

## Determining Hotspots of Road Accidents Using Spatial Analysis

S.Sarifah Radiah Shariff\*<sup>1</sup>, Hamdan Abdul Maad<sup>2</sup>, Nursyaza Narsuha Abdul Halim<sup>3</sup>,  
Zuraidah Derasit<sup>4</sup>

<sup>1</sup>Malaysia Institute of Transport (MITRANS), Universiti Teknologi MARA Shah Alam, Malaysia

<sup>2,4</sup>Centre for Statistics and Decision Science Studies, Faculty of Computer & Mathematical Sciences, Universiti  
Teknologi MARA Shah Alam, Malaysia

<sup>3</sup>Digi Telecommunications, Shah Alam, Malaysia

---

### Article Info

#### Article history:

Received Aug 11, 2017

Revised Nov 10, 2017

Accepted Nov 27, 2017

#### Keywords:

Predictive Accuracy

Road Accidents Hotspots

Spatial Analysis

---

### ABSTRACT

Road accidents continuously become a major problem in Malaysia and consequently cause loss of life or property. Due to that, many road accident data have been collected by highway concessionaries or build-operate-transfer operating companies in the country meant for coming up with proper counter measures. Several analyses can be done on the accumulated data in order to improve road safety. In this study the reported road accidents cases in North South Expressway (NSE) from Sungai Petani to Bukit Lanjan during 2011 to 2014 period is analyzed. The aim is to determine whether the pattern is clustered at certain area and to identify spatial pattern of hot spots across this longest controlled-access expressway in Malaysia as hotspot represents the location of the road which is considered high risk and the probability of traffic accidents in relation to the level of risk in the surrounding areas. As no methodology for identifying hotspot has been agreed globally yet; hence this study helped determining the suitable principles and techniques for determination of the hotspot on Malaysian highways. Two spatial analysis techniques were applied, Nearest Neighborhood Hierarchical (NNH) Clustering and Spatial Temporal Clustering, using CrimeStat® and visualizing in ArcGIS™ software to calculate the concentration of the incidents and the results are compared based on their accuracies. Results identified several hotspots and showed that they varied in number and locations, depending on their parameter values. Further analysis on selected hot spot location showed that Spatial Temporal Clustering (STAC) has a higher accuracy index compared to Nearest Neighbor Hierarchical Clustering (NNH). Several recommendations on counter measures have also been proposed based on the details results.

Copyright © 2018 Institute of Advanced Engineering and Science.  
All rights reserved.

---

### Corresponding Author:

S.Sarifah Radiah Shariff,  
Malaysia Institute of Transport (MITRANS),  
Universiti Teknologi MARA Shah Alam,  
Malaysia.  
Email: radiah@tmsk.uitm.edu.my

---

## 1. INTRODUCTION

One of the major causes of death in Malaysia is due to road accidents. With the rapid development in Malaysia's economy, the issue of road safety becomes increasingly important. According to Malaysian Institute of Road Safety Research (MIROS), in 2014 the total number of road accidents was 476,196 with the number of road deaths is 6,676 cases seriously injured is 4,432 cases and slightly injured is 8,598 cases. Malaysia has been ranked the 8<sup>th</sup> in fatalities of road crashes in the Mortality from Road Crashes in 193 Countries Report. Malaysia Road Transports Department (RTD) reported that in 2014, Malaysia lost RM nine billion due to road deaths as sudden death may result in lost of natural asset. On average, more than ten

billion ringgit has losses due to road accidents every year. The highest fatalities by the age group are among young person aged 16 to 25 years old [1]. A study [2] on Malaysia road accident situation stated that the psychological sufferings are often intense, lasting and even permanent. The victims may generate somatic illnesses which worsen this psychological distress, creating a vicious circle.

There are several methods being used to determine the hotspots for road accidents in the previous studies. Among which are a statistical package, CrimeStat that offers various algorithms in determining the spatial pattern that exist within the data [3-6] also used Geographical Information System (GIS) to visualize the identified hotspots to further strengthen the validity of the results. In this study, the focus was on the accidents pattern along the North-South Expressway (NSE). The NSE is the longest controlled-access expressway in Malaysia with the total length of about 772 km (480 mi) running from Bukit Kayu Hitam in Kedah near the Malaysian-Thai border (connects with Phetkasem Road (Route 4) in Thailand) to Johor Bahru at the southern portion of Peninsular Malaysia and to Singapore. The expressway links many major cities and towns in western Peninsular Malaysia, acting as the 'backbone' of the west coast of the peninsula. It provides a faster alternative to the old Federal Route 1, thus reducing travelling time between various towns & cities.

