Detection and Classification of Road Accident Black Zones Using Exploratory Spatial Data Techniques

Harizi Riadh, Ouni Fedy and Mriahi Rafaa,
GEF2A laboratory, Higher Institute of Management of Tunis, Tunis University, Tunisia

Abstract: The road accidents are realized in dangerous locations which qualified as “black spots” or “black zones”. Actually, their technical and analysis methods identification become an important issues for the road safety research (Exploratory Spatial Data Analysis). In this paper, we identify the spatial concentration places of the road accidents “black zones” using the Spatial Data Analysis Method (Moran’s I and Getis-Ord ). Our results show that the black zones number detected is higher than the claimed one’s by public authorities (20 black zones against 9 black spots). Furthermore, this method allows a real classification of these areas. Thus, the black zones are classed by their danger degree. We show that black zones length detected by our methods is less important than the one’s detected by public authority of road safety (2000 meters against 7000 meters). This implies that it is possible to reduce the insecurity cost by 71%. Moreover, our proposed classification gives a priority order in terms of the public expenditure required to secure infrastructure.

Keywords: Road safety; Road accidents; Black zones; Spatial autocorrelation, Moran’s I; Getis-Ord.

I. INTRODUCTION

The development of worldwide road traffic is the direct outcome of the continuous increase of the global vehicle fleet. Indeed, the global vehicle fleet was raised from 1,031,284,909 vehicles in 2007 to 1,789,251,397 vehicles in 2011 (Comité des Constructeurs Français d’Automobiles, 2012). Alongside the development of motorization and traffic, road accidents, have increased considerably. For these reasons the current research focuses on two aspects: (i) solving technical problems related to car models and (ii) maintenance of road infrastructure failures by adapting routes to modern security measures. These generate heavy investment and a need to detect hazardous road locations to fit the smooth flow of traffic. The identification of accidents black spots has grown considerably according to the techniques of spatial econometrics and Geographic Information System (GIS). In practice, many techniques have been developed to detect these high-risk areas (Gundogdu, 2010; Elvick, 2008; Gundogdu et al., 2008; Steenberghen et al., 2004; Hauer, 1996; Fotheringham et al., 1993). Researchers have used various techniques, but most of them are complementary: the coefficient of risk, probabilistic distributions, the score of priority, the accident rate per space-time, the severity index (Vanhauf et al., 2006).

In recent years, researchers have introduced the local indicator of spatial autocorrelation into their analyses. These new techniques have the advantage of considering the geographical aspect of spatial concentrations of the phenomena, which can provide a local index of dangerousness (Flahaut et al., 2002, Hauer (1996); Thomas (1996); Syllock et Smith 1985.).

This article aims to identify and classify the accident black spots. First, we present the most frequent definitions of a black spot or a black area of road traffic. Then, we identify the black areas on a road section and compare the results with those reported by official agencies. Our method is based on exploratory spatial data analysis (ESDA). Finally, we show the relevance of this method since it gives more details to black areas, to classify them in order of dangerousness and reduce investments of insecurity.

II. PRACTICES AND FORMULATION OF THE BLACK SPOT CONCEPT

Black spots do not have a universal definition. These are sites of dangerous areas with high rate of accidents (Geurts et al., 2005; Elvick, 2007). Theoretical and empirical definitions of black spots or black zones can be categorized into six definitions.

A. Definition Based on a risk coefficient

In Austria the black points are defined as being any location which satisfies two criteria: (1) Three accidents or more accidents are recorded along a period of three years, (2) the coefficient of risk should not exceed 0.8 (Austrian Code of Roads Planning, Building and Maintenance, 2002). The value of the coefficient of risk is calculated according to expression (1).

\[ R_x = U/0.5 \times (7 \times 10^{-5} \times AADT) \]  

Where,

- \( R_x \) is a critical value of the risk,
- \( AADT \) is an average daily traffic measured by number of vehicles per 24 hours, and
- \( U \) is a number of bodily accidents in three years.

B. Definition based on a probability distribution

The identification of hazardous road sections in Denmark is based on a more detailed classification; it is made according to different types of road networks and different types of intersections (Vistisen, 2002). For
National Highway (NH), a distinction is made between sections of roads, roundabouts and other intersections. So to identify a dangerous road site, a test based on Poisson distribution is used. A site is considered as dangerous if five bodily accidents along a period of five years are recorded.

