Singapore Expressway Traffic Analysis: An Interactive Web-Based Application for Visualizing Spatial Points Patterns of Road Accidents and Traffic Congestions Along Expressways in Singapore

GWEE Wei Ling, TAN Ming Kwang, TAN Zhi Chong

Abstract – Road accidents and traffic congestions are common occurrences as a society progresses. There are many problems associated with such incidents. In Singapore, there are few readily available tools which allows users to analyse and make insights of such incidents. Thus, the authors aim to design and develop a web analytical tool using R Shiny that allows users to analyse traffic data along Singapore's expressway. The analysis methods used in this application include Kernel Density Estimation to detect hotspots, K Function to detect clustering patterns and Multitype K Function to determine spatial correlation between two events.

Keywords – Singapore, Traffic, Heavy Traffic, Accidents, Kernel Density Estimation, Network Kernel Density Estimation, K Function, Multitype K Function, R Shiny

1 INTRODUCTION

As a city-state that is scarce of land, Singapore faces the challenge of a growing vehicle population. Road accidents and traffic congestions are inevitable in any countries and are often the results of increased mobility in today's society. Such occurrences pose a huge problem to the safety of road users and leads to negative impact on the health, social and economic progress of the country.

Therefore, it is important to understand where and why such incidents normally happen so that the authorities can take effective measures to reduce the occurrences of such events. This calls for the need to analyse the spatial and temporal patterns of the traffic incidents, which can be achieved by applying Geographical Information Systems (GIS) [1].

GIS technology has played a crucial role in helping many traffic authorities to analyse traffic accidents and identify accident hotspots [2]. A wide range of traffic analysis has been performed by researchers in recent years. Some researchers used GIS to perform simple linear analyses including the number of crashes, number of deaths, while others used it for spatial analyses and identification of high risk areas [3].

2 OBJECTIVES

The research is motivated by a lack of publicly available and easy to use web enabled application that allows user to analyse traffic data in Singapore.

The team found that the only way for an average user to identify traffic accident hotspots is through the Land Transport Authority's Black Spot Programme found in the OneMotoring website. Under this programme, a list of accident prone areas is published in the website. However, as this information is displayed in a text format, people who are less familiar with the road might be unable to recognise those locations. This diminishes the effectiveness of the analysis.

Moreover, there are currently no known web application that allows the average user to analyse the patterns of road accidents and traffic congestions in Singapore.

Therefore, the team aims to incorporate GIS technology and spatial analysis methods to create a web analytical tool application that allows transport planners, businesses and the average user to analyse the patterns of traffic incidents along Singapore's expressway.

More specifically, the tool allows user to:

1. Identify road accidents hotspots and traffic congestions hotspots
2. Analyse spatial patterns of road accidents and traffic congestions
3. Analyse the spatial relationship between traffic cameras and road accidents / traffic congestions

3 RELATED WORKS

Geospatial Information Systems (GIS) based spatial analysis of traffic incidents is not new and has been used by researchers for many years.

Shafabakhsh, Famili and Bahadori applied GIS Technology and spatial statistical analysis methods to study the traffic accidents patterns in the road networks of Mashhad, Iran. The approach taken by the researchers was to utilise a combination of traditional Kernel Density Estimation, nearest neighbour analysis and K function. Kernel Density Estimation was used to identify the hotspots along the traffic network. While nearest neighbour analysis and K function were used to determine the presence or absence of clusters of accidents. The analysis was performed using ARCMAP and SANET 4th edition software [4].

A similar study was conducted on traffic accidents in Brazil on highway BR 277. The researchers used four techniques to identify and analyse accident hotspots. Kernel Density Estimation and wavelet analysis were used to identify the accident the hot spots. In addition, built environmental analysis and principal component analysis were performed to analyse the patterns of the accident occurrences [5].

Benedek, Ciobanub and Man analysed the traffic hotspots along the road networks of Cluj-Napoca in Romania and also discussed the social background of road traffic crash occurrences. Kernel Density Estimation was also employed to determine hotspots of traffic accidents along the road network [3].

As seen above, Kernel Density Estimation is a frequently used technique to determine traffic accident hotspots. Another similarity of the research identified is that the analysis was done using ARCGIS, a proprietary software created by ESRI, with a package called 'SANET'. While ARCGIS is a comprehensive spatial analysis tool, it has a steep learning curve for the average user.

