BIVARIATE MORAN’S I AND LISA TO EXPLORE THE CRASH RISKY LOCATIONS IN URBAN AREAS

Ali Akbar Matkan¹, Matin Shahri², Mojgan Mirzaie³

1 Associate professor of remote sensing & GIS center, faculty of earth sciences, Shahid Beheshti University, Tehran, Iran
a-matkan@sbu.ac.ir

2 MSc of remote sensing & GIS center, faculty of Geomatics, Arak University of Technology, Arak, Iran
matinshahri@gmail.com

3 MSc of remote sensing & GIS center, faculty of earth sciences, Shahid Beheshti University, Tehran, Iran
m_mirzaie87@yahoo.com

ABSTRACT:
Reducing the number and severity of urban crashes has always been a primary concern of urban planners and safety specialist. Identifying the critical locations is an essential step in urban management since the safety improvements are prioritized to hazardous regions. The spatial essence of crash data, particularly the presence of spatial autocorrelation reveals that crash occurrences are not only inclined to cluster in the same locations but within particular time intervals. This study aims to detect spatial-temporal dependencies among crash occurrences using bivariate Moran’s I and LISA (Local Indicator of Spatial Association). Employing the yearly number of crashes aggregated in 253 TAZs (Traffic Analysis Zone) in Mashhad, Iran over four successive years (2006 – 2009), indicated that both bivariate Moran’s I and LISA yielding significant patterns of spatial-temporal autocorrelation of crashes. The results of this analysis help the safety officials and urban planners to sufficiently allocate the limited resources such as budget and time by prioritizing the riskiest locations.

KEY WORDS
Spatial-temporal autocorrelation; bivariate Moran’s I; LISA; Urban planning
INTRODUCTION

The intra-city crashes linking with the increasing number of urban transportation means have always been a certain concern of traffic engineers and safety specialists. Analysis of geographic data such as traffic crashes demands more attention because they display different characteristics than aspatial data. The most important issue is the presence of spatial influence among neighboring locations over space or time which complicates the related analyses (Yamada & Rogerson, 2003). Geographic Information System (GIS) as a powerful instrument to manage, manipulate and display the spatial data, has recently attracted the attention of specialists in different fields of transportation studies; particularly safety sciences. Developing the spatial crash prediction models (Aguero-Valverde & Jovanis, 2006; Eksler & Lassarre, 2008; El-Basyouny Karim & Sayed, 2009; Levine, Kim, & Nitz, 1995; Miaou, 2003; Quddus, 2008) and investigating for the clustering of crash data in certain locations (spatial) and certain time (temporal) (Eckley & Curtin, 2012; Erdogan, Yilmaz, Baybura, & Gullu, 2008; Li, Zhu, & Sui, 2007; Plug, Xia, & Caulfield, 2011; Wang & Abdel-Aty-Mohamed, 2006) are the good example of the case. Detecting the spatial and temporal crash patterns over time and space plays a crucial role in analyzing the clustering of crash data since any errors in detecting the clusters might lead to identifying the relatively dangerous regions as safe or inversely the relatively safe regions as dangerous and consequently insufficient assignment of budgets and resources (Cheng & Washington, 2005; Montella, 2010).

Several studies can be referred in which the crash behaviors were examined by developing micro-level models at intersections, segments or corridors (Anastasopoulos, Tarko, & Mannering, 2008; Cunto & Saccomanno, 2009; de Smith, Goodchild, & Longley, 2007). Data availability in micro-level explanatory analyses is almost restricted and it leads the researchers to develop macro-level models in which the data are aggregated over areal units e.g. Traffic Analysis Zones (TAZ) (de Guevara, Washington, & Oh, 2004; Lovegrove & Sayed, 2006), census blocks (Wier, Weintraub, Humphrey, Seto, & Bhatia, 2009), county (Noland & Oh, 2004) and ward (Noland & Quddus, 2004).