**C. Definition based on a priority score or on a severity index**

We can define a location of dangerous road using a priority score calculated using the formula (2).

$$S = L_i + (3 \times S_i) + (5 \times D_i)$$

(2)

Where,

- $L_i$ is a total number of slight injuries,
- $S_i$ is total number of serious injuries (victims who are hospitalized more than 24 hours), and
- $D_i$ is a total number of fatal accidents.

This definition is retained in Flanders. The score is calculated for each location recording three or more accidents in three successive years (with bodily injury) and the black spot is retained when the score is superior or equal to 15 (Geurts et al., 2005).

The definition used in Flanders takes into account the severity of accidents, giving important weight to fatal accidents at the expense of accidents procreating serious and slight injuries. For example, the priority score for a location where there were two fatal accidents and an accident causing serious injuries is 13. For instance; In Portugal, the severity index (3) is based on the number of deaths, serious injuries and that of the slight injuries (The National Laboratory of Civil Engineering, 1997).

$$I_G = (100 \times N_d) + (10 \times N_{bg}) + N_{bl}$$

(3)

Where,

- $N_d$ is a number of deaths,
- $N_{bg}$ is a number of serious injuries, and
- $N_{bl}$ is a number of slight injuries.

In this case, a black spot is a section of road with a minimum length of 200 meters, which record five accidents which $I_G$ is superior to 20.

**D. Definition based on the number of accidents per space-time**

In other countries like Germany, Hungary and Norway, an area is classified as dangerous if the number of serious accidents is realized during a given period in a specific location. In Germany, we consider at least five accidents over a range of 1 to 3 years on a stretch of road, Hungary (4 accidents in 3 years) and Norway (4 or more accidents per 100 meters for 5 years).

In Belgium, according to police, black spots are road segments with a length of 100 meters (1 hectometer) which count three accidents with bodily lesions during one year (Flahaut, 2001).

**E. The crash frequency (CF)**

The road segments and dangerous road intersections are classified in a regressive order according to the number of accidents along a given period. Observations with a higher order than the fixed threshold will be classified as high-risk areas (Ghulam, 2009.). This method can be applied as follows:

- Determination of the average frequency of accidents by calculating the average annual number of accidents for the three previous years.
- Classification sites by using several criteria (presence or lack of signs, number of lanes, average daily traffic, type of area, etc...)
- Determination of the critical frequency of accidents.

Comparison between the frequency of accidents and the critical frequency in all selected locations: If the critical frequency is superior or equal to the normal frequency of accidents, the segment will be considered hazardous. The calculation of $CF$ is as follows:

**F. The crash rate ($C_r$)**

This method takes into account traffic volume $v$ and the number of accidents occurring during a period ($a$). The crash rate ($C_r$) on a segment, recorded in 100 million vehicles / km, is (Powers et Carson, 2004) (5).

$$CR = \frac{a}{v}$$

(5)

The crash rate is multiplied by a million and divided by 365 days; it is summarized as follows (6).

$$C_r = \frac{C \times 10^6}{(365 \times T \times v)}$$

(6)

Where,

- $C$ is the number of reported accidents,
- $T$ is a period of time (number of days) for which accidents are counted, and
- $v$ is an average daily traffic volume (vehicles per day).

The $C_r$ has the advantage of being able to compare areas with similar characteristics but on roads with higher levels of exposure to different risks (Wang, 1989).

**III. DATA, METHODS AND ANALYSIS**

**A. Description of the study area**
The city of Sousse in Tunisia connects the northern and southern parts of the country (Sfax and Gabes), and Sahel in the center (Kairouan and Sidi Bouzid). Its proximity to the A1 motorway crossing the west of the city, and the railway line turns it into a thoroughfare (North-South Mobility). The presence of Monastir airport within 15 km and its commercial port strengthens its role as a communication hub in the entire country.

Figure 1: Road network of the region of Sousse (in yellow, NH-1)

The huge demand of mobility is increasing traffic on the NH-1 which crosses the city of Sousse. The NH-1 has a path nearly parallel to the motorway (A1), it serves all cities along the coast (Fig. 1).

B. Statistical analysis

The NH-1 is the most deadly infrastructure in the region of Sousse. According to the Tunisian National Guard, (TNG, 2010), it recorded the highest number of accidents in the region of Sousse with 55 out of 178 accidents, 80 out of 295 injuries and 22 out of 57 dead (Table 1). These accidents are recorded in hot spots identified by the National Observatory for Information, Training, Documentation and Studies on Road Safety (NOITDSRS). According to the latter institution, a hot spot means any part of the road having a length of 1000 meters recording 10 serious and fatal accidents during a period of 5 years. Thus, in Tunisia the result is as follows (Fig. 2): the hot spots are detected mainly in the regions of Nabeul, Sfax, Sousse, Medenine and Tunis.