4 ANALYSIS METHODS

4.1 Kernel Density Estimation along a Linear Network

As mentioned in the previous section, Kernel Density Estimation is one of the most commonly used technique to determine hotspots of point events. There are two known forms of Kernel Density Estimation: Planar Kernel Density Estimation (PKDE) and Network Kernel Density Estimation (NKDE). PKDE determines the hotspots of point patterns using Euclidean distance between events. This is not suitable for determining the hotspots of traffic crashes because they can only occur along a road network. On the other hand, NKDE is an extension of PKDE which adds in a linear network constraint, such as a road network, which makes it more relevant for our study. Therefore, the team have decided to use NKDE to determine the traffic accident hotspots along the road network.

NKDE calculates the density of a point events on a linear unit, such as a road network, rather than on a 2D homogenous area unit which is used in PKDE. The formula to calculate NKDE is shown below [6].

$$\hat{\lambda}_z(s) = \frac{1}{\sigma_z(s)} \sum_{i=1}^{n} \frac{1}{\tau^2} k\left(\frac{s - s_i}{\tau}\right)$$

NKDE can be applied in R using the density.lpp function from the spatstat package.

4.2 Ripley's K Function

The Ripley's K Function is a spatial analysis method used to describe how point patterns are distributed over an area of interest. It allows analyst to determine if the point patterns are dispersed, clustered or randomly distributed over an area of interest. The K function counts the number of point patterns found within a given distance, d of a point pattern. The number of point patterns within the circle of radius d is counted. The same step is repeated for all point patterns and the distance d is varied [7].

This gives us the K function of the observations after applying the formula below

$$\widehat{K}(h) = \frac{R}{n^2} \sum \sum_{i \neq j} \frac{I_h(d_{ij})}{W_{ij}}$$

where n is the sample size, A is the area of the plot, W_{ij}

The K function is the observations are then compared to the K function of point patterns that one would expect to find when the points are in complete spatial randomness. This is done by running Monte Carlo Simulations. If the K function of the observation is larger than the K function of the observation, it means that clustering exists. On the other hand, if the K function of the observation is smaller than the K function of the observations, it means that dispersion effect exists. Otherwise, if the K function of the observations is similar to the K function of the simulated points, it means that the observations are in complete spatial randomness.

Similar to Kernel Density Estimation, it exists in 2 forms: planar K function and network K function. The K function described above is the planar K function, which is unsuitable for the study. Therefore, the network K function will be used instead.

The network K function is an extension of the planar K Function and it considers the locational constraint by the network as well as the distance measurement constraint [9]. Similarly, when running the Monte Carlo simulations of the network K function, the random points will only be plotted along the road network and not on a planar 2D space.

4.3 Multitype K Function

The network K function can be extended to perform multitype network K function. The difference between this and the network K function is that, instead of counting the number of type i point patterns at a radius d from each type i point, the number of type j point patterns are counted. This step is then repeated for each point. The remaining computation is the same as normal network K function [8].

The network K function and multitype network K function can be applied using the linearK and linearKcross function respectively using the spatstat package.

5 DATA COLLECTION

There are three types of data that is required for the analysis - traffic incidents data, road network data and road cameras data.

The road accidents and traffic congestions data points were obtained from Land Transport Authority (LTA) DataMall. LTA DataMall is the only source of traffic data in Singapore. However, there are no historical traffic incident data available on the website, only real-time traffic data is available.

To obtain the traffic incident data, the calling of application programming interface (API) is required. As such, the team wrote a script that periodically calls the API. The inbuilt Windows Scheduler was used to automate and trigger the script hourly. This was done for a period of 5 weeks.

The road network data and road cameras data are publicly available and are downloaded from OpenStreetMap and data.gov.sg respectively.

6 DATA CLEANING

6.1 Initial Cleaning of Traffic Incidents Data

After five weeks of data collection, 7000 records of traffic incidents in Singapore were generated. The raw data is in the JSON format. The team converted the file to a CSV format for ease of manipulation.

The data contains several types of traffic incidents such as heavy traffic, accidents, roadwork, etc. As the only required data for the analysis are accident and heavy traffic data points, the other types of traffic incidents are filtered out. After filtering, around 2000 records were left.

