Determining Spatial Patterns of Road Accidents at Expressway by Applying Getis-Ord Gi* Spatial Statistic

Norhafizah Manap, Muhammad Nazri Borhan, Muhamad Razuhanafi Mat Yazid, Mohd Khairul Azman Hambali, Asyraf Rohan

Abstract**:** ***This paper was using Getis–Ord (Gi*) spatial statistics to identify hot spots on the controlled-access expressway. The application of the method was demonstrated through a case study by using the reported road accident cases in North-South Expressway (NSE). The method successfully identified the clusters of accidents from more than 47,359 accident records from 2016 to 2019. 25 hotspot locations were identified at this study area represents 26.81% of reported cases with the lengthiest hotspot is 31.2 km and the shortest is 300m. The largest and the second largest means of z score of hotspots were identified near to well-known high populated and busy city Kuala Lumpur with scores of 6.17243 and 6.074437. The largest z score means the more intense clustering at the location will be and statistically significance to reject the null hypothesis. This study also found that the accident hotspots tend to occur at the location where the continuous traffic flow is disturbed. There are 16 hotspot locations were identified which is equivalent to 64% from the total hotspots that occur at the location where were the existing of interchange, exit ramp, slip road, rest area or lay by spotted at the area. The interference of traffic flows including diverge and merge activities will affect the speed consistency and which if often, leads to sideswipe and rear accidents. By using GIS, the location of hotspots can be analyzed meticulously at the location. It can help in determining effective countermeasures based on the analysis of the causal factors***

Keywords**:** ***spatial analysis, Getis-Ord Gi, controlled access highway, accidents, interchange***

I. INTRODUCTION

Identifying and coping with accidents prone location is vital for understanding the causes of crashes and to determine effective countermeasures based on the analysis of the causative factors. The foremost negative result with rapid development is road accidents that cost loss of lives. Without action, annual road traffic deaths are expected to become the seventh leading reason for death by 2030 [1]. However, the death caused by road traffic accidents in Malaysia is at the leading fourth in 2018 [2]. As a result of this appalling issues, action ought to be taken to mitigate it. One of the strategies to improve road network safety level is to identify safety deficient area in the highway network. One of the most important problems that traffic official face is where and how to implement precautionary measures and provisions so that they can have the most significant impact for traffic [3]. There are several methods being used to determine the hotspots for road accidents within the previous studies. Some of the researchers using statistical package CrimeStat that offers numerous algorithms in determining the spatial pattern that exist within the data and using Geographical Information System (GIS) to visualize the identified hotspots to strengthen the validity of the results and some others using spatial statistic tools that supported by Arcgis [4][5][6]. GIS is a very vital and comprehensive management tool for traffic safety. Since 1990's, GIS technologies have been used more frequently for such studies due to the availability of low cost GIS with user-friendly interfaces [3]. GIS has the ability to hold a massive amount of data that can be easily stored, shared and managed. This study will use spatial statistics tool that calculates Getis Ord Gi* statistical analysis to identify crashes hotspot at the study area.

II. STUDY AREA AND DATA COLLECTION

North-south expressway was chosen as a study area. This expressway is a controlled access highway which has total length of 772km. The data were collected from Malaysian Highway Authority (MHA). The NSE is divided into 2 main routes which are the northern routes E1 (the upper part) with 458km length and the southern routes E2 (the lower part) with 310km length. This study was covered all along the NSE. The data given are accidents data including accidents location per 100m, accidents severity, vehicle types, report number, highway type for three consecutive years (June 2016 to May 2019). The location of North-South Expressway as shown in Figure 1.

III. METHODOLOGY

The main objective of this study is to analyze the hotspots of the study area in an effort of

Revised Manuscript Received on October 22, 2018.

