Accident Analysis and Prevention 95 (2016) 236–249

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Land use and traffic collisions: A link-attribute analysis using Empirical Bayes method

Bruce Zi Yang*, Becky P.Y Loo

Department of Geography, The University of Hong Kong, Hong Kong, China

ARTICLE INFO

Article history:
Received 4 March 2016
Received in revised form 30 May 2016
Accepted 4 July 2016
Available online 25 July 2016

Keywords:
Public health
Hazardous road locations
Land uses

ABSTRACT

Road traffic collisions represent one of the major public health problems among the leading causes of deaths globally. This paper examines several approaches in detecting hazardous road locations, and discusses the spatial distribution of these locations as well as their relationships with different land uses in Hong Kong. Two most commonly used methodologies in detecting hazardous road locations are used: the hot spot and hot zone methodologies. Both methodologies are performed using raw collision count, excess collision count and Empirical Bayes (EB) estimations. The EB estimation uses land use characteristics near the road network in defining the reference groups. Finally all the approaches are compared by a test to assess their stability. The results show that for different hazardous road location detection methodologies, the best fit estimation methods on sites are different. The results confirm some land use impacts in previous studies, and suggest some further patterns on road safety. The findings are useful in understanding the complex interrelationships between land use and road safety, and in facilitating planners and policy makers to build safer cities.

1. Introduction

Road safety is a major public health issue since traffic collisions represent one of the leading causes of death globally (World Health Organization, 2013). As an important step in improving road safety, the detection of hazardous road locations has been studied for several decades. Currently, the most commonly used methods of hazardous road location detection are the hot spot and hot zone methodologies. The hot spot methodology, which is widely used by traffic and/or police authorities around the globe, considers road segments or junctions individually. If the number of traffic collisions at a site is above the critical value, a hot spot is detected (Elvik, 2008b; Maher and Mountain, 1988; Meuleners et al., 2008). Although the thresholds vary in different countries, the ways to define hot spots are quite similar: find the most dangerous sites on the road network that has a collision measure (usually collision count or collision rate) above a critical value.

Apart from the hot spot methodology, an alternative hot zone methodology (Flahaut, 2004; Loo, 2009) has been found to be useful in detecting hazardous road locations in recent years. The hot zone methodology considers the network or spatial continuity as well as the collision measure to identify a hazardous location: multiple sites can form a hot zone if there are at least two continuous segments with high collision counts, though lower than the traditional critical value used for detecting hot sites (Loo, 2009; Loo and Yao, 2013; Yao et al., 2015). But research in the hot zone methodology is not as rich as that in the hot spot methodology (Yao et al., 2015).

There is no consensus on which of these two methodologies is better, but it is believed that they are complementary and have respective advantages for different research purposes. For instance, Loo (2009) found that the hot zone methodology is more powerful in detecting hazardous road locations on expressways and in rural/suburb areas with fewer intersections. In contrast, the hot spot methodology may be more suitable for dense urban areas with small street blocks and frequent road intersections.

Regardless of the approach used, it is understood that pure collision counts cannot represent the dangerousness of a site very well since there is randomness in the occurrence of traffic collisions in each year. Thus, many other collision measures have been proposed to represent “dangerousness”. They include the collision rate, excess collision count, and Empirical Bayes (EB) estimation. Among them, the EB estimation (Hauer, 1997, 2001) was considered to be more reliable because it makes use of both the performance on the site and that on similar sites (that is, the reference population) to mediate the randomness of collision counts in a specific year (Elvik,

* Correspondence to: Room 1023, 10th Floor, The Jockey Club Tower C, Centennial Campus, Pokfulam Road, Hong Kong.
E-mail addresses: bzyang@connect.hku.hk (B.Z. Yang), bpyloo@hku.hk (B.P.Y Loo).

http://dx.doi.org/10.1016/j.aap.2016.07.002
0001-4575/

2008a; Li et al., 2008; Miranda-Moreno et al., 2007; Montella, 2010; Persaud and Lyon, 2007).

Up to date, most previous studies have only considered the engineering aspects of the road segment, such as the type of junctions, segment length or traffic volume, in calculating the expected collision count from the reference population in EB estimation. For example, Cheng and Washington (2008) used a safety performance function based on traffic volume; Persaud et al. (2010) used a function of both the road length and traffic volume. When people's activities are also considered, the land use categories could also be helpful in defining the reference population since land use is found to be among the most important environmental factors affecting road safety, especially in relation to pedestrian collisions (Dissanayake et al., 2009; Graham and Glaister, 2003; Loukaitou-Sideris et al., 2007; Priyantha Wedagama et al., 2006; Pulugurtha and Repaka, 2008; Pulugurtha and Sambhara, 2011; Wier et al., 2009).

In the 1990s, some scholars started to directly examine the relationship between land use and road safety. For example, Levine et al. (1995a,b) observed that traffic collisions are more likely to occur in employment centers than residential areas in Honolulu, Hawaii. Petch and Henson (2000) found out that the risk of children involved in a pedestrian-vehicle collision is hugely different in urban and rural areas, with the former being substantially higher. They also suggest that sub-district level analysis is essential to understand the distribution of child pedestrian-vehicle collisions. Kim and Yamashita (2002) further suggest that the largest proportion of vehicle–vehicle collisions was found near residential lands, with the second largest proportion in commercial areas. But when the rate of traffic collisions per acre of land is considered, the rate for commercial use is almost 4 times that for residential use. Although the study linked 10 years of collision data to only 1 year of land use information, the results are still enlightening since land development was quite slow in the study area. Wang and Kockelman (2013) argue that it is the mixture of residential and commercial land uses that are related to pedestrian-vehicle collisions. Elias and Shiftan (2014) found two zones in their study area to be the most dangerous for pedestrians: one is with mixed land use of residential, commercial and public buildings; the other is characterized by high residential density and low socio-economic status. In Hong Kong, Ng et al. (2002) examined 27 land use variables and their relationship with traffic risk, including area of different land use categories, number of facilities, etc.

Meanwhile, there are three major problems in previous road safety studies related to land use. The first one is that most previous studies are based on zonal analysis only. Zonal analysis, however, aggregates the collision statistics and removes the variability within the same zone. Aggregated traffic collision data at the zonal level can be integrated with land use data more easily (as both features are polygons in a GIS environment). But it does not consider the distribution of road networks and does not provide detailed location information for further hazardous road location detection. As all spatial heterogeneity within the zone is removed, a wrong impression that the zone is uniformly dangerous (or safe) may result. The larger the zone is, the greater the problem is. The second one is the representation of land use and its linkage to road safety: land use is described in a specific or highly selective manner, including the number of driveways of various types (Ivan et al., 2000), land use entropy (Wang and Kockelman, 2013), employment density, commercial, parks, number of schools, etc. (Miranda-Moreno et al., 2011) or part of an index (Cottrill and Thakuriah, 2010) instead of comprehensively based on all land use categories. The linkage between land use and collisions has similar problems: for example, only land uses near home (Elias and Shiftan, 2014) or at the nearest junction of each collision (Kim and Yamashita, 2002) are used in respective studies. We would argue that these are not good ways to link land uses to collisions because the former can only influence part of the trips, that is, home-based trips only; and the latter ignores the fact that traffic on the road network is affected not just by the closest land use to the nearest junction of a traffic collision. The third problem is that, there are some research papers on understanding the relationship between land uses and traffic collisions through modeling the former as explanatory factors for the latter, but there are very few studies on applying the knowledge about the land use impact on road safety in detecting hazardous road locations.

This paper tries to address these three research gaps. For the first one, we use the link-attribute approach based on the road network (that is, lines or one-dimensional space) instead of zonal analysis (that is, areas or two-dimensional space) is used to measure the land use attributes. For the second one, a systematic examination of all land use categories in Hong Kong and their combinations are considered. For the third one, the land use impact is incorporated into the EB method that is used in the estimation of hazardous road locations. Taking advantage of multiple years of road collision and land use data in Hong Kong, this paper will discuss the results and patterns of both hot spots and hot zones identified using different methods.

The organization of paper is straightforward. After the Introduction, a description of the methodology will then be given to illustrate our research design. Finally, the results of the analysis, spatial patterns and sensitivity tests will be presented and discussed.

2. Methodology

2.1. Study area and data

Three districts in Hong Kong are chosen as the study area: Wan Chai (on Hong Kong Island), Yau Tsim Mong (in Kowloon) and Tai Po (in the New Territories). Some basic statistics in these three districts are shown in Table 1. According to the numbers in the table, there are several differences among the three districts. First, according to the area, road network length and population, the developing intensity in the three districts are different. Yau Tsim Mong has the largest population, longest road network length and the least area. This is because the entire district is in the urban area of Hong Kong. Tai Po has the largest area (3.3% of total land area in Hong Kong), which is about three times the area of the other two districts combined, but also has the shortest road network length. This is because it is outside the urban areas in Hong Kong in the New Territories. The second difference is the possible vulnerable population in each district. Here we summarize the percentages of children (age under 14) and seniors (age over 65) in the total population in each district. The percentage of children in the total population in each district does not vary a lot from the average percentage in Hong Kong (11.6%). But the percentage of seniors varies a lot. Wan Chai has over 15% of population that are over 65 years old, but Tai Po only has 10.2%. Another noteworthy difference is that the demographics of the population in three districts are different. For example, the proportion of non-student population aged 20 and over having attained post-secondary education in Tai Po (24.1%) is close to the average in Hong Kong (25.9%). But the proportion in Yau Tsim Mong is much higher (31.8%) while the proportion in Wan Chai is even higher (46.2%). The median monthly household income in Tai Po (HKD 22,340) and Yau Tsim Mong (HKD 22,070) are slightly above the median in Hong Kong (HKD 20,500), but much lower than the figure in Wan Chai (36,150). The average domestic household sizes in three districts are roughly similar but there are more larger families living in Tai Po.

