Accident Analysis and Prevention 128 (2019) 185–196

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Pattern of road traffic crash hot zones versus probable hot zones in Tunisia: A geospatial analysis

Fedy Ouni[b], Mounir Belloumi[a,b,*]

[a] College of Administrative Sciences, Najran University, BP. 1988 Najran, Saudi Arabia
[b] Faculty of Economic Sciences and Management, University of Sousse, Sahloul 4, BP 526 Sousse, Tunisia

ARTICLE INFO

Keywords:
Traffic crashes
Spatial autocorrelation approach
Prediction accuracy index
Hot zones
Probable hot zones

ABSTRACT

Focusing on how hot zones mapping can predict spatial patterns of crashes and how different mapping approaches compare can help to better inform their application in practice. This study examines the stability of the performance of two spatial autocorrelation measures on the basis of a Road Safety Risk Index (RSRI) through the comparison of the results for three regions (North-West, Center-East, and Center-West) and for three time periods (2002–2005, 2006–2009 and 2010–2013) in Tunisia. Our study differs from others in that it discusses the identification of probable hot zones and enhances the capability to examine a given highway by determining "dangerous probable lengths", which aims to anticipate the traffic crashes in the future. The identified hot zones and probable hot zones exhibit different regional and temporal characteristics. There are clearly some outstanding spatial clusters of crashes covering specific locations. In both Northwest and Center-West regions, the majority of the identified hot zones and probable hot zones predominantly occur along mainly highways characterized by a dominant rural character. In the Center-East region, both hot zones and probable hot zones are mostly spread northeast and south-west more precisely in NH1 and NH2 where many urban activities are taking place. Spatial autocorrelation indices per region address the diversity within the regions and provide us with useful insights that can be translated into safety policies in Tunisia.

1. Introduction

Like many countries, the extent of road safety in Tunisia is considerable. Tunisia owns a few numbers of vehicles compared to other countries. It is recognized as being unlucky regarding crash numbers. According to the National Observatory for Information, Training, Documentation and Studies on Road Safety in Tunisia (National Observatory of observation, training, documentation and study on road safety, Tunisia. NOITDSRS, 2015), the number of crashes in 2015 was 8373, with one death for five crashes and three injuries for two crashes. Road traffic crashes linked to the rise in motor vehicles using highways have changed into a pertain concern of traffic planners (Mohaymany et al., 2013). Investigation of crash hotspots represents an important criterion in road safety management simply due to the fact that any errors in detecting high-risk segments result in identifying the really risky segments as safe or vice versa and, therefore can lead to inadequate allocation of financial resources (Mohaymany et al., 2013).

In recent years, the concerns about the efficient way to locate crash hotspots have been the main topic amongst road safety researchers and transport specialists around the world (Young and Park, 2014). The existing literature has proven that different lengths characterize crash hotspots (e.g. 100 m in Belgium (Flahaut et al., 2003), 250 m in Austria (Elvik, 2007), 100 m in Germany (Elvik, 2008), 100 m in Hong Kong (Lai and Chan, 2004; Loo, 2009), 100 m from segment location and 50 m for intersection in Flanders (Geurts, 2006), 100 m in Norway (Elvik, 2007), 100 m in Hungary (Elvik, 2007), 200 m in Portugal (ANSR,2010), and 300 m to 1000 m in Croatia (Zovak et al., 2014)). Others fundamental parameters in hotspots identification method are the crash number threshold and period (e.g. 3–5 crashes over 5 years in New Zealand (LTNZ, 2006), 3 or more crashes over 3 years in Belgium (Flahaut et al., 2003), 4 crashes over 5 years in Denmark (Vistisen, 2002), 5 injury crashes over the past 3 years or 3 fatal or serious injury crashes over 5 years in Germany (Elvik, 2008), 3 or more serious crashes over 3 years in Hungary (Elvik, 2007), and 4 or more serious crashes over 5 years in Norway (Elvik, 2007)). In Tunisia, according to National Observatory of observation, training, documentation and study on road safety, Tunisia. NOITDSRS (2010), a hotspot is any part of the road having 1000 m of length and recording 10 serious and fatal crashes during 5 years.

The majority of the available successful methodologies of hotspots

* Corresponding author.
E-mail address: mounir.balloumi@gmail.com (M. Belloumi).

https://doi.org/10.1016/j.aap.2019.04.008
Received 6 March 2018; Received in revised form 1 April 2019; Accepted 13 April 2019
Available online 30 April 2019

identification developed in some nations around the world, particularly the advanced methods such as empirical Bayes approach (Lord and Persaud, 2004; Lord and Park, 2008) where safety performance functions have to be formulated, cannot be used in Tunisian context. That is why it is recommended to effectively evaluate and find an appropriate method for identification and prioritization of crashes hotspots via available limited details. In all previous prediction models, just the reported number of crashes in a specific time is used and spatial features are therefore modeled as a constant within a given time (Anderson, 2009). The spatial dependence of crashes is therefore ignored. Spatial analysis methods treat the road network as a continuous entity with an infinite number of spatial units (Thomas, 1996). This study assumes that the crashes that occur in neighboring spatial units are spatially dependent. The majority of previous scientific studies are based on hypothetical or selected highways rather than the entire road network (Flahaut et al., 2003). This paper is the first that addresses a methodological issue of identifying road crashes hot zones and probable hot zones in an entire road network systematically. Focusing on how hot zones mapping can predict spatial patterns of crashes and how different mapping approaches compare will help to better inform their application in practice. The study investigates the methodological issues of identifying road crashes hot zones and probable hot zones in Tunisia. It examines the stability of the performance of two spatial autocorrelation measures based on RSRI through the comparison of the results for three regions (North-West, Center-East, and Center-West) and for three time periods (2002–2005, 2006–2009 and 2010–2013). The RSRI serves as a reliability indicator and permits to assess the hot zone method for both precision and consistency. The reason behind the comparison is that if a location is identified as a hot zone within periods, it is less likely to be a false hot zone. The authors are unaware of any identical study that has conducted a spatial autocorrelation approach with a particular attention to hot zones and probable hot zones. The present study first introduces the probable hot zone terminology in road safety research and enhances the capability to examine a given highway by determining “dangerous probable lengths”, which aims to anticipate the traffic crashes in the near future. This study has been designed as a contribution to the enhancement of road safety situation in Tunisia. To our knowledge, this study reflects one of the first attempts in Tunisia that uses exploratory spatial data analysis techniques for examining road traffic crashes. From a policy viewpoint, this kind of analysis can certainly help public authorities, transportation safety professionals in Tunisia to develop appropriate safety measures that can possibly reduce the number of injuries and fatalities in those critical locations. The adopted methodology is transferable to other countries around the world and other types of events for traffic safety where crashes are spatially and temporarily analyzed in a network space.

The remainder of this work is organized as follows. Section 2 presents a relevant literature review on GIS-based hotspots and hot zones identification. Sections 3 and 4 discuss the data and methodology used, whereas Section 5 presents the results and discussion. Finally, the authors close the paper by a conclusion followed by a brief discussion concerning limitations and perspectives for future research directions in Section 6.

2. Literature review

Areas of the concentrated crashes are often referred to as crash hotspots (Anderson, 2007), sites with promise (Hauer, 1996), hot pieces (Gundogdu, 2010), and road traffic crashes vulnerability area (Benedek et al., 2016). The detection of clustering patterns of road traffic crashes has been facilitated by both the effective employment of GIS into transportation research area and by the opportunity given by Global Positioning System (GPS) in terms of spatial accuracy of localization of road traffic crashes (Bíl et al., 2013). The primary reason behind employing spatial techniques for detection of crash hotspots rather than classical statistical techniques is the fact that crashes are a spatial phenomenon. Classical statistical techniques neglect the geographical relationship between the different locations (Ouni and Belloumi, 2018). Yet, it appears reasonable that the structure of the road network can play a crucial role in determining dangerous locations (Moons et al., 2009). Spatial methods, which are employed for identification of clustering patterns of road traffic crashes, produce two kinds of results. The first one is identifying global clustering tendency of crashes within road section, which includes Quadrat methods (Nicholson, 1999), the Nearest Neighbor methods (Levine et al., 1995; Nunn and Newby, 2015), and K-function (Yamada and Thill, 2004; Ouni and Belloumi, 2018). The second result is identifying local clustering tendency of crashes within the road section. It includes Kernel density estimation approach (KDE) (Flahaut et al., 2003; Anderson, 2009; Loo et al., 2011; Prasannakumar et al., 2011; Mohaymany et al.,2013; Xie and Yan, 2013; Bíl et al., 2013; Yu et al., 2014; Erdogan et al., 2015; Thakali et al., 2015; Harirforoush and Bellalite, 2016; Ulak et al., 2017), and spatial autocorrelation approach including local Moran and local Getis-ord indices (Flahaut, 2002; Flahaut et al., 2003; Steenberghen et al., 2004; Erdogan, 2009; Moons et al., 2009; Gundogdu, 2010; Songchitruksa and Zeng, 2010; Truong and Somenahalli, 2011; Prasannakumar et al., 2011; Blazquez and Celis, 2013; Loo and Yao, 2013; Manepalli and Bham, 2013; Xie and Yan, 2013; Young and Park, 2014; Yu et al., 2014; Erdogan et al., 2015; Choudhary et al., 2015; Scott et al., 2016; Moradi et al.,2016; Aghajani et al., 2017; Soltani and Askari, 2017; Blazquez et al., 2018). There are two types of spatial autocorrelation measures such as global measures and local measures based on whether the methods use the spatial autocorrelation significance test globally or locally within the study area. Flahaut et al. (2003) suggested that spatial autocorrelation approach is a powerful spatial clustering method. According to (Fotheringham et al., 2000), global measures are characterized by an average value of a spatial pattern in a data set through study area and these methods may not exactly characterize the real situation. Therefore, they have a less relevance contribution. Local measures of spatial autocorrelation can disclose specific locations of the statistically significant clusters of points (Ulak et al., 2017). It is necessary that global measures rely on local measures in order to identify the individual contribution of each location. The local measures of spatial autocorrelation have the benefit of considering the geographical component of spatial concentrations of the phenomena, which often provides a local index of danger (Thomas, 1996; Flahaut, 2002). In local measures of spatial autocorrelation, crashes are generally attributed to line features, namely links (Loo and Yao, 2013). Roads are in turn split into smaller segments generally known as basic spatial units (BSUs). In our study, the total number of victims (including the number of people killed or injured) is treated as attribute values of these BSUs. In our case, a hotspot is identified when one BSU has an attribute value greater than the threshold value, whereas a hot zone is identified when more than one individual neighboring hotspots taken together to become a single hot zone. Loo (2009) first compared the hot zone methodology with the black site (hotspot) methodology in Hong Kong. A hot zone includes more than one contiguous road segment with high crash numbers.