***** Correspondence Author

Norhafizah Manap*, Department of Civil and Structural Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Malaysia

Muhammad Nazri Borhan, Department of Civil and Structural Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Malaysia

Muhamad Razuhanafi Mat Yazid, Department of Civil and Structural Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Malaysia

Mohd Khairul Azman Hambali, Department of Project Management, Malaysian Highway Authority, Kajang, Malaysia

Asyraf Rohan, Department of Project Management, Malaysian Highway Authority, Kajang, Malaysia

obtaining information required to help decision makers for taking appropriate measures in order to prevent and reduce road accidents. The data have been geocoded and analyzed by using Arcmap 10.3. The crashes location were mapped by Rectified Skewed Orthomorphic (RSO) projection with 0.001m of x,y tolerance. The data were then converted to weighted point data before the following analysis steps were done. Moran I was used in this study to measure spatial autocorrelation based on both feature locations and feature values. It will evaluate whether the pattern expressed is clustered, dispersed or random. Moran I has value from -1 to +1, which -1 shows perfect dispersion, 0 shows perfect randomness and +1 indicates perfect clustering opposites with perfect dispersion [7]. The Moran Index is expressed as (1):

$$I = \frac{N \sum_{i=1}^{N} \sum_{j=1, j \neq 1}^{N} \omega_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S_0 \sum_{i=1}^{N} (x_i - \bar{x})^2} \quad \forall_i = 1, \dots, n; \forall_j = 1, \dots, n \tag{1}$$

The distance threshold of 1000m were selected in this study to make ease the decision maker to mitigate the problem at hotspot location. Getis-Ord statistic were then used to identify accidents hotspots. This test assesses whether the clusters of crashes are statistically significant. The resultant z value expressed either high or low values of neighbouring features. Features with a high value may not be a statistically significant hot spot. To be a statistically significant hotspot, a feature will have a high value and be surrounded by other features with high values as well. Positive z score indicates a hotspot, the larger z scores the more intense the clustering while negative z score indicates cold spot and the smaller the z score indicates the more intense the clustering of low values (cold spot) [8]. The Getis-Ord local statistic is given as (2):

$$G_i^* \frac{\sum_{j=1}^{n} \omega_{i,j} x_j - \bar{X} \sum_{j=1}^{n} \omega_{i,j}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} \omega_{i,j}^2 \left(\sum_{j=1}^{n} \omega_{i,j} \right)^2 \right]}{n-1}}} \tag{2}$$

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features (3) and;

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2} \tag{3}$$

The G_i* Statistic is a z score so no further calculation is required.

IV. RESULT AND DISCUSSION

A. Accident Data

In this paper, road accident data of NSE for three consecutive years have been investigated; the total number of accidents, fatalities, and injuries are shown in Table-I. Total of 29891 crashes were recorded and 47359 vehicles involved in the crashes in 3 years at the study area. From the amount, 900 are fatal accidents which is equal to 3%, 4348 (14.5%) are severe injury, 4434 (14.83%) are slight injury and the remaining 20230 (67.7%) are property damage. This study was using the entire crashes data in determining the hotspot area.

Table-I : Total Accidents at NSE

Total Accidents	Total Vehicles	Fatal	Severe Injury	Slight Injury
29891	47359	900	4348	4434

