes are interrelated with the increase of population and number of vehicles. Vachal and Malchose (2009) said that, in North Dakota, traffic crash fatalities is one of the main causes for young deaths in the state.

Figure 1.2 illustrates the total number of traffic crash fatalities for states from 1980 to 2011. The total number of fatalities was aggregated according to the National Highway Traffic Safety Administration's Fatality Analysis Reporting System (FARS) published data. During this period, there were 1,343,632 fatalities in 48 states, with California, Texas, and Florida accounting for 342,746 fatalities, or 25.5% of the total fatalities. The fewest fatalities were in North Dakota, South Dakota, Nebraska, Wyoming, Montana, Idaho, Utah, and Nevada. Based on the total number of traffic crash fatalities, North Dakota ranked 46th among 48 states, which accounts for 3,439 (0.26%) fatalities from 1980 to 2011.

However, normalized fatality rate based on VMTs or per people change state ranking dramatically specifically for the study area. The total number of VMT (in millions) was found to be 76,384,256 in 48 states from 1980 to 2011. California and Texas accounted for a total of 14,586,193 million VMT, which is 19.10 percent of total vehicle miles traveled. Based on total vehicle miles traveled, the highest number of fatality are in Montana, New Mexico, Arkansas, Louisiana, Mississippi, West Virginia, and South Carolina, which ranges from 2.28 to 2.65 fatalities per 100 million vehicle miles traveled. The lowest number of fatalities per 100 million vehicle miles traveled is found in such as Washington, Minnesota, Virginia and so on. North Dakota is 30th among 48 states, accounting for 1.60 fatalities per 100 million vehicle miles traveled from 1980 to 2011.The highest number of fatalities per population is in Wyoming and Mississippi, which ranges from 753 to 904 fatalities per 100,000 peoples. North Dakota ranked 17^{h} among 48 states, accounting for 503 fatalities per 100,000 peoples from 1980 to 2011.

Therefore, it can be seen that the traffic crash fatalities is in critical situation in North Dakota and merits for further research. In this regards, the specific goals of this research are set to:

- Detect the traffic accident/incident hot spot in North Dakota using Geographic Information System (GIS).
- Cluster traffic accident/incident hot spots using ordinary Kriging (GIS) and Single Linkage Method (SAS) in North Dakota.

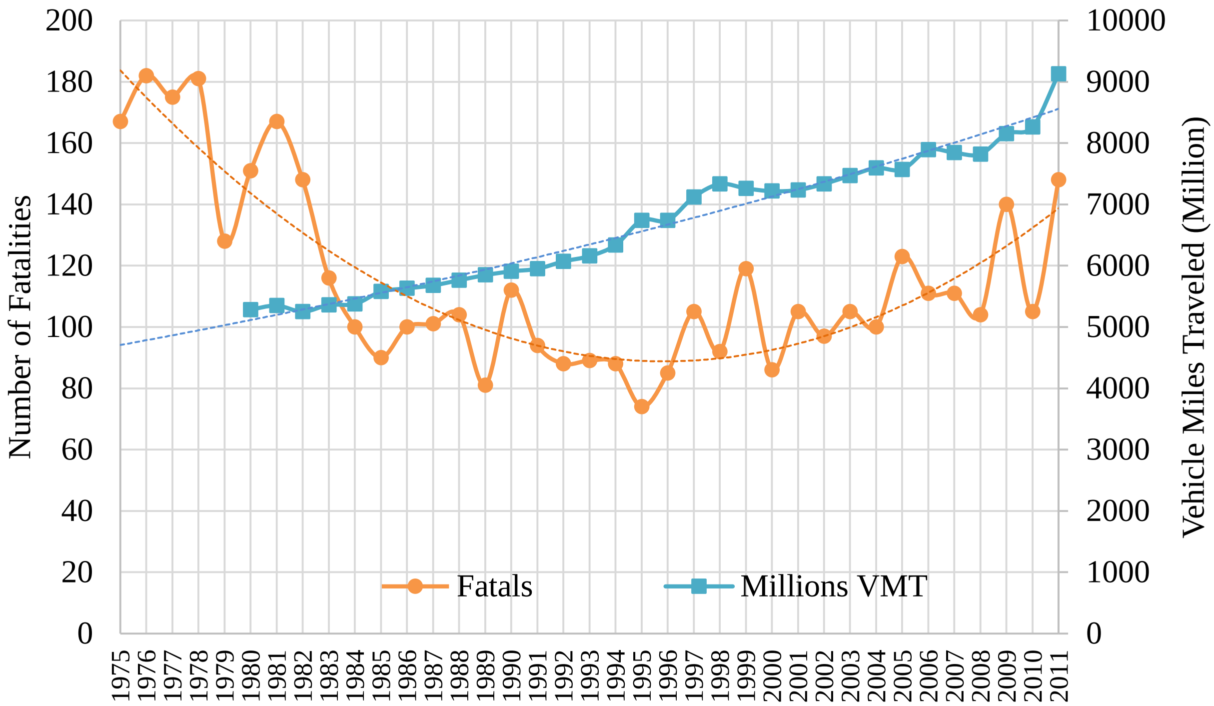


Figure 1.1 Trends of Fatalities and VMT in North Dakota (1975-2011).

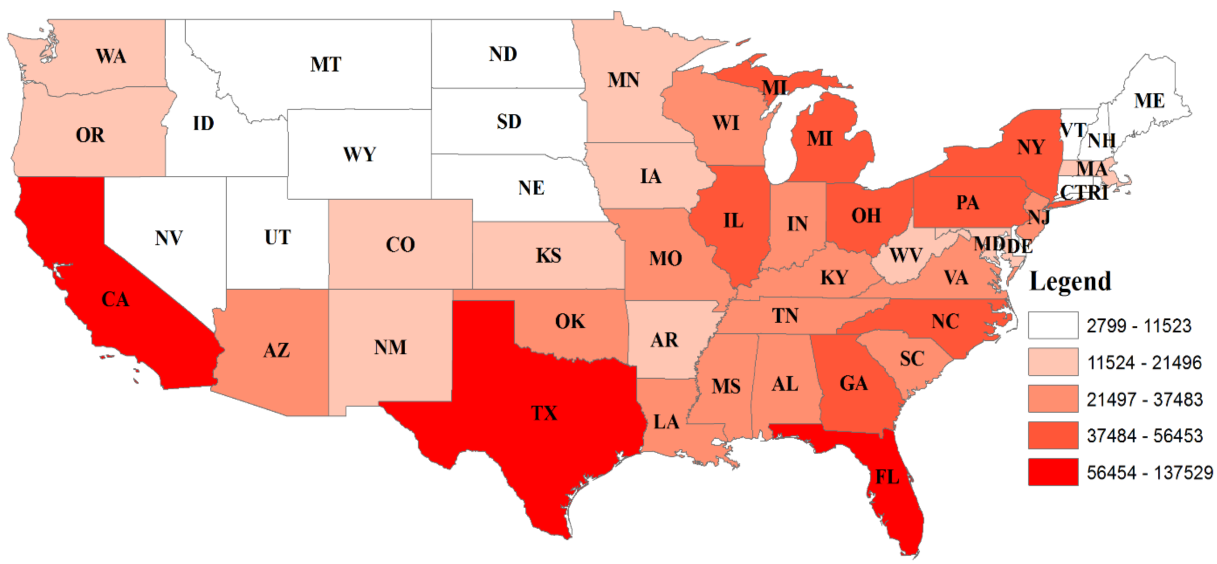


Figure 1.2 Total Numbers of Fatalities (1980-2011).