Table 1: Number of accidents by road type in Sousse during 2009

<table>
<thead>
<tr>
<th>NUMBER AND TYPE OF ROADS</th>
<th>ACCIDENTS NUMBER</th>
<th>INJURIES</th>
<th>DEATHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH-1</td>
<td>55</td>
<td>80</td>
<td>22</td>
</tr>
<tr>
<td>NH-2</td>
<td>24</td>
<td>36</td>
<td>12</td>
</tr>
<tr>
<td>NH-12</td>
<td>14</td>
<td>24</td>
<td>3</td>
</tr>
<tr>
<td>TOTAL 1</td>
<td>93</td>
<td>140</td>
<td>37</td>
</tr>
</tbody>
</table>

Figure 2: Number of accidents on the road network in Tunisia 2010

[Source: The NOITDSRS (2010) and representation of the authors]

C. Methodology

The first step is to study the spatial concentration of road accidents: identifying areas of spatial concentration of road accidents, or hot spots (hazardous areas). This is an exploratory study; no hypothesis is still posed. Among the most used statistical methods, we find the spatial autocorrelation method which is based on the following hypothesis: what happens in a given geographic area depends on what happens in neighboring places (Tobler, 1970). The spatial autocorrelation takes into account the relative position of places, in relation to each other. Thus, two nearby places are more similar than two distant ones. This method is based on indices measuring global and/or local spatial autocorrelation. More recent works have developed local measures of spatial autocorrelation (Getis et Ord, 1992, Ord et Getis, 1995; Anselin, 1995).
1. Global Autocorrelation

This is the global study of the existence of a close relationship between the frequency of accidents and the geographic space. Used measures decompose the overall index in order to identify the individual contribution of each location. For each these, they measure successively the spatial “dependence-association” between value $X_j$ taken by the same variable $X$ in places characterized by a spatial proximity (Griffith, 1992). The variable $X$ is the number of accidents per road segment. In a first step, we measure the global spatial autocorrelation throughout the study area (NH-1), allowing to test and assess if the overall study area has a spatial autocorrelation or not. The most common measures are:

The global Moran’s index ($I_m$) and the General Getis-Ord index ($G$).

1- The ($I_m$) coefficient measuring simultaneously the Spatial autocorrelation according to the location of the areas and their assigned values and assessing whether the model is clustered, dispersed, or random (Upton and Fingleton, 1985).

$$I_m = (n/s) \times \left[ \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2} \right]$$

(7)

Where,

- $w_{ij}$ represent some weightings reflecting different relationships of proximity,
- $s_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$,
- $z_i = X_i - \bar{X}$,
- $z_j = X_j - \bar{X}$,
- $X_i$ is the value of variable $X$ in place $i$,
- $X_j$ is the value of variable $X$ in place $j$,
- $\bar{X}$ is an average value of $X_i$,
- $n$ is a number of places, and

$(i, j)$ represent different localities.

2- The $G$ measures the degree of aggregation of high or low values i.e. it measures the concentration of high or low values for a particular study area and it checks if the model is clustered, dispersed, or random.

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} x_i x_j}{\sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j}$$

(8)

$w_{ij}$ represent some weightings reflecting different relationships of proximity.

- $X_i$ is a value of variable $X$ in the place $i$,
- $X_j$ is a value of variable $X$ in the place $j$, and

$(i, j)$ represent different localities.

The general Getis-Ord index ($G$) allows, in particular, locating the place in space where a phenomenon occurs in order to study the spatial dimension and location of several phenomena at once. This measure compares the different attributes of the sites studied, insofar as these places are comparable.

2. Local Autocorrelation

Several works have allowed developing local measures of spatial autocorrelation (Getis et Ord, 1992, Ord et Getis, 1995; Anselin, 1995). Each spatial unit $i$ (100 meters in our case) is characterized by a value of local spatial autocorrelation, LISA (Local Indicator of Space Association). LISA is a decomposition of ($I_m$) from measures identifying the individual contribution of each place of the study area. In other terms, LISA is an indicator of the extent to which the value of an observation is similar to or different from its neighboring observations. LISA is defined as follows (9):

$$I_i = z_i w_{ij} z_j$$

(9)

The Local Getis-Ord index ($G_i$) can statistically identify significant spatial clusters of high values (black areas or dangerous sections in our case) and low values (non-hazardous sections). It creates a new output feature class with a Z-score and p-value for each road hectometer.