Spatial autocorrelation measures the level to which the extent of the value of a variable at a certain location relates to the same value in neighboring locations. When the level of interaction exceeds the expected level, the nearby locations have similar values and the autocorrelation is said to be positive. Conversely when the interaction is negative, the high values of variables are proximate with the low values and the spatial autocorrelation is negative. If data are located in space so that no relationship exists between the nearby values, the data exhibit zero spatial autocorrelation. Univariate Moran’s I and LISA are the most common techniques to explore the existence of spatial autocorrelation among samples (Anselin, 1995; Anselin, Sridharan, & Gholston, 2007) and its adapted form known as bivariate Moran’s I and LISA explains the spatial pattern formed by two different variables. Spatial-temporal autocorrelation is a special case in which the correlation of a variable in reference to spatial location of the variable within a time interval is assessed, i.e. correlation of a variable with itself over space and time (Anselin, Syabri, & Smirnov, 2002). While both univariate and bivariate Moran’s I aim to measure similarities and dissimilarities of spatial data, they are found to be less useful in case of uneven spatial clustering. In fact, the global univariate and bivariate Moran’s I might wrongly show that there is no relationship among the samples while there might be strong correlation in different parts of the study area hidden. Thus to explore extra information on the location of the clusters, whereas univariate and bivariate LISA are used to detect local autocorrelation by indicating the exact location of clusters (Anselin, et al., 2002).
Khan et al. (2008) employed univariate Moran’s I indicator to investigate if neighboring counties in Wisconsin display similar trends of ice-related crashes or not (Khan, Qin, & Noyce, 2008). Erdogan in 2009 employed the spatial autocorrelation indicators Moran’s I, Geary C and Getis Ord to examine the dependency of the crash rates over the provinces in Turkey. The spatial distribution of provinces with high rates of crashes indicated cluster pattern and were detected with significance of p-value <0.05 based on spatial autocorrelation analyses (Erdogan, 2009).

Investigating the spatial-temporal correlation among neighboring samples, has been reviewed by different researchers as well. In the study conducted by Wang et al. in 2006 the spatial and temporal correlation for longitudinal data and intersection clusters long corridors for the rear-end crashes at 476 signalized intersections in Florida counties were investigated and the results of temporal analysis indicated the correlation of crash occurrences for years 2000 and 2002 (Wang & Abdel-Aty-Mohamed, 2006). Li et al. (2007) presented a GIS-based Bayesian approach to investigate the spatial and temporal patterns of intra-city motor vehicle crashes. The spatial analysis according to the day of the week and different years (1996-2000) suggested the stability of high-risk segments (Li, et al., 2007). In 2008, Erdogan et al. employed GIS-methods to explore the crash hotspots in city of Afyonkarahisar in Turkey to examine the hotspots conditions. The temporal analysis revealed the seasonal correlation of crashes in the summer and winter in crossroad of the villages, small cities and slippery regions (Erdogan, et al., 2008).

Eckley & Curtin in 2012 investigated the spatial and temporal clustering of fatal crashes in eastern Fairfax County, Virginia and injury incidents in Franklin County, Ohio on a network was investigated using Knox method (Eckley & Curtin, 2012). Knox index seems a simple comparison of the relationship between incidents in terms of distance and time, but each individual pair is compared in terms of distance and time interval. Employing Knox as micro-level method requires detailed numerous data the collection of which depends on the study area.

In the present study, Mashhad the second most populated city in Iran will be considered. In recent years, the city has faced with the challenging problem of rapid socio-economic growth and the increasing number of vehicles. In this research we aim to explore the spatial-temporal autocorrelation of crashes aggregated in 253 Traffic Analysis Zones (TAZs) in Mashhad for four successive years from 2006-2009. The bivariate Moran’s I and LISA which had been particularly applied in the field of health (Astutik, Rahayudi, Iskandar, Fitriani, & Murtini, 2011; Hu & Ranga Rao, 2009; Uthman, 2008) and ecology (Stéphane, Sonia, & François, 2008) will be employed in the field of safety. Among the methods developed in previous studies in the field of safety, no study was found to simultaneously investigate spatial-temporal dependency of crashes over spatial units in urban areas using bivariate Moran and LISA. Combination of GIS and statistical analyses contribute to characterize the spatial influences of crashes and provide a quick view of the relatively critical areas that need more attention from transportation authorities.

METHODOLOGY

This study is based on a database including the total number of crashes available for Mashhad during four years from 2006 to 2009, updated by Mashhad Transportation and Traffic Organization (MTTO, 2007). Several comprehensive studies have been carried out and accordingly the city has been divided into 253 homogenous TAZs. Primarily, the TAZs’ borders were examined for any topological gaps or
 overlays and then the geometrical errors were modified and eliminated. Crash data were then assigned the TAZs using spatial join and were grouped and allocated to the TAZs' centroid. Since no GPS is used in reporting the crashes in Iran, the geographical locations of crash points are not very exact. On the other hand, the TAZ's borders are not accurately correspondent to the road network centerline (particularly by considering the highway width). To investigate the case an analysis in ArcGIS indicated that for each year of study, very less than 0.5% of crashes corresponded to the TAZs’ borders for which we eliminated from analysis. Other researchers suggested different methods. For example (Fotheringham & Wenger, 2000) pointed that crash locations were almost located with approximately 20m error by considering the street width which was large enough to overlap with one or more zone. Such inference can be true in present analysis either, since the spatial autocorrelation measures the similarity of crash occurrences in neighboring TAZs whether are assigned to a TAZ or its neighbors. In the study by (Lovegrove & Sayed, 2006), the crashes were single-counted and allocated to zones based on their geocoded locations. (Abdel-Aty, Siddiqui, Huang, & Wang, 2011; Hadayeghi, Shalaby, & Persaud, 2010; Naderan & Shahi, 2010) also developed the TAZ-level safety models based, but no discussion was done over the issue of assigning the crashes located on TAZ borders. In macro-level analysis, census blocks or regions can also be employed as spatial units, however since this research was considered as a pre-analysis to develop a spatial crash prediction model and in which we aimed to employ travel demand as an underlying factor affecting the crash occurrence, TAZ was selected. It is worth to note that in macro-level models, data are assigned to zones’ centroid rather than their specific location which might lead to the strange effect of ecological fallacy (O’Sullivan & Unwin, 2002), however choosing the homogeneous areas such as TAZs outweighs such bias (Hadayeghi, et al., 2010; Levine, et al., 1995).

GLOBAL BIVARIATE SPATIAL AUTOCORRELATION

The global bivariate Moran’s I statistic quantifies the spatial dependency between two variables $x_i, x_k$ (in this paper number of total crashes in two different time of study) in a same location $i$ (Anselin, et al., 2002). This yields a counterpart of a univariate Moran-like spatial autocorrelation defined as follows:

$$ I_{kl} = \frac{z_k w z_i}{n} $$

(1)

Where n is the number of observations, $z_k = [x_k - \overline{x_k}] / \sigma_k$ and $z_i = [x_i - \overline{x_i}] / \sigma_i$ have been standardized such that the mean is zero and standard deviation equals one. $w$ is the row-standardized spatial weight matrix. The weight matrix defines the neighbor set for each observation with non-zero elements for neighbor and zero for the others. The significance of this bivariate spatial correlation can be assessed typically by means of a randomization (or permutation) approach. In this case, the observed values for one of the variables are randomly reallocated to locations (centroids of TAZs) and the statistic is recomputed for each such random pattern. The resulting empirical reference distribution provides a way to evaluate how extreme the observed statistic is relative to what its distribution would be under spatial randomness and will produce a Moran’s I scatter plot. The Moran’s I scatter plot visualizes a spatial autocorrelation statistic as the slope of the regression line with the spatial lag (a weighted average of the value of variable in the neighboring locations) on the horizontal axis and the original variable on the vertical axis (using the variables in standardized form) (Anselin, et al., 2002). The same analysis is true for bivariate Moran’s I scatter plot. The slope of the linear regression
through this scatter plot equals the statistic in equation.1. In this paper, yielding an interpretation of the spatial lag as an inverse distance neighboring values will be used for the spatial weight matrix. Since the z variables are standardized, the sum of squares used in the denominator of equation.1 is constant and equals to n, no matter whether z_k or z_l are used. Therefore, the focus will be on the linear association between a variable z_k at a location i (z_{ki}), and the corresponding spatial lag for the other variable, [w z_l]_i. This concept was derived from bivariate spatial correlation and therefore centers on the extent to which values for one variable z_k observed at a given location k show a systematic (more than likely under spatial randomness) association with another variable z_l observed at the neighboring locations of i (Wartenberg, 1985).