2. LITERATURE REVIEW

In the multivariate fuzzy road network environments, well advanced predictions for location and time of traffic crashes resulting in fatalities are not easy to quantify over the short run. However, over a long period, the distribution of accidents through identification of hot spot pattern might form a spatial patterns. Consequentially, a cause-and-effect relationship could be inferred based on the traffic crash hot spots, which will provide more information about fuzzy road networks. Moreover, the formation of spatial patterns over land use can be identified through different methods. Clustering of counties' boundaries could be developed together so that policy makers, implementers, and, eventually, law enforcement agencies could share in the benefits for further betterment. From that point of view, an extensive literature review was conducted in order to achieve the highest interests of this research. There are many relevant articles on traffic crash hot spot clustering but nothing was found specifically for North Dakota. Thus, it will be new and unique research for the North Dakota region.

Traffic accidents and crime occurrence are well-defined threats to public safety (Kuo et al., 2013). They said that using data-driven procedures, police departments would assign constraint resources efficiently in order to help crime and traffic crash safety, which substantially reduces the crime and crashes in the hot spot areas. Mitra (2008) said "identification of crash "hot spots," "black spots," "high risk", or "high collision concentration locations" as standard practice in departments of transportation throughout the united states to ensure efficient allocation of safety dollars in reducing crash frequencies and severities. There are fairly good numbers of literature focused on methods for identifying "hotspots" utilizing advanced statistical methods."

USDOT (2013) cited "Some of life's greatest lessons come when we learn from our mistakes, and in transportation, where safety is of paramount importance, that maxim is all the more true." Despite the reality, USDOT and FHWA are strongly bound to reduce highway fatalities and injuries. Consequentially, according to FHWA (2012), highway fatalities and injuries has been reduced substantially from 2007 to 2010 because highway safety programs had a positive effects on the nation. According to traffic safety facts, it might be inferred that after gathering traffic fatality crash data since 1975, traffic fatalities declined for seven consecutive years (NHTSA, 2010).

Contrary to the above statistics, the literature review revealed some interesting facts about traffic crash fatalities on U.S. highways. Spear (2008) mentioned that state reports compiled by the National Highway Traffic Safety Administration reveal that every highway crash involved a fatality. FBI (2003) said by some estimates, 40% of Americans would be associated with an alcohol-related crash at some time in their lives. The U.S. Department of Transportation is standardizing practices in order to detect traffic crash hot spots/black spots (Cheng and Washington, 2005). Unsurprisingly, traffic crashes become heterogeneous in frequency and rate due to the heterogeneity of transportation road network variables. Chen et al. (2013) studied the safety countermeasures and crash reduction in New York City. Traffic fatalities and injuries have resulted in a serious health problem globally. Their studies documented that developing countries are ahead of the United States on traffic safety, though New York City resembles an ideal model in the United States. NDDOT is working with traffic crashes by challenging the problem strategically. The creation of new safety division in 2009 makes traffic safety as a top priority for continuous improvement in North Dakota (NDDOT, 2012).

Several methods have been developed in order to identify the possible traffic crash clustered hot spots for a certain region. Xie and Yan (2013) stated that, according to Moon et al. (2009), in the four stages of traffic safety management, traffic crash hot spot is a key item. Anderson (2009) also reiterated that traffic crash hot spot identification is the first primary stage for proper safety management. Ramli (2011) said highway safety studies are frequently using a statistical or crash forecasting model. The most influential attributes with their relationship between crashes and crash variables can be determined. Li et al. (2007) uses a geographic information system and spatial temporal pattern analysis of intra-city motor vehicle crashes in Texas, and identified and ranked roadway components based on risk factors related to crashes. Traffic crash hot spots/black spots identification and clustering are widely used; these include hot zone identification with GIS-based post-network screening analysis (Young and Park, 2014), applying linear analysis methods to GIS-supported procedures for preventing traffic accidents: Case study of Konya (Gundogdu, 2010), GIS-based Economic Cost Estimation of Traffic Accidents in St. Louis, Missouri (Yang et al., 2013), Spatio-Temporal Clustering of Road Accidents: GIS Based Analysis and Assessment (Prasannakumar et al., 2011), Kernel density estimation and K-means clustering to profile road accident hot spots (Anderson, 2009), Geographical information systems aided traffic accident analysis system case study: the City of Afyonkarahisar (Erdogan et al., 2008).

3. STUDY AREA AND DATA

The main study area for this research was North Dakota, a state in the U.S. midwest regions. The number of fatalities data were aggregated according to the National Highway Traffic Safety Administration's Fatality Analysis Reporting System (FARS) published data; the number of VMT were summarized according to the U.S. Department of Transportation – Federal Highway Administration (Office of Highway Information) released data. Population data were collected from Fact Monster web portal, where the actual data source was the U.S. Census Bureau. The GIS shape files were downloaded from ESRI and ND GIS Hub. FARS data were stored on a yearly basis and were geoprocessed into ArcGIS environments and statistical SAS language to achieve different outcomes of the research object. Yearly VMT data collected from Federal Highway Administration data were combined and organized using Microsoft Excel.

3.1 Traffic Fatality Data

The latitude and longitude coordinate details for each fatal crash event in the FARS data set were only available from 2001 to 2011. The crash data for 2001 to 2011 were spatially joined with the North Dakota boundary shape file and created a choropleth map to compare the North Dakota counties. Figure 3.1 illustrates the fatal crashes. From 2001 to 2011, the map indicated that six counties (Ward, Cass, Grand Forks, McKenzie, Rolette, and Mountrail) had a total number of fatalities greater than 50. Ward County had the highest number of fatalities with 84, presented in Table 3.1.