Table 1
Statistics about the Study Areas.

District	Tai Po		Yau Tsim Mong		Wan Chai		Whole Hong Kong
		Out of HK		Out of HK		Out of HK	
Area (square km)	36.44	3.30%	4.67	0.42%	8.88	0.80%	1104.39
Road Network Length (km)	90.21	3.48%	107.47	4.14%	97.87	3.77%	2594.27
No. of BSUs	1016	4.12%	1312	5.32%	1181	4.79%	24,651
Population	296,853	4.20%	307,878	4.35%	152,608	2.16%	7,071,576
	Out of the District		Out of the District		Out of the District		
Male/Female	1/1.13		1/1.16		1/1.21		1/1.14
Children (Age under 14)	10.20%		12.30%		10.20%		11.60%
Seniors (Age over 65)	10.70%		14.50%		15.60%		13.30%
Proportion of non-student population aged 20 and over having attained post-secondary education	24.10%		31.80%		46.20%		25.90%
Median monthly domestic household income (HK$)	22,340		22,070		36,150		20,500
Average domestic household size	3.1		2.7		2.7		2.9

Table 2
Consolidated Land Use Categories.

Original Land Uses (15 categories)	Consolidated Land Uses (9 categories)
Commercial	Commercial
Government, Institutional & Community (GIC)	Government, Institutional & Community (GIC)
Agriculture Conservation Area Green Belt	Green Space
Industrial or Storage	Industrial or Storage
Residential A (High Density Development)	Residential A (High Density Development)
Residential B (Median Density Development)	Residential B (Median Density Development)
Residential C (Low Density Development) Residential Village	Residential C (Low Density Development)
Residential Public Housing	Residential Public Housing
Construction Site/Vacant Open Space Other Specified Uses Utilities	Utilities

Table 3
Land Use Patterns in the Three Districts.

Land Uses	Tai Po	Yau Tsim Mong	Wan Chai
Commercial	0.69%	16.63%	4.94%
GIC	6.57%	18.74%	8.14%
Industrial or Storage	3.14%	0.25%	0.00%
Green Space	62.28%	0.00%	58.07%
Residential A (High Density)	0.73%	20.59%	2.98%
Residential B (Median Density)	1.19%	1.78%	5.68%
Residential C (Low Density)	14.35%	0.00%	7.34%
Residential Public Housing	1.95%	2.97%	0.00%
Utilities	9.11%	39.03%	12.85%

The land use data come from the Land Utilization Map and Outline Zoning Plans provided by the Planning Department, and are digitized manually. All land uses other than woodland and grassland without human daily activities and road network are considered. Unlike the broad zonal analysis in previous studies, the land use impact on road safety is examined at micro level in this study (100 m buffer from the road segment). Given the tedious nature of producing the land use data, this study cannot cover the whole territory of Hong Kong and only three districts with a variety of land use types are considered. In addition, it is recognized that the impact of land use on road safety is highly localized. The local level analysis will allow sufficient attention to be given to the hazardous road locations identified for road safety improvements.

In order to ensure that the number of land use categories is reasonable and appropriate for the EB method, some consolidations are made from the original land use categories, which are shown in Table 2. For example, Residential Village is merged into "Residential C" since they are low-density residential areas. Agriculture, Conservation Area, and Green Belt land uses are merged as "Green Space" since they are all distributed outside developed areas and are mostly covered with green plants. Open Space, Utilities, and Other Specified Uses are merged as "Utilities". They are not merged with "Industrial and Storage" because these "Utilities" land uses are distributed across developed areas while the "Industrial and Storage" land use is concentrated in an industrial estate in Tai Po.

The consolidated land use patterns of the three districts are shown in Table 3. According to the table, land uses in Tai Po clearly exhibit a new town pattern: very little high-density residential areas and most of them are in the town center; a lot of low-density residential areas all over the district; little commercial land uses; and plenty of green space. As for the land uses in Yau Tsim Mong, it has a clear urban center pattern: much area is devoted to commercial, high-density residential and Government, Institutional & Community land uses; no green space; and a lot of utilities. The land uses in Wan Chai is different from the other two: In the northern part of the district it is like an urban center, and in the southern part it is like a suburb area. It is hoped the different land use scenarios in these three districts in Hong Kong would give us more insights on the land use and road safety relationship.

The collision database TRADS are recorded by the Hong Kong Police Force, organized by the Transport Department of Hong Kong SAR Government, and geovalidated to the road network by the method proposed by Loo (2006). In this paper, a collision refers to an incident reported to the Hong Kong Police Force, involving personal injury occurring on roads in the Territory, in which one or more vehicles are involved (Transport Department of Hong Kong SAR Government, 2016). The road network information used in this study is from the Lands Department.

2.2. Defining "dangerousness" of road segments

There are many existing methods of defining the critical thresholds in detecting hazardous road locations (hot spots and hot zones). In this study, we tried three different methods: the pure collision count (CC), excess collision count (EX) and the Empirical Bayes (EB) methods. The CC method uses the pure collision count as the single measure of road safety. The threshold to determine hot spots and hot zones are entirely based on historical collision records. The EX method is based on the difference in collision number between the actual collision count and the predicted count. The actual collision count is calculated from road safety database and the predicted count is based on a linear regression model trained from empirical data using traffic volume as the independent variable. The EX method recognizes that some sites may have high collision count but it is not considered dangerous because it has high traffic volume. The aim is simply to select hazardous road locations based on the potential for reduction (differences away from the regression line) rather than for predicting traffic collisions based on traffic volume. For the later, a non-linear and more sophisticated model needs to be built (Loo and Anderson, 2015).

When the hot spot approach is used, the EB method was found to be having better performances in terms of stability in comparison with the collision frequency, collision rate, and potential for reduction approaches (Cheng and Washington, 2008; Montella, 2010). The main advantage is that the EB method can deal with the regression-to-mean (RTM) bias better. When the RTM bias exists, the "dangerous" sites with extra random collisions will turn to have lower collision numbers in the next study period since the "positive" randomness is gone. This reduction will happen regardless of any safety remedial treatment puts in place. Without considering this, the estimation of effectiveness of safety interventions will be over-estimated (Maher and Mountain, 2009). By using the performance from the reference population, the EB method can mediate the randomness across similar sites (that is, in the same reference population, in this case, performances of road segments with similar surrounding land uses). The fundamental idea is to calculate a weighed combination of both the performance of the reference population (the average collision count) and the actual collision count in each road segment or basic spatial unit (BSU) of analysis according to the characteristics of the land use in its nearby areas. The equation of the EB estimation can be shown as follows:

$$E\left\{x|X\right\} = aE\{x\} + (1-a)X \quad (1)$$

where E{x|X} is the collision estimate of a specific BSU; X is the collision number on the specific BSU; E{x} is the average collision number from the reference population; a is a parameter to represent the weights of the two parts, which is defined by:

$$a = 1/\left[1 + \left(VAR\{x\}/E\{x\}\right)\right] \quad (2)$$

where VAR{x} is the variance of the reference population.

In this study, we incorporate the land use information into the EB method: the definition of reference population groups. The road network is first segmented into shorter segments called BSUs. The standard length of BSUs is set to be 100 m since we need road segments to be short enough to capture the land use variations, while long enough to make the detection meaningful on the road network. After the network segmentation and dissolving short BSUs (Loo and Yao, 2013), the BSU layer is obtained and intersected with the traffic collision layers from 2007 to 2009 and 2010–2012 to get the number of collisions on each BSU in the two study periods of three years. Using aggregated three-year data can also remove some random fluctuations year by year and partially address the RTM bias. Finally, buffer zones are generated for the BSU layer; these zones are then intersected with the land use layer to characterize each BSU. The buffer distance is set to 100 m to capture the land use impact right near the BSUs. 500 m was also considered and tested but the total number of land use classes (15) is less than that using 100 m as buffer distance (19). More importantly, the numbers in each land use class has sharp differences: about half of the BSUs are defined as miscellaneous land uses, and 4 land use classes have less than 1% of the total BSUs. Therefore, 100 m is chosen as the buffer distance for further analysis.

Next, the proportion of each land use category within the 100 m-buffer zone is calculated. The land use with the largest percentage is marked as LU1, and the land use with the second largest percentage is labelled as LU2. Based on the above, each road segment can be classified as Single Land Use, Dual Land Use or Miscellaneous Land Use. The thresholds are tested according to the empirical data:

a Single Land Use

$P_{LU1} - P_{LU2} \geq 30\%$

- Dual Land Use

$P_{LU1} + P_{LU2} \geq 70\%$

$P_{LU1} - P_{LU2} < 30\%$

- Miscellaneous Land Use

$P_{LU1} + P_{LU2} < 70\%$

$P_{LU1} - P_{LU2} < 30\%$

There are reasons why we need to carefully define single, dual and miscellaneous land uses and choose the thresholds. At one extreme, we have simply considered one single "largest" land use type for each BSU. In this case, a BSU with multiple land use categories such as A (55%), B (40%) and C (5%) would only be represented by land use A. This is not desirable. Alternatively, we have tried to consider each BSU as either having a single land use type or more than one land use type. In this case, most BSUs would fall into the multiple land use categories. Through setting the thresholds for different land use types, the underlying patterns related to specific mixture of land uses can be examined. The threshold of 70% and 30% are decided based on the empirical data: thresholds of 80% and 10% create large number of single major land use and very few "two major land uses", and the number of reference groups is quite small, while thresholds of 50% and 40% do not generate meaningful "two major land uses" and distinct groups. For example when $P_{LU1} = 49\%$ and $P_{LU2} = 11\%$ ($P_{LU1} - P_{LU2} = 38\%$), it is not quite convincing to characterize the BSU as having two major land uses. After a pilot test, 70% and 30% are found to be the most appropriate thresholds.

Then BSUs with Single Land Use are marked as "LU1" land use. BSUs with Dual Land Use are marked as "LU1 and LU2" land use. BSUs with Miscellaneous Land Use are marked as "Miscellaneous" land use. Given the specific land uses (see Table 2 above) and some additional merges, a total of 19 final classes are used. They are called the referenced land use groups. The additional merges are necessary because of two reasons: 1. some classes have very few BSUs. If the number is less than 1% of the total number of BSUs, they will be merged into similar classes. Otherwise, there will be too few cases in the same reference category. 2. "LU1 and LU2" (with $P_{LU1} > P_{LU2}$) and "LU2 and LU1" (with $P_{LU1} < P_{LU2}$) are considered as the same. These referenced land use groups used in the EB method are shown in Table 4.

The process of developing the reference population of the EB approach is illustrated in Fig. 1.

2.3. Hot spot and hot zones

Next we proceed to consider the definition of hazardous road locations, including both hot spots and hot zones.

Table 4
Land Use Categories for the EB Analysis.

	BSU classes after merge	BSU classes before merge
Single Land Use	Commercial	Commercial
	Green Space	Green Space
	Industrial or Storage	Green Space and Industrial or Storage
		Industrial or Storage
		Industrial or Storage and GIC
		Industrial or Storage and Residential C
		Industrial or Storage and Utilities
		Residential C and Industrial or Storage
		Utilities and Industrial or Storage
	Residential A	Residential A
	Residential B	Residential B
	Residential C	Residential C
	Residential Public Housing	Residential Public Housing
	Utilities	Utilities
Dual Land Uses	Commercial and GIC	Commercial and GIC
		GIC and Commercial
	Commercial and Residential	Commercial and Residential A
		Residential A and Commercial
		Residential Public Housing and Commercial
	Commercial and Utilities	Commercial and Utilities
		Utilities and Commercial
	GIC	GIC
		GIC and Green Space
		Green Space and GIC
	GIC and Residential	GIC and Residential A
		GIC and Residential B
		GIC and Residential C
		GIC and Residential Public Housing
		Residential A and GIC
		Residential B and GIC
		Residential C and GIC
		Residential Public Housing and GIC
	GIC and Utilities	GIC and Utilities
		Utilities and GIC
	Residential and Green Space	Green Space and Residential A
		Green Space and Residential B
		Green Space and Residential C
		Green Space and Residential Public Housing
		Residential A and Green Space
		Residential B and Green Space
		Residential C and Green Space
		Residential Public Housing and Green Space
	Residential and Utilities	Residential A and Utilities
		Residential B and Utilities
		Residential C and Utilities
		Residential Public Housing and Utilities
		Utilities and Residential A
		Utilities and Residential B
		Utilities and Residential C
		Utilities and Residential Public Housing
	Utilities and Green Space	Green Space and Utilities
		Utilities and Green Space
Mixed Land Uses	Miscellaneous	Miscellaneous
	Mixed Residential	Residential A and Residential B
		Residential A and Residential Public Housing
		Residential B and Residential A
		Residential B and Residential C
		Residential B and Residential Public Housing
		Residential C and Residential B
		Residential C and Residential Public Housing
		Residential Public Housing and Residential A
		Residential Public Housing and Residential B
		Residential Public Housing and Residential C

2.3.1. *Hot spot methodology*

To recall, the definition of a hot spot (or black spot) might be different but the idea is similar: define the population of all possible locations on the road and find the most dangerous ones according to the critical threshold of a collision measure as hot spots. In this paper, a hot spot is defined as any BSU having a collision measure that is equal to or above the respective collision measure of the BSU ranked at top 5% among all BSUs. If there are ties at 5%, they will all be treated as hot spots. The three collision measures considered are the pure collision count, excess collision count, EB estimation.

2.3.2. *Hot zone methodology*

Instead of detecting the hazardous road locations solely on safety performances, the hot zone methodology also considers network or spatial continuity. Two steps are necessary to define a hot zone: first find all BSUs that having a collision measure that are equal to or above a critical threshold; then identify any two or more

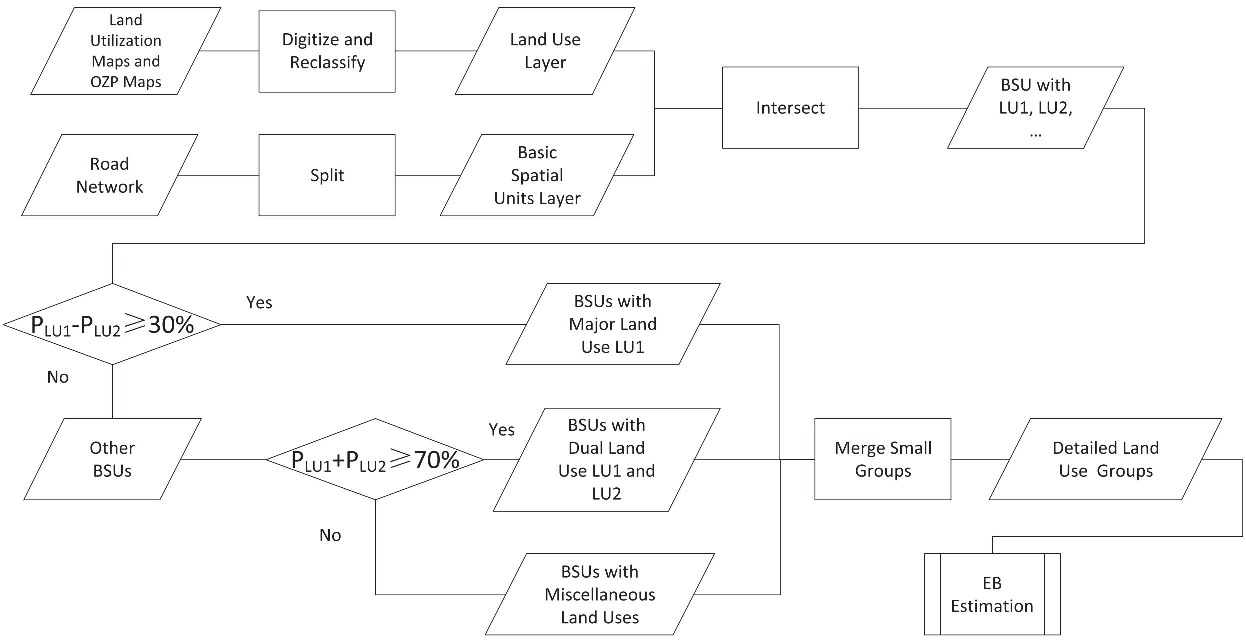


Fig. 1. Processes of Incorporating Land Use in the EB Method.

continuous BSUs within the group. Since a hot zone has at least two BSUs, the critical collision measure for hot zones is set lower than the values used for the hot spot methodology. Accordingly, the collision measure of the BSUs ranked at top 10% of all BSUs will roughly capture about twice as many potential BSUs for detecting the same number of hot zones. If there are ties at 10%, they will all be treated as candidates for forming hot spots. Similarly, the three collision measures considered are the pure collision count, excess collision count, EB estimation.