## 2. RESEARCH METHOD

This study first examined monthly reported road accidents involving all types of vehicles that occurred along North-South Expressway (NSE) starting from Sungai Petani to Bukit Lanjan from 2011 to 2014. The data was obtained from PLUS Expressway Berhad that includes accident locations (in terms of kilometer), the month, date, day, time, vehicle involved, causes of accident, collision type and injury type. Assuming that the accident can occur at any spot along the expressway, we divided the length, along Sungai Petani to Bukit Lanjan, into six (6) sections as described in Table 1.

Table 1. Location and Section of Study Area

Section	Location	Distance
N3	Sungai Petani to Jawi	54.90 km
N4	Jawi to Cangkat Jering	56.80 km
N5	Cangkat Jering to Ipoh	55.40 km
C1	Ipoh to Bidor	64.40 km
C2	Bidor to Tanjung Malim	59.70 km
C3	Tanjung Malim to Bukit Lanjan	60.30 km

Descriptive analysis on road accidents at each section is presented to observe the pattern. In order to identify the hotspots, two methods from CrimeStat, Nearest Neighbor Hierarchical (NNH) Clustering Analysis and Spatial Temporal Clustering Analysis are considered and compared. These methods automatic collect the surface location of the target and solved using CrimeStat, a spatial statistics program for the analysis of crime incident locations [7].

### 2.1. Nearest Neighbor Analysis (NNA)

Nearest Neighbor Analysis (NNA) produces a calculation called the Nearest Neighbor Index (NNI) which is the ratio of the observed distance divided by the expected distance. If the index value shows less than 1, it means that the pattern of incident exhibits clustering and if the index is greater than 1, it means that the incident trend is random or toward dispersion. In NNA, the Z score values measure the statistical significance whether the patterns of incident are randomly or not randomly distributed.

NNI is calculated as follows:

$$\text{Nearest Neighbor Index} = \frac{\text{Avg Nearest Neighbor}}{\text{Expected Avg Nearest Neighbor}}$$

Where:

$$\text{Average Nearest Neighbor} = \frac{\text{Distance}}{\text{Number of Accidents}}$$

$$\text{Expected Average Nearest Neighbor} = \frac{1}{2} \sqrt{\frac{\text{Area}}{\text{Number of Points}}}$$

## 2.2. Nearest Neighbor Hierarchical (NNH) Clustering

The next step is to identify the density of road accidents occurrence or the hot spots using the Nearest Neighbor Hierarchical (NNH) Clustering. In NNH, a threshold value and minimum number of accidents,  $n_{\min}$  within a cluster need to be predetermined. Using CrimeStat software, 1 kilometer has been set up as a threshold distance, and the minimum number per clusters,  $n_{\min}$  is 200. At the same time, NNH results are given with their severity value, which is the number of accidents located in the boundary of a cluster. The advantage to this technique is that it can identify small geographical environments where there are concentrated incidents. This can be useful for specific targeting, either by police deployment or community intervention.

## 2.3. Spatial Temporal Clustering Analysis (STAC)

STAC is one of the widely used methods in detecting hotspots in the entire study area based on its shape. In most previous studies it is suggested to use rectangular if the analysis area has most regular pattern, while the triangular pattern is most suitable for the area which generally has an irregular pattern. STAC places a circle on every node, counts the number of accidents which falling within each circle and ranks the circle in descending order. The X and Y coordinates of any node with at least two incidents within the search radius are recorded, along with the number of data points found for each node.

## 2.4. Comparison of Hot Spot Techniques

In order to observe the accuracy of each method, Prediction Accuracy Index (PAI) that is commonly used in crime hot spot mapping is adapted. According to [7] this method was developed in order to determine the differences between each method in capturing or predicting hotspots location. Finding 100 per cent of future events in 100 per cent of the area would give a PAI value of 1. If the hit rate and the area percentage fall by an equal measure, the value would be computed as 1 also. Thus, the greater the number of accidents in a hotspot area that is smaller in size to the whole study area, the higher the PAI value. The PAI is easy to calculate, considering the number of accidents that fall into the area determined as hotspots against the size of the hotspot and the size of the study area. The higher the PAI the more accurate the method is.

Prediction Accuracy Index is calculated as follows:

$$\text{Prediction Accuracy Index (PAI)} = \frac{\left(\frac{n}{N}\right) \times 100}{\left(\frac{a}{A}\right) \times 100}$$

Where:  $n$  is the total number of accidents found in that particular cluster,  
 $N$  is the total number of accidents in the study area,  
 $a$  equals to total area of hotspots and  
 $A$  equals to the entire area in the study region.

## 3. RESULTS AND ANALYSIS

Figure 1 shows, from 2011 to 2014 sections C3 (Tanjung Malim to Bukit Lanjan) recorded 30.5% of total accidents, the highest number of road accident compared to the other sections. This section remains the record of having the highest cases of road accident from 2011-2014. While, the lowest number of road accident is section N3 which is from Sungai Petani to Jawi which contributed only 6.7% of the total number of accident in all sections. The severity rating of the vast majority of road accident across North South Expressway is described in Figure 2.

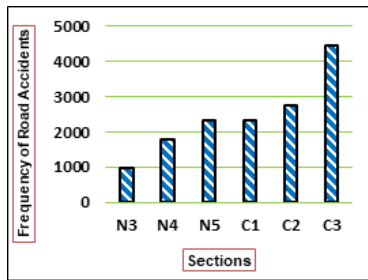


Figure 1. Effects of selecting different switching under dynamic condition

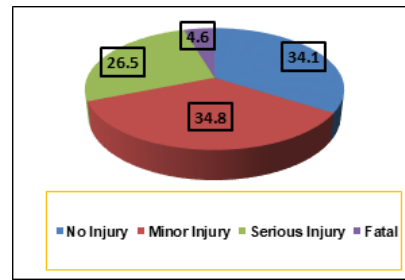


Figure 2. Effects of selecting different switching under dynamic condition

**3.1. Randomness Pattern Identification**

As mentioned earlier, 1 kilometer have been set up as a threshold distance while the minimum number per clusters is 200, using CrimeStat software. In Table 2, it shows that the road accident cases across North South Expressway (NSE) are having a clustered distribution because the Nearest Neighbor Index is 0.00506 which is less than 1. The p-value also shows that it less than  $\alpha = 0.05$ , thus we reject null hypothesis of the occurrence of road accidents across NSE was random. This goes to the conclusion that the road accidents exhibit a clustering pattern.

Table 2. Nearest Neighbor Index Result

Test Statistic (Z) : - 237.6788		P Value One Tail : 0.0001	
Mean Nearest Neighbor Distance (m)	3.8432	Expected Nearest Neighbor Distance (m)	758.807
		Nearest Neighbor Index	0.00506

**3.2. Hotspots Determination**

According to [8], choice of threshold limits is important, which basically defines the scope of the clustering analysis. Hence, using CrimeStat software in conducting this Nearest Neighbor Hierarchical Clustering across North South Expressway, 1 kilometer have been set up as a threshold distance with the minimum number per clusters is 200. Table 3 shows that, there are 12 hot spots with the highest number of road accident are 349 cases.

Due to its nature of detecting the hotspots based on pattern or shape, there is a need to determine the search radius and the minimum number of accidents per cluster. After doing the sensitivity analysis, the search radius allowed in this study have been set as 0.5 kilometer with  $n_{min} = 100$  based on triangular pattern for scan type. As shown in Table 3, there are 14 hotspots found with the highest number of road accidents are 220 cases.

The number of hot spots distribution within the six sections is summarized in Table 4. Both methods show that the most hotspots are found in section C3 (Tanjung Malim to Bukit Lanjan).

Table 3. Hot spots Identified based on NNH and STAC

NNH				STAC			
Cluster	Mean X (longitude)	Mean Y (latitude)	Frequency (No. Of Accident)	Cluster	Mean X (longitude)	Mean Y (latitude)	Frequency (No. Of Accident)
1	100.9703	4.69366	349	1	100.672	4.9269	220
2	101.57	3.2301	349	2	101.5668	3.23364	218
3	100.6718	4.92755	328	3	100.4891	5.19155	161
4	101.023	4.69206	298	4	101.542	3.43634	148
5	101.5579	3.37062	298	5	100.983	4.68848	133
6	101.545	3.43102	295	6	101.5603	3.37508	130
7	101.5202	3.66175	241	7	101.5549	3.36426	129
8	101.5841	3.193	234	8	101.1708	4.45563	127
9	101.2082	4.40789	223	9	101.5557	3.38886	114
10	101.5468	3.30654	222	10	101.5766	3.22311	112
11	100.8173	4.77865	213	11	101.0037	4.68782	104
12	101.4904	3.69159	210	12	101.553	3.35176	102
				13	101.026	4.68611	101
				14	101.5657	3.24579	101

Table 4. Number of Hot Spot by Sections using NNH

Section	Number of Hot Spots	
	Using NNH	Using STAC
N3 (Sungai Petani to Jawi)	0	1
N4 (Jawi to Cangkat Jering)	1	1
N5 (Cangkat Jering to Ipoh)	3	3
C1 (Ipoh to Bidor)	1	1
C2 (Bidor to Tanjung Malim)	1	0
C3 (Tanjung Malim to Bukit Lanjan)	6	8

### 3.3. Prediction Accuracy

The performance of the two methods in identifying the hotspots is measured using PAI and summarized in Table 5. Since the PAI values for STAC are higher, it can be concluded that STAC performs better compared to NNH for all injury types and fatalities. STAC also managed to identify two more spots for possible hotspots compared to NNH.

Table 5. Comparison Between NNH and STAC

		Area (sq metre)	Frequency	Hit Rate	PAI
All Injury Type	NNH	1.39	3260	22.29	163.69
	STAC	0.29	1900	12.99	461.57
Fatalities	NNH	0.80	198	1.35	17.28
	STAC	0.51	77	0.53	1054.58

## 4. CONCLUSION

Overall, the pattern of road accident on North South Expressway (NSE) resulted in clustering pattern which means that the accident was grouped in one location. For both methods used in determining the hotspots, the most hotspots are found in C3 section which was from Tanjung Malim to Bukit Lanjan. This is also parallel with descriptive statistics result on the highest road accidents occurrence. It is also observed that the types of road along the section is 3 road lane. Based on the results obtained, it is important for the concerned bodies to take a remedial action at selected hot spots. This includes adding more caution signage or implements specific interventions to prevent road accident.

Both applied methods can be used are practical in determining the hotspot based on choosing an accurate parameter value such as much smaller threshold value. Further analysis on more suitable parameter values to be used should be done for better results [9, 10].

It is advisable to further research to add more of hot spot technique methods such as Kernel Density Estimation (KDE), K-Means Clustering, Getis-Ord Statistics and Moran's Index. Besides that, the study can be divided into another sub group such as collision type, injury type or vehicles involved. Compared to other studies which consider the network design such as roundabout and intersections, this study cannot take any design type in consideration since the highway is a straight road network. However, the number of lanes or road types could be used to understand the relationship between the hot spots and roads. Speed status of the particular road segments can be used when analyzing the hot spots.

## ACKNOWLEDGEMENT

We would like to acknowledge Research Management Institute (RMI) of Universiti Teknologi MARA, Malaysia, for funding this research through the Grant No: 600-IRMI/DANA 5/3/LESTARI (0039/2016).

## REFERENCES

- [1] Shamsuddin, S., Anill, M., Othman C., Asmah, H. S., & Tarmizi, I. Spatial and Temporal Pattern of Road Accidents and Casualties in Peninsular Malaysia. *Jurnal Teknologi*, 2015. 76(14), 57–65.
- [2] Kareem, A. Review of Global Menace of Road Accidents with Special Reference to Malaysia- A Social Perspective. *The Malaysian Journal of Medical Sciences*, 2003. 10(2).
- [3] Prasannakumar, V., Vijith, H., Charutha, R., & Geetha, N. Spatio-Temporal Clustering of Road Accidents : GIS Based Analysis and Assessment. *Procedia - Social and Behavioral Sciences*, 2011. 21, 317–325.
- [4] Truong, L., & Somenahalli, S. Using GIS to identify pedestrian-vehicle crash hotspots and unsafe bus stops. *Journal of Public Transportation*, 2011. 14(1), 99–114.

- [5] Xie, K., Ozbay, K., & Yang, H. Spatial analysis of highway incident durations in the context of Hurricane Sandy. *Accident Analysis & Prevention*, 2015. 74, 77–86.
- [6] Karpagavalli, S., & Balamurugan, V. Online Marine Satellite System Using GIS. *Indonesian Journal of Electrical Engineering and Computer Science*, 2017. 8(2).
- [7] Chainey, S., Tompson, L., & Uhling, S. The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. *Security Journal*, 2008. 21(2).
- [8] Levine, N. *CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations* (version 3.3). 2010, Ned Levine & Associates: Houston, TX/ National Institute of Justice: Washington, DC.
- [9] MM Othman & Musirin, I. A novel approach to determine transmission reliability margin using parametric bootstrap technique. *International Journal of Electrical Power & Energy Systems*, 2011. 33 (10), 1666-1674.
- [10] Awal, L.J., H. Mokhlis, M. K. Rahmat, Shilpa, S., Albatsh, F. & Ismail, B. Fault Distance Identification Using Impedance and Matching Approaches on Distribution Network. *Indonesian Journal of Electrical Engineering and Computer Science*, 2017. 8(3).