$G_i$ is defined as follows (10 and 11):

$$G_i = \frac{(\sum_{j=1}^{n} w_{ij} x_j - \bar{X} \sum_{j=1}^{n} w_{ij})}{s} \sqrt{\frac{n \sum_{j=1}^{n} w_{ij}^2 - (\sum_{j=1}^{n} w_{ij})^2}{n-1}}$$

(10)

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

(11)

$$s = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n}}$$

(12)

$w_{ij}$ represent some weightings reflecting different relationships of proximity.

- $X_i$ is a value of variable $X$ in the place $j$, and
\[(i, j)\] represent different localities.

We will use these measures in “two-way roads” on the NH-1 to give more details on the black spots already detected by the TNG, and we will prove that it is possible to restrict black areas from 1000 meters to 100 meters, which allows to reduce and minimize the costs of insecurity.

\section*{IV. RESULTS}

\subsection*{A. Previous results and insecurity costs}

In Tunisia, corporal road accident is identified by the TNG and the Tunisian Civil Security (TCS). These institutions process information and then publish the black spots to the public. Based on these data, national and regional authorities finally intervene to put right the infrastructure limitations (roads, signs…). To secure the NH-1, the Ministry in charge spent an average of 89,000 TND: 103,500 TND in 2007, 73,500 TND in 2008 and 90,500 TND in 2009. (General Office of Infrastructure - Ministry of Equipment, 2011).

\subsection*{B. Study results}

The selected network is spread over 65 km of the NH-1. 320 accidents recorded between 2006 and 2010 on this section. TNG and TCS identify the location of an accident by the nearest Kilometer point. We will consider the distance of 100 meters as the unit of spatial reference for the length of road segments studied where accidents’ data are spatially available.

To measure the spatial autocorrelation, we used the Euclidean distance that corresponds to the straight line distance between two geographical points. When the scale of analysis decreases, the correlation increases monotonically and the variance decreases (Thomas, 1996). Local measures of spatial autocorrelation indicate the individual contribution of each point in the global autocorrelation measures. For this reason, we chose to study small sections of 100 meters in length. Instead, the Tunisian authorities retain longer distances (1000m), which subsequently induces massive investments which are spread over the 1000 meters reported. Our aim is to reduce these distances to better detect the black zones and subsequently reduce relative insecurity costs.

Results show that there is a tendency of clustering: positive value of Global Moran’s close to 1 (Fig. 3) and critical value of Getis-Ord equal to 1.36, (Fig. 4). The black zone is identified only if the product of two values of Z score is positive.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{fig3.png}
\caption{Result of the global Moran index (Global autocorrelation)}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{fig4.png}
\caption{Result of general index of Getis-Ord (Global autocorrelation)}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{fig5.png}
\caption{Normal distribution of p-value and Z score}
\end{figure}

Thus, a black area is a combination of road sections presenting a high number of accidents. Therefore, only the positive local measures resulting
from the product of two positive values of Z score are retained as evidence of danger to the identification of black zones and evaluation of the intensity of their hazardous nature (table 2).

Table 2: Dangerosity Classification of Blacks Zones in NH-1 by Tow Methods

<table>
<thead>
<tr>
<th>L INDEX</th>
<th>Z-SCORE</th>
<th>P-VALUE</th>
<th>L INDEX</th>
<th>Z-SCORE</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>128.700 - 128.800</td>
<td>22,326,099</td>
<td>3,728,303</td>
<td>0.00001</td>
<td>128.400 - 128.500</td>
<td>2,717,307</td>
</tr>
<tr>
<td>115.900 - 116.000</td>
<td>23,387,699</td>
<td>3,728,303</td>
<td>0.00007</td>
<td>128.200 - 128.300</td>
<td>2,717,307</td>
</tr>
<tr>
<td>118.200 - 118.300</td>
<td>15,651,599</td>
<td>3,728,303</td>
<td>0.00059</td>
<td>128.00 - 128.100</td>
<td>2,717,307</td>
</tr>
<tr>
<td>117.900 - 118.100</td>
<td>116.900</td>
<td>3,728,303</td>
<td>0.00016</td>
<td>128.100 - 128.200</td>
<td>2,717,307</td>
</tr>
<tr>
<td>118.200 - 118.300</td>
<td>14,563,199</td>
<td>3,728,303</td>
<td>0.00097</td>
<td>128.300 - 128.400</td>
<td>2,717,307</td>
</tr>
<tr>
<td>117.600 - 118.000</td>
<td>990,132</td>
<td>3,728,303</td>
<td>0.01338</td>
<td>128.400 - 128.500</td>
<td>2,717,307</td>
</tr>
<tr>
<td>150.600 - 150.700</td>
<td>591,001</td>
<td>3,728,303</td>
<td>0.017907</td>
<td>128.500 - 128.600</td>
<td>2,717,307</td>
</tr>
<tr>
<td>150.800 - 150.900</td>
<td>5,509,409</td>
<td>3,728,303</td>
<td>0.04994</td>
<td>128.600 - 128.700</td>
<td>2,717,307</td>
</tr>
</tbody>
</table>