2.4. Stability tests

After the generation of hot spots and hot zones, statistical tests are used to compare the stability among different approaches. Among many tests available (Cheng and Washington, 2008; Montella, 2010), we choose the Method Consistency Test (MCT) to measure the stability here. The purpose of MCT is to check whether the hazardous locations detected in two consecutive periods are the same. MCT simply considers the degree of overlap of the dangerous locations. It considers neither the relative dangerousness among these detected sites nor the number of collisions on these sites. Previous studies have looked at the number of sites commonly detected as hot spots in two consecutive time periods. In this study, a normalization process, however, is necessary. This is because when there are ties at the critical value to define hot spots, the number of hot spots detected in different study periods would be different. More importantly, depending on the underlying pattern of spatial contiguity of dangerous road sections, the number of hot zones will vary even if the critical values to define the hot zones are the same. The normalization equation is given as follows:

$$MCT = \frac{count\,(M \cap N)}{\min\,(m,n)} \quad (3)$$

where M represents the BSUs detected as hot spots/hot zones in the first study period; N represents all the BSUs detected in the second study period; m and n are the number of BSUs detected in the two consecutive study periods. If the hazardous BSUs detected in the two study periods are the same, the test result should be 1. If the BSUs detected in two study periods are totally different, the test result should be 0. In other words, the test result is between 0 and 1. The larger the value is, the higher is the stability of the approach tested.

3. Results and discussion

3.1. Descriptive statistics on land use and collisions

The maximum (Max) and average (Ave), as well as the Standard Deviation (SD), of collisions happened on BSUs in each land use class during 2007–2009 and 2010–2012 are shown in Table 5. The minimum collision number is 0 for all groups and hence is not reported. Some general observations can be made. Firstly, previous studies seem to suggest that road segments near commercial land use tend to have more collisions than residential or other land uses. This is also true in Hong Kong. “Commercial” and all of its combinations with other land uses have very high average collision counts: In both study periods, the top three dangerous land use groups according to the average collision counts are “Commercial and Residential”, “Commercial and GIC” and “Commercial”. And these three groups are the only land use groups with an average collision number above 4. Within these three groups, the single land use “Commercial” is slightly less dangerous than the other two. During 2007–2009, the “Commercial” land use group has an average collision number of 4.53 but the number in the other two land use groups are all above 5. During 2010–2012, BSUs near single commercial land use has an average collision number of 4.29, while the numbers in the other two commercial land use groups are 4.55 and 5.66 respectively. Yet, commercial land use is not necessarily more dangerous when mixed with other land uses. For instance, “Commercial and Utilities” has 3 and 2.72 average numbers of collisions in the two study periods, which are all lower than the single land use category of “Commercial”.

Secondly, residential land uses have also different safety performances. High-density residential (“Residential A”, “Residential Public Housing”) land uses seem to be much more dangerous than low-density residential land uses (“Residential B”, “Residential C”). The two high-density residential land use groups have an average collision number over 1.7 but the numbers in the two low-density residential referenced land use groups never exceed 0.8. Also, the high-density residential land use groups have higher variability

Table 5
Road Safety Statistics by Land Use Class.

	Land Uses	No. of BSUs in each class	Collisions happened on BSUs in each land use class					
			2007–2009			2010–2012		
			Max	Ave	SD	Max	Ave	SD
	Single Land Uses							
1	Commercial	378	32	4.53	5.71	41	4.29	5.53
2	GIC	185	20	1.74	2.63	17	1.8	2.98
3	Green Space	392	58	0.78	3.6	53	0.9	3.68
4	Industrial or Storage	94	22	1.68	3.01	20	2	3.38
5	Residential A (High Density)	377	45	2.85	4.57	59	2.75	5.1
6	Residential B (Median Density)	39	5	0.62	1.29	7	0.79	1.52
7	Residential C (Low Density)	209	16	0.48	1.48	27	0.45	2.09
8	Residential Public Housing	59	8	1.78	2.5	10	2.02	2.5
9	Utilities	294	37	2.13	4.47	31	2.49	4.71
	Dual Land Uses							
1	Commercial and GIC	92	26	5.16	5.44	30	4.55	5.69
2	Commercial and Residential	137	54	5.22	7.24	64	5.66	9
3	Commercial and Utilities	128	28	3	4.98	24	2.72	4.48
4	GIC and Residential	163	19	2.74	3.75	23	2.75	4.06
5	GIC and Utilities	152	59	3.69	7.64	42	3.22	6.08
6	Residential and Green Space	177	17	1.36	2.56	14	1.44	2.52
7	Residential and Utilities	267	36	2.76	4.91	37	2.78	5.36
8	Utilities and Green Space	42	8	1.52	2.14	14	1.81	3.2
	Miscellaneous Land Uses							
1	Miscellaneous	249	63	3.29	6.95	42	3.65	7.18
2	Mixed Residential	75	12	2.33	2.66	10	2.21	2.56

(SD all over 2.5) than the two low-density residential referenced land use groups (SD all below 2.1). The "Mixed Residential" group has an average collision number over 2.2. Compared to the other residential land uses, the average collision number is higher for "Residential Public Housing" than "Residential A" in both study periods.

Thirdly, the "Government, Institutional & Community" single land use is comparatively safe with less than or equal to 1.8 average collisions. But "Commercial and GIC" (with over 4.5 collisions), "GIC and Residential" (with over 2.7 collisions) and "GIC and Utilities" (with over 3.2 collisions) land use groups all seem to be quite dangerous. These three dual land-use groups seem to be even more dangerous than the land use groups of "Commercial", "Mixed Residential", and "Utilities" respectively, suggesting that "GIC" land use is more dangerous when mixed with others.

Fourthly, the "Industrial or Storage" land use group, which is always believed to be dangerous due to heavy trucks, has only 1.68 and 2 average collisions in the two study periods. The number is not high compared to many of the other land use groups except for the low-density residential groups (less than 1), "Green Space" (less than 1), "Residential and Green Space" (less than 1.5) and "Utilities and Green Space" (less than 1.9). Maybe the heavy trucks are more likely to be related to collision severity rather than the number of collisions. Also, this may be due to the fact that all three districts are not major industrial areas. Hence, many combinations of the "Industrial or Storage" land use with other land uses are also merged into this single land use category. This is also reflected to the high variability of average collisions (high SD above 3) in both study periods.

Finally, the safety performance of roads with "Miscellaneous" land uses has very high variability (large SD of 6.95 and 7.18 in the two study periods respectively). The values of the SD are even higher than the average collision counts. When only the average collisions is considered, BSUs with "Miscellaneous" land uses seem to be more dangerous than both high-density "Residential A" and "Mixed Residential" land uses. But it is still safer than the top 3 commercial-related land use groups. In other words, mixed land use may not necessarily be safer or more dangerous than single or dual land uses.

3.2. Land use and hazardous road locations

Following the methodological procedures described above, about 180 hot spots and 50–80 hot zones (involving 210–270 BSUs) are identified in each three-year study period. The number of BSUs identified as hot spots/hot zones by the three methods are shown in Table 6. It is noticeable that using the hot spot methodology the number of sites detected has very little difference across different methods and in the two study periods. When using the hot zone approach, some patterns emerged. Regardless of the methods used, there are fewer hot zones in the 2007–2009 period compared to the 2010–2012 period, but these hot zones are longer. The CC method always detects the highest number of hot zones with the largest number of BSUs involved. This may suggest that the pure collision count method is too crude and it captures some of the less dangerous sites as hot spots. The EB method always detects the lowest number of hot zones, but the average length (represented by the average number of BSUs within a hot zone) of these hot zones is always longer than the other two methods. This underlines the methodological strengths of hot zones in identifying hazardous locations along a continuous stretch of the road network.

The locations and hot spots commonly detected in both study periods are shown in Fig. 2(a)–(c) and hot zones commonly detected are shown in Fig. 3(a)–(c). Generally the distribution of hot spots is scattered over the districts but the patterns of hot zones seem more concentrated. The spatial patterns of hot spots and hot zones identified in both the 2007–2009 and 2010–2012 periods by the three different methods are quite similar. This may be due to the use of three-year aggregated collision data. Aggregating the data already removes some of the random effect and partially addressed the RTM bias. Thus the advantage of EX and EB methods is not clearly shown in detecting the underlying patterns over the two study periods.

These maps, without showing all hot spots and hot zones identified in each period, are not intended for comparing different methods but for facilitating policy makers to identify and further investigator hazardous road locations commonly detected by all methods. Based on the overlaps of the maps, 115 BSUs are detected as hot spots in both study periods using all three methods. Among

Table 6
Number of Hot Spots/Hot Zones Detected.

	Hot Spots		Hot Zones					
	2007–2009	2010–2012	2007–2009			2010–2012		
	Number of hot spots	Number of hot spots	Number of hot zones	Total number of BSUs involved	Average number of BSUs in hot zones	Number of hot zones	Total number of BSUs involved	Average number of BSUs in hot zones
CC	182	188	57	272	4.77	75	262	3.49
EX	176	176	55	219	3.98	66	212	3.21
EB	179	181	48	234	4.88	65	230	3.54

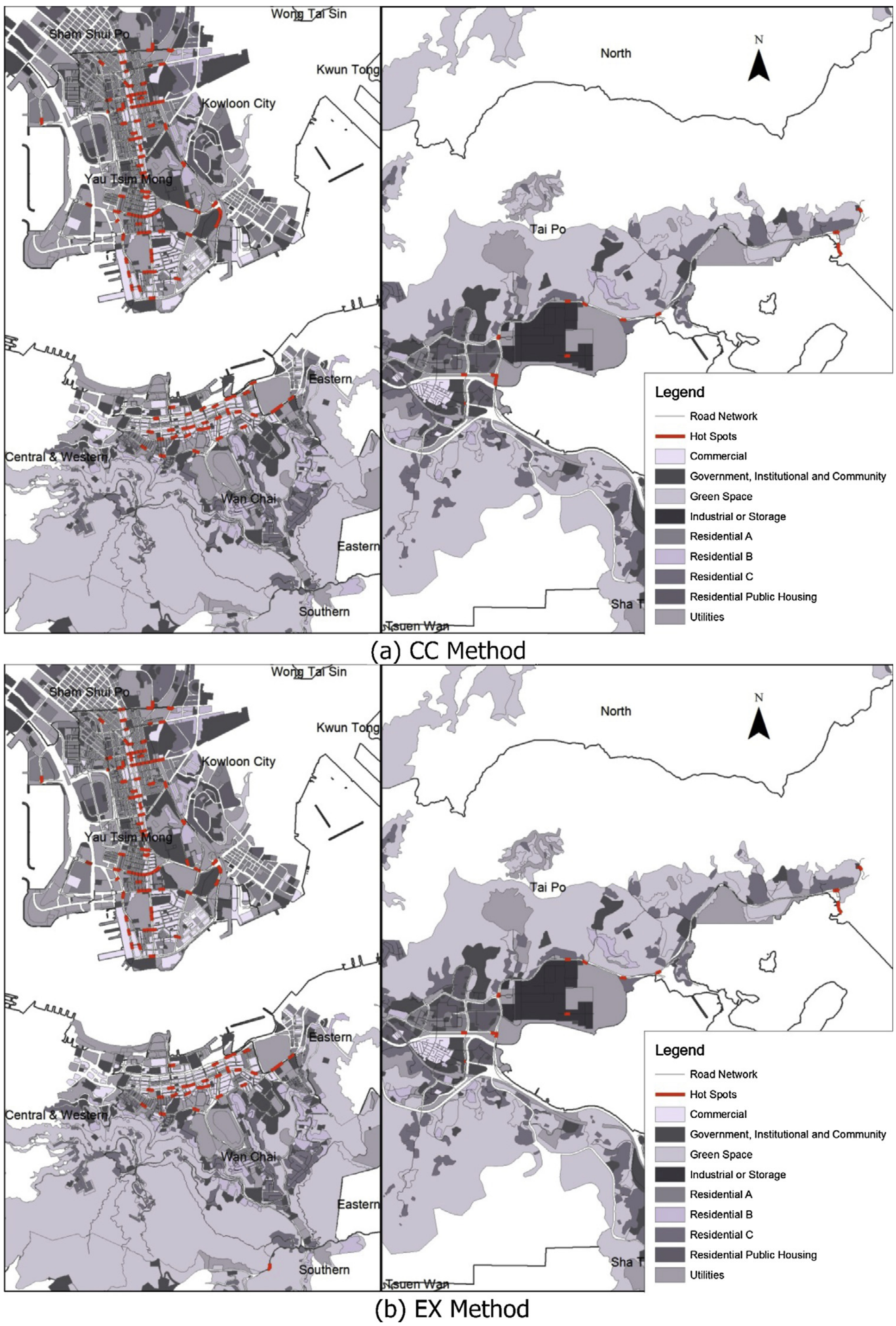


Fig. 2. Spatial Patterns of Hot Spots by Different Methods.

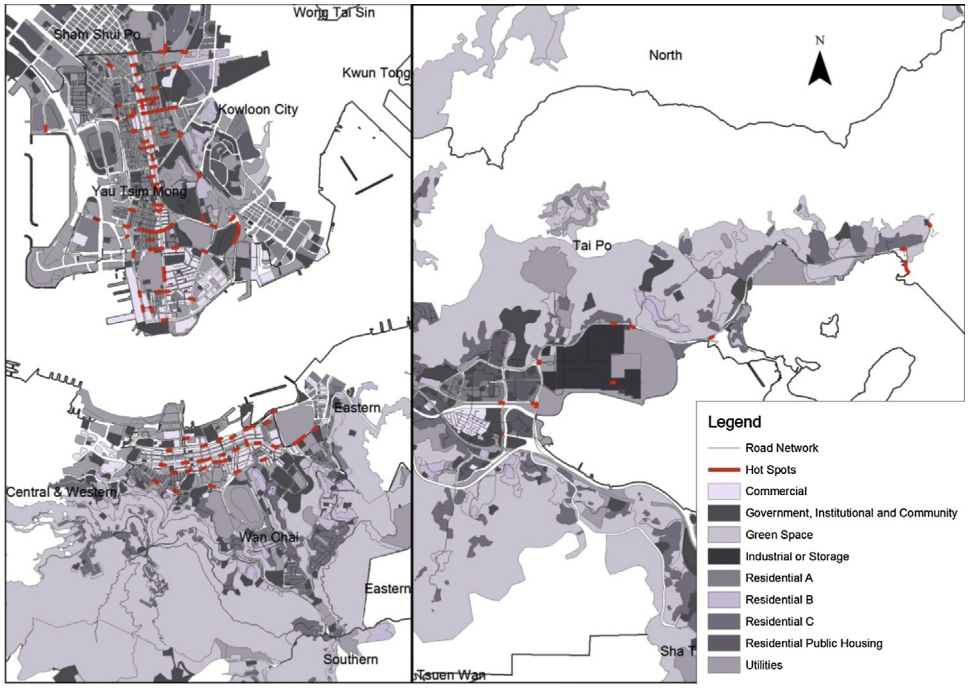


Fig. 2. (*Continued*)

them, 14 of them are in Tai Po, 74 are in Yau Tsim Mong and the other 27 are in Wan Chai. As for 135 BSUs commonly detected as hot zones by the three methods, 66 are in Yau Tsim Mong, 50 are in Wan Chai and 19 are in Tai Po. In other words, Yau Tsim Mong has over 60% of the hot spots and 48% of the hot-zone BSUs. At this stage, it is not clear how the different land use patterns play a role but it is clear that Yau Tsim Mong is having the highest population and road densities, with the highest shares of commercial and GIC land uses (Table 3).

Second, from these maps we can see that there is a clear urban-suburb difference in the distribution of hot spots. In the suburban area (Tai Po), most of the hot spots are at junctions of main roads, except for one inside the Tai Po Industrial Estate. But in Yau Tsim Mong, many hot spots are found on the lower-level roads that connect major roads. In Wan Chai, where the northern part is with high-density commercial and residential land uses as urban area while the southern part is with some low-density residential and green spaces, hot spots and hot zones are mainly on roads that are in the northern part of the district. These roads are in east-west direction, connecting the waterfront areas of the Hong Kong Island. The differences among the three districts might due to the design of road network since the road network is less dense in suburban areas and the connectivity is also lower. But also, the land uses in different districts may matter as well. The issue will be discussed in the next section (3.3).

Finally, it is also noteworthy that the major arterial road in Yau Tsim Mong is a north-south direction road called Nathan Road in the center of the district. Basically all hot spots and hot zones are allocated in the north-south direction. But only two hot spots and four BSUs within hot zones are actually located on this road. The other hot spots and hot zones are mainly on the minor roads that are intersecting with Nathan Road. These roads are narrower than Nathan Road, and they don't necessarily have higher traffic volume than Nathan Road. One of the common characteristics is that, these BSUs are all located near commercial or high-density residential areas. This is probably another reason why we need to look further into the land use impact on road safety.

3.3. *Land use and hot spots/hot zones*

Next we analyze the patterns of hot spots and hot zones across different road segment classes with different land use characteristics. The percentages of BSUs detected as hot spots/hot zones in each referenced land use group are shown in Fig. 4. It is obvious that the percentage of BSUs detected as hot spots/hot zones in Tai Po is quite low compared to the percentages in the other districts. In Tai Po, there is no land use group with over 10% of BSUs detected in any of the study periods by using either the hot spot or hot zone methodology. Neither is there a land use group having consistently above 5% of the BSUs detected as hot zones. As for the other two districts, the percentage of BSUs detected as hot zones seem generally higher than the percentage of BSUs detected as hot spots.

Again, BSUs with nearby commercial land use are more likely to be hazardous road locations. When all the three districts are considered (first row in Fig. 4), "Commercial and Residential" is the most dangerous referenced land use group among all other land uses. It has the highest percentage of BSUs detected as hot spots and/or hot zones. In 2007–2009, 1 in every 6–7 BSUs within "Commercial and Residential" is detected as dangerous. In 2010–2012, 1 in every 4–5 BSUs within "Commercial and Residential" is detected as dangerous. The "Commercial" and "Commercial and GIC" land use categories also have 7–23% BSUs identified as hot spots and/or hot zones. BSUs within these three commercial-related land use groups are the most likely to be detected as hot spots/hot zones. When the three districts are considered separately, there is a difference between Tai Po and the other two districts. There are not many BSUs in the commercial-related land use groups in Tai Po, and they are not having high shares (less than 4% in both study periods) of hazardous locations. On contrary, "Commercial", "Commercial and GIC" and "Commercial and Residential" in both Yau Tsim Mong and Wan Chai have quite high percentages of BSUs detected as hot spots/hot zones. In these three commercial-related groups, the percentage of BSUs detected as hot zones is always higher than the percentage of the BSUs detected as hot spots. This is suggesting that in the urban context, the commercial land uses might be also related with continuous dangerous road segments rather than individual dangerous locations.

The safety performance of mixed land uses is again very variable. The "Miscellaneous" land use group is the fourth dangerous according to the percentage of BSUs detected as hot spots and/or hot zones in all three districts (first row of Fig. 4). But in the three districts, the situations actually vary. In Tai Po, it is the most dangerous land use group using the hot spot methodology, and the second dangerous land use group using the hot zone methodology. But in Wan Chai, it ranked at the lower half of different land use groups that have BSUs detected as hot spots/hot zones. In Yau Tsim Mong, it seems quite dangerous using the hot spot methodology (ranked second), but not as dangerous as the commercial-related land use groups using the hot zone methodology (ranked fourth). This suggests that mixed land use does not necessarily make the road safer or more dangerous, but it depends on other contexts.

Residential land uses are not among the most dangerous land use groups, unless it is mixed with commercial land use ("Commercial and Residential"). Generally, high-density residential ("Residential A") land use is more dangerous than low-density residential land uses according to the percentage of BSUs detected as hot spots and/or hot zones. This finding is consistent with the results using the average collision count in each land use group. The land use group "Residential A" has 8.11% of the BSUs detected

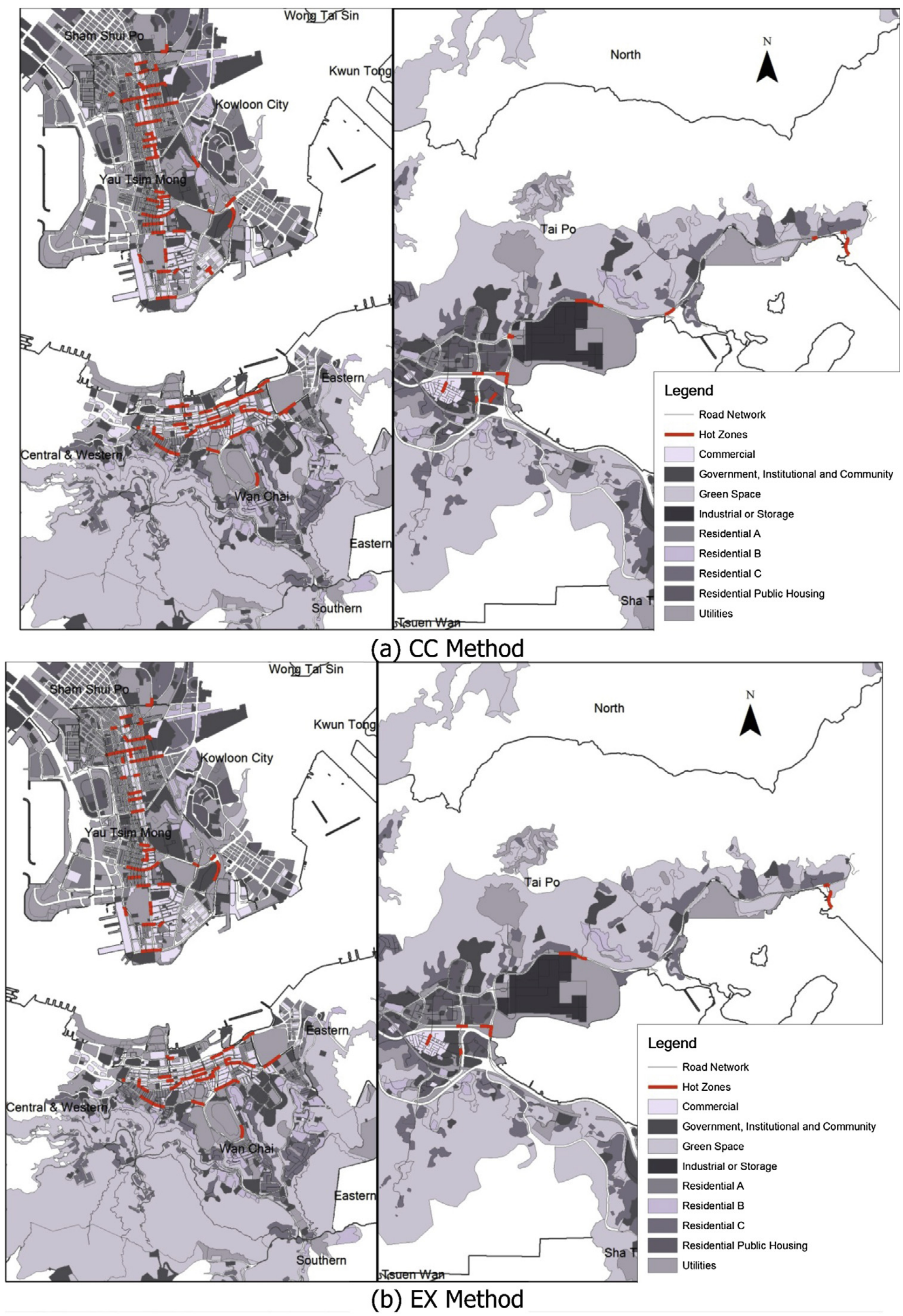
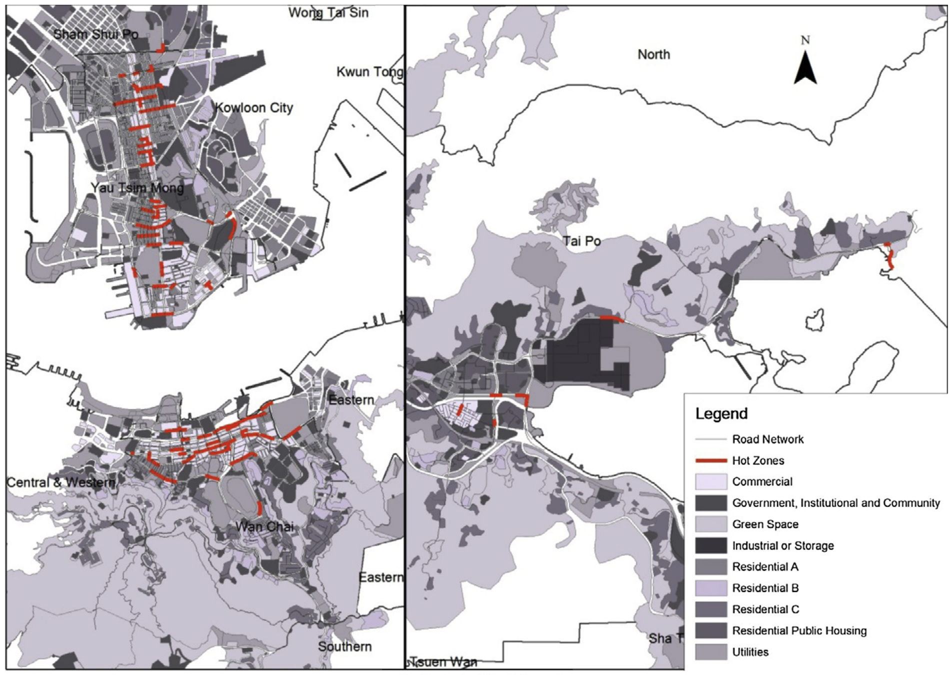


Fig. 3. Spatial Patterns of Hot Zones by Different Methods.



(c) EB Method

Fig. 3. (*Continued*)

as hot spots in Tai Po during 2007–2009. This is the highest percentage among all the referenced land use groups in Tai Po, but it can be due to some randomness since there are not so many BSUs detected as hot spots during 2010–2012. The percentages of BSUs detected as hot zones are also low.

The GIC land use has very different safety performances when combined with different land uses. In Tai Po, the pure "GIC" group ranked the second most dangerous among all the land use groups when using the hot zones methodology. But it is not so using the hot spot methodology. Also, it seems not dangerous in other districts. Instead, the combination of GIC land use and commercial land use ("Commercial and GIC") look dangerous in Yau Tsim Mong and Wan Chai, but not in Tai Po. In other words, the impact of GIC land use on road safety also seems to depend largely on local circumstances like building access design. This land use category in Hong Kong also seems to be capturing a wide range of functions that make consistent observations difficult.

3.4. Different methods to generate hot spots and hot zones

In this paper, three methods are used to generate hot spots/hot zones, but we don't know which one is better yet. Thus, our final section compares the results of the three methods by the Method Consistency Test. The results of the comparisons, using both yearly data and the aggregated three-year periods, are shown in Table 7. It is obvious that the stability of the three methods using three-year data is higher than the results using one-year data (Values of MCT being higher and closer to one). This confirms that the aggregation of collision data is a very important step in detecting hazardous road locations regardless of the approach (hot spot versus hot zones) and the method (CC, EX versus EB) used.