It is essential to note that hotspots and hot zones located with statistical significance are the primary concern to find better results. A major weakness of KDE methods is that it calculates the density of crashes in a neighborhood around those crashes without statistical significance (Anderson, 2009). Very few studies have attempted to compare different hotspots and hot zones identification methods within a study area. Erdogan et al. (2015) found that using different hotspot detection methods leads to different results.

Flahaut et al. (2003) applied the local Moran index and the KDE method for identifying crash hot zones in Belgium. About 510 reported crashes between 1992 and 1996 were carried out. The results suggest that both methods lead to quite similar results. Steenberghen et al. (2004) demonstrated the effectiveness of GIS and point pattern techniques for identifying road crash hot zones in Belgium using both KDE and spatial autocorrelation approach. Spatial autocorrelation approach using local Moran index gives the best results for the identification of clusters on rural roads, whereas KDE appears to perform better result in an urban area. Pulugurtha et al. (2007) identified 29 high pedestrian crash zones in Las Vegas metropolitan area by using KDE, sum-of-the-ranks and crash score methods. Results of rankings of high pedestrian crash zones were relatively consistent with little to no variation when the sum-of-the-ranks and the crash score methods were used. Anderson (2007) compared three spatial statistical methods for distinguishing road crash hotspots in London using KDE, network analysis and area-wide analysis. He found that KDE is able to quickly and visually identify hotspots from large data sets and therefore produce a satisfactory result. Pulugurtha and Vanapalli (2008) developed a GIS based methodology by identifying and ranking hazardous bus stops in high auto-pedestrian collision concentration areas using crash rate and crash frequency methods. Their results revealed that using crash rate method carries out better results than crash frequency method as it accounts for pedestrian exposure. Zhang (2010) found that hotspot analysis using local Getis-Ord index carries out better results than central feature and point density function in identifying the hazardous road locations. This result is explained by the fact that point density function offers an overall view of where crashes frequently occur without any statistical significance while hotspot analysis using local Getis-Ord index identifies the locations where crashes frequently happen with statistical significance. Truong and Somenahalli (2011) used the KDE and spatial autocorrelation approach to identify the pedestrian vehicle crash hotspots and unsafe bus stops in Adelaide metropolitan area, Australia. They found that three pedestrian vehicle crash hotspots were identified at intersections while ten pedestrian vehicle crash hotspots were identified at mid-block location. The majority of pedestrian vehicle crash hotspots were located near the intersections of Main North Road. Choudhary et al. (2015) compared local Getis-Ord index and KDE using GIS for hotspot identification between 2009 and 2016 for Varanasi city. By using three conceptualization of spatial relationships, results revealed the superiority of local Getis-Ord index over KDE when used with inverse distance and inverse distance squared. Erdogan et al. (2015) examined crash hotspots for the Turkish highways based on the crash rate, the crash frequency, crash severity ranking, KDE, and local spatial autocorrelation measures. Results clearly showed that using KDE with local spatial autocorrelation methods could identify hotspots more accurately. Recently, researchers have begun to consider measures that can be used to assess the effectiveness of hotspot mapping approaches. Chainey et al. (2008) firstly introduced a predictive accuracy index (PAI) as a measure to compare prediction capabilities of crime mapping methods. Successfully applied in a number of other spatial crimes analysis (Van Patten et al., 2009), it has been recently used in spatial crashes analysis (Thakali et al., 2015; Harirforoush and Bellalite, 2016). Harirforoush and Bellalite (2016) proposed a two-step integrated method for identifying crash hotspots on a roadway network in the city of Sherbrooke. Their findings indicated that based on the PAI measure, the network KDE results based on potential hotspots are better for determining hotspots than the results based on aggregated crash data. Therefore the PAI measure was used for the case study. However to avoid confusion with Chainey's PAI, the measure was renamed the Road Safety Risk Index (RSRI).

In general, despite a large number of studies on the topic of crash hotspots and hot zones identification, only Gundogdu (2010) investigated the occurrence of potential or probable hotspots. To our knowledge, no previous studies discussed the identification of probable hot zones. This paper tries to fill this gap. The probable hot zones methodology is a relatively new direction that derives from the spatial interaction between contiguous probable hotspots.

3. Description of data and the study area

To test the applicability of spatial autocorrelation approach, three sets of data are used: (1) Collisions data obtained from NOITDSRS in Tunisia, (2) Highways data obtained from ministry of equipment, housing and territorial development (MEHTD) in Tunisia, and (3) administrative map obtained from DIVA-GIS.[1] The NOITDSRS is the only national source to offer complete information on collision circumstances, vehicles implicated and resulting casualties. Its reliable single database concerns crashes that can be used for longitudinal research in Tunisia. Data on crashes are usually reported and registered by police and National Guard, and they are centralized at the NOITDSRS in Tunisia. All the data included in this study are stored as shape files and shown as a road traffic-crash map in an ArcGIS 10.2 platform.

3.1. Study area

The study area for this research is Tunisia; a developing country situated in the Eastern part of North Africa. In 2016, Tunisia has an estimated population of about 11 134 588 inhabitants, an area of 162 155 km^2, and a population density being 65 inhabitants/ km^2. The urban population represents 67.7% of the total population (National Statistics Institute, 2014). As shown in Fig. 1, the study area centers on three regions of Tunisia: The North-West region (including the governorates of Kef, Beja, Siliana and Jendouba), the Center-East region (including the governorates of Sousse, Monastir, Mahdia and Sfax) and the Center-West region (including the governorates of Kairouan, Kasserine and Sidi Bouzid). The North-West region is ranked among the most rural regions in the country, with only 41% of the region's inhabitants living in urban areas (National Statistics Institute, 2014). The geographical position of the Center-East region has significantly contributed to making it as an important link of interconnection with its national and international environment by a developed basic infrastructure in terms of road, maritime and air transport. The national road number 1, the Tunis-Msaken freeway, the railway network, the Monastir and Sfax airports, and the Sousse and Sfax commercial ports compose this infrastructure. As an important economic region, the Center-West regionis part of the homogeneous space constituted by the inland regions between the coastal regions to the east and the Tunisian-Algerian border to the west.

3.2. Road network and crash data

The data cover a 12 years period from 2002 until 2013. This work is restricted only to crashes with corporal injuries. The total number of crashes between 2002 and 2013 was 13242, consisting of 23,086 with injuries and 4200 with fatalities. The period under study is long enough to limit random fluctuations (Flahaut et al., 2003). The study is originally applied to all numbered highways.

The collected highway data consist of four different functional classifications of roads provided by MEHTD in Tunisia. The proposed functional hierarchy includes freeways, national highways (NH), regional highways (RH) and local highways (LH) but excludes municipal highways and railways. In the road system, the stone markers named also kilometer-post, are placed every 1 km along each road so that any location of a crash in the system is identified by a route ID (road number) and a measure, which specifies a reference marker on the specified road using the dynamic segmentation method (Yamada and Thill, 2007). The collected crashes' data used in our research include the route name and the stone markers information identical as GIS layer of highways. They are employed for locating the crashes on highway map using a linear referring tool that is available in ArcGIS 10.2. Another reason for choosing these selective highways is that not all of the highways in Tunisia have a kilometer-post. As mentioned above, the analysis of hotspots in Tunisia should be based on 1-kilometer (km) long road segments (National Observatory of observation, training,

[1] http://www.diva-gis.org.