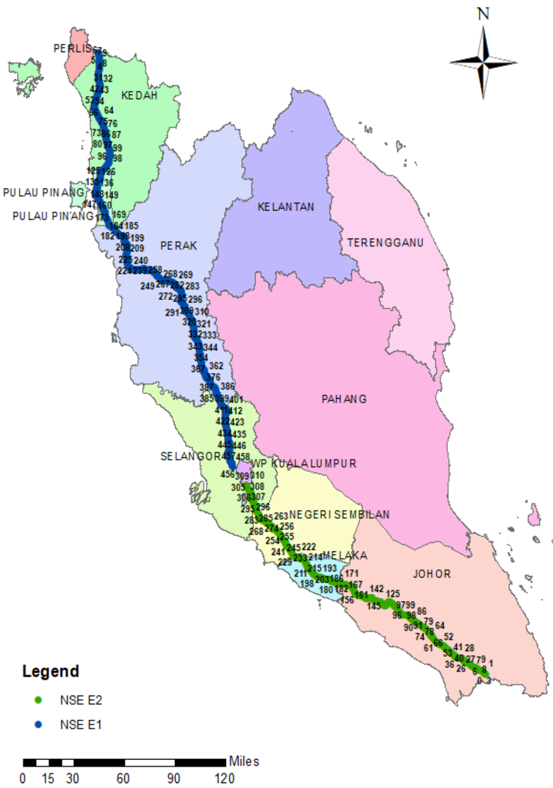


Fig.1 The location of North-South Expressway with KM markers

B. Weighted Point Data

Figure 2 shows the weighted point data of the crashes. The weighted point data is a combine of coincident point which hold the sum of all incidents at each unique location. The largest weighted point data for this study is 154 crashes at a unique location and the smallest is one. The weighted point data shows the total amount of crashes at the location, however, as mentioned earlier, a feature with high value may not be a statistically significant hot spot because a hot spot is determined by high value feature surrounded by other features with high values as well.

C. Spatial Autocorrelation

Moran I is a spatial autocorrelation used to determine patterns in complicated data set. It values indicates whether the features location and features values are clustered, disperse or random. The Moran Index, z score and p value were computed by using the Spatial Autocorrelation (Global Moran's I) with 1000m threshold distance and the result as shown in Figure 3. It is apparent from the Figure 3, the z score has a high value which is 62.252 larger than 2.58 and there is less than 1% likelihood that this clustered pattern could be the result of random chance. The Moran Index also has a positive value (0.246041) indicates clustering.