6.2 Cleaning of Road Network

The road network data downloaded from OpenStreetMap contains all the road networks in Singapore. As the scope of the analysis is on Singapore's expressway, the irrelevant road networks need to be filtered out. Using the road type attribute of the road network, the researchers were able to extract out the expressway road networks using QGIS software.

6.3 Cleaning of Traffic Cameras Data

There are different types of traffic cameras along Singapore's road network. Each type of traffic camera is found in a separate file. Using the expressway road network, the team extracted traffic cameras that are found along the expressway. Next, this is repeated for the different types of traffic cameras. Lastly, the different traffic cameras are combined into a single SHP file. All of the manipulation was done using QGIS software.

6.4 Cleaning of Traffic Incidents Data

Similarly, the expressway road network is used to extract the traffic incidents that are found along the Singapore expressway. This was also done using the QGIS software.

7 SYSTEM ARCHITECTURE

As multiple statistical analyses are needed in this project, the team have chosen to use the R programming language as a tool for analysis. This is because it is a powerful statistical programming language with numerous spatial analysis packages required for the analysis. In addition, R offers a convenient way to build an interactive web application with minimal coding.

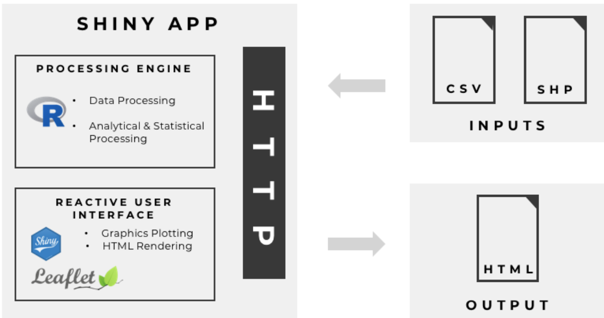

Fig 1. Multi-tier Architecture

The project is built with a multi-tier architecture where different tiers depict the different layers of interaction.

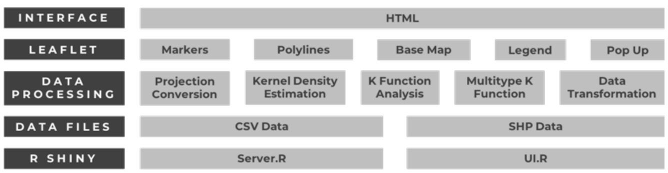

Fig 2. Architecture Layers

The key layer of the architecture is the R Shiny layer as it acts as an integration tool that links the server side computation to the visualization of the analysis to display on the interface. R Shiny performs all the data processing such as data transformation, Kernel Density Estimation, K Function and Multitype K Function on the server side, which then converts the output to HTML and displays to the users. The leaflet plugin allows us to plot the spatial points of the events and overlay the output of the Kernel Density Estimation onto the Singapore map. Users are then able to interact with the map and view the visualization. The web application also allow users to upload both CSV files and SHP files which are used as inputs for the analysis.

8 APPLICATION DESIGN

The application consists of two tabs – Upload Data Tab and the Main Tab.


Fig 3. Application Overview

8.1 Upload Data Tab

The Upload Data Tab allows user to upload their own data for analysis. The data that are required are the traffic data and the road network data. The traffic data must be in the CSV Format. The structure of the data needs to be in the same structure as the data released by LTA's Datamall. The road network data must be uploaded in the SHP File format. It should be in the projected coordinate system - SVY21.

8.2 Main Tab

The Main Tab consists of a side panel on the left and a map. A dark theme is used as it allows the analysis to stand out more. This directs the user's attention to the KDE output and the K Function analysis. The overall design is set as a clean and minimal interface to reduce unnecessary distraction.

Most of the user interface is covered by the map, with a simple side menu for the analysis selection as well as a top menu to toggle between different tabs to show the map, or to upload data.

8.2.1 Side Panel

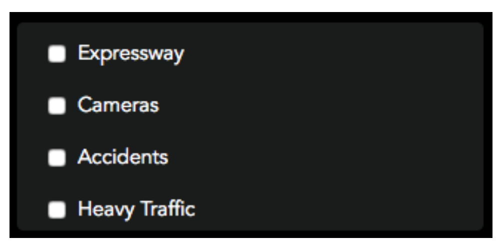

Fig 4. Side Panel

The Side Panel allows user to control the map visualisation as well as to select and view the analysis.

Users are able to toggle the different point features to display onto the map. The point features are generated from the files uploaded by users, such as Accident points as well as Heavy Traffic points.

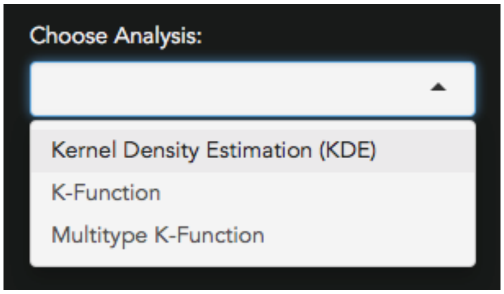

Fig 5. Type of Analysis

Users are allowed to choose between the three different analyses and depending on their selection, the side menu will display specific elements pertaining to the chosen analysis.

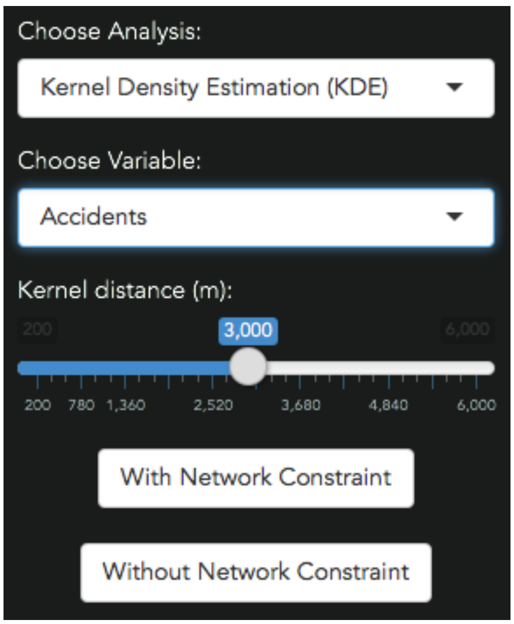

Fig 6. Kernel Density Estimation (KDE) Selection

Upon selection of the Kernel Density Estimation analysis, users will be able to select their desired variable (accidents or heavy traffic) as well as the kernel distance. The kernel distance will determine the noise level of the map, a lower kernel distance will result in a more localised pattern on the map while a larger kernel distance will provide a more generalised pattern. Lastly, users will have to select if they want to include or exclude the road network constraint.

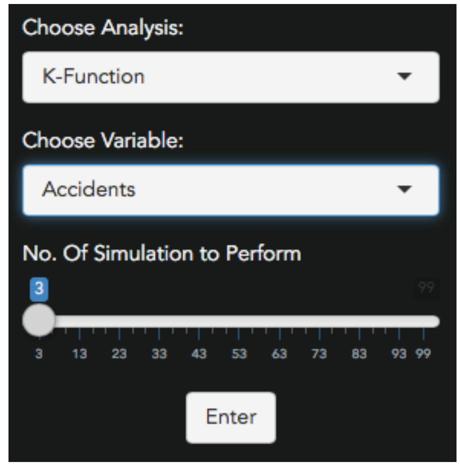

Fig 7. K-Function Selection

For both K Function and Multitype K Function, the user is able to select the variable to analyse, as well as the number of simulations to perform. The larger the number of simulations, the longer the processing time is needed. As the K Function algorithm is a computationally intensive algorithm, users are given the option to zoom in to perform their analysis on a smaller area so as to speed up the analysis. The analysis area is bounded by the window of the map.

8.2.2 Map View


Fig 8. Map View

The map view shows the point features, road network and the KDE output. User can hover over and click on the point features as well as the road network to display more information about the feature.

9 RESULTS

The authors generated two KDE plots with a kernel distance of 1.5km for both the accidents and heavy traffic.

By observing the KDE plot of the accidents, the team discovered a few expressways which are more prone to accidents, namely Seletar Expressway (SLE), Pan-Island Expressway (PIE) and Tampines Expressway (TPE). The exact location can be seen in the figure below. More specifically, the locations of the observed accident hotspots are along SLE between Upper Thomson Road exit and Woodlands Ave 12 Exit, PIE between Thomson Road Exit and BKE , PIE between Bedok North Rd Exit and KPE Exit and TPE between Tampines Ave 7 Exit and Elias Road Exit.

Fig 9. Kernel Density Estimation Output for Accidents

On the other hand, based on the KDE output for the heavy traffic, it is observed that the Central Expressway (CTE), PIE SLE and TPE are more heavily congested as compared to the other expressways. The exact locations are shown in the image below. More specifically, the areas are SLE between Mandai Road Exit and Woodlands Ave 2 Exit, PIE between Stevens Road Exit and BKE, TPE between Punggol Road Exit and Jalan Kayu Exit, TPE between Tampines Ave 7 Exit and Elias Road Exit and CTE between Ang Mo Kio Avenue 3 Exit and Moulmein Road Exit.

Fig 10. Kernel Density Estimation Output for Heavy Traffic

To further understand if the point patterns are indeed clustered from a statistical perspective, the team performed the K Function analysis. 50 simulations were selected and the analysis window was set as the area of the hotspots identified from the KDE output.

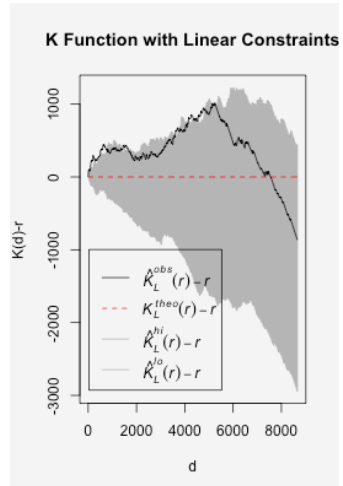

Fig 11. K Function on Accidents – PIE

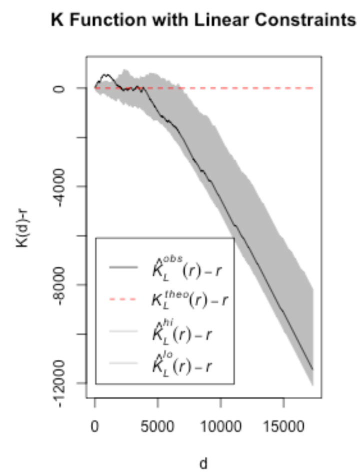

Fig 12. K Function on Accidents – BKE

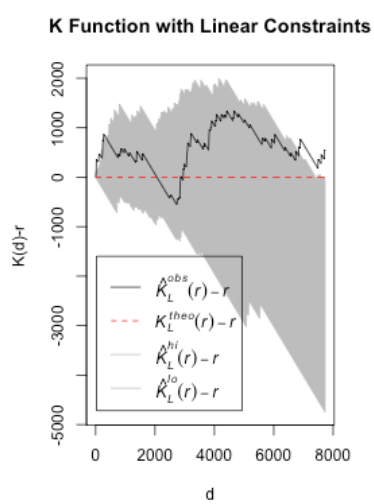

Fig 13. K Function on Accidents – TPE

As seen from the K Function analysis graph of the traffic accidents, it appears that the accidents points at the 4 hotspots are randomly distributed, contrary to what was observed in the KDE output. This is because the K Function of the observed accident points are between the envelope of the simulated point's K Function. This suggests that the observed accident points show a behavior which is similar to the behavior of the randomly distributed points of the simulations. Thus, from a statistical point of view, the accident points are not clustered.

Similarly, a K Function analysis was performed on heavy traffic using 50 simulations. The figure below shows the K Function output of the heavy traffic points. In contrast to accidents, the results show clear signs of clustering. The K function line of the observed points are all observed to be above the envelope of the simulated point's K function. This suggests that the heavy traffic points show a clustering behavior.

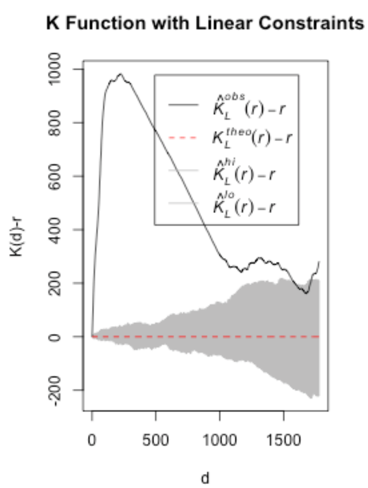

Fig 14. K Function on Heavy Traffic – PIE

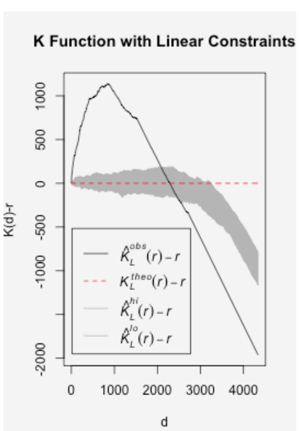

Fig 15. K Function on Heavy Traffic – CTE

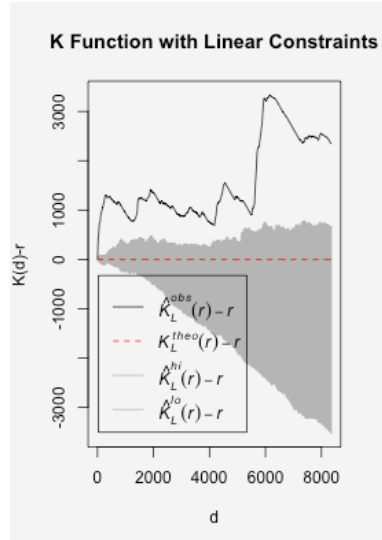

Fig 16. K Function on Heavy Traffic – SLE

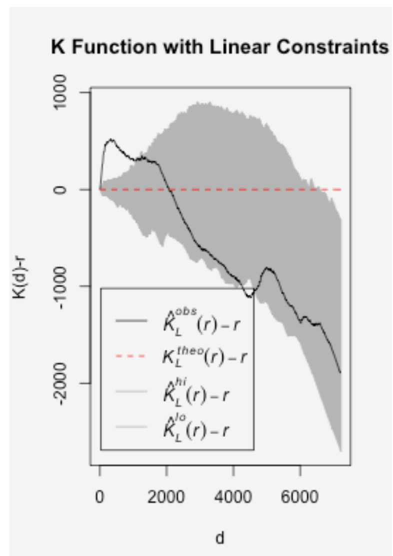

Fig 17. K Function on Heavy Traffic – TPE

As it was proven statistically that the heavy traffic points were clustered, the team would like to find out if this clustering could be correlated with other features along the road, such as the traffic cameras. Out of the 4 hotspots identified, two speed cameras are found to be situated in one of the hotspot along the CTE. Thus, a Multitype K Function analysis between heavy traffic and traffic cameras with 50 simulations was performed in the area along CTE. The figure below shows the output of the multitype K Function.

Fig 18. Multiple K Function on Heavy Traffic – CTE

Based on the multitype K Function output, it appears that there are some signs of clustering between 500m to 800m from the speed cameras.

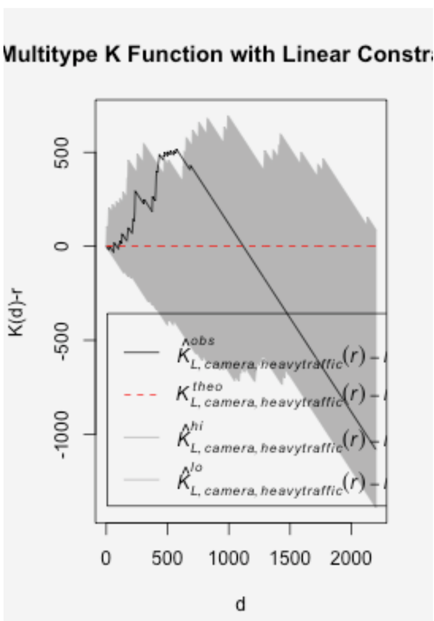

Fig 19. Multiple K Function on Heavy Traffic with Speed Camera

10 DISCUSSIONS ON ANALYSIS

From the analysis, the team could conclude that there are no signs of clustering of accidents along each expressway. This means that accidents are not a result of any particular features along the road. Road accidents that occur along Singapore's expressway are more likely due to other reasons such as lack of focus of the driver or drink driving.

However, this does not deny the fact that certain expressways tend to have a higher occurrence of accidents than other expressways. The higher risk expressways include SLE, PIE and TPE. Further analysis will be required to explain why certain expressways tend to have higher rate of accidents than others.

On the other hand, the team was able to detect clustering patterns of heavy traffic along the congested expressways. These expressways are SLE, TPE, PIE and CTE. However, the cause of the clustering could not be identified.

Using the Multitype K function, an additional variable - speed cameras, were brought in to analyse if there is any correlation between the clustering of heavy traffic and the speed cameras. The results suggest that there is some correlation between the speed cameras and the heavy traffic hotspots along CTE. Based on this analysis alone, it is insufficient to conclude that the heavy traffic along CTE is caused by the speed cameras. More ground work and analysis needs to be conducted to establish a causation of the speed cameras.

11 LIMITATIONS

The scope of the team's analysis is limited by the quantity of the data set that was collected.

As this project timeline was relatively short, only 5 weeks of live data could be collected. While the data is sufficient for spatial analysis, the size of the data set made it impossible to analyse the events across both space and time. If a filtering is done by hours of the day or days of the week, the remaining data that is available will be inadequate for a spatial analysis.

Given this limitation, the team is unable to answer further questions about the clusters of accidents. For example, what time does a particular accident hotspot becomes more prone to accidents?

12 TOWN HALL FEEDBACK

During the poster presentation that took place on 11 April 2018, the team received many valuable feedback on how the application could be improved.

One of the industry professional was interested to explore the patterns between Electronic Road Pricing(ERP) and the heavy traffic. This would allow the team to analyse the effectiveness of the ERP gantry in reducing heavy traffic. However, to perform such analysis, the team will have to filter the point events during the hours of operations of the ERP. As mentioned in the previous section, that feature unavailable at the moment given the size of the dataset available. It would be an interesting feature to include into the analysis once the team is able to acquire more data.

A common question that was received was how is the application useful when Google Maps already has a live traffic feature that can allow users to determine heavy traffic hotspots. The misconception is that the application created by the team aims to serve the same purpose as Google Maps. While Google Maps is useful in allowing road users to make adjustments to their travel plans, it lacks the capability to allow users to identify recurring hotspots, which is required by urban planners to identify problem areas and delivery companies to plan optimal routes.

13 FUTURE WORKS AND CONCLUSION

Moving forward, the team could collect more data for analysis. With a larger dataset available, the scope of the analysis could be broadened to allow both spatial and temporal analysis. In addition, the team could explore the correlation between traffic incidents and other features along the roads, such as road bends and ERP.

In conclusion, the team has demonstrated how GIS could be used to allow people to gain a better understanding of traffic incidents. Based on the analysis, four heavy traffic hotspots have been identified. This will prove useful to the traffic authorities in making the road better for everyone.

14 ACKNOWLEDGEMENTS

The team would like to thank Dr Kam Tin Seong for his kind support and guidance throughout the IS415 coursework.

REFERENCES

[1] Cheng, W., Washington, S., (2008). New criteria for evaluating methods of identifying hotspots. Transportation Research Record 2083, 76 - 85.

[2] DeepthiJayan, K., Ganeshkumar, B., (2010). Identification of accident hot spots: a GIS based implementation for Kannur District, Kerala. International Journal of Geomatics and
Geoscience 1 (1), 51 - 59.

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

[4] 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

[5] Andrade, L. D., Vissoci, J. R., Rodrigues, C. G., Finato, K., Carvalho, E., Pietrobon, R., . . . Carvalho, M. D. (2014). Brazilian Road Traffic Fatalities: A Spatial and Environmental Analysis. PLoS ONE, 9(1).

[6] Xie, Z., & Yan, J. (2008). Kernel density estimation of traffic accidents in a network
space. Computers, Environment and Urban Systems, 32(5), 396–406.

[7] How Multi-Distance Spatial Cluster Analysis: Ripley's k-function (Spatial Statistics) works. (n.d.). Retrieved April 14, 2018, from http://resources.esri.com/help/9.3/arcgisdesktop/com/gp_toolref/spatial_statistics_tools/how_multi_distance_spatial_cluster_analysis_colon_ripley_s_k_function_spatial_statistics_works.htm

[8] Ripley's K Function and Pair Correlation Function. (n.d.). Retrieved April 14, 2018, from http://wiki.landscapetoolbox.org/doku.php/spatial_analysis_methods:ripley_s_k_and_pair_correlation_function

[9] Yamada, I., & Thill, J. (2004). Comparison of planar and network K-functions in traffic accident analysis. Journal of Transport Geography, 12(2), 149-158. doi:10.1016/j.jtrangeo.2003.10.006