Fig. 1. Study area and road network.

documentation and study on road safety, Tunisia. NOITDSRS, 2010). Roads have been in turn split into 1 km long fixed segments generally known as basic spatial units for depth spatial analysis. In our study, the total number of victims (including the number of people killed or injured) is treated as attribute values of these BSUs. A hotspot is identified when one BSU has an attributed value greater than the threshold value, whereas a hot zone is identified when more than one individual neighboring hotspots are taken together to become a single hot zone. The locations of crashes are mapped with World geodetic system 1984 (WGS84) projection. Many previous studies are based on hypothetical or selected highways rather than the entire road network (Flahaut et al., 2003). This study differs completely from those studies by the use of a large crashes' database, and a detailed geocoded road network consists of approximately 7428 km of numbered road, divided in about 93 km of a freeway, 2279 km of NH, 3309 km of RH and 1747 km of LH. The highway network considered here encompasses 7428 BSUs. Given that the resolution of the data depends upon on reference markers (Yamada and Thill, 2007), BSUs are also used as reference points for which the spatial autocorrelation measures are to be calculated.

4. Methodology

The first step of this study is to analyze the spatial concentration of road crashes by identifying areas of spatial concentration of crashes. This is an exploratory study; no assumption is still asked (Flahaut, 1999). Among the most commonly used statistical methods, the authors find the spatial autocorrelation method, which is based on the hypothesis of what happens in a given geographic location depends on what happens in neighboring locations (Tobler, 1970). Spatial autocorrelation reflects the comparative placement of the emplacements, with reference to each other. Therefore, two nearby places are alike more than two remote ones. In a foremost measure, the authors evaluate global spatial autocorrelation in order to evaluate if the overall study area receives a spatial autocorrelation or not. The Global Moran's index (I_{Moran}) and the General Getis-Ord index ($G_{Getis\text{-}Ord}$) are the most commonly used global statistics for measuring spatial autocorrelation by translating a non-spatial correlation to a spatial context (Erdogan et al., 2015). The null hypothesis, H_0, is generally defined as "no cluster exists". Testing H_0 based on an overall statistical value does not provide a significant assessment for particular locations that could lead to many errors in local cluster detection (Rogerson and Yamada, 2009). The authors consider here both indexes of I_{Moran} and $G_{Gestis\text{-}Ord}$. The I_{Moran} coefficient is given as follows:

$$I_{Moran} = \frac{n}{S_0}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}Z_iZ_j}{\sum_{i=1}^{n} Z_i^2},\ \forall\ i \neq j \quad (1)$$

Where n is the total number of BSUs; the coefficients $w_{i,j}$ represent some weighting that indicates the different relationships of proximity between BSUs i and j; z_i is a deviation of an attribute value for BSU i from its mean; zj is a deviation of an attribute value for BSU j from its mean;

Table 1
Results of I_{Moran} and $G_{Getis\text{-}Ord}$.

Regions		I_{Moran}					$G_{Getis\text{-}Ord}$				
		I_{Moran}	Z-score	p-value	Level of agregation	The null[a] hypothesis of randomness	$G_{Getis\text{-}Ord}$	Z-score	p-value	Level of agregation	The null hypothesis of randomness
North-West	2002-2005	0.05	5.05	0.000	Clustered	Rejected	0,0005	2. 81	0.04	High clustered	rejected
	2006-2009	0.07	5.39	0.000	Clustered	Rejected	0,0007	5.45	0.000	High clustered	rejected
	2010-2013	0.06	5.002	0.000	Clustered	Rejected	0,0006	3.48	0.000	High clustered	rejected
Center-East	2002-2005	0.03	3.50	0.000	Clustered	Rejected	0,0005	1.96	0.04	High clustered	rejected
	2006-2009	0.04	3.16	0.001	Clustered	Rejected	0,001	2.76	0.005	High clustered	rejected
	2010-2013	0.02	1.89	0.05	Clustered	Rejected	0,0007	1.06	0.28	Random	accepted
Center-West	2002-2005	0.01	1.30	0.19	Random	Accepted	0,0007	0.22	0.82	Random	accepted
	2006-2009	0.04	2.82	0.004	Clustered	Rejected	0,0007	1.90	0.05	High clustered	rejected
	2010-2013	0.04	4.84	0.000	Clustered	Rejected	0,0007	3.63	0.000	High clustered	rejected

[a] The null hypothesis of randomness: H_0: There is no clustering of high or low values within the specified distance of location i, test statistic is close to zero; H_a: There is clustering of high or low values within the specified distance of location i. A significant positive value implies a clustering of high values, and a significant negative value indicates a clustering of low values.

and s_0 represents the aggregate of all spatial weights ($s_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij} z_i z_j$).

The value of I_{Moran} ranges from -1 (perfect dispersion) to +1 (perfect correlation). A zero value indicates a completely random spatial pattern. For the statistical hypothesis test, I_{Moran} can be transformed into standardized Z-scores, for which values greater than the positive significance threshold or smaller than the negative significance threshold indicate significant spatial autocorrelation.

The standardized Z-score (Z_{Moran}) for I_{Moran} coefficient is computed as follows:

$$Z_{Moran} = \frac{I_{Moran} - E[I_{Moran}]}{\sqrt{V[I_{Moran}]}} \tag{2}$$

Where $E[I_{Moran}]$ is the expected value of I_{Moran} under the null hypothesis and $V[I_{Moran}]$ is the variance of I_{Moran} under the assumption of normally distributed data.

The $G_{Getis\text{-}Ord}$, developed by Getis and Ord, compares the different properties of the sites studied, insofar as these situations are similar (Getis and Ord, 1992; Ord and Getis, 1995). The $G_{Getis\text{-}Ord}$ is computed as follows:

$$G_{Getis-Ord} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}\, x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j}, \forall i \neq j \tag{3}$$

Where, n is the total number of BSUs indexed by i and j, w_{ij} represent some weighting reflecting different relationships of proximity between feature i and j, x_i and x_j are attribute values for BSUs i and j. The standardized Z-score ($Z_{Getis\text{-}Ord}$) for $G_{Getis\text{-}Ord}$ is computed as follows:

$$Z_{Getis-Ord} = \frac{G_{Getis-Ord} - E[G_{Getis-Ord}]}{\sqrt{V[G_{Getis-Ord}]}} \tag{4}$$

Where $E[G_{Getis-Ord}]$ is the expected value of $G_{Getis\text{-}Ord}$ and $V[G_{Getis-Ord}]$ is the variance $G_{Getis\text{-}Ord}$.

While global spatial autocorrelation is useful to test the global clustering tendency of road crashes, it cannot detect where exactly road crashes aggregate across the study area. Several works have developed local measures of spatial autocorrelation (Getis and Ord, 1992; Anselin, 1995; Ord and Getis, 1995). Local measures of spatial autocorrelation allow accidentologists to locate clusters of high crashes neighborhoods throughout a study area within a user-defined distance. They can be easily incorporated into GIS visualization tools because they provide geo-referenced information at a local level, thus considered as a powerful tool for exploratory spatial data analysis (Unwin, 1996). Only local Moran index (L_{Moran}) and local Getis-Ord index ($L_{Gestis\text{-}Ord}$) are used here. There are two main properties that L_{Moran} and $L_{Gestis\text{-}Ord}$ must fulfill in order to be considered as local measures as defined by Anselin (1995). Firstly, for each observation, both indices give indications of a possible grouping of similar values in its neighborhood. Secondly, the sum of the local index for all observations is proportional to the corresponding global index. The L_{Moran} is given as follows:

$$L_{Moran} = \frac{(n-1)(x_i - \bar{x})\sum_{i=1}^{n} w_{ij}(x_j - \bar{x})}{\sum_{j=1}^{n} w_{ij}(x_j - \bar{x})^2}, \ \forall\ i \neq j \tag{5}$$

Where w_{ij} is a 0–1, x_i is an attribute value for BSU i, x_j is an attribute value for BSU j, $\bar{x}$ is the mean of the corresponding attribute and n equating to the total number of BSUs.

The $L_{Gestis\text{-}Ord}$ applied locally is considered as follows:

$$L_{Getis-Ord} = \frac{\sum_{j=1}^{n} w_{ij}x_j - \bar{X}\sum_{j=1}^{n} w_{ij}}{S\sqrt{\frac{n\sum_{j=1}^{n} w_{ij}^2 - (\sum_{j=1}^{n} w_{ij})^2}{n-1}}}, \ \forall\ i \neq j \tag{6}$$

Where xj is the attribute value of location j, $w_{i,j}$ is a 0–1 contiguity matrix, n is the total number of BSUs, $\bar{X}$ is the mean value of x_j and S is the standard deviation of xj values.

If L_{Moran} is utilized as a measure of safety, the riskiest area would be

Table 2
Statistics on hotspots, hot zones, probable hotspots and probable hot zones.