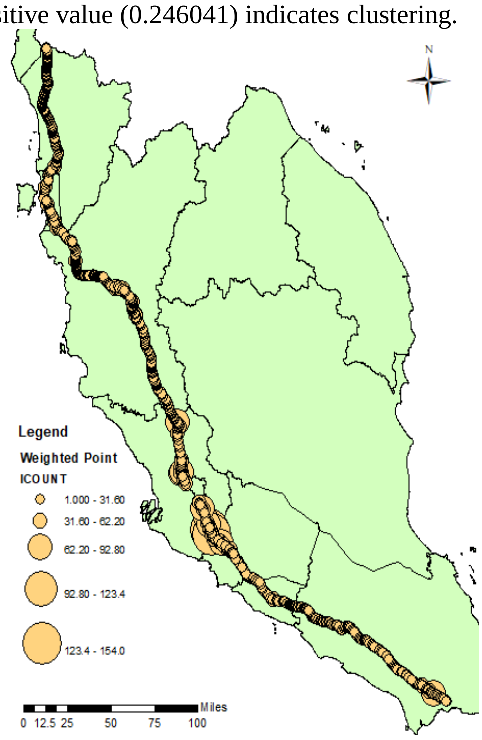


Fig.2 Weighted Point Data of Crashes

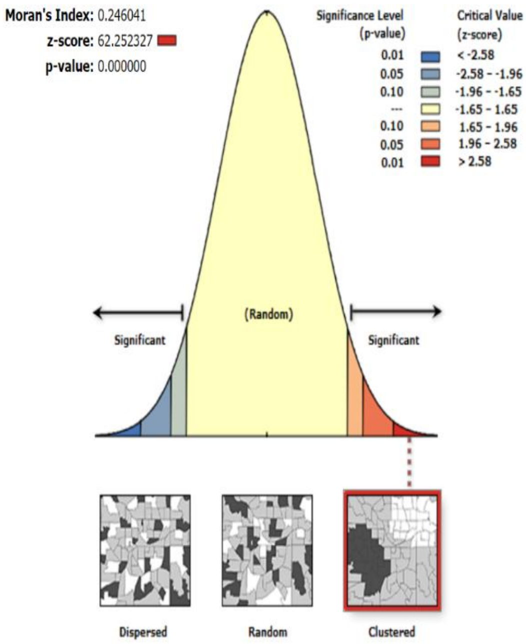


Fig.3 Spatial Autocorrelation Report

A. Hotspot Analysis

According to the previous Moran I spatial autocorrelation analysis, the distance threshold of 1000 meters associated with the high value of z score and low p value in the previous step was chosen for the Getis-Ord Gi∗ analysis. The map of the hotspot analysis for the study area is presented in Figure 4 while the details of the finding of the hotspots are distributed in Table-II. There were 25 hotspots have been identified and marked in pink. The hotspots represent 26.8% of reported cases and 20.56% from the total of fatal cases. The total length of hotspots is 87.1 km equal to 11.28% of the total length of NSE. The highest length of hotspot is at KM 286.4-255.2 with 31.2 km length, and 4864 reported cases and the shortest is at KM 254.7-254.4 with 300 m length and 21 reported cases, both are located at NSE E2. However, the mean of z value at KM 286.4-255.2 which is 6.035978 with a maximum z value of 12.179435 is not the largest value of z score. The largest mean value of z score for this hotspot study is 6.17243 located at KM 442.0-458.9 with a length of 16.9 km and hold of 2672 reported cases. The second largest mean value of z score is located near to the first which is at NSE 2 KM 305.7-301.8 with the total length of 3.9 km and mean z value of 6.074437. Both are located near the center of Malaysia, Kuala Lumpur separated by the city itself because NSE is an interurban highway. This location is well known highly populated city with busy traffic. The largest z score means the more intense the clustering at the location will be and statistically significance to reject the null hypothesis.

After thorough analysis at every location of hotspot, this study also found that the accident hotspots tend to occur at the location where the continuous traffic flow is disturbed. There are 16 (64%) hotspot locations identified have whether interchange, exit ramp, slip road, rest area or lay by at the location as shown in Table-III. It also identified some locations have more than one infrastructure that will lead to conflict at the hotspot area as examples at NSE 1 KM 442.0-458.9 and NSE 2 KM 286.4-255.2.

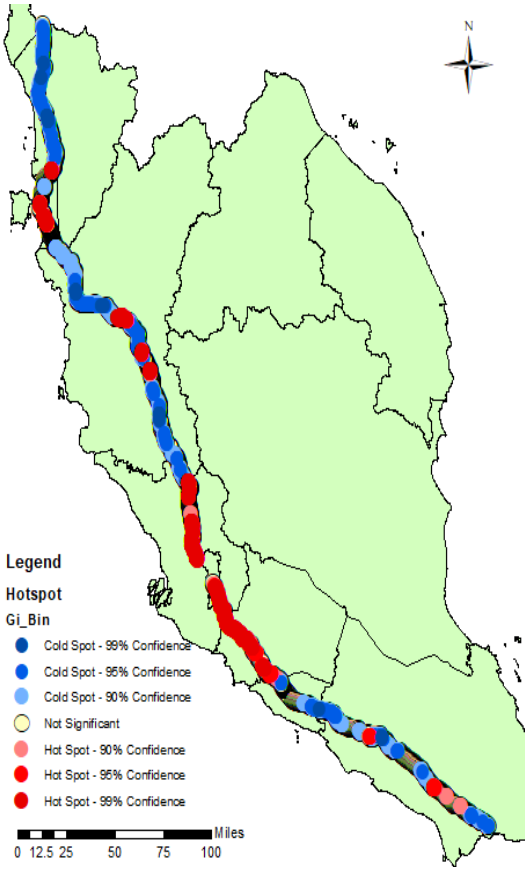
The interference of traffics flow including diverge and merge activities can affect the speed consistency and which if often, leads to side swipe and rear accidents. By using Arcmap and portable base map, the location of hotspot can be analyzed meticulously at the location, the hotspot points fall at the exact location at the map and display the hotspots area surrounding. Figure 5 shows the hotspots at KM 409.7-410.7 and KM 433.5-477.8. It apparently presents that the hotspot points fall onto the trumpet interchange at both locations.

Table-III also illustrates the number of accidents and fatal accidents per 100 m at the hotspot locations. As can be seen from the table, the highest amount of crashes per 100 m is 17 spotted at KM 293.7 - 294.7 followed by 16 number of crashes located at KM 309.7 - 310.9, KM 409.7 - 410.7, KM 442.0 - 458.9 and KM 286.4 - 255.2.