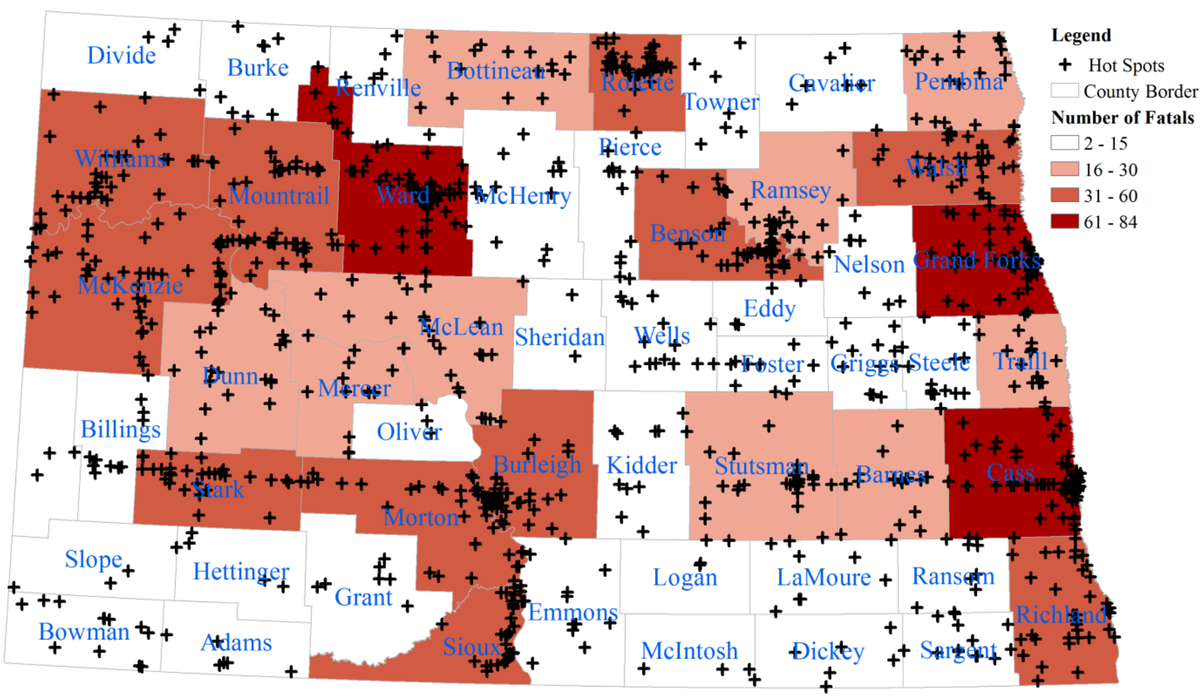


Figure 3.1 Fatalities of North Dakota Counties (2001-2011).

Cass County was the second highest in fatalities with 79. Studies revealed there is a big difference in fatalities among counties in North Dakota because just within these 11 years, Eddy, Golden Valley, and Oliver indicate less than three fatalities for/within the 11 years.

Table 3.1 The Number of Fatalities (2001-2011)

County	Number of Fatalities	County	Number of Fatalities
Ward	84	Steele	15
Cass	79	Billings	14
Grand Forks	70	Bowman	14
McKenzie	66	Griggs	14
Rolette	64	Pierce	12
Mountrail	59	Nelson	11
Burleigh	50	Foster	10
Benson	48	Kidder	10
Morton	42	Burke	9
Richland	42	LaMoure	9
Sioux	41	Ransom	9
Williams	39	Grant	8
Walsh	37	Cavalier	7
Stark	36	Dickey	7
Stutsman	32	Hettinger	7
McLean	30	Renville	7
Barnes	27	Towner	7
Dunn	27	Adams	6
Ramsey	26	Divide	4
Pembina	21	Logan	4
McHenry	20	Sheridan	4
Bottineau	19	Slope	4
Wells	19	Eddy	3
Mercer	17	Golden Valley	3
Trail	17	McIntosh	3
Emmons	15	Oliver	3
Sargent	15	State Total	1,246

4. RESEARCH METHOD

In this paper, we are attempting to identify the existence of clusters of traffic crash fatalities using SAS Programming and visualize the resulting clusters using ArcGIS ordinary Kriging method. Actually single linkage clustering was performed in order to verify the ordinary Kriging method. In order to achieve the outcomes, we have followed the following steps.

1. Examining the distribution of data and transformation
2. Identifying global trends and Anisotropy of data
3. Conducting hot spot analysis using ordinary Kriging
4. Testing cluster methods and ranking the methods
5. Clustering counties with the number of fatalities
6. Developing severity index

4.1 Examining the Distribution of Data

Initially, the distribution of the data was examined in order to fit the best model to use for prediction. It is clear from the histogram presented in Figure 4(a) that the total fatalities by North Dakota county were skewed right. The quantile-quantile (QQ) plot in Figure 4(b) did not indicate a straight line, thus, the data were not normally distributed. Therefore, it is necessary to transform the data before creating any surface mapping. Nevertheless, it is clear that data are unimodal (one hump) and skewed to the right. Thereafter, a log transformation of the data in Figures 4(c) and 4(d) further improves and indicates the normal distribution of the data.

4.2 Identify Global Trends and Anisotropy of Data

The transformed data indicated the existence of global trends or local variation. Each vertical stick in Figure 4.1 showed each location, representing county, and the height of each stick represents the scale measurement of fatality. Best fit polynomial lines were also added into Figure 4.1. The green and blue lines were not flat, which represents the existence of directional variation. Initially, the green lines start at a point, then increase over distance when it moves over the *x*-axis and, similarly, blue lines maximize at the middle. Therefore, it can be inferred that a strong trend in fatalities is seen over all the edges and center region of the state. The possible reason might be high density of the population of those areas. The semivariogram presented in Figure 4.2 shows that adjacent data points are statistically correlated. Since the distance is increasing, the likelihood of the observations becomes smaller. Therefore, directional influence might be considered for the interpolated surface creation.