Regions		Hotspots		Hot zones		Probable hotspots	Probable hot zones
		L_{Moran}	$L_{Getis\text{-}Ord}$	L_{Moran}	$L_{Getis\text{-}Ord}$	$L_{Getis\text{-}Ord}$	$L_{Getis\text{-}Ord}$
North-West region	Total number of BSUs (2002-2005)	18	27	3	7	61	14
	Total length(2002-2005)	18 Km	27 Km	7 Km	22 Km	61 Km	39 Km
	Total causalities (2002-2005) Injuries/Death	170/19	239/37	68/6	188/25	342/43	245/28
	Total number of BSUs (2006-2009)	25	25	9	7	58	15
	Total length(2006-2009)	25 Km	25 Km	24 Km	17 Km	58 Km	45 Km
	Total causalities (2006-2009) Injuries/Death	161/27	165/19	143/20	130/18	299/44	179/37
	Total number of BSUs (2010-2013)	21	36	5	9	53	11
	Total length(2010-2013)	21 Km	36 Km	14 Km	32 Km	53 Km	35 Km
	Total causalities (2010-2013) Injuries/Death	166/21	346/42	132/9	280/40	339/47	201/28
Center-East region	Total number of BSUs (2002-2005)	24	21	5	6	40	7
	Total length(2002-2005)	24 Km	21 Km	14 Km	18Km	40 Km	23Km
	Total causalities (2002-2005) Injuries/Death	183/38	200/39	108/20	182/34	306/60	213/39
	Total number of BSUs (2006-2009)	18	14	4	3	49	10
	Total length(2006-2009)	18 Km	14 Km	13 Km	11 Km	49 Km	37 Km
	Total causalities (2006-2009) Injuries/Death	189/20	170/23	155/17	153/19	264/49	156/26
	Total number of BSUs (2010-2013)	14	50	3	11	58	13
	Total length(2010-2013)	14 Km	50 Km	7 Km	31 Km	58 Km	51 Km
	Total causalities (2010-2013) Injuries/Death	115/33	271/59	53/6	175/44	303/68	161/41
Center-West region	Total number of BSUs (2002-2005)	8	4	3	2	55	10
	Total length(2002-2005)	8 Km	4 Km	6Km	4 Km	55 Km	23 Km
	Total causalities (2002-2005) Injuries/Death	116/7	65/1	61/5	65/1	183/55	81/20
	Total number of BSUs (2006-2009)	12	10	3	3	43	6
	Total length(2006-2009)	12 Km	10 Km	9 Km	9 Km	43 Km	13 Km
	Total causalities (2006-2009) Injuries/Death	127/23	86/20	87/14	82/20	162/40	54/11
	Total number of BSUs (2010-2013)	26	41	3	6	38	7
	Total length(2010-2013)	26 Km	41 Km	9 Km	21 Km	38 Km	18 Km
	Total causalities (2010-2013) Injuries/Death	155/34	338/62	91/9	190/25	180/40	86/21

the area with the largest local Moran value, and the least risky area would be the area with the smallest local Moran value. If $L_{Getis\text{-}Ord}$ is utilized as a measure of safety, the riskiest area would be the area with the largest Z-score value, and the least risky area would be the area with the smallest Z-score value. In any type of analysis of spatial autocorrelation, one of the most basic issues is that of the conceptualization of spatial relationship among the BSUs, or the construction of the spatial weight matrix w_{ij}. The w_{ij} indicates how spatial relationships are defined among BSUs such as inverse distance, fixed distance band, and inverse distance squared. In our case, the authors considered a fixed distance band of 1 km, which means that each BSU is analyzed in the context of neighboring BSUs. The neighboring BSUs within the specified distance band receive a weight of one and neighboring features beyond the distance band are weighted to zero and have no influence on calculations.

5. Results and discussion

I_{Moran} and $G_{Getis\text{-}Ord}$ were computed along with p-value and z–score within three successive periods, 2002–2005; 2006–2009; and 2010 – 2013. Results of I_{Moran} and $G_{Getis\text{-}Ord}$ are summarized in Table 1. Both I_{Moran} and $G_{Getis\text{-}Ord}$ showed that the spatial distribution of crashes was highly clustered. A positive z-score indicates the existence of adjacent BSUs with similar values. High z-scores associated with very small p-values indicate spatial clustering of high values. A very low (negative) z-score associated with very small p-value indicates a spatial aggregation of low values. When more z-scores are high (or low), the aggregation is more intense. A z-score near zero indicates the absence of spatial aggregation. In the Center-East region, the case's distribution of period between 2002 and 2005 produces an I_{Moran} value of 0.03, which is an indication that road crashes patterns are spatially auto-correlated and thus clustered. A calculated z-score of $3.50 \geq 2.58$ at 1% significance level indicates the rejection of the null hypothesis of randomness, which means that there is clustering of high or low values within the specified distance of BSUs. In the Center-West region, the results revealed no significant spatial autocorrelations for a number of crashes between 2002 and 2005 as characterized by I_{Moran}, the corresponding z-score and with a p-value > 0.01. Similarly, spatial clustering pattern of road traffic crashes within three periods in all regions were clearly observed with the exception of crashes occurred in the Center-West region between 2002 and 2005, which shows a random tendency. In Center-West region, the case's distribution of period between 2002 and 2005 produces a $G_{Getis\text{-}Ord}$ value of 0.0007, a calculated z-score of 0.22 and a p-value superior to 0.01 indicating that the null hypothesis of randomness can be accepted.

While global spatial autocorrelation is useful to test the global clustering tendency of road crashes within the study area, it cannot detect where exactly road crashes aggregates. To overcome this, local measures of spatial autocorrelation are used. The L_{Moran} and $L_{Getis\text{-}ord}$ methods are computed along for three regions (Northwest, Center-East, and Center-West) and for three periods (2002–2005, 2006–2009 and 2010–2013). In terms of practical interpretation, it should be mentioned that not all segments and their computed L_{Moran} and $L_{Getis\text{-}Ord}$ are used in our empirical case. In the case of L_{Moran}, the appropriate cases for hotspots detection are those high-high cases. The high-high cases for locations (i) should satisfy two conditions: (1) $(\boldsymbol{x_i} - \bar{\boldsymbol{x}} > 0)$ and (2) $\left(\sum_{j=1, j\neq 1}^{n} (\boldsymbol{x_i} - \bar{\boldsymbol{x}} > 0\right)$ (Moons et al., 2009; Xie and Yan, 2013). The L_{Moran} is computed only for these high-high segments. In the case of $L_{Getis\text{-}Ord.}$, the categorization of hotspots is very straightforward. A positive value of $L_{Getis\text{-}Ord}$ suggests a high-high cluster or hotspot; contrariwise, a negative value of $L_{Getis\text{-}Ord}$ suggests a low-low cluster or cold spot. $L_{Getis\text{-}Ord}$ is computed only for the high-high clusters. In this study, hotspots and hot zones were determined based on the total number of victims (including the number of people killed/injured). Without a doubt, the total number of victims is one of the ranking criterion for hotspots selection for detailed diagnosis. In general, despite a large number of studies on the topic of hotspots identification, only Gundogdu (2010) who investigated the occurrence of potential or probable hotspots, yet a focus on identifying probable hot zones is lacking. According to $L_{Getis\text{-}ord}$, the probability value of a being a hotspot must be greater than the threshold value of 1.645 obtained from a normal distribution at 95% confidence level (Gundogdu, 2010).

Therefore, the threshold value between 1.002 and 1.645 is chosen to highlight probable hotspots (Gundogdu, 2010). Similar to hot zone methodology, more than one individual neighboring spatially related probable hotspots took together to become a single probable hot zone.

Detected hotspots, hot zones, probable hotspots, and probable hot zones are presented in detail for each region in Table 2. Results are presented based on L_{Moran} and $L_{Getis\text{-}Ord}$ along with their corresponding frequencies, total length and total injuries/deaths in order to facilitate a comparison.

According to L_{Moran}, the North-West region recorded the highest number of hotspots with 64 hotspots followed by Center-East region with 56 hotspots. According to $L_{Getis\text{-}Ord}$, the Center-East region recorded the highest number of hotspots with 85 hotspots followed by Northwest and Center-West regions with 82 and 55 hotspots, respectively. The authors can notice that the majority of hotspots occurred in the Center-East region. The authors can explain this by their importance in terms of concentration of population (among the most densely populated regions of Tunisia). In addition, the particular geographical position of the Center-East region has significantly made it an important link of interconnection with its national and international environment. It has a basic infrastructure in terms of road, maritime and air transport (the NH1 National road, the Tunis-Msaken freeway, the railway network, the Monastir and Sfax airports, and the Sousse and Sfax commercial ports) (Ouni and Belloumi, 2018). Thus, it has an increased traffic volume through their principal highways. Generally, the increase in traffic volume and the level of congestion with irregular driving behavior increase the increased frequency of severe crashes. A finding is consistent with previous studies based on the relationship between the frequency of crashes and traffic volume (Anastasopoulos, 2016). L_{Moran} allows identifying 165 hotspots and 38 hot zones with a total length of 165 km and 103 km respectively within the study area. $L_{Getis\text{-}Ord}$ allows identifying 222 hotspots and 54 hot zones with a total length of 222 km and 165 km respectively within the study area. Regarding the retained total length of hotspots, 62.42% and 74.32% of the identified hotspots are considered as hot zones based on L_{Moran} and $L_{Getis\text{-}Ord}$ respectively. According to L_{Moran}, the North-West region retained the highest total length of hot zones with 45 km out of 132 km. Similarly, according to $L_{Getis\text{-}Ord}$, the Northwest region retained the highest total length of hot zones with 71 km out of 205 km. In the Center-East region, the authors have identified 56 hotspots according to L_{Moran} stretching over 1 km each resulting in 487 injuries and 97 fatalities and 85 hotspots according to $L_{Getis\text{-}Ord}$ resulting in 641 injuries and 121 fatalities. In case of hot zones identification, L_{Moran} retains a dangerous total length of 34 km resulting in 316 injuries and 43 fatalities. It corresponds to nine injuries and more than one death per km. $L_{Getis\text{-}Ord}$ retains a dangerous total length of 60 km resulting in 510 injuries and 97 fatalities, which corresponds to more than eight injuries and more than one death per km. In the North-West region, L_{Moran} retains a dangerous total length of 45 km resulting in 343 injuries and 49 fatalities. It corresponds to more than seven injuries and one death per km. $L_{Getis\text{-}Ord}$ retains a dangerous total length of 71 km resulting in 598 injuries and 83 fatalities, which corresponds to more than eight injuries and more than one death per km. The results demonstrate that $L_{Getis\text{-}Ord}$ identified more and longer hotspots and hot zones than L_{Moran}.