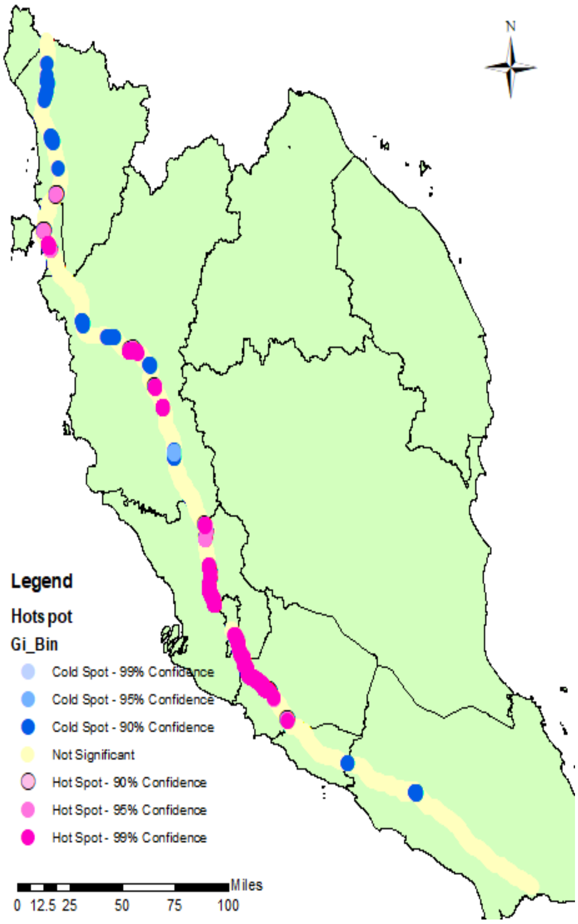
However, the highest number of crashes were not significantly indicating the highest number of death, the result apparently shows that the fatal cases at KM 293.7 - 294.7 has only 0.2 fatal cases per 100 m. The fatal cases per 100 m at the hotspot locations were found between 0.0 to 0.9 cases with the highest number of fatal cases is at KM 107.3 - 108.0.

It is also found that even though the highest fatal cases are located at interchange, but at the same time several hotspot locations with interchange spotted at the area score 0 fatal such as at KM 398.4 - 400.3 and KM 428.8 - 429.4. The fatal cases for each interchange are between 0.0 – 0.9 cases, and the number of accidents per 100 m for each interchange are between 7 to 16 cases.

It apparently shows that other factors might be also contributed to the occurrence of the accidents. Further analysis can be done to measure other factors that contribute to those locations becoming accident prone location.



Before FDR correction



After FDR correction

Fig.4 Hotspot Analysis by Getis Ord-Gi

Table-II : Z Score of hotspot locations

Location	Highway	KM	Length	Total Accidents	Mean z score	Max z score
1	NSE E1	107.3-108.0	700	75	2.611208	2.795878
2		135.5-136.1	600	70	2.639021	2.795878
3		146.1-146.8	700	43	3.442085	4.290993
4		147.0-149.1	2100	315	4.199705	5.467585
5		150.3-151.0	700	88	2.91806	3.219442
6		255.7-259.3	3600	463	4.164452	6.527722
7		259.6-260.7	1100	116	3.171013	3.755114
8		262.6-264.2	1600	213	3.564558	4.724985
9		293.7-294.7	1000	172	3.152484	3.545259
10		309.7-310.9	1200	194	3.82534	4.151137
11		398.4-400.3	1900	234	3.150795	3.480096
12		404.5-404.9	400	21	2.804271	3.090786
13		409.7-410.7	1000	157	2.789596	3.122552
14		428.8-429.4	600	42	3.112387	3.675587
15		430.0-430.6	600	63	3.027956	3.489776
16		433.5-437.8	4300	509	3.552355	4.925289
17		442.0-458.9	16900	2672	6.172428	16.04762
18	NSE E2	305.7-301.8	3900	601	6.074437	9.377401
19		300.7-295.5	5200	756	5.114163	8.171874
20		294.0-291.8	2200	261	3.450347	4.522713
21		291.6-288.0	3600	547	5.731711	9.116746
22		286.4-255.2	31200	4864	6.035978	12.17944
23		254.7-254.4	300	21	2.584097	2.632969
24		249.5-248.9	600	69	3.140314	3.480096
25		231.8-230.7	1100	132	2.977793	3.38235