Generally, the results of the MCT test using different methods are quite close when three-year data are used. Nonetheless, some differences are discernable. The CC method has the best results in four out of five periods using the one-year data, and it also has the best result using three-year data. This makes the CC method the most reliable among the three methods when the hot spot methodology is used. When the hot zone methodology is used, the EB method has the best result in three out of five periods using the one-year data, and also has the best result using three year data. So the EB method seems to be the most reliable among the three methods when the hot zone methodology is used. The EX method always has worst results in terms of stability.

The advantage of the CC method in the hot spot methodology and the advantage of the EB method in the hot zone methodology may be due to the different purposes of the two methodologies. The hot spot methodology targets specific extremely dangerous locations on the road. Collision count, therefore, is a straightforward indicator, and yields the most consistent results over the years. However, the use of collision count for the purpose also has many other disadvantages (such as not controlling for traffic volume) that we mentioned earlier. The hot zone methodology emphasizes the network continuity of highly dangerous locations. These locations may not be the most dangerous ones with the highest collision counts, but they represent a hidden danger along a continuous stretch of the road network. Hence, the CC method may not be as effective in detecting these dangerous stretches of roads. For the purpose of identifying hot zones, the EB method is superior.

4. Conclusion

In this study, the link-attribute approach is used in detecting hazardous road locations in Hong Kong. It incorporates land use information into the estimation for further engineering or planning improvements. Both hot spot and hot zone methodologies are performed, and they give quite different patterns. The different advantages of the two methodologies result in different choice of the most appropriate estimation method for them. According to the stability test, the most appropriate method for hot spot methodology is using the collision count directly, because it is a straightforward indicator. But the most appropriate method for hot zone methodology is the EB method. Both the results of stability test and the average lengths of hot zones detected by different methods prove this.

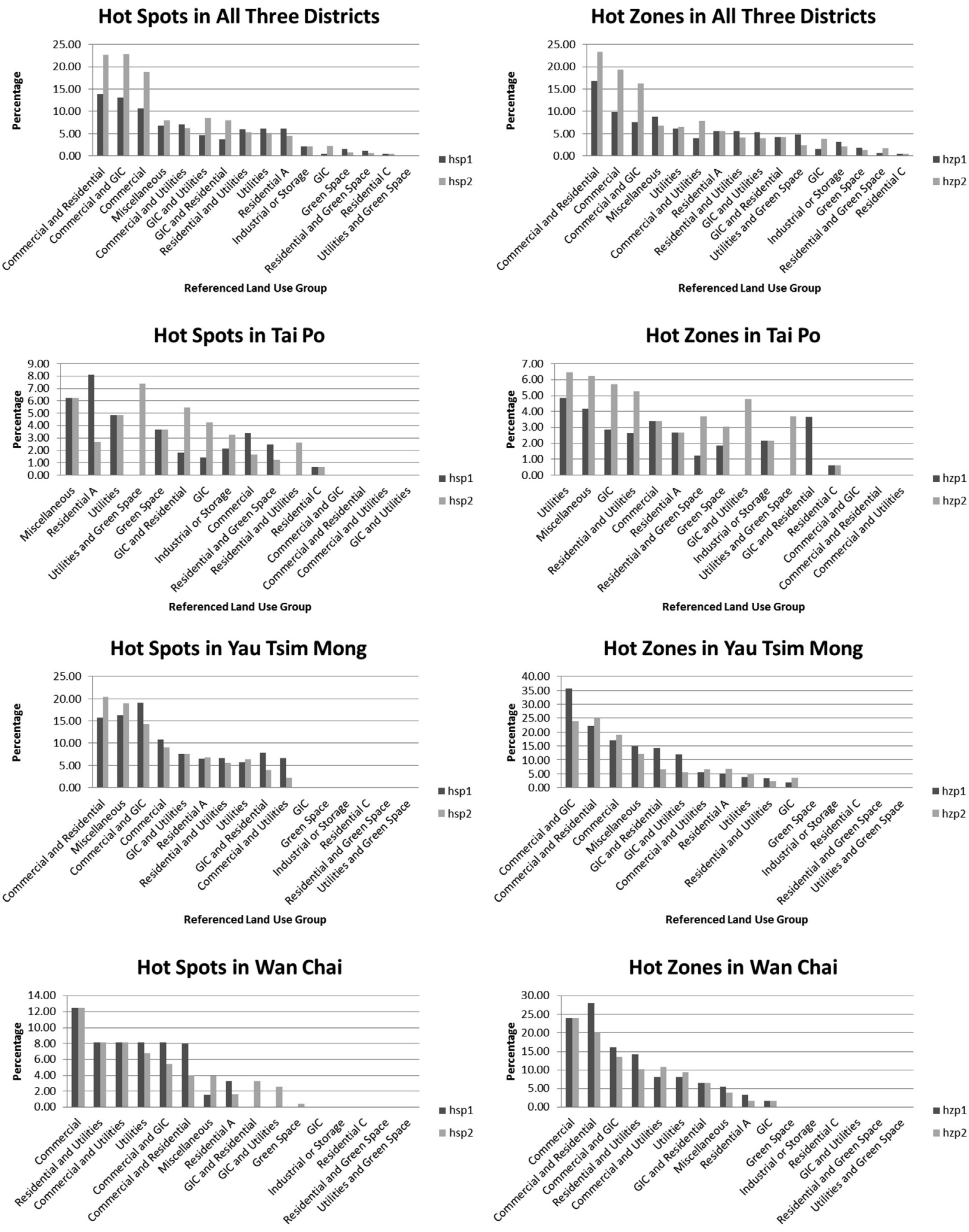


Fig. 4. Percentage of BSUs Detected as Hot Spots/Hot Zones in Each Land Use Group.

Regardless of the choice of methodology, or the method used for estimation, there is clearly an urban-suburban difference in the hazardous road locations in Hong Kong. The percentages of hot spots/hot zones detected are much higher in the two districts in the urban area. The types of roads that have hazardous road locations on them also vary in the three districts. These differences may due to many reasons, one of which is examined in this paper–land use.

The evidences from Hong Kong confirm some of the existing knowledge about the impact of land use on safety. For instance,

Table 7
Comparisons of MCT results.

(a) Hot Spots

	2007–2008	2008–2009	2009–2010	2010–2011	2011–2012	2007–2009 vs 2010–2012
CC	0.525	0.545	0.617	0.615	0.568	0.676
EX	0.494	0.500	0.551	0.585	0.523	0.670
EB	0.528	0.506	0.579	0.601	0.559	0.670

(b) Hot Zones

	2007–2008	2008–2009	2009–2010	2010–2011	2011–2012	2007–2009 vs 2010–2012
CC	0.544	0.578	0.619	0.604	0.596	0.676
EX	0.501	0.538	0.598	0.595	0.573	0.656
EB	0.562	0.569	0.611	0.607	0.601	0.691

commercial land use seems to be more dangerous since all three commercial-related land use groups have higher average crash count and higher percentage of hazardous road location detected. But the results in this study further suggest that the commercial land uses might be related with continuous dangerous road segments rather than individual dangerous locations, since the percentage of BSUs detected as hot zones is always higher than the percentage of the BSUs detected as hot spots.

Evidences also suggest that the safety impact of some land uses needs further research. To name a few, the results suggest that (i) high-density residential is much more dangerous than low-density residential area; (ii) residential land use is not among the most dangerous, unless it is mixed with commercial land use; (iii) the mixture of land use is not necessarily safer or more dangerous, but the different contexts need to be considered; (iv) “Government, Institutional & Community” could be very dangerous if the road segment has other major surrounding land use(s). On the one hand, these findings help us to understand the situation in Hong Kong, or other places like Hong Kong, better. On the other hand, they underline the importance of local contexts and land-use mixes in understanding the impact of land use on road safety.

To conclude, this study makes a contribution in understanding the land use impact in road safety, as well as in some exploration of the methodologies used for hazardous road locations detection. Yet, there are also limitations. More comprehensive land use data with larger study areas and/or more detailed attributes such as intensity will give more interesting and insightful results. But due to the limited manpower and time it cannot be produced manually right now. There may be a relationship between land use type and type of traffic collisions. Nonetheless, when different types of traffic collisions are analyzed in detail, more specific factors like underreporting need to be considered. For instance, Loo and Tsui (2007) found that the traffic collision records of Hong Kong are very accurate for fatal traffic injury but much less so for slight traffic injury involving cyclists and children. More methods for estimation on road segments could be also used or modified. For example, if the estimation is conducted under a cost-effectiveness criterion, the excess collision count method might have better stability. But this requires substantial cost estimation efforts and is not discussed in this paper.

Obviously, land use is only one factor that affects the occurrence of traffic collisions. Other social, policy and engineering factors, such as traffic arrangements, police enforcements, road user characteristics, infrastructure design, etc., may also be important in affecting road safety performance. It is hoped that this study can contribute to the discussion of road safety and make policy makers more sensitive to the link between land use and road safety in design and planning.

Acknowledgement

This research project is funded by the Research Grant Council General Research Fund (GRF) (HKU749210).

References

Cheng, W., Washington, S., 2008. New criteria for evaluating methods of identifying hot spots. J. Transp. Res. Board 2083 (1), 76–85, http://dx.doi.org/10.3141/2083-09.
Cottrill, C.D., Thakuriah, P.V., 2010. Evaluating pedestrian crashes in areas with high low-income or minority populations. Accid. Anal. Prev. 42 (6), 1718–1728.
Dissanayake, D., Aryaija, J., Wedagama, D.M.P., 2009. Modelling the effects of land use and temporal factors on child pedestrian casualties. Accid. Anal. Prev. 41 (5), 1016–1024, http://dx.doi.org/10.1016/j.aap.2009.06.015.
Elias, W., Shiftan, Y., 2014. Analyzing and modeling risk exposure of pedestrian children to involvement in car crashes. Accid. Anal. Prev. 62 (0), 397–405, http://dx.doi.org/10.1016/j.aap.2013.06.035.
Elvik, R., 2008a. The predictive validity of empirical Bayes estimates of road safety. Accid. Anal. Prev. 40 (6), 1964–1969, http://dx.doi.org/10.1016/j.aap.2008.07.007.
Elvik, R., 2008b. A survey of operational definitions of hazardous road locations in some European countries. Accid. Anal. Prev. 40 (6), 1830–1835, http://dx.doi.org/10.1016/j.aap.2008.08.001.
Flahaut, B., 2004. Impact of infrastructure and local environment on road unsafety: logistic modeling with spatial autocorrelation. Accid. Anal. Prev. 36 (6), 1055–1066, http://dx.doi.org/10.1016/j.aap.2003.12.003.
Graham, D.J., Glaister, S., 2003. Spatial variation in road pedestrian casualties: the role of urban scale, density and land-use mix. Urban Stud. 40 (8), 1591–1607, http://dx.doi.org/10.1080/0042098032000094441.
Hauer, E. (1997). Observational before–after studies in road safety: estimating the effect of highway and traffic engineering measures on road safety. Oxford, OX, U.K.; Tarrytown, N.Y., U.S.A.: Pergamon.
Hauer, E., 2001. Overdispersion in modelling accidents on road sections and in empirical Bayes estimation. Accid. Anal. Prev. 33 (6), 799.
Ivan, J.N., Wang, C., Bernardo, N.R., 2000. Explaining two-lane highway crash rates using land use and hourly exposure. Accid. Anal. Prev. 32 (6), 787–795, http://dx.doi.org/10.1016/S0001-4575(99)00132-3.
Kim, K., Yamashita, E., 2002. Motor vehicle crashes and land use: empirical analysis from Hawaii. J. Transp. Res. Board 1784, 73–79, http://dx.doi.org/10.3141/1784-10.
Levine, N., Kim, K.E., Nitz, L.H., 1995a. Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns. Accid. Anal. Prev. 27 (5), 663–674, http://dx.doi.org/10.1016/0001-4575(95)00017-T.
Levine, N., Kim, K.E., Nitz, L.H., 1995b. Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. Accid. Anal. Prev. 27 (5), 675–685, http://dx.doi.org/10.1016/0001-4575(95)00018-U.
Li, W., Carriquiry, A., Pawlovich, M., Welch, T., 2008. The choice of statistical models in road safety countermeasure effectiveness studies in Iowa. Accid. Anal. Prev. 40 (4), 1531–1542, http://dx.doi.org/10.1016/j.aap.2008.03.015.
Loo, B.P.Y., Anderson, T., 2015. Spatial Analysis Methods of Road Traffic. CRC Press (ISBN 9781439874127 – CAT# K13435).
Loo, B.P.Y., Tsui, K.L., 2007. Factors affecting the likelihood of reporting road crashes resulting in medical treatment to the police. Inj. Prev. 13 (3), 186–189, http://dx.doi.org/10.1136/ip.2006.013458.
Loo, B.P.Y., Yao, S., 2013. The identification of traffic crash hot zones under the link-attribute and event-based approaches in a network-constrained environment Computers. Environ. Urban Syst. 41 (0), 249–261, http://dx.doi.org/10.1016/j.compenvurbsys.2013.07.001.
Loo, B.P.Y., 2006. Validating crash locations for quantitative spatial analysis: a GIS-based approach. Accid. Anal. Prev. 38 (5), 879–886, http://dx.doi.org/10.1016/j.aap.2006.02.012.

Loo, B.P.Y., 2009. The identification of hazardous road locations: a comparison of the blacksite and hot zone methodologies in Hong Kong. Int. J. Sustainable Transp. 3 (3), 187–202, http://dx.doi.org/10.1080/15568310801915583.

Loukaitou-Sideris, A., Liggett, R., Sung, H.-G., 2007. Death on the crosswalk: a study of pedestrian-automobile collisions in Los Angeles. J. Plann. Educ. Res. 26 (3), 338–351, http://dx.doi.org/10.1177/0739456x06297008.

Maher, M.J., Mountain, L.J., 1988. The identification of accident blackspots: a comparison of current methods. Accid. Anal. Prev. 20 (2), 143–151, http://dx.doi.org/10.1016/0001-4575(88)90031-0.

Maher, M.J., Mountain, L.J., 2009. The sensitivity of estimates of regression to the mean, Accid. Anal. Prev. 41 (4), 861–868, http://dx.doi.org/10.1016/j.aap.2009.04.020.

Meuleners, L.B., Hendrie, D., Lee, A.H., Legge, M., 2008. Effectiveness of the black spot programs in western Australia. Accid. Anal. Prev. 40 (3), 1211–1216, http://dx.doi.org/10.1016/j.aap.2008.01.011.

Miranda-Moreno, L.F., Labbe, A., Fu, L., 2007. Bayesian multiple testing procedures for hotspot identification. Accid. Anal. Prev. 39 (6), 1192–1201.

Miranda-Moreno, L.F., Morency, P., El-Geneidy, A.M., 2011. The link between built environment, pedestrian activity and pedestrian-vehicle collision occurrence at signalized intersections. Accid. Anal. Prev. 43 (5), 1624–1634, http://dx.doi.org/10.1016/j.aap.2011.02.005.

Montella, A., 2010. A comparative analysis of hotspot identification methods. Accid. Anal. Prev. 42 (2), 571–581, http://dx.doi.org/10.1016/j.aap.2009.09.025.

Ng, K., Hung, W., Wong, W., 2002. An algorithm for assessing the risk of traffic accident. J. Saf. Res. 33 (3), 387–410, http://dx.doi.org/10.1016/S0022-4375(02)00033-6.

Persaud, B., Lyon, C., 2007. Empirical Bayes before-after safety studies: lessons learned from two decades of experience and future directions. Accid. Anal. Prev. 39 (3), 546–555, http://dx.doi.org/10.1016/j.aap.2006.09.009.

Persaud, B., Lan, B., Lyon, C., Bhim, R., 2010. Comparison of empirical Bayes and full bayes approaches for before–after road safety evaluations. Accid. Anal. Prev. 42 (1), 38–43.

Petch, R.O., Henson, R.R., 2000. Child road safety in the urban environment. J. Transp. Geogr. 8 (3), 197–211, http://dx.doi.org/10.1016/S0966-6923(00)00006-5.

Priyantha Wedagama, D.M., Bird, R.N., Metcalfe, A.V., 2006. The influence of urban land-use on non-motorised transport casualties. Accid. Anal. Prev. 38 (6), 1049–1057, http://dx.doi.org/10.1016/j.aap.2006.01.006.

Pulugurtha, S., Repaka, S., 2008. Assessment of models to measure pedestrian activity at signalized intersections. J. Transp. Res. Board 2073, 39–48, http://dx.doi.org/10.3141/2073-05.

Pulugurtha, S.S., Sambhara, V.R., 2011. Pedestrian crash estimation models for signalized intersections. Accid. Anal. Prev. 43 (1), 439–446, http://dx.doi.org/10.1016/j.aap.2010.09.014.

Transport Department of Hong Kong SAR Government, 2016. Terminology, Accessible at http://www.td.gov.hk/filemanager/en/content_4757/terminology_e.pdf. (accessed 13.05.16).

Wang, Y., Kockelman, K.M., 2013. A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. Accid. Anal. Prev. 60 (0), 71–84, http://dx.doi.org/10.1016/j.aap.2013.07.030.

Wier, M., Weintraub, J., Humphreys, E.H., Seto, E., Bhatia, R., 2009. An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. Accid. Anal. Prev. 41 (1), 137–145, http://dx.doi.org/10.1016/j.aap.2008.10.001.

World Health Organization, 2013. Global Status Report on Road Safety 2013: Supporting a Decade of Action. WHO, Geneva, Switzerland.

Yao, S., Loo, B.P.Y., Lam, W.W.Y., 2015. Measures of activity-based pedestrian exposure to the risk of vehicle-pedestrian collisions: space-time path vs. potential path tree methods. Accid. Anal. Prev. 75 (0), 320–332, http://dx.doi.org/10.1016/j.aap.2014.12.005.