In addition, $L_{Getis\text{-}Ord}$ allows identifying 93 probable hot zones with a total length of 284 km resulting in 1376 injuries and 251 fatalities. The North-West region recorded the highest number of probable hot zones followed by Center-East and Center-West regions with 40, 30 and 23 respectively. In addition, the Northwest region recorded the highest total length of probable hot zones with 119 km followed by Center-East region with 111 km.

Once the crash hotspots or hot zones are identified, they are assessed with the RSRI. The RSRI serves as a reliability indicator and permit us to assess the hot zone method for both precision and consistency. The RSRIis a ratio of the proportion of crashes occurring within the identified hotspot or hot zone to the proportion of the whole study area covered by it (Harirforoush and Bellalite, 2016). The RSRI varies from 0 to positive infinity.The authors have made a slight modification in the denominator applying the total length of the road segment rather of using its area. This is acceptable because a linear 1-D feature rather than a 2-D feature (Thakali et al., 2015) represent much better highway network. Both measured $RSRI_{Hotspot}$ and $RSRI_{Hot\ zone}$ are constructed. The both measured RSRI are calculated for each period and each region individually as follows:

$$RSRI_{HotSpot} = \frac{\frac{n}{N}*100}{\frac{m}{M}*100} \tag{7}$$

$$RSRI_{Hot\,Zone} = \frac{\frac{n'}{N}*100}{\frac{m'}{M}*100} \tag{8}$$

Where n is the number of victims in the identified hotspots; n' is the number of victims in the identified hot zones; N is the total number of victims within study area; m is the length of BSUs in the hotspot; m' is the length of BSUs in the hot zones; and M is the total length of BSUs within the study area. A larger RSRIvalue means that a method performs better in locating a high potential of crashes in an area (Harirforoush and Bellalite, 2016). Table 3 presents a comparison between L_{Moran} and $L_{Getis\text{-}Ord}$ in terms of RSRI.

For the convenience of analysis, the authors focus only on $RSRI_{Hot\ zone}$ values. The reason behind the comparison is that if a location is identified as a hot zone within periods, it is less likely to be a false hot zone. In the Center-East region, $RSRI_{Hot\ zone}$ value for L_{Moran} is found to be 7.93 between 2002 and 2005, while a comparative value of 10.41 is found for $L_{Getis\text{-}Ord}$. In the Center-East region, $RSRI_{Hot\ zone}$ value for L_{Moran} is found to be 9.53 between 2010 and 2013, while a comparative value of 7.83 is found for $L_{Getis\text{-}Ord}$. In most cases, the magnitude of the difference between L_{Moran} and $L_{Getis\text{-}Ord}$ is very marginal. Introducing the RSRI lets it practical to make comparisons amongst hot zone mapping techniques per regions in their capability to predict potential crash events. When comparing $RSRI_{Hotspot}$ and $RSRI_{Hot\ zone}$ for the hotspot and hot zone mapping techniques, there is no technique that excelled for both measures. Hence, this study recommended using a combination of L_{Moran} and $L_{Getis\text{-}Ord}$ in both hotspot and hot zone analyses. The combined strength of both L_{Moran} and $L_{Getis\text{-}Ord}$ is expected to result in effective hot zones discovery. Comparable outcomes from L_{Moran} and $L_{Getis\text{-}Ord}$ can certainly provide well-performing interpretations among the clustering patterns. Using only one method will often provide unreliable results. The L_{Moran} and $L_{Getis\text{-}Ord}$ methods for identifying hot zones take each one a slightly different approach to the task. Based on the identified hotspots, hot zone locations are derived. The definition of neighborhood and contiguity governs the structure of the hot zone. Thus, local risk factors vary gradually and continuously between neighboring spatial units (Yu et al., 2014). Fig. 2 delineates the hot zones identified within Center-West, Center-Est and North-West regions for 3 successive periods, 2002–2005, 2006–2009, and 2010 – 2013. This provides a more realistic picture of hot zones distribution.

Closer examination of hot zone maps reveals some outstanding spatial clusters of crashes covering specific locations. Since each method has identified distinct and similar clustering patterns. According to Fig. 2, it is obvious that identified hot zones in different regions exhibited different regional and temporal characteristics, which have to be useful to public authorities and police force in targeting their road safety enhancement measures and to develop a surveillance approach over space and time at the local level. In the Center-East region, the map shows clear hot zones that were mostly spread northeast and south-west of the region especially in the governorate of Sousse and Sfax. Some hot zones were also portrayed in the central part of the region. Several important spatial features are discernible. All these hot zones are found along National Highways and Regional Highways, more precisely in NH1 and NH2, where more urban activities are taking

Table 3
Performance comparisons of L_{Moran} and $L_{Getiss\text{-}Ord}$.

	Number of victim in hotspots	Number of victim in hot zones	Total victims	Length of BSUs in Km (hotspots)	Length of BSUs in Km (hot zone)	Total length of segment in Km	RSRI (Hotspot)	RSRI (Hot zone)
L_{Moran}								
North-West (2002-2005)	189	74	3663	18	7	2839.88	8.15	8.21
North-West (2006-2009)	188	163	2553	25	24	2839.88	8.36	7.55
North-West (2010-2013)	187	141	3559	21	14	2839.88	7.10	8.05
Center-East (2002-2005)	221	128	3520	24	14	3053.14	7.98	7.93
Center-East (2006-2009)	209	172	2642	18	13	3053.14	13.42	15.31
Center-East (2010-2013)	148	59	2754	14	7	3053.14	11.73	9.35
Center-West (2002-2005)	123	66	2905	8	6	1538.2	8.14	5.82
Center- West (2006-2009)	150	101	2376	12	9	1538.2	8.09	7.26
Center- West (2010-2013)	189	100	3314	26	9	1538.2	3.37	5.15
$L_{Getiss\text{-}Ord}$								
North-West (2002-2005)	276	213	3663	27	22	2839.88	7.93	7.51
North-West (2006-2009)	184	148	2553	25	17	2839.88	8.18	9.69
North-West (2010-2013)	388	320	3559	36	32	2839.88	8.60	7.98
Center-East (2002-2005)	239	216	3520	21	18	3053.14	9.88	10.41
Center-East (2006-2009)	193	172	2642	14	11	3053.14	15.94	18.08
Center-East (2010-2013)	330	219	2754	50	31	3053.14	7.34	7.83
Center-West (2002-2005)	66	66	2905	4	4	1538.2	8.73	8.73
Center- West (2006-2009)	106	102	2376	10	9	1538.2	6.86	7.33
Center- West (2010-2013)	400	215	3314	41	21	1538.2	4.52	4.76

place. These findings were consistent with those of Soltani and Askari (2017), which reported that the majority of crash hotspots were located in main urban arteries. According to the hierarchy of road network, absolutely no hot zones are found in local Highways as compared to National and Regional Highways. This finding likely reflects that the smaller and local roads tend to be less dangerous (Steenberghen et al., 2010). The NH1 has a path nearly parallel to the motorway (A1). It is known to carry heavy traffic volume that other regional and local highways are connected to and links all cities along the coast. In North-West region, Fig. 2 shows that hot zones are identified everywhere in the region, mainly in the center and northwestern part of the region, which correspond to the governorates of Kef and Jendouba, respectively. The majority of these hot zones predominantly occur along major highways characterized by a dominant rural character more precisely in NH6 and NH7. The North-West region is ranked among the most rural regions in the country, with only 41% of the region's inhabitants living in urban areas (National Statistics Institute, 2014). This is not surprising, given that the majority of the identified hot zones are more at rural areas often did not have sidewalks, separate footpaths for pedestrians, an absence of public lighting, an absence of road signs and speed bump. In such condition, the traffic rule violations by drivers can pose a higher risk to the community. Other explanations are probably linked to the effect of higher speed limits, aggressive driving behaviors, longer emergency response time and lack of medical facilities in rural areas. In Center-West region, most of the hot zones are found in the northeastern and central part of the region, especially in the governorates of Kairouan and Sidi Bouzid and only one hot zone is observed in the center-western part. The highest numbers of hot zones are observed at the central part of the region particularly in road junction between NH14 and LH 889, which link many of the residential areas to this highway.

Fig. 3 delineates the probable hot zones. $L_{Getis\text{-}ord}$ provides a statistically robust and consistent method of detecting probable hotspot and probable hot zones within the study area.

A comparison between hot zones maps and probable hot zones maps, both similarities and discrepancies are observed. In Center-East region, clear probable hot zones that are mostly spread northeast and south-west of the region especially in the governorates of Sousse and Sfax. There is only one probable hot zone identified in the region of Monastir, which is a sparsely populated governorate. The discovered probable hot zones in this region could be due to the development of major commercial activities in northeastern part especially in NH1 in the region of Sousse. Advertising posters and restaurants limit the visibility and legibility of the road with the existence of left and right turns of the road. It is known that such activities attract heavy traffic. Additionally, notable probable hot zones are observed at the south-western part of the region particularly in highly congested road junction between NH1 and RH 124, which link many of the residential areas to this Highway. In North-West region, it is obvious that the probable hot zones are not randomly distributed along North-West highways. The occurrences appear to be regular at segments located in rural major arterial roads where there are curved roads, U-turns and poor pavement conditions. This suggests that a priority for traffic safety enhancement should be put along the rural hot zones. In the Center-West region, notable probable hot zones are identified in the central part of the region, especially in the governorate of Sidi Bouzid. All these probable hot zones are found along national highways and local highways characterized by a dominant rural character, more precisely in NH3, NH13, and LH887. The NH 13 is a very prominent road that is connected to other regional and local highways and it is linked to many of the land-locked areas in the Center-West region. Some probable hot zones are also portrayed in the eastern part of the region especially in NH2 in the governorate of Kairouan. Not surprisingly, the highlighted governorate overlaps the more densely populated area in Center-West region.

Despite the multitudes of research initiatives on the topic of hot

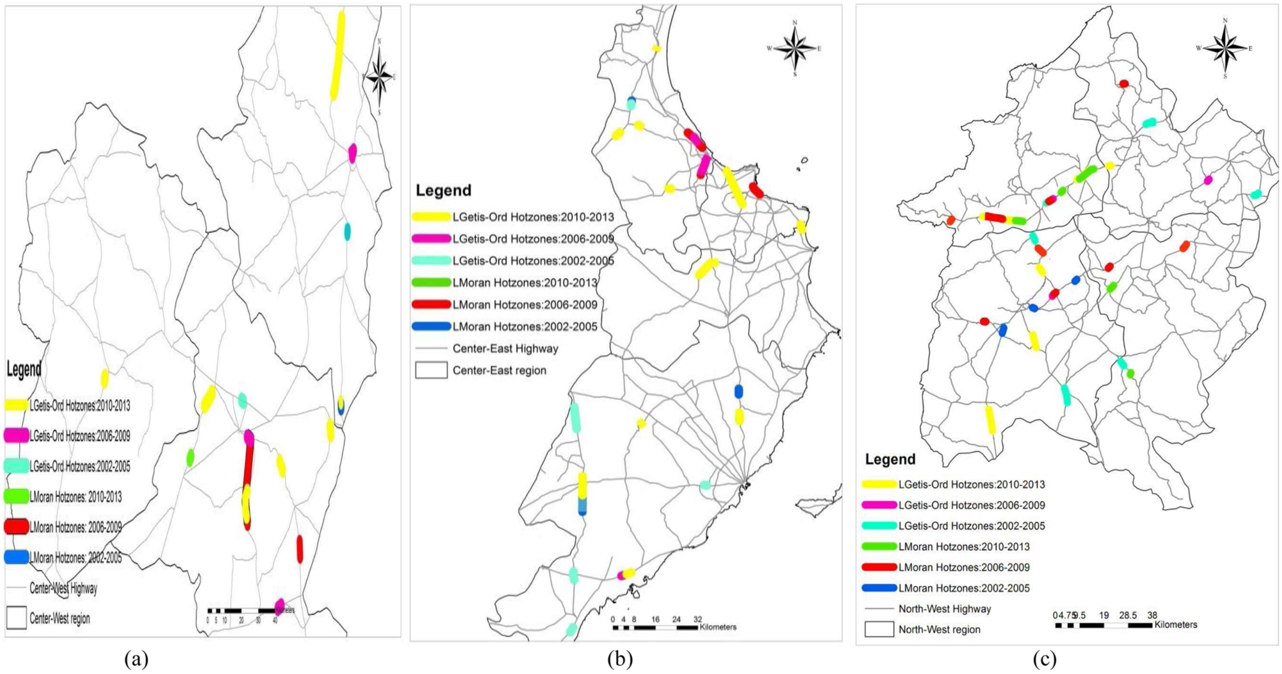

Fig. 2. Crashes hot zones: (a) Center-West region, (b) Center-Est region, (c) North-West region.

zones identification, yet still surprising, the absence of researches dealing with the length of hot zones and probable hot zones. The present research contributes to filling that gap. Table 4 portrays the variability of the length of the identified hot zones and probable hot zones within the study area. In this table, the length of identified hot Zones (L_{HZ}) and the length of identified probable hot zones (L_{PHZ}) are grouped into four lengths: 2 km, 3 km, 4 km, and > 4 km. As shown in Table 4, the length of identified hot zones is not fixed. As previously mentioned, L_{Moran} allows identifying 38 hot zones with a total length of 103 Km. Among the 38 hot zones, 22, 8 and 8 stretching over 2, 3 and 4 km, respectively. $L_{Getis\text{-}Ord}$ allows identifying 54 hot zones of which 22, 21, 4 and 7 stretching over 2, 3, 4 and > 4 km, respectively. The shares of L_{HZ} = 2 km were 57.89% in case of L_{Moran} and 40.7% in the case of $L_{Getis\text{-}Ord}$. The shares of L_{HZ} = 3 km were declined to 21.05% and 38.88% in case of L_{Moran} and $L_{Getis\text{-}Ord}$, respectively. $L_{Getis\text{-}Ord}$ allows identifying 93 probable hot zones with a total length of 284 km of which 41, 33, 14 and 5 stretching over 2, 3, 4 and > 4 km, respectively.

The share of L_{PHZ} = 2 km are 44.08% against 35.48% and 15.05% for L_{PHZ} = 3 km and 4 km respectively. The Northwest region records 40 probable hot zones with a total length of 119 km of which 14, 16, 8

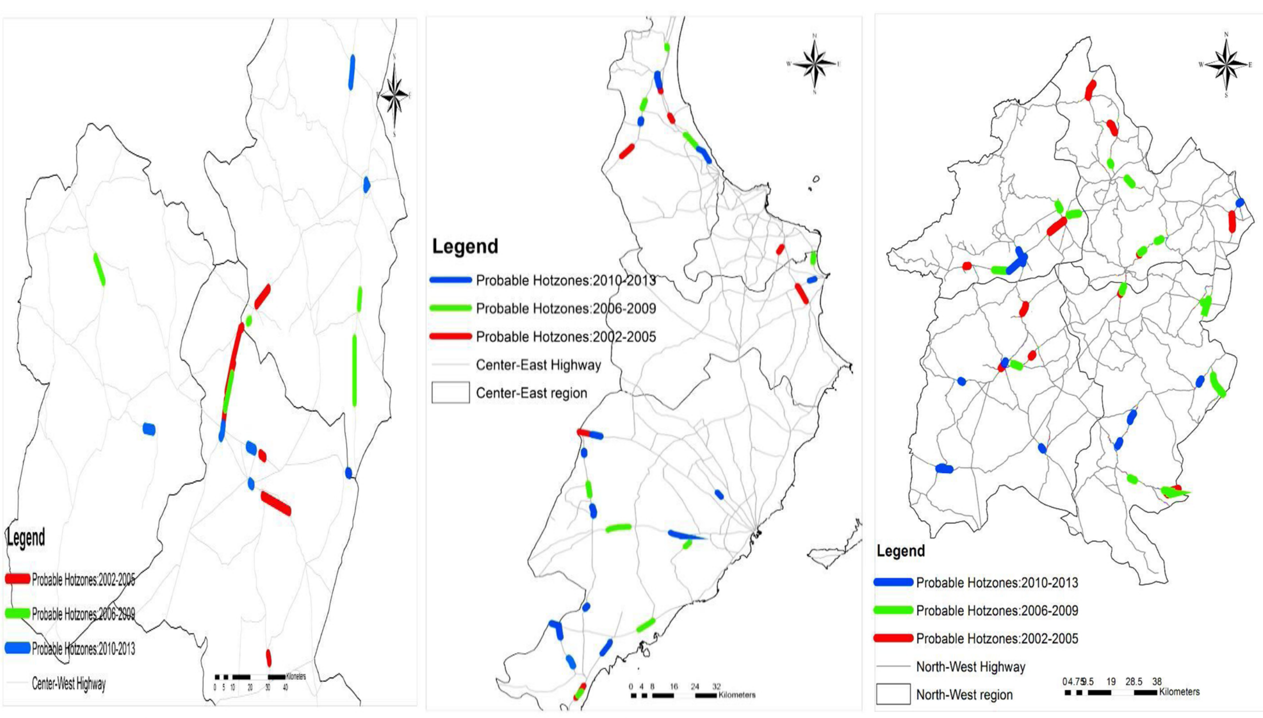

Fig. 3. Crashes probable hot zones: (a) Center-West region, (b) Center-Est region, (c) North-West region.

Table 4
Characteristics of the length of identified hot zones and probable hot zones.

Regions	Length of identified hot zones (L_{HZ})								Length of identified probable hot zones (L_{PHZ})			
	L_{Moran}				$L_{Getis\text{-}Ord}$				$L_{Getis\text{-}Ord}$			
	L_{HZ} = 2 km	L_{HZ} = 3Km	L_{HZ} = 4 km	L_{HZ} > 4 Km	L_{HZ} = 2 km	L_{HZ} = 3Km	L_{HZ} = 4 km	L_{HZ} > 4 Km	L_{PHZ} = 2 km	L_{PHZ} = 3Km	L_{PHZ} = 4 km	L_{PHZ} > 4 Km
North-West (2002-2005)	2/3	1/3	0/3	0/3	1/7	5/7	0/7	1/7	3/11	5/11	2/11	1/11
North-West(2006-2009)	7/9	1/9	1/9	0/9	4/7	3/7	0/7	0/7	5/14	7/14	2/14	0/14
North-West (2010-2013)	3/5	0/5	2/5	0/5	5/9	1/9	1/9	2/9	6/15	4/15	4/15	1/15
Center-East (2002-2005)	2/5	2/5	1/5	0/5	2/6	2/6	2/6	0/6	2/7	2/7	2/7	1/7
Center-East (2006-2009)	1/4	1/4	2/4	0/4	1/3	1/3	0/3	1/3	4/10	3/10	2/10	1/10
Center-East (2010-2013)	2/3	1/3	0/3	0/3	5/11	4/11	1/11	1/11	6/13	4/13	2/13	1/13
Center-West (2002-2005)	3/3	0/3	0/3	0/3	2/2	0/2	0/2	0/2	7/10	3/10	0/10	0/10
Center- West (2006-2009)	1/3	1/3	1/3	0/3	0/3	3/3	0/3	0/3	5/6	1/6	0/6	0/6
Center- West (2010-2013)	1/3	1/3	1/3	0/3	2/6	2/6	0/6	2/6	3/7	4/7	0/7	0/7

and 2 stretching over 2, 3, 4 and > 4 km, respectively. The Center-West region records 23 probable hot zones with a total length of 54 km of which 15 and 8 stretching over 2 and 3 km respectively.

A fascinating characteristic of spatial autocorrelation approach is the variability of the length of the identified hot zones and probable hot zones, which is not the case for other geospatial approaches (Flahaut et al., 2003). The spatial autocorrelation approach makes it possible a more suitable adaptation to the local spatial structure for a given highway by providing two kinds of benefits: (1) the dangerousness of a hot zone, and (2) its length (Flahaut et al., 2003). This strategy allows us to understand the dangerous nature of Tunisian highways more clearly and appropriately. This is particularly relevant in the context of Tunisia where each studied region has unique characteristics in population density, built environment, demographics, culture, etc. In other words, each region requiring a model and insights specific to its characteristics. Spatial autocorrelation indices per region address the diversity within the regions and provide us with useful insights that can be translated into safety policies in Tunisia.

6. Conclusions and future research directions

Focusing on how hotspot mapping can predict spatial patterns of crashes and how different mapping approaches compare will help to better inform their application. This study investigates the methodological issues of identifying road crashes hot zones and probable hot zones in Tunisia. In order to avoid false negative locations, authors use at least two spatial autocorrelation measures to identify clusters that will be subsequently used to target public health interventions at the local level to ensure that the hazardous road section is as accurately defined as possible. Global measures of spatial autocorrelation using I_{Moran} and $G_{Getis\text{-}Ord}$ to assess whether crash pattern has a tendency to cluster in space and in time are employed. Results indicate that the spatial pattern is clustered as there is strong spatial dependence across the study area. While global measures cannot detect where exactly road crashes aggregate across the study area, local measures of spatial autocorrelation are employed based on L_{Moran} and $L_{Getis\text{-}ord}$. To assess the performance of hot zone analysis techniques, RSRI is considered to make comparisons amongst L_{Moran} and $L_{Getis\text{-}ord}$ in their capability to predict potential crash events. Hence, this study highly recommends using a combination of L_{Moran} and $L_{Getis\text{-}ord}$ in hot zone analysis. The results demonstrate that $L_{Getis\text{-}Ord}$ identifies more and longer hotspots and hot zones than L_{Moran}. The originality of this study is that it discusses the identification of probable hot zones and hence enhances the capability to examine a given highway by determining "dangerous probable lengths". This leads to anticipating the traffic crashes in the future.

After visual assessment of both hot zones and probable hot zones maps, it is noteworthy that some probable hot zones appear repeatedly in the same location for different periods. This indicates that these locations will be soon as hot zones and not just temporary probable hot zones that may have been appeared by chance. In both North West and Center-West regions, the majority of the identified hot zones and probable hot zones predominantly occur along major's highways characterized by a dominant rural character. In the Center-East region, both hot zones and probable hot zones are mostly spread northeast and south-west more precisely in NH1 and NH2 where urban activities are taking place. This can be explained by pedestrian crossing anyhow, anarchic businesses located throughout these areas, schools built without taking into account the safety of children helping to create an urban dynamic that has resulted in an increased displacement of pedestrians and daily traffic volume at this level, increasing thus the risk of crashes. Furthermore, the design of these old roads does not comply with actual safety requirements.

From a policy viewpoint, a clearer understanding of the spatial pattern of crashes could guide decision makers in implementing effective measures to improve road safety by adopting proper land use practices, building efficient transport systems, and formulating appropriate traffic policies and laws. Tunisia's public authorities could use these findings to evaluate if some land-use practices help to reduce the overall regional traffic demand (Ouni and Belloumi, 2018). Generally, there are apparently some considerable differences in both hot zones and probable hot zones pattern in rural areas versus urban areas. This is not surprising, given that the majority of crashes are more numerous at rural areas because they do not have sidewalks, separate footpaths for pedestrians, an absence of public lighting, an absence of road signs and speed bump. Appropriate actions are needed such as development and implementation of more targeted design of pedestrian facilities, targeted road safety campaigns and improvement of communication with regard to traffic priorities. This research suggests also correcting an engineering defect that was judged responsible for a high frequency of crashes at a specific location including a redesign of major highways, adopting reduction speed limits in identified hot zones and area-wide traffic calming.

The crash data used in this study are subject to several limitations. Firstly, the data are restricted to all numbered highways in Tunisia excluding local streets due to data limitation. Secondly, the road number and the distance to a stone marker determine the location of crashes on the digital network and crashes occurred at intersections have no exceptions. The case study assigned them to the nearest stone marker from the related intersections, which might misrepresent both L_{Moran} and $L_{Getis\text{-}Ord}$ results in the neighborhood of intersections. Further locations reference including the use of GPS can potentially enhance the current location approaches (Bíl et al., 2013). Thirdly, under-reporting of data about crashes is an issue in Tunisia. Implementing an effective and sustainable road-accident information system should be a top priority in Tunisia (Ouni and Belloumi, 2018).

This study can be extended more deeply in many directions. Firstly, an investigation is required to recognize the factors that might affect crashes in hot zones, such as roadway features or average daily traffic volume. Secondly, more attention should be given to specific crashes such as vulnerable road users crashes, vehicles crashes, and heavy trucks crash. Finally, the effectiveness of spatial autocorrelation approach can be bench marked with the results from other hot zones identification approach in Tunisian context.

References

Aghajani, M.A., Dezfoulian, R.S., Arjroody, A.R., Rezaei, M., 2017. Applying GIS to identify the spatial and temporal patterns of road accidents using spatial statistics (case study: Ilam Province, Iran). Transp. Res. Procedia 25, 2126–2138.

Anastasopoulos, P.C., 2016. Random parameters multivariate tobit and zero-inflated count data models: addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis. Anal. Methods Accid. Res. 11, 17–32.

Anderson, T., 2007. Comparison of spatial methods for measuring road accident 'hot-spots': a case study of London. J. Maps 3 (1), 55–63.

Anderson, T., 2009. Kernel density estimation and K-means clustering to profile road accident hot spots. Accid. Anal. Prev. 41 (3), 359–364.

Anselin, L., 1995. Local indicators of spatial association-LISA. Geogr. Anal. 27 (2), 93–115 1995.

ANSR, 2010. Sinistralidade rodoviária. Autoridade Nationale de Segurança Rodoviária.

Benedek, J., Ciobanu, S.M., Man, T.C., 2016. Hot spots and social background of urban traffic crashes: a case study in Cluj-Napoca (Romania). Accid. Anal. Prev. 87, 117–126.

Bíl, M., Andrášik, R., Janoška, Z., 2013. Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. Accid. Anal. Prev. 55, 265–273.

Blazquez, C., Celis, M., 2013. A spatial and temporal analysis of child pedestrian crashes in Santiago. Accid. Anal. Prev. 50, 304–311.

Blazquez, C.A., Picarte, B., Calderón, J.F., Losada, F., 2018. Spatial autocorrelation analysis of cargo trucks on highway crashes in Chile. Accid. Anal. Prev. 120, 195–210.

Chainey, S., Tompson, L., Uhlig, S., 2008. The utility of hotspot mapping for predicting spatial patterns of crime. Secur. J. 21 (1-2), 4–28.

Choudhary, J., Ohri, A., Kumar, B., 2015. Spatial and statistical analysis of road accidents hot spots using GIS. Third Conference of Transportation Research Group of India. pp. 1–12.

Elvik, R., 2007. State-of-the-art Approaches to Road Accident Black Spot Management and Safety Analysis of Road Networks. Transport økonomisk institutt.

Elvik, R., 2008. A survey of operational definitions of hazardous road locations in some European countries. Accid. Anal. Prev. 40, 1830–1835.