*Based on reported cases

Table-III : Severity and infrastructures at hotspot locations

Location	Highway	KM	Fatal	Severe Injured	Slight Injured	Property damages only	Accidents per 100m	Fatal per 100m	Infrastructures
1	NSE E1	107.3-108.0	6	14	15	40	11	0.9	Sungai Petani Exit Interchange
2		135.5-136.1	1	1	18	50	12	0.2	Split Road at seberang jaya
3		146.1-146.8	2	7	11	23	6	0.3	-
4		147.0-149.1	15	48	52	200	15	0.7	-
5		150.3-151.0	3	10	13	62	13	0.4	Tambun Utara Toll Plaza
6		255.7-259.3	5	40	77	341	13	0.1	-
7		259.6-260.7	3	3	15	95	11	0.3	-
8		262.6-264.2	3	9	44	157	13	0.2	Vista Point Ipoh Rest Stop
9		293.7-294.7	2	31	33	106	17	0.2	-
10		309.7-310.9	6	40	27	121	16	0.5	-
11		398.4-400.3	0	16	42	176	12	0.0	Ulu Bernam Rest Area and Tanjung Malim Toll Plaza Interchange
12		404.5-404.9	0	2	7	12	5	0.0	-
13		409.7-410.7	5	16	26	110	16	0.5	Lembah Beringin Toll Plaza Interchange
14		428.8-429.4	0	8	6	28	7	0.0	Bukit Beruntung Interchange
15		430.0-430.6	1	6	10	46	11	0.2	-
16		433.5-437.8	20	65	55	369	12	0.5	Sungai Buaya Toll Plaza Interchange
17		442.0-458.9	75	491	368	1738	16	0.4	Plaza Tol Rawang Interchange , Plaza Tol Sungai Buloh (in bound and out bound) Interchange, Sungai Buloh Rest Area
18	NSE E2	305.7-301.8	13	133	103	352	15	0.3	UPM Inbound Toll Plaza Interchange, Kajang Toll Plaza Interchange
19		300.7-295.5	24	152	114	466	15	0.5	Bangi Outbound Toll Plaza Interchange
20		294.0-291.8	10	57	50	144	12	0.5	Southville City Toll Plaza Interchange
21		291.6-288.0	10	113	79	344	15	0.3	Putera Mahkota Inbound Toll Plaza Interchange
22		286.4-255.2	142	775	643	3304	16	0.5	Elite Toll Road Interchange, Nilai Outbound Toll Plaza Interchange, Seremban North and South Bound Rest Area, Bandar Ainsdale Toll Plaza Interchange, Seremban Toll Plaza Interchange , Port Dickson Toll Plaza Interchange, Senawang Toll Plaza Interchange
23		254.7-254.4	0	3	2	16	7	0.0	-
24		249.5-248.9	2	6	10	51	12	0.3	Senawang North Bound Lay by
25		231.8-230.7	9	18	22	83	12	0.8	Pedas Linggi Lay by