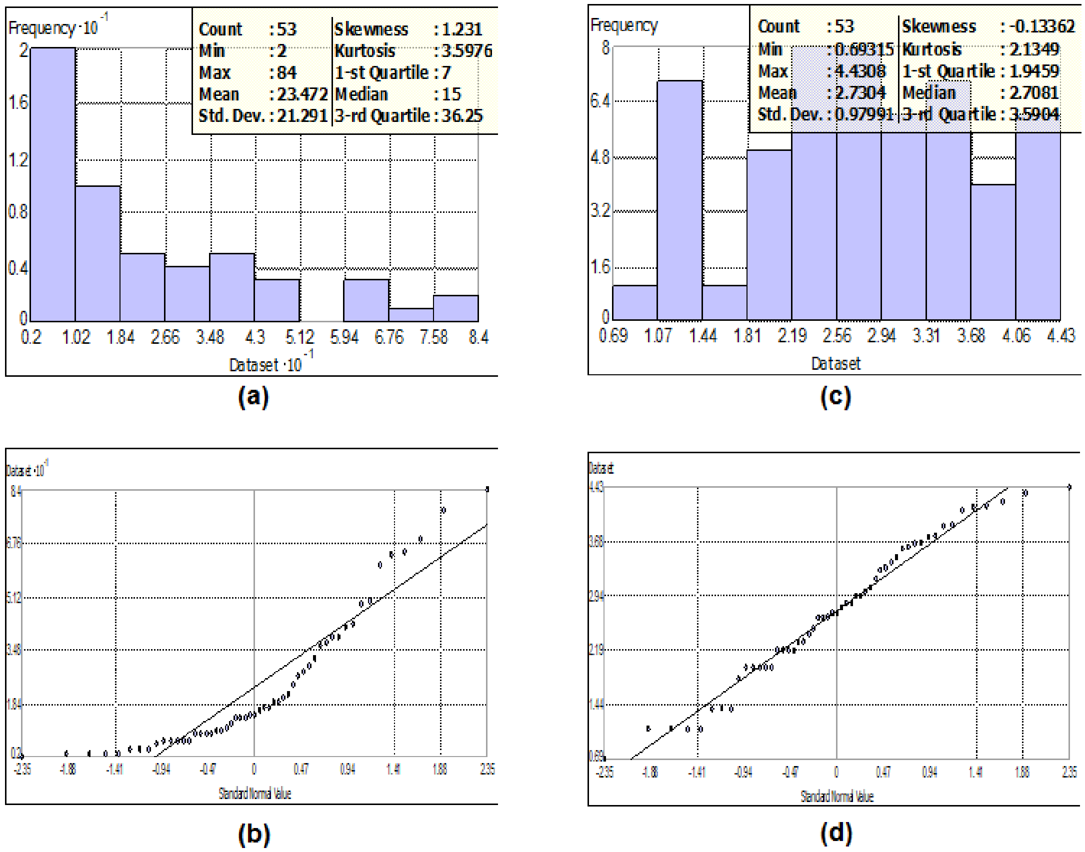


Figure 4.1 Examining the Distribution of Data and Transformation.

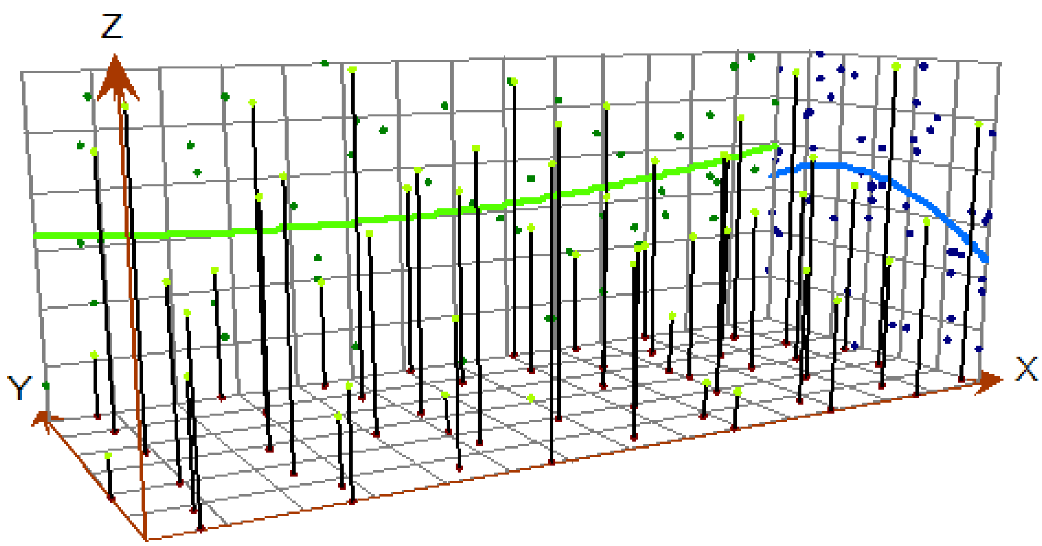


Figure 4.2 Global Trends of County Log Fatalities throughout North Dakota.

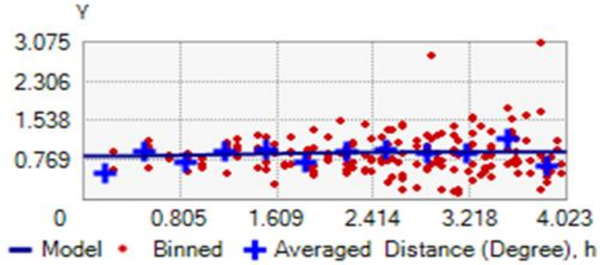


Figure 4.3 Semivariogram of County Fatalities.

4.3 Hot Spot Analysis Using Ordinary Kriging

This research study adopted the ordinary kriging methodology. Hill et al. (2009) described that ordinary Kriging can be used in order to identify the appropriate unbiased linear stationary model when the fixed mean is not known. Savelieva (2005) stated that the well-known ordinary Kriging can be used as a true candidate in order to enhance the decision-making process. Xie and Yan (2013) said, "There are many relevant studies in the literature on traffic accident hot-spot analysis." Manepalli and Bham (2011) suggested that ordinary Kriging performed better than the empirical Bayes (EB) method to forecast the traffic crash frequencies and the severity index. Wang and Kockelman (2009) performed forecasting on network data by interpolating Texas traffic count data, and in this study, they found that Kriging serves weight better than any other spatial extrapolation method. Other data that can investigated smoothly using the Kriging method include Annual Average Daily Traffic (AADT) estimation, pavement situations, traffic speeds, intensity of populations, generation of trip rates, valuation of lands, and household income data. Bolstad (2012) said, "Kriging is unique and powerful because we use the observed change in spatial autocorrelation with distance to estimate values at our unknown locations."

Figure 4.4 presents the clustered traffic fatalities in North Dakota. The map was created with ordinary Kriging considering a log transformation of fatality data. Global trends for third order polynomial and anisotropy/directional influence were also considered in the prediction map. The ordinary Kriging performed linear predictions for each county with weighted average of neighborhood data points. Ordinary Kriging reveals that some of the red zones such as Cass, Grand Forks, Ward, Mountrail, Richmond, and McKenzie show critical density of fatal traffic accidents, while Bowman, Emmons, Logan, LaMoure, and Stutsman indicate cold spots with low fatal crash density. Figure 4.4 revealed that some counties such as Rolette and Burleigh, were not identified as high fatality prone areas with ordinary Kriging though those points seem to be critical areas of interest. The possible reason for not including those counties might be subsequent prediction errors of the model.

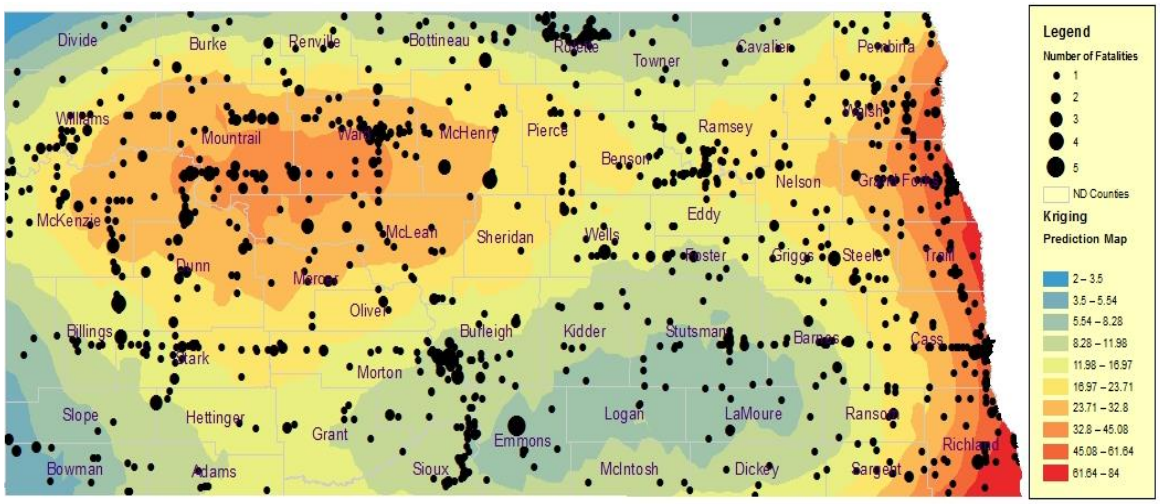


Figure 4.4 Hot Spot Analysis using Ordinary Kriging.

The prediction plot (Figure 4.5) indicates that with the auto correlation and good kriging, the model (blue lines) has been deviated from a 45-degree straight line. Therefore, the significance of ordinary kriging provides well-informed prediction advantages and may require further investigation of those locations with statistical univariate clustering analysis.

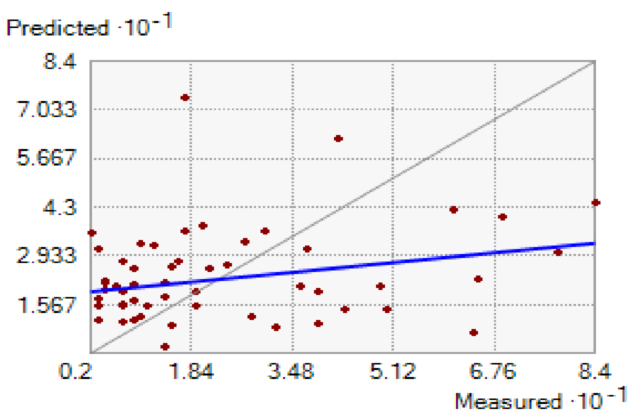


Figure 4.5 Fitted Prediction Plot.

4.4 Selection of Clustering Method

After comparing 12 available methods, the single linkage method proved to be the most suitable one to compute the number of clustering because all other 11 methods showed unstable statistics for cubic clustering criterion, Pseudo-*F*, and *t* values. Rencher (2001) defined that in the single linkage method, the distance between two clusters A and B is defined as the minimum distance between a point in A and a point in B (where distance is the Euclidean distance) or some other distance. This approach is also called the nearest neighbor method.

Table 4.1 Cluster History

Number of Clusters	Frequency	Semi-partial R-Square	R-Square	Approximate Expected R-Square	Cubic Clustering Criterion	Pseudo F Statistic	Pseudo t-Squared	Norm Minimum Distance	Tie
15	237	0.0020	.993	.996	-11	11E3	99.2	0.0717	Tie
14	303	0.0060	.987	.995	-23	6357	271	0.0717	Tie
13	108	0.0002	.987	.994	-20	6804	.	0.0717	Tie
12	56	0.0001	.987	.993	-16	7381	.	0.0717	Tie
11	147	0.0003	.986	.992	-12	7945	1136	0.0717	Tie
10	213	0.0014	.985	.990	-10	7990	824	0.0717	-
9	124	0.0009	.984	.988	-6.5	8460	1032	0.1076	-
8	171	0.0015	.983	.985	-3.0	8811	1557	0.1435	Tie
7	337	0.0146	.968	.980	-11	5515	1735	0.1435	Tie
6	154	0.0009	.967	.973	-4.5	6427	.	0.1435	-
5	640	0.1254	.842	.960	-34	1458	2649	0.1794	Tie
4	218	0.0035	.838	.938	-24	1895	444	0.1794	-
3	730	0.0705	.768	.889	-20	1816	330	0.2152	-
2	948	0.3602	.407	.750	-29	756	1473	0.3228	Tie
1	1102	0.4075	.000	.000	0.00	.	756	0.3228	-

4.5 Selection of Number of Clusters

Using the single linkage clustering procedure, a hierarchical cluster of observation was initially determined. Table 4.1 provides the results portraying the last 15 generations of the cluster history for the single linkage method. Table 4.1 also includes the number of clusters, clusters joins, and some statistical parameters.

The cubic clustering criterion, Pseudo-*F*, and *t* statistics are used to determine the number of cluster for the data. SAS (2008) discussed that the CCC and PSF are not suitable in order to identify the number of clusters in a single linkage method because the method has a tendency to shear the tails of the distribution. In that case, PST2 could be utilized for this purpose. Initially, any cluster with a large PST2 value should be selected; the number of clusters used in the analysis could be one greater than the initial number of clusters.

Figure 4.6 plots all three statistics based on the number of clusters. From Table 4.1 and Figure 4.6, it is evident that for these data, there are large/peak values of PST2 for clusters 2, 4, 5, and 11. However, in the last column of Table 4.1, based on SAS output (2008), ties exist for levels 2, 5, and 11, which means at these three levels, estimation of number of clusters would be indeterminate and changing the order of the observation could change the clusters. Therefore, at level four, the number of clusters might be suggested as five. Using five clusters, however, resulted in a lack of separation between the clusters.

At level three in Table 4.1, minimum distance is not ties. The schematic plots in Figure 4.6 displayed that all four clusters at level three are apart and well separated. From the plot, is it shown that there are some outliers for the first cluster. Since our initial goal is to identify accident prone counties, low accident prone counties with outliers could be neglected for this limited scope of work. Therefore, with due consideration, it was assumed that undertaking four clusters will be more appropriate and could be validated with Kriging method. Grouping into four clusters, the R-square value at level three was approximately 77%. Thus, it can be interpreted that 77% of the proportion of the variation can be explained by the cluster analysis.

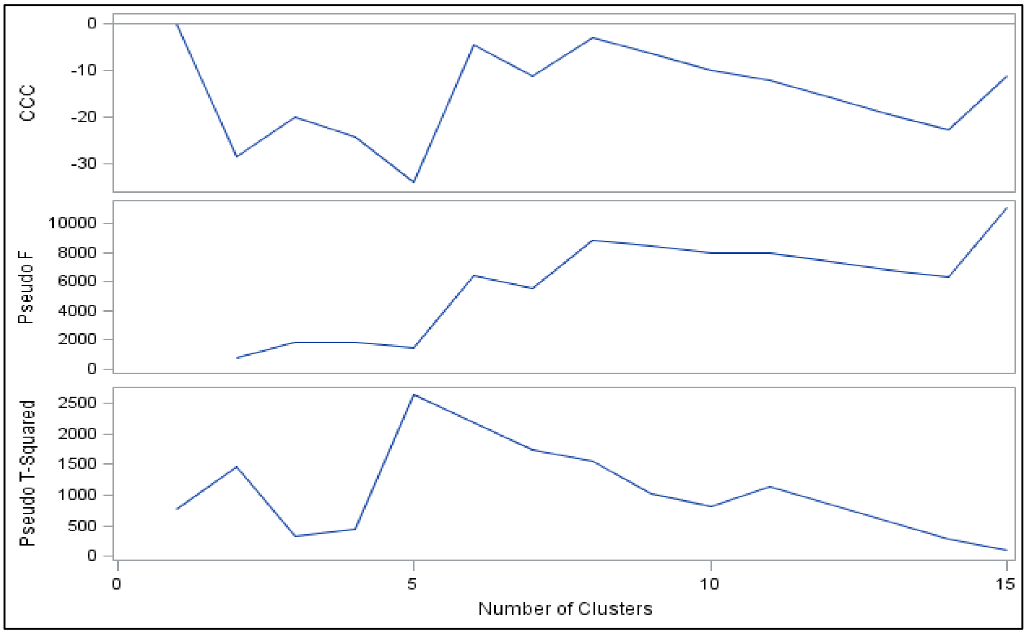


Figure 4.6 Criteria for Number of Clusters (Output using SAS).

```
90 +
      |
      |
      |                                  +-----+
80 +                                     *--+--*
      |
      |
      |
70 +                              |
      |                        +-----+
      |                        *--+--*
      |                        +-----+
60 +                              |
      |
      |
      |
50 +                 *--+--*
      |              +-----+
      |
      |       |
40 +          |
      |       |
      |       |
      |       |
30 +          |
      |       |
      |       |
      |       |
20 +       +-----+
      |    |     |
      |    *--+--*
      |    |     |
10 +       |     |
      |    +-----+
      |       |
      |       |
 0 +
           ---+---   ---+---   ---+---   ---+---
CLUSTER       1         2         3         4
```

Figure 4.7 Schematic Plots of Four Clusters.

4.6 Developing Severity Index

The results of the hierarchical clustering procedure have been presented visually using a tree diagram/Dendogram in Figure 4.8. The Dendogram displays all the stages of the hierarchical procedure and the distances at which clusters were merged. As the number of branches increases from the root to the left, the R-square approaches one. Grouping into four clusters, with R-square value equal to 77%, accounts for over half of the variation. Therefore, only four clusters are required to explain over three-fourths of the variation.

The clustered counties shown in Figure 4.4 are summarized in Table 4.2, where four clusters were categorized based on a Likert scale with four categories; the Likert categories used in the map include low, medium, high, and severe fatalities. We have included the Likert scale in order to facilitate a better understanding for general acceptance by the audience so that traffic crash fatalities can be categorized based on a unique severity index. The scale was designed such that low, medium, high, and severe severity index comprises cluster one, two, three, and four, respectively. Both Figures 4.8 and 4.9 displayed that a majority of the traffic crash fatalities, shown as severe in the severity index, are occurring principally in only two counties: Ward and Cass. It makes sense that both counties have high population densities and are hubs for business in North Dakota. The high severity indexed counties included Grand Forks, McKenzie, Mountrail, and Rolette. The medium severity indexed counties included Benson and Burleigh. The other 45 counties were clustered in the low severity indexed counties.

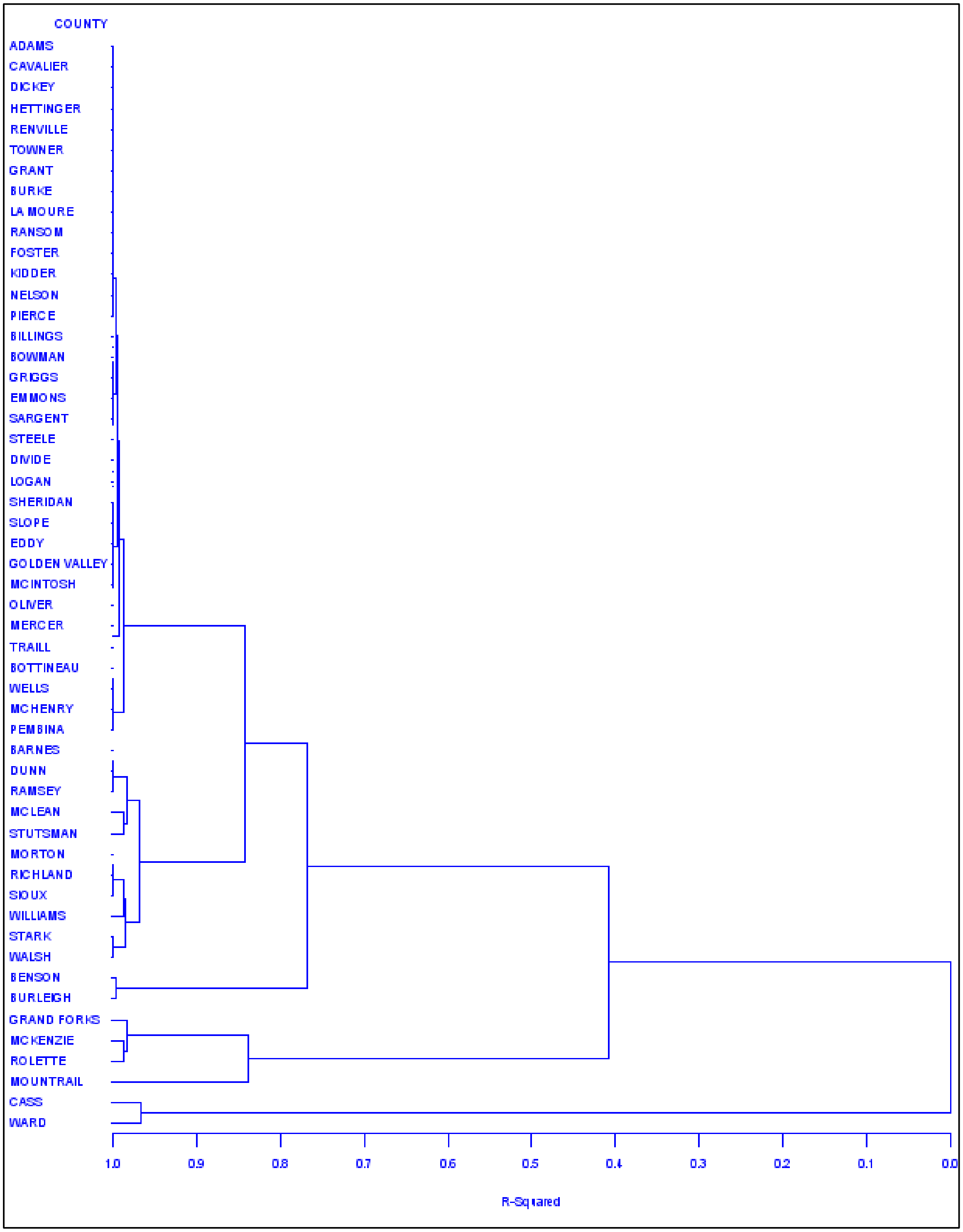


Figure 4.8 Tree Diagram of Clusters versus R-Square Values.

Table 4.2 List of Clustered Counties

Low			Medium			High			Severe		
County	**No. of Fatals**	**Cluster**	**County**	**No of Fatals**	**Cluster**	**County**	**No. of Fatals**	**Cluster**	**County**	**No. of Fatals**	**Cluster**
Adams	6	1	Benson	48	2	Grand Forks	70	3	Cass	79	4
Barnes	27	1	Burleigh	50	2	McKenzie	66	3	Ward	83	4
Billings	14	1				Mountrail	59	3			
Bottineau	19	1				Rolette	64	3			
Bowman	14	1									
Burke	9	1									
Cavalier	7	1									
Dickey	7	1									
Divide	4	1									
Dunn	27	1									
Eddy	3	1									
Emmons	15	1									
Foster	10	1									
Golden Valley	3	1									
Grant	8	1									
Griggs	14	1									
Hettinger	7	1									
Kidder	10	1									
LaMoure	9	1									
Logan	4	1									
McHenry	20	1									
McIntosh	3	1									
McLean	30	1									
Mercer	17	1									
Morton	42	1									
Nelson	11	1									
Oliver	3	1									
Pembina	21	1									
Pierce	12	1									
Ramsey	26	1									
Ransom	9	1									
Renville	7	1									
Richland	42	1									
Sargent	15	1									
Sheridan	4	1									
Sioux	41	1									
Slope	4	1									
Stark	36	1									
Steele	15	1									
Stutsman	32	1									
Towner	7	1									
Trail	17	1									
Walsh	37	1									
Wells	19	1									
Williams	39	1									

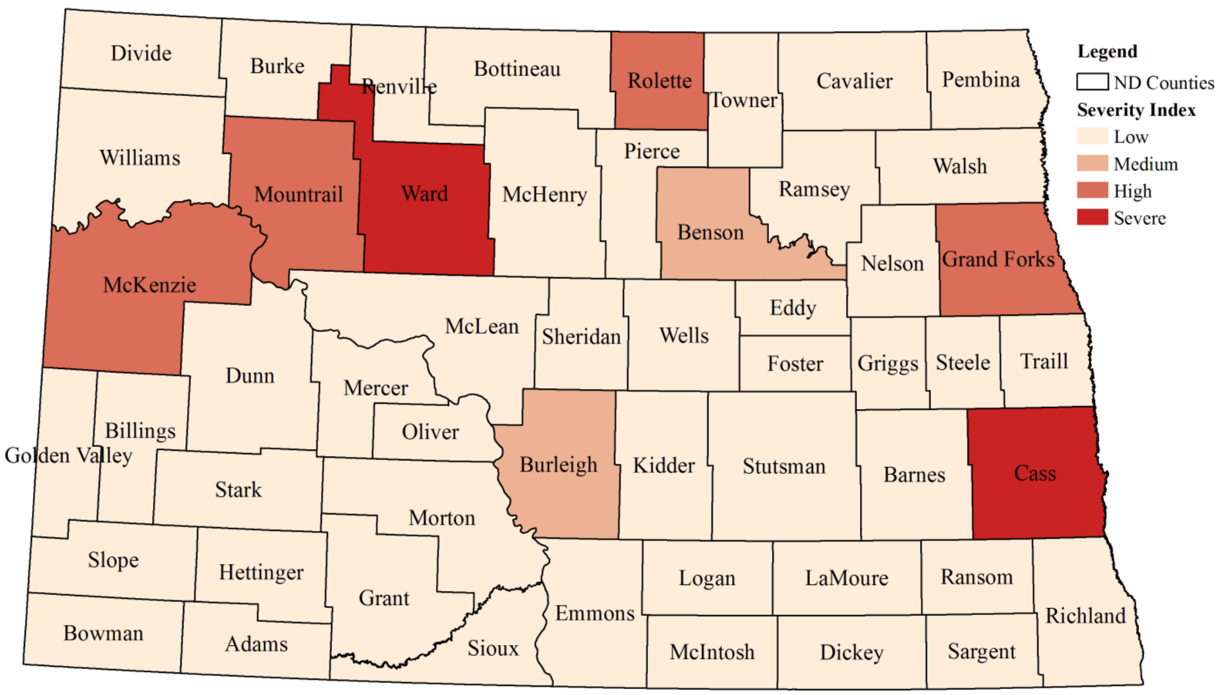


Figure 4.9 Severity Index Mapping of ND Counties.

5. RESULTS, DISCUSSIONS, AND CONCLUSIONS

The peace garden state, North Dakota, the fastest growing in economy parallel to the increasing growth of its population, now has to face the spatial challenge of some counties such as Cass and Ward regarding road safety because of their heavy development. The exponential increase in traffic fatalities in North Dakota, concurrent with the fastest economic growth in the state's history, aggregated with widely spreading traffic accident hot spots, may eclipse the all-time high set in 1976. State transportation agencies and law enforcement agencies are working strategically toward implementing zero fatalities on North Dakota roadways. The research work performed, based on 37 years (1975–2011) of traffic crash fatality data, discovered significant facts and features and yielded critical threshold zones for fatality prone counties/areas in North Dakota.

In order to implement the strategic safety plan, it is vital to identify traffic crash locations with their spatial relationship. Therefore, a geographic information system (GIS) was used to identify the hot spot locations. But before using GIS, the extensive segregated data were aggregated into statistical programming language, which eventually generated summary tables and outputs that have been presented in these studies. Out of 37 years of data, only 11 years of data included the coordinates for fatal crashes. The state department of transportation usually includes five years of crash hot spots in its yearly crash summary report. Therefore, we have tried to include all years and see if there is any spatial pattern among those hot spots. Based on total traffic fatalities of 1,343,632 in 48 states during the period of 1980-2011, North Dakota ranked 46th among all states, which is only 0.26% of total traffic fatalities. But if we normalized the fatalities based on a million vehicle miles travelled (VMT) or peoples, North Dakota ranked 30th among 48 states, which accounts for 1.60 fatalities per 100 million VMT and North Dakota ranked 17^{h} among 48 states, accounting for 503 fatalities per 100,000 peoples during 1980 to 2011. Though the normalized fatalities rate increased the ranking position for this state, interestingly, we found that the majority of traffic crashes are apparently clustered and self-centered based on some counties and interstates 29 and 94. But we cannot draw any inferences based on this apparent visualization. Therefore, the ordinary Kriging was used to predict and interpolate the fatalities in the counties. The ordinary Kriging moderately identified the hot spot region but failed to identify some of the regions/counties such as Rolette and Burleigh. Thereafter, we used a single linkage clustering technique and identified the four best possible clusters among the counties.