Figures 6 and 7 allow exploring the spatial autocorrelation. They show what the local index of Moran and Getis-Ord have a positive correlation with the average number of accidents per hectometer and the relationship between these two variables is not linear. It is rather exponential. The calculation of danger index based on the spatial autocorrelation method according to Getis-Ord seems to be very effective for the identification of accidents’ black areas. For calculating local Moran index and Getis-Ord index, the number of neighbors taken into account is in relation to spatial structure observed around each hectometer. Thus, there is a great correspondence between the results obtained by applying the two methods. Knowing that the hectometer is the smallest spatial unit in which accidents are localized, it is to aggregate a neighboring hectometer in geographical areas in order to identify hazardous and non-hazardous areas. An association with high value of accidents between neighboring hectometers will be considered as forming a black area.
As mentioned above, according to the “National Observatory of Road Safety” a black point is any part of road having a length of 1 km and recording 10 serious and fatal accidents during a period of 5 years. On The 63 km studied, the TNG has identified 7 black spots stretching over 1000 meters each, while our method allowed identifying 20 black zones of 100 meters in length each (table 3).

Theoretically, among the components of the cost of insecurity we find the following solutions: improvement of lighting, creation of refuges for the pedestrians in the middle of the road, installation of signals and roundabouts, enlargement of the angle of deviation, etc. These solutions are necessary to make the road network safer. Our results, compared to those published by the Tunisian public institutions, show that:

- The public services identify 9 black spots on two sections of the NH-1: 7 black spots on the section 115.00 - 128.00 and 2 black spots on the section 149.00 - 159.00.
- They reported 2 black spots 9,500 km far from the last black zones detected by our approach.
- They have retained a dangerous total length of 7000 meters while our method retains only 2000 meters. Therefore, our solution provides a fairly high number of black zones (20 zones against 9) on a smaller geographical area (2000 m of length against 7000 m), which represents an important opportunity to lower the costs of insecurity (or future investment to secure these sections) of 71.42% and will thus be more limited in space.

$$100\% - \left(\frac{2000}{7000} \times 100\right) = 100\% - 28.57\% = 71.42\%$$

National institutions declared 9 black spots without specifying their level of danger. We have provided a classification for these areas according to danger degrees (Figures 8 and 9). Indeed, both figures 8 and 9 show that the most dangerous zones are located between milepost 115.00 - 118.00 and 128 - 130.00 and the danger is at its extreme when the vehicle arrives at the milepost 11,128.00.

**CONCLUSIONS**

The literature review around theoretical and empirical definition of black spots or black zones of the road allowed us to present six different definitions (based on a risk coefficient, probability distribution, priority score, the number of accidents by space-time, severity score and the frequency of accidents). These definitions differ from one country to another, but all have one thing in common: the repetition over time and geographical space of a serious event (the accident), so we talk about black zones of road traffic accidents rather than black spots.

We used the spatial autocorrelation method to detect these areas on the NH-1 in the region of Sousse in Tunisia (the local Moran index and local Getis-Ord index). Our results, compared with the results of Civil Security and the Tunisian National Guards, seem to be more satisfactory for these reasons:

- The detected zones are now spread over 100 meters rather than 1,000 meters, the number of black zones detected (20 zones) is higher than the number claimed by the Tunisian authorities (9 zones). In addition, the detected zones are much more specific on space according to their danger degree.
- The total zones detected is shorter (2000 meters against 7000 meters) while the number of detected areas twice as high as the one claimed by the foregoing bodies, which implies an outstanding reduction of investments to improve road infrastructure (cost reduction of insecurity nearly 71%). interventions would therefore be more accurate and relevant.
Thus, the spatial autocorrelation method appears to be an adequate method of identifying hazardous road areas. This would probably encourage public authorities to use the most sophisticated techniques to solve this problem.

Finally, road accidents could have been avoided if management actions, prevention, awareness in road safety and control were carried out in priority to these areas. We hope the information and the details mentioned in this article could help public decision makers in road safety especially with the considerable increase of the fleet in Tunisia.

References