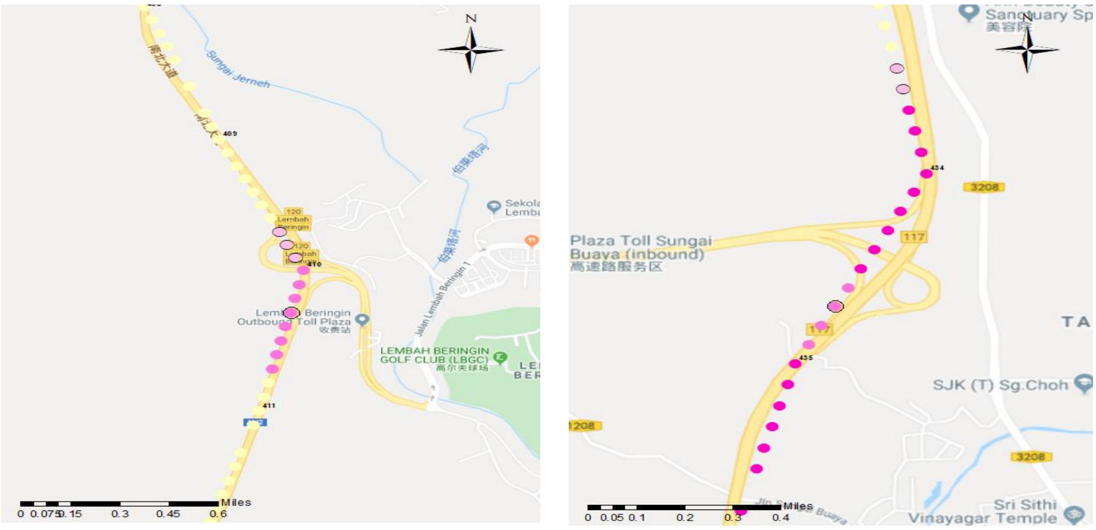


Fig. 5 Hotspot points on the Trumpet Interchange at KM 409.7-410.7 and KM 433.5-477.8

V. CONCLUSION

By using Geoinformation system the hotspot analysis can be done thoroughly at the specific location. The result found 25 locations of hotspot at this study area with the total length of hotspots are 87.1 km with 12698 reported cases. The result also illustrates that accidents tend to occur at the location where the continuous traffic flow is disturbed, 64% crash hotspot at NSE is spotted occur whether at interchange, exit ramp, slip road, rest area or lay by area. The hotspot location with highest fatal accidents is identified occurred at interchange, however, at the same time, there is also hotspot accident at the interchange scores 0 fatal accident. It apparently shows that, there are some other factors that contribute to the causal of fatal accident. This study also found that, hotspot locations with highest number of crashes are not indicating highest number of death. Further analysis can be done to understand the causes of accident hotspot area and countermeasures can be determined to mitigate the accidents.

ACKNOWLEDGMENT

The authors appreciatively acknowledge the Malaysian Highway Authority for providing the accidents data. The author also would like to thank Universiti Kebangsaan Malaysia (UKM) for their financial support under Research University (RU) Grant FRGS/1/2019/TK08/ UKM/02/1.

REFERENCES

1. WHO 2017. Global Status Report On Road Safety. Retrieved from : https://www.who.int/violence_injury_prevention/publications/road_traffic/Media_brief_all_factsheets_web_rev_nov_2017.pdf?ua=1
2. Jabatan Keselamatan Jalan Raya Malaysia. 2018. Buku Statistik Keselamatan (November). Retrieved from http://www.jkjr.gov.my/ms/maklumat_keselamatan/statistik/Statistik---Statistic/BUKU-STATISTIK-KESELAMATAN-JALAN-RAYA-(KEMASKINI-19-11-2018)/lang,ms-my
3. Khan, G., Qin, X., Noyce, D.A., 2006, Spatial Analysis of Weather Crash Patterns in Wisconsin, 85th Annual Meeting of the Transportation Resarch Board, Washington, USA, (www.topslab.wisc.edu/publication)
4. Fedy Ouni and Mounir Belloumia. Pattern of road traffic crash hotzones versus probable hotzones inTunisia: A geospatial analysis, Accident Analysis & Prevention. 128 (2019) 185–196. 2019
5. Radhiah Shariff and Zuraidah Derasit. Determining Hotspots of Road Accidents Using Spatial Analysis. Indonesian Journal of Electrical Engineering and Computer Science. Vol. 9 2018, No. 1, pp. 146~151.
6. S.Erdogan,I.Yilmaz,T.Baybura,andM.Gullu,"Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar," Accident Analysis & Prevention, vol.40,no.1,pp.174–181,2008.
7. Esri Resources (n.d.) Moran I (Spatial Statistic). S.S. Retrieved on October 31, 2019 from http://resources.esri.com/help/9.3/arcgisengine/java/gp_toolref/spatial_statistics_tools/how_spatial_autocorrelation_colon_moran_s_i_spatial_statistics_works.htm
8. Esri Resources (n.d.) Getis-Ord Gi statistic. S.S. Retrieved on October 15, 2019 from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/hot-

Retrieval Number: C10041183S319/2019©BEIESP
DOI:10.35940/ijrte.C1004.1183S319

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication