

GIS-BASED ANALYSIS TO DETECT ROAD ACCIDENT HOTSPOTS USING NETWORK KERNEL DENSITY ESTIMATION

Narong Pleerux*

Received: October 07, 2019; Revised: July 17, 2020; Accepted: August 12, 2020

Abstract

Thailand has one of the highest numbers of road accidents globally. A study of the location, duration, and cause of accidents provided critical information to plan and resolve tribulations. In this study, spatiotemporal analysis by employing geographic information system techniques by using network kernel density estimation as a tool along with a network toolbox was performed to analyze the density of road accidents in the Sri Racha district, Chon Buri province. Data for 2012-2017 were obtained from the road accident data center. Three areas with high accident densities-Laem Chabang City municipality, Sri Racha municipality, and Bowin subdistrict-were investigated because these areas are considered the economic, industrial, and transportation centers of Sri Racha, respectively. Surveillance guidelines for areas with high accident densities can be drawn from the results of this study, and they include planning approaches to prevent road accidents during specific high-risk hours.

Keywords: NKDE, SANET, hotspot, geoinformation technology

Introduction

Road accidents have become a rising concern in several countries as they often result in fatalities, injuries, and losses to property (Anderson, 2009; ICMR, 2009; Xie and Yan, 2013) in developing countries (Satria and Castro, 2016). Most road accidents are attributed to careless driving because of drivers and pedestrians often neglect traffic rules (Aghajania *et al.*, 2017). Therefore, understanding the scenario of road accidents, including how, where, and when they occur, is essential to minimizing accidents and increasing road safety (Benedek *et al.*, 2016). In particular, understanding spatiotemporal aspects related to accidents will help reduce them effectively. Mohaymany *et al.* (2013) indicated that it is difficult to achieve accurate

results because the data collection process is very time consuming when a large amount of information is involved. Over the past years, several studies investigated facts about accidents by employing different approaches. Harirforoush and Bellalite (2016) investigated the occurrence of accidents using accident hotspot measurements. Ouni and Belloumi (2018) performed a similar study using spatiotemporal analysis; many other studies investigated the relationship between real-time traffic and road safety (Zheng *et al.*, 2010; Xu *et al.*, 2013; Yu and Abdel-Aty, 2014).

Thailand is listed among countries with the highest road accidents. The ratio of fatalities related to road accidents was estimated as 24,237 people,

Faculty of Geoinformatics, Burapha University, Saensuk, Mueang, Chon Buri 20131, Thailand, narong_p@buu.ac.th

* Corresponding author

i.e., 36.2 per 100,000, for 2018 (WHO, 2018). The Sri Racha district in Chon Buri province was selected as a case study owing to its economic importance because the main deep-sea port of the country used for international shipping (Laem Chabang) and other industrial estates are located here. People from different parts of the country come into this city to earn a living; consequently, this has led to an increase in travel and transportation. In 2018, road accidents in Sri Racha district were divided into 3,716 injuries and 121 fatalities (ThaiRSC, 2019).

In this study, the geographic information system (GIS) was used to analyze spatiotemporal accident data and display the result of road accidents (Deepthi and Ganeshkumar, 2010). The effectiveness of GIS has been proved in many previous studies; for example, it was used for analyzing the density and distribution of road accidents, and for accident hotspot evaluation (Zheng *et al.*, 2010; Mohaymany *et al.*, 2013; Satria and Castro, 2016). Further, spatiotemporal statistical techniques in GIS have also been used for accident data. The most widely used technique is kernel density estimation (KDE), and it is useful and easy to understand. Thus, it is always applied to analyze the distribution and hotspots of road accidents (Pulugurtha *et al.*, 2007; Plug *et al.*, 2011; Bil *et al.*, 2013; Hashimoto *et al.*, 2016). Further, KDE is divided into two types: planar KDE (PKDE) and network KDE (NKDE) (Benedek *et al.*, 2016). PKDE is a statistical nonparametric analysis commonly used in road accident studies because it calculates accident density within a homogeneous 2D space using the Euclidean distance method between accident points (Delso *et al.*, 2018). However, in a real scenario, accidents occur on a 1D roadway (Yamada and Thill, 2004); therefore, if a 2D model is used for the analysis, the results may be inaccurate. The NKDE was thus developed to analyze accident points along with the road network (Xie and Yan, 2008).

The density of road accidents between the years 2012–2017 in Sri Racha district were analyzed using the NKDE method in the SANET Toolbox (Okabe *et al.*, 2015). The remainder of this paper is organized as follows. Section 2 presents the study area, data, and methods. Section 3 presents the results and discussion. Finally, Section 4 presents the conclusions.

Materials and Methods

Study Area

The study area is Sri Racha, a district in the Chon Buri province located on the eastern coast of Thailand with an area of approximately 616.40 km²

and a population of 51,197 people. Sri Racha is an industrial and transportation center that consists of many industrial estates. Laem Chabang Port is the 20th largest port in the world, and it has helped Sri Racha achieve rapid economic development and growth. Consequently, the number of vehicles in Sri Racha has increased, which has led to an increase in the number of road accidents.

Road accident database

Between 2012–2017, about 20,003 accidents were recorded in Sri Racha (Figure 1). Accident parameters included the date, time, and location of the accident. The accident data were collected from the road accident data center (ThaiRSC, 2019). The collected data were then managed into a form of GIS shapefile and grouped into spatiotemporal road accident data. The final data were divided into various groups: weekdays (Monday–Friday), weekends (Saturday–Sunday), and accidents involving injuries and fatalities; further, those in 24 h were divided into six subgroups based on times: 02:01–06:00, 06:01–10:00, 10:01–14:00, 14:01–18:00, 18:01–22:00, and 22:01–02:00.

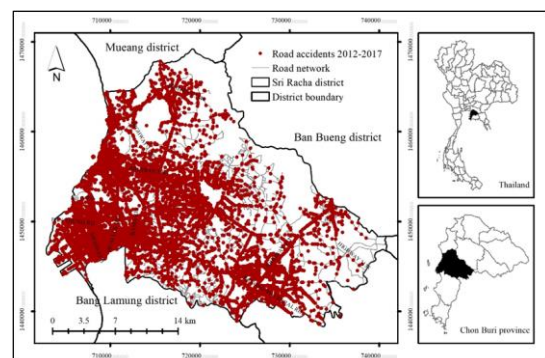


Figure 1. Study areas in Sri Racha district, Chon Buri province, and the distribution of accidents in 2012–2017

Linear Regression and Pearson's Correlation Coefficient

Linear regression is used to analyze the relationship between the number of road accidents and the number of injuries and fatalities from 2012–2017, and this relationship is expressed as

$$Y = aX + B \quad (1)$$

where X is the total number of road accidents, and Y is the number of injuries or fatalities.

The Pearson's correlation coefficient was calculated to determine the size and direction of the linear relationship between the number of road

accidents and the number of injuries and fatalities. The coefficient is given by

$$r = \frac{\sum_{i=1}^n \frac{x_i y_i}{n} - \bar{X} \cdot \bar{Y}}{S_x S_y} \quad (2)$$

where r is Pearson's correlation coefficient; $\bar{X}$ and $\bar{Y}$ are the averages of x and y , respectively; and S_x and S_y are respectively the standard deviations of x and y .

Network Kernel Density Estimation

SANET is a toolbox used to analyze incidents that occur along the network or parallel to the network such as road accidents. This research used SANET Standalone Beta 1.0 (Okabe *et al.*, 2015), which consists of NKDE tools to perform hotspot or density analysis of road accidents. The NKDE equation of SANET is shown as (Okabe *et al.*, 2009):

$$K_y(x) = \begin{cases} k(x), & \text{for } -h \leq x \leq 2d-h \\ k(x) - \frac{n-2}{n} k(2d-x), & \text{for } 2d-h \leq x \leq d \\ \frac{2}{n} k(x), & \text{for } d \leq x \leq h \end{cases} \quad (3)$$

where $k(x)$ is the basic kernel function, x is a point located on the network, y is the center of the kernel function, d is the shortest path from y to x (meters), h is the bandwidth (meters), and n is the node degree. In this analysis, the bandwidth value was set at 1,000 m and the cell size was equivalent to 50 m.

Results and Discussion

Spatiotemporal Road Accident Database

Between 2012-2017, 20,003 road accidents have been reported in the Sri Racha district. The number of accidents increase between 17:00-17:59 until 20:00-20:59, which records the highest number of road accidents (1,570). During these periods, heavy traffic congestion occurred because of factory workers leaving from various industrial estates, and workers leaving from Laem Chabang Port, students, and local people. Then, the trend of accidents began to decline to a minimum during 04:00-04:59 (261 accidents) because it was night, when most people were already asleep. The number of accidents increased again in the morning, especially between 07:00-07:59 (1,542 accidents) because people travel to work and school during this period. The traffic density then decreased until late afternoons, as people were still at work, school, and home; this resulted in a decreased number of accidents as shown in Figure 2.

When classifying the accidents on a day-by-day basis, it was found that their occurrence was higher on weekends than that on weekdays. The occurrence on Sunday was the highest (3,423 accidents), followed by Saturday (3,119 accidents). Meanwhile, the number of accidents during the weekdays was between 2,590-2,787 accidents. The increased accidents during weekends compared to that during weekdays are attributed to the tourist destinations in Sri Racha. This is because Sri Racha is one of the ways to Pattaya and Koh Lan, which are global tourist attractions. Therefore, during both weekends, Thai and foreign tourists travel through Sri Racha to reach Pattaya and Koh Lan. These were the major causes of increased accidents during the weekends as illustrated in Figure 2.

Correlation of road accidents, injuries, and fatalities

Figure 3 shows the relationship between the number of accidents and the number of injuries. Further, the figure shows the number of accidents and the number of fatalities using linear regression and Pearson's correlation coefficient.

Figure 3(a) shows the relationship between the number of accidents and the number of injuries in the same direction when the number of accidents increased; the number of injuries increased as well. This was reflected by the coefficient (R^2) value (0.9995), which implies that the possibility an accident could cause injuries is 99.95%. Based on the analysis of Pearson's correlation coefficient, it showed that the R^2 value for the number of accidents and injuries was equal to 1. This was highly correlated in a positive direction, and thus, it could be interpreted that the increase in accidents results in an increase in the number of injuries.

Figure 3(b) shows the number of accidents and the number of fatalities that were slightly related as indicated by the R^2 value, 0.2582. The accidents that occurred resulted in 25.82% fatalities. The analysis of Pearson's correlation coefficient showed that the

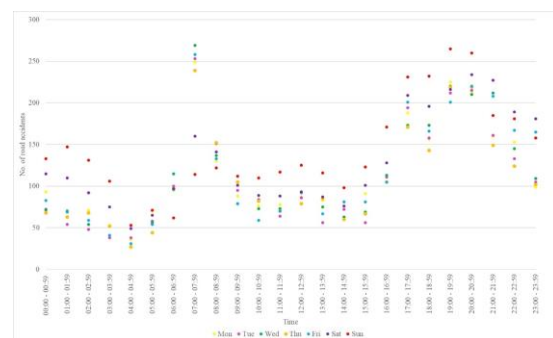


Figure 2. Number of hourly and daily road accidents during 2012-2017 in Sri Racha district

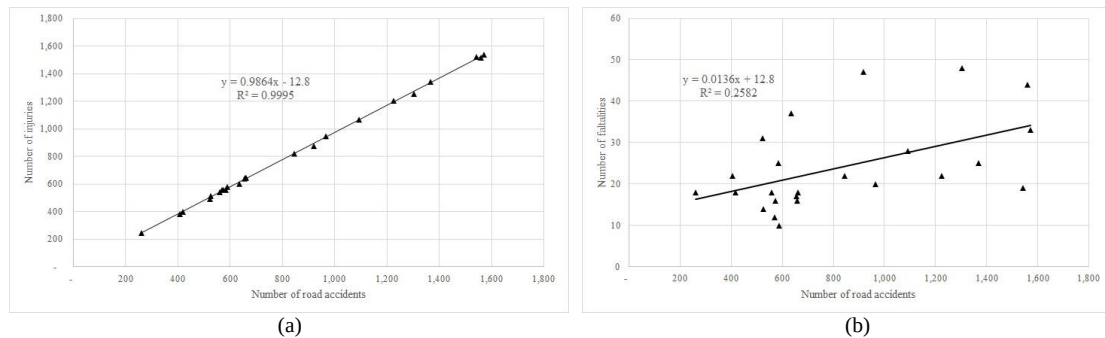


Figure 3. Linear regression of the (a) number of road accidents vs. number of injuries and (b) number of road accidents vs. number of fatalities

R^2 value of the number in accidents and fatalities was correlated at an average level of the R^2 value of 0.5080 and in a positive direction. Consequently, a rise in accidents increased the number of fatalities.

The analysis of linear regression and Pearson's correlation coefficient to examine the relationship between the number of accidents and the number of injuries and fatalities showed that the occurrence of accidents could cause injuries and fatalities. If the number of accidents increased, it would result in an increase in the number of people injured. However, the degree of injury was different in each case. Despite the positive relationship between road accidents and fatalities, the interpretation of such a relationship needs to be considered. From the evening until the night (17:00-21:59), there were many accidents and fatalities because these periods were rush hours, and there were several commuters; late at night (23:00-03:59), the number of accidents decreased, but the number of fatalities remained high. This can be attributed to several reasons such as darkness when driving in the night, drowsiness, and intoxication. In the mornings (07:00), there was another rush hour and very high traffic volume, which resulted in a high number of accidents; however, the number of fatalities were less compared to that in other periods.

Road Accident Hotspots

Road accident hotspots were analyzed using NKDE. The data were collected for 2012–2017 (20,003 instances) and divided into spatiotemporal parameters. The first group was divided into four subgroups: weekdays ($n = 13,461$); weekends ($n = 6,542$); injuries ($n = 19,423$); and fatalities ($n = 580$). Considering that accidents from the hotspots map indicate accidents in each group, three areas were found to have a high accident density: Laem Chabang City municipality, Sri Racha municipality, and Bowin subdistrict (Figures 4a-d). These areas are important in terms of economy, industry, and transportation. The Laem Chabang City

municipality is located southwest of the district and adjacent to the coastline; it is known for its transportation and logistics center. In addition, it is the location of the deep-sea port for international shipping; however, there are many shipping and logistics companies and a large population, most of which works at the port. Consequently, this area has high traffic congestion throughout the day because workers and trucks travel to the port. The high NKDE values were presented at Highway No. 3 (Sukhumvit Road) in front of Laem Chabang Port and other roads such as Laem Thong Road and Mueang Mai Klang Road.

The Sri Racha municipality is the center of the Sri Racha district and consists of government agencies, shops, hotels, and hospitals. Further, there is a pier crossing to Ko Sichang, which is an important tourist attraction in the district. In the morning and evening, these areas experience heavy traffic for long periods, especially during weekends because several tourists pass through the city. Further, it is a gateway to Pattaya and Koh Lan, which are other famous tourist destinations in Thailand. Both Thai and foreign tourists travel considerably, and this increases traffic during weekends and holidays, and consequently, the rate of accidents. The area around the clock tower has the highest accident density, and it includes Highway No. 3 (Sukhumvit Road).

Further, the Bowin subdistrict is another area with a high accident density; there are many large industrial estates such as WHH Chon Buri Industrial Estate 1, Eastern Seaboard Industrial Estate, Rojana Bowin Industrial Estate, Hemaraj Eastern Seaboard Industrial Estate 2, and Amata City Rayong Industrial Estate in this subdistrict. Several workers live in these areas, and therefore, the traffic is considerably heavy. This is caused by the high rate of movement of the workers to various industrial estates, transportation of goods, and production supplies at all times, in addition to the daily commute of local people. Further, it was found that

accidents occurred frequently in these areas, especially on Highway No. 33, which led to WHH Industrial Estate Chon Buri 1 and Rural Highway No. 4013 (Rayong).

Figure 4(d) shows the NKDE map of fatalities, which shows other areas with a high density of accidents. Area no.1, Nong Kho village, is a large community with many housing estates and high population density. The high NKDE values of fatalities occurred at the intersection of Highway No. 3241, which is close to the Nong Kho Reservoir. Area no. 2, the intersection of the National Highway No. 7 (Motorway) and Highway No. 3241, has a considerable number of shops and houses; further, it is near famous tourist destinations such as Sri Racha Tiger Zoo and Nong Kho Reservoir (at distances of 3 and 5 km, respectively). This area experiences heavy traffic when tourists visit these tourist destinations. Finally, Area no. 3, the upper part of Si Racha district around the intersection of the National Highway No. 7 (Motorway) and Highway No. 3144, is close to tourist centers such as Bang Phra Reservoir and Bang Phra Reservoir No Hunting Wildlife Area. The large number of hotels, resorts, and restaurants lead to heavy traffic in this area.

The NKDE was analyzed based on the occurrence of the accidents in 24 h. The road accident data were divided into six subgroups:

02:01–06:00 ($n = 1,537$), 06:01–10:00 ($n = 3,887$), 10:01–14:00 ($n = 2,300$), 14:01–18:00 ($n = 3,675$), 18:01–22:00 ($n = 5,654$), and 22:01–02:00 h. ($n = 2,950$). Figures 5(a-f) show that the subgroups with high accident densities were in the same areas, including Laem Chabang City municipality, Sri Racha municipality, and Bowin subdistrict. Figure 5(e) shows that the accident density from 18:01–22:00 had the highest number of accidents compared to other periods. Therefore, the density of accidents during these periods was found at the high level, followed by periods 22:01–02:00 and 06:01–10:00. The number of accidents during these periods was quite high, especially during 18:01–22:00 and 06:01–10:00. These were peak hours that included commuting students, workers, and the general public, in addition to the transportation of goods.

The analysis of accident hotspots using the NKDE method revealed that areas with high density include the Laem Chabang City municipality, Sri Racha municipality, and Bowin subdistrict. This was consistent with the accident density-analysis using the KDE method, indicating that the high-density areas were in these three areas (Pleerux, 2020). In conclusion, the occurrence of accidents in Sri Racha during 2012-2017 using the hotspot analysis by NKDE and KDE methods, were similar for both spatial patterns and the density distribution that appeared in the same area. The high-density

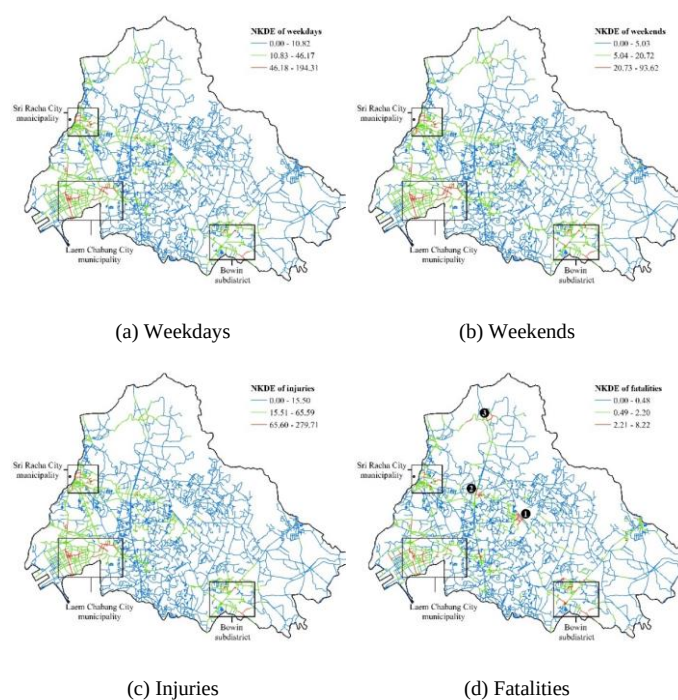


Figure 4. NKDE of road accidents in Sri Racha district: (a) on weekdays, (b) weekends, (c) involving injury, and (d) fatality

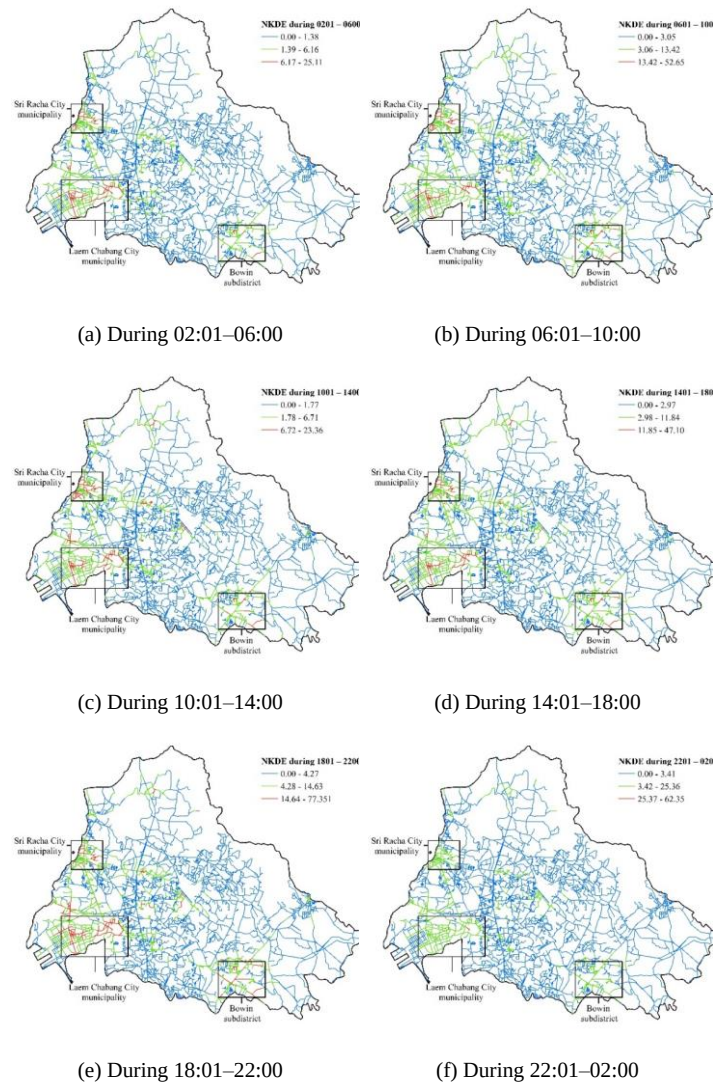


Figure 5. NKDE of road accidents in Sri Racha district during (a) 02:01–06:00, (b) 06:01–10:00, (c) 10:01–14:00, (d) 14:01–18:00, (e) 18:01–22:00, and (f) 22:01–02:00

areas were the same, irrespective of whether it was analyzed with NKDE or KDE.

Conclusions

Thailand is a country where accidents are a cause of concern, and therefore, identifying the location, period, and cause of the accident is very important. These can help solve the problem and reduce the number of road accidents. Sri Racha district was chosen as a case study, and the NKDE technique in the SANET Toolbox was applied to analyze the density of accidents on the road networks. The road accident data were divided into spatiotemporal groups. The results showed that high-density areas were found in three main areas: Laem Chabang City

municipality, Sri Racha municipality, and Bowin subdistrict. These areas are regarded as economic, transport, and logistics centers, and there are many large industrial estates in these areas. Further, these areas are densely populated and include local people and workers who have migrated from different areas. This migration is a major cause of traffic and road accidents in these areas.

This study employs GIS to study road accidents; the GIS can store and display information about the location and time of the accidents. The NKDE, the tool in SANET, was used to analyze the density of accident occurrence on the road network. The results are highly accurate and help traffic agencies identify areas with high accident risks, thereby leading to timely resolutions and a decrease in budget and time spent in operations.

There are some implications drawn from the research. First, NKDE is a technique used to determine the hotspots or density of road accidents. However, other methods can be used to study road accidents such as KDE, Getis Ord Gi*, and Ripley's K Function. Second, the road accident data used in this research were obtained from ThaiRSC; however, it does not contain all information about accidents that occur in the investigated areas. Therefore, to acquire complete information, the accident data need to be collected from other relevant agencies. Third, the road accident data still lack several important parameters such as gender, age, and causes of accidents. This information is useful and can be analyzed to determine a guideline to resolve the accident problem.

References

- Aghajania, M. A., Dezfoulianb,c, R. S., Arjroodyb, A. R., and Rezaeid, M. (2017). Applying GIS to identify the spatial and temporal patterns of road accidents using spatial statistics (case study: Ilam province, Iran). *Transp Res Proc.*, 25:2,126-2,138.
- Anderson, T.K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Anal Prev.*, 41:359-364.
- Benedek, J., Ciobanub, S.M. and Man, T.C. (2016). Hotspots and social background of urban traffic crashes: a case study in Cluj-Napoca (Romania). *Accident Anal Prev.*, 87:117-126.
- Bíl, M., Andrásik, R. and Janoska, Z. (2013). Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. *Accident Anal Prev.*, 55:265-273.
- Deepthi, J.K. and Ganeshkumar, B. (2010). Identification of accident hot spots: a GIS based implementation for Kannur district, Kerala. *Int. J. Geomat. Geosci.*, 1(1):51-59.
- Delso, J., Martín, B. and Ortega, E. (2018). A new procedure using network analysis and kernel density estimations to evaluate the effect of urban configurations on pedestrian mobility. The case study of Vitoria-Gasteiz. *J. Transp. Geogr.*, 67:61-72.
- Harirforoush, H. and Bellalite, L. (2016). A new integrated GIS-based analysis to detect hotspots: a case study of the city of Sherbrooke. *Accident. Anal. Prev.*, (In Press).
- Hashimoto, S., Yoshiki, S., Saeki, R., Mimura, Y., Ando, R. and Nanba, S. (2016). Development and application of traffic accident density estimation models using kernel density estimation. *J. Traffic. Transp. Eng. Engl. Ed.*, 3(3):262-270.
- Indian Council of Medical Research, (ICMR). (2009). Development of a feasibility module for road traffic injuries surveillance. *JCMR Bull.*, 39(10-12):41-50.
- Mohaymany, A.S., Shahri, M. and Mirbagheri, B. (2013). GIS-based method for detecting high-crash-risk road segments using network kernel density estimation. *Geo. Spat. Inf. Sci.*, 16(2):113-119.
- Okabe, A., Satoh, T. And Sugihara, K. (2009). A kernel density estimation method for networks, its computational method and a GIS-based tool. *Int. J. Geogr. Inf. Sci.*, 23(1):7-32.
- Okabe, A., Okunuki, K., Satoh, T., Shiode, S., Sugihara, K., Kawabata, M., Takahashi, A. and Morioka, W. (2015). SANET spatial analysis along networks. Available from: sanet.csis.u-tokyo.ac.jp. Access date: December 5, 2019.
- Ouni, F. and Belloumi, M. (2018). Spatio-temporal pattern of vulnerable road user's collisions hot spots and related risk factors for injury severity in Tunisia. *Transport Res F-Traf.*, 56:477-495.
- Pleerux, N. (2020). Geographic Information System-based analysis to identify the spatiotemporal patterns of road accidents in Sri Racha, Chon Buri, Thailand. *CAST*, 20(1):59-70.
- Plug, C., Xia, J.C. and Caulfield, C. (2011). Spatial and temporal visualisation techniques for crash analysis. *Accident. Anal. Prev.*, 43:1,937-1,946.
- Pulugurtha, S.S., Krishnakumar, V.K. and Nambisan, S.S. (2007). New methods to identify and rank high pedestrian crash zones: an illustration. *Accident. Anal. Prev.*, 39:800-811.
- Road Accident Data Center (ThaiRSC). (2019). Statistic of injuries and fatalities all provinces. Available from: www.thairsc.com. Access date: May 10, 2019.
- Satria, R. and Castro, M. (2016). GIS tools for analyzing accidents and road design: A review. *Transp Res Proc.*, 18:242-247.
- World Health Organization (WHO). (2018). Road traffic deaths: Data by country 2013. Available from: <http://www.who.int>. Access date: April 19, 2019.
- Xie, Z. and Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. *Comput. Environ. Urban. Syst.*, 32(5):396-406.
- Xie, Z. and Yan, J. (2013). Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. *J. Transp. Geogr.*, 31:64-71.
- Xu, C., Tarko, A.P., Wang, W. and Lui, P. (2013). Predicting crash likelihood and severity on freeways with real-time loop detector data. *Accident Analysis and Prevention* 57:30-39.
- Yamada, I. and Thill, J.-C. (2004). Comparison of planar and network K-functions in traffic accident analysis. *J. Transp. Geogr.*, 12:149-158.
- Yu, R. and Abdel-Aty, M. (2014). Using hierarchical bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data. *Accident. Anal. Prev.*, 62:161-167.
- Zheng, Z., Ahn, S. and Monsere, C.M. (2010). Impact of traffic oscillations on freeway crash occurrences. *Accident. Anal. Prev.*, 42:626-636.