The computed clusters revealed there are low to severe traffic fatality crash prone counties in North Dakota. It might be concluded that the single linkage clustering studies significantly support or are in agreement with the Kriging prediction map for traffic crash fatalities hot spot detection and clustering for counties. Both methods agreed that traffic crash fatalities distribution showed spatial clustered patterns. It is strongly believed that the outcomes of this research and the analyzed clusters will provide abundant information that can create an avenue for proper planning, implementation, and decision making of traffic safety management strategies and administration.

Further research might focus on geospatial and clustering analysis of road network components and root causes of traffic fatalities in North Dakota, including vehicle types and weather conditions present when accidents occurred. Factors affecting the spatial distribution on those locations might also be investigated. Research on sates adjacent to North Dakota might be analyzed in order to oversee any existence of spatial relationships with adjacent state boundaries.

REFERENCES

Anderson, T.K. (2009). “Kernel density estimation and K-means clustering to profile road accident hot spots.” *Accident Analysis and Prevention* 41, 359–364.

Bolstad, P. (2012). “GIS Fundamentals, A first text on geographic information systems.” Fourth Edition, Eider Press, White Bear Lake, MN.

Chen, L., C. Chen, R. Ewing, C.E. McKnight, R. Srinivasan, and M. Roe (2013). “Safety countermeasures and crash reduction in New York City-Experience and lessons learned.” *Accident Analysis and Prevention* 50, 312– 322.

Cheng, W. and S.P. Washington (2005). “Experimental evaluation of hotspot identification methods.” *Accident Analysis and Prevention* 37, 870–881.

Erdogan, S., I. Yilmaz, T. Baybura, and M. Gullu (2008). “Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar.” *Accident Analysis and Prevention* 40 (1), 174–181.

FBI (Federal Bureau of Investigation) (2003). FBI Law Enforcement Bulletin 72(1).

FHWA (Federal Highway Administration-U.S. Department of Transportation). (2012).<http://safety.fhwa.dot.gov/provencountermeasures/pc_memo.htm> (Accessed December 2013).

Forbes (Forbes, LLC.) (2012). <http://www.forbes.com/sites/morganbrennan/2012/12/20/north- dakota-leads-list-of-americas-fastest-growing-states/> (Accessed December 2013).

Gundogdu, I. B. (2010). “Applying linear analysis methods to GIS-supported procedures for preventing traffic accidents: Case study of Konya.” *Safety Science* 48(6), 763–769.

Hill, J., F. Hossain, and A.C. Bagtzoglou (2009). “Zonal management of arsenic contaminated ground water in Northwestern Bangladesh.” *Journal of Environmental Management* 90, 3721– 3729.

Kuo, P., D. Lord, and T.D. Walden (2013). “Using geographical information systems to organize police patrol routes effectively by grouping hotspots of crash and crime data.” *Journal of Transport Geography* 30, 138-148.

Larsen, M. (2010). “Philadelphia Traffic Accident Cluster analysis using GIS and SANET.” < penniur.upenn.edu/uploads/media.../larsen-capstone-musa.original.pdf> (Accessed December 2013).

Li, L., L. Zhu, and D.Z. Sui (2007). “A GIS-based Bayesian approach for analyzing spatial– temporal patterns of intra-city motor vehicle crashes.” *Journal of Transport Geography* 15, 274– 285.

Manepalli, U. R. R., and G.H. Bham (2011). “Crash Prediction: Evaluation of Empirical Bayes and Kriging Methods.” Proc., 3rd International Conference on Road Safety and Simulation, 2011, Indianapolis, IN.

Mitra, S. (2008). “Enhancing Road Traffic Safety: A GIS Based Methodology to Identify Potential Areas of Improvement.” California Polytechnic State University, San Luis Obispo, CA.

NDDOT. (2012). “2012 North Dakota Crash Summary.” North Dakota Department of Transportation, Safety Division, Bismarck, ND.

NHTSA (National Highway Traffic Safety Administration) (2010). <http://wwwnrd.nhtsa.dot.gov/Pubs/811363.pdf> (Accessed December 2013).

Ramli, M.Z.B. (2011). "Development of Accident Prediction Model by Using Artificial Neural Network (ANN)." M.S. thesis, Universiti Tun Hussien Onn, Malaysia.

Prasannakumar, V., H. Vijith, R. Charutha, and N. Geetha (2011). "Spatio-Temporal Clustering of Road Accidents: GIS Based Analysis and Assessment." *Proc., Social and Behavioral Sciences* 21, 317–325.

Rencher, A.C. (2001). "Methods of Multivariate Analysis." Second Edition, John Wiley & Sons, Inc., New York, NY.

SAS (SAS Institute Inc.) (2008). "SAS/STAT 9.2 User's Guide." Cary, NY.

Savelieva, E. (2005). "Using ordinary kriging to model radioactive contamination data." *Journal of Applied GIS* 1(2).

Spear, B.D. (2008). "Using GIS to analyze crash data in the I-95 corridor." *Proc., 21st Symposium of Geo-spatial Information Systems for Transportation*, Houston, TX.

USDOT (United States Department of Transportation-Research and Innovative Technology Administration) (2013). <http://www.rita.dot.gov/sites/default/files/rita_archives/rita_publications/horizons/2008_11_12/html/improving_operational_safety.html> (Accessed December 2013).

U.S. Census Bureau. (2012). <https://www.census.gov/newsroom/releases/archives/population/cb12-250.html> (Accessed December 2013).

Vachal, K. and D. Malchose (2009). "What can we learn about North Dakota's youngest Drivers from their Crashes?" *Accident Analysis and Prevention* 41, 617-623.

Wang, X. and K.M. Kockelman (2009). "Forecasting Network Data: Spatial Interpolation of Traffic Counts Using Texas Data." Transportation Research Board, 2105, 100-108.

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

Yang, S., S. Lu, and Y. Wu (2013). "GIS-based Economic Cost Estimation of Traffic Accidents in St. Louis, Missouri." *Proc., 13th COTA International Conference of Transportation Professionals, Social and Behavioral* Sciences 96, 2907-2915.

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