Erdogan, S., 2009. Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. J. Safety Res. 40 (5), 341–351.

Erdogan, S., Ilçi, V., Soysal, O.M., Kormaz, A., 2015. A model suggestion for the determination of the traffic accident hot spots on the Turkish highway road network: a pilot study. Boletim de Ciências Geodésicas 21 (1), 169–188.

Flahaut, B., 1999. Concentration spatiale des accidents de la route : méthode d'identification des zones noires basée sur l'autocorrélation spatiale : application et étude de sensibilité. Eur. J. Geogr. 4 èmes Rencontres de Théo Quant, Besançon, France 11-12 février 1999.

Flahaut, B., 2002. Identifier les zones noires d'un réseau routier par l'autocorrélation spatiale locale. Analyses de sensibilité et aspects opérationnels. Revue internationale de Géomatique 12 (2), 245–261.

Flahaut, B., Mouchart, M., San Martin, E., Thomas, I., 2003. The local spatial autocorrelation and the kernel method for identifying black zones: a comparative approach. Accid. Anal. Prev. 35 (6), 991–1004.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2000. Quantitative Geography: Perspectives on Spatial Data Analysis. Sage 2000.

Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. 24 (3), 189–206 1992.

Geurts, K., 2006. Ranking and Profiling Dangerous Accident Locations Using Data Mining and Statistical Techniques. Doctoral Dissertation. Faculty of Applied Economics, Hasselt University.

Gundogdu, I.B., 2010. Applying linear analysis methods to GIS-supported procedures for preventing traffic accidents: case study of Konya. Saf. Sci. 48, 763–769.

Harirforoush, H., Bellalite, L., 2016. A new integrated GIS-based analysis to detect hot-spots: a case study of the city of Sherbrook. Accid. Anal. Prev.

Hauer, E., 1996. Identification of sites with promise. Transp. Res. Record: J. Transp. Res. Board (1542), 54–60.

Lai, P., Chan, W., 2004. GIS for road accident analysis in Hong Kong. Int. Assoc. Chin. Prof. Geogr. Inf. Sci. 10 (1), 58–67.

Levine, N., Kim, K.E., Nitz, L.H., 1995. Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns. Accid. Anal. Prev. 27 (5), 663–674.

Loo, B.P., 2009. The identification of hazardous road locations: a comparison of the blacksite and hot zone methodologies in Hong Kong. Int. J. Sustain. Transp. 3 (3), 187–202.

Loo, B.P., Yao, S., 2013. The identification of traffic crash hot zones under the link-attribute and event-based approaches in a network-constrained environment. Comput. Environ. Urban Syst. 41, 249–261.

Loo, B.P., Yao, S., Wu, J., 2011. Spatial point analysis of road crashes in Shanghai: a GIS-based network kernel density method. Geoinformatics, 2011 19th International Conference on. pp. 1–6.

Lord, D., Park, P.Y.J., 2008. Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on empirical Bayes estimates. Accid. Anal. Prev. 40 (4), 1441–1457.

Lord, D., Persaud, B.N., 2004. Estimating the safety performance of urban road transportation networks. Accid. Anal. Prev. 36 (4), 609–620.

LTNZ, 2006. Road Safety Issues: Christchurch City Land Transport, New Zealand. Christchurch, NZ. .

Manepalli, U., Bham, G., 2013. Identification of crash-contributing factors: effects of spatial autocorrelation and sample data size. Transp. Res. Rec.: J. Transp. Res. Board 2386, 179–188.

Mohaymany, A.S., Shahri, M., Mirbagheri, B., 2013. GIS-based method for detecting high-crash-risk road segments using network kernel density estimation. Geo-Spatial Inf. Sci. 16 (2), 113–119.

Moons, E., Brijs, T., Wets, G., 2009. Identifying hazardous road locations: hot spots versus hot zones. Trans. Comput. Sci. 6, 288–300.

Moradi, A., Soori, H., Kavousi, A., Eshghabadi, F., Jamshidi, E., Zeini, S., 2016. Spatial analysis to identify high-risk areas for traffic crashes resulting in death of pedestrians in Tehran. Med. J. Islam. Repub. Iran 30, 450.

National Observatory of observation, training, documentation and study on road safety, Tunisia. NOITDSRS, 2010. Repartition Des Accidents Routiers Par Gouvernorats. Rapport Annuel, 2010.

National Observatory of observation, training, documentation and study on road safety, Tunisia. NOITDSRS, 2015. Repartition des accidents routiers par gouvernorats. Rapport annuel, 2015.

National Statistics Institute, 2014. General Census of Population. Tunisia. http://www.ins.nat.tn.

Nicholson, A., 1999. Analysis of spatial distributions of accidents. Saf. Sci. 31 (1), 71–91.

Nunn, S., Newby, W., 2015. Landscapes of risk: the geography of fatal traffic collisions in Indiana, 2003 to 2011. Prof. Geogr. 67 (2), 269–281.

Ord, J.K., Getis, A., 1995. Local spatial autocorrelation statistics. Distributional issues and applications. Geogr. Anal. 27 (1995), 286–306.

Ouni, F., Belloumi, M., 2018. Spatio-temporal pattern of vulnerable road user's collisions hot spots and related risk factors for injury severity in Tunisia. Transp. Res. Part F Traffic Psychol. Behav. 56, 477–495.

Prasannakumar, V., Vijith, H., Charutha, R., Geetha, N., 2011. Spatio- Temporal clustering of roads accidents: GIS based analysis and assessment. Procedia Soc. Behav. Sci. 21, 317–325.

Pulugurtha, S.S., Vanapalli, V.K., 2008. Hazardous bus stops identification: an illustration using GIS. J. Public Trans. 11 (2), 4.

Pulugurtha, S.S., Krishnakumar, V.K., Nambisan, S.S., 2007. New methods to identify and rank high pedestrian crash zones: an illustration. Accid. Anal. Prev. 39 (4), 800–811.

Rogerson, P., Yamada, I., 2009. Statical Detection and Surveillance of Geographycal Clusters.

Scott, M., Sen Roy, S., Prasad, S., 2016. Spatial patterns of off-the-system traffic crashes in

Miami–Dade County, Florida, during 2005–2010. Traffic Inj. Prev. 17 (7), 729–735.
Soltani, A., Askari, S., 2017. Exploring spatial autocorrelation of traffic crashes based on severity. Injury 48 (3), 637–647.
Songchitruksa, P., Zeng, X., 2010. Getis-Ord spatial statistics to identify hot spots by using incident management data. Transp. Res. Rec.: J. Transp. Res. Board (2165), 42–51.
Steenberghen, T., Dufays, T., Thomas, I., Flahaut, B., 2004. Intra-urban location and clustering of road accidents using GIS: a Belgian example. Int. J. Geogr. Inf. Sci. 18 (2), 169–181.
Steenberghen, T., Aerts, K., Thomas, I., 2010. Spatial clustering of events on a network. J. Transp. Geogr. 18 (3), 411–418.
Thakali, L., Kwon, T.J., Fu, L., 2015. Identification of crash hotspots using kernel density estimation and kriging methods: a comparison. J. Mod. Transp. 23 (2), 93–106.
Thomas, I., 1996. Spatial data aggregation: exploratory analysis of road accidents. Accid. Anal. Prev. 28, 251–264.
Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. Econ. Geogr. 46 (sup1), 234–240.
Truong, L.T., Somenahalli, S.V.C., 2011. Using GIS to identify pedestrian-vehicle crash hot spots and unsafe bus stops. J. Public Trans. 14 (1).
Ulak, M.B., Ozguven, E.E., Spainhour, L., Vanli, O.A., 2017. Spatial investigation of aging-involved crashes: a GIS-based case study in Northwest Florida. J. Transp. Geogr. 58, 71–91.
Unwin, D.J., 1996. GIS, spatial analysis and spatial statistics. Prog. Hum. Geogr. 20 (4), 540–551.
Van Patten, I.T., McKeldin-Coner, J., Cox, D., 2009. A micro spatial analysis of robbery: prospective hot spotting in a small city. Crime Mapp.: J. Res. Pract. 1 (1), 7–32.
Vistisen, D., 2002. Models and Methods for Hot Spot Safety Work. PhD Dissertation. Department for Informatics and Mathematical Models, Technical University of Denmark, Lyngby.
Xie, Z., 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.
Yamada, I., Thill, J.C., 2004. Comparison of planar and network K-functions in traffic accident analysis. J. Transp. Geogr. 12 (2), 149–158.
Yamada, I., Thill, J.C., 2007. Local indicators of network-constrained clusters in spatial point patterns. Geogr. Anal. 39 (3), 268–292.
Young, J., Park, P.Y., 2014. Hot zone identification with GIS-based post-network screening analysis. J. Transp. Geogr. 34, 106–120.
Yu, H., Liu, P., Chen, J., Wang, H., 2014. Comparative analysis of the spatial analysis methods for hotspot identification. Accid. Anal. Prev. 66, 80–88.
Zhang, Y., 2010. Hotspot Analysis of Highway Accident Spatial Pattern Based on Network Spatial Weights. A&M University, Texas.
Zovak, G., Brčić, D., Šarić, Ž., 2014. Analysis of Road Black spots identification method in Republic of Croatia. IX International Conference Road Safety in Local Communities.