SPACE-TIME AUTOCORRELATION

Space-time autocorrelation is a special case of mentioned bivariate spatial autocorrelation. Instead of employing different variables, z_k and z_l could be the same variable observed but in two instants of time, t and \( t' \). In this case, the bivariate Moran’s I computes the relationship between the spatial lag, \( w z_t \), at time t and the original variable, z, at time \( t' \). Therefore, this statistic quantifies the influence that a change in a spatial variable z, which operated in time \( t' \) at an individual location k, exerts over its neighborhood at the time t (\( w z_t \)). Hence, it is possible to define the global Moran’s I space-time autocorrelation statistic as follows:

\[
I_{t,t'} = \frac{z_t^t \cdot w z_{t'}}{n} \tag{2}
\]

where, as in the last case, the variable z is also standardized.

LOCAL MORAN’S I, LISA

The global Moran’s I does not give any information on where the clusters exist. It is likely that the global statistics wrongly show that there is no relationship among the data while there might be a strong correlation in different parts of the study area. Local Indicators of Spatial Association (LISA) provides a measure of association for each spatial unit (e.g. TAZ) and helps to identify the type of spatial correlation. The numerator in equation.2 can be decomposed into the contributions of the individual observations (Anselin, et al., 2002). The bivariate LISA can be defined as:

\[
I_{id} = z_i^t \sum_j w_{ij} z_j^t \tag{3}
\]

This statistics gives an indication of the degree of linear association (positive or negative) between the value for one variable at a given location i and the average of another variable at neighboring locations (such as j,) (spatial lag). Greater similarity than indicated under spatial randomness suggests a spatially similar cluster of the two variables. However, in case of spatial temporal analysis, the first and second variables are the attribute under study at specific location but in different times. The results of LISA is a map which helps to determine the nature of spatial-temporal autocorrelation so that they can be categorized into four groups: two categories of positive spatial correlation, or spatial clusters (high-high and low-low) which relate to values physically surrounded by neighboring TAZs with similar values and two categories of negative spatial correlation, or spatial outliers (high-low and low-high) which relate to values whose neighbor TAZs are dissimilar. Inference can be based on a permutation or randomization approach. The spatial analyses in this
Bivariate Moran’s I and LISA to explore the crash risky locations in urban areas
Ali Akbar Matkan, Matin Shahri, Mojgan Mirzaie

paper are implemented through the GeoDa software (Anselin, 2005) which provides a very user-friendly environment to implement Spatial Data Analysis (SDA) methods.

RESULTS

RESULTS OF GLOBAL BIVARIATE MORAN’S I

As described earlier, the purpose of employing global bivariate Moran’s I analysis is to indicate the variations of spatial distribution of data patterns and in case where the correlation of one variable is investigated for different time intervals, this leads to the spatial-temporal autocorrelation. To this end, the bivariate Moran’s I indicator was examined for number of crashes occurred in every TAZ in Mashhad for one year of study as the original variable and the spatial lag of the crash for the next year as the second variable; therefore the results imply the spatial-temporal autocorrelation (e.g. the crashes in the year 2006 as the original variable and spatial weighted in the year 2007 in [Fig. 1a]). The results of global bivariate Moran’s I and spatial weighted in year 2007 have been illustrated in [Fig. 1a].

The results of global bivariate Moran’s I have been indicated in Table 1 with significance values based on a permutation approach. As evident, the significant and positive spatial autocorrelation across all spatial weights can be inferred or in simple terms, that crash occurrence in Mashhad have some kind of organized spatial-temporal patterns. Such inference is based on Moran’s I indicator and corresponding p-value < 0.05. The systematic variation of crash occurrences across space and time would suggest that spatial-temporal pattern should be taken into account for safety analyses. [Fig. 1] also depicts the corresponding global Moran’s I scatter plot. The slopes of the regression line in six scatter plots equals to Moran's I indicator in Table 1, differentiate from zero which is an indicator of significant spatial-temporal autocorrelation among crashes. The significant autocorrelation of crashes justifies that the crashes follow an organized cluster pattern in Mashhad. It is also worth to note that the analysis can be repeated if the vertical and horizontal axes in scatter plot are inversed i.e. the crashes in year 2007 as the original variable on the vertical axis and spatial weighted in year 2006 on the horizontal axis in [Fig. 1a], although the results will be approximately the same.

<table>
<thead>
<tr>
<th>Original variable</th>
<th>Spatial lag</th>
<th>Bivariate Moran's I</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAZ-Level crashes in 2006</td>
<td>TAZ-Level crashes in 2007</td>
<td>0.2348</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>TAZ-Level crashes in 2006</td>
<td>TAZ-Level crashes in 2008</td>
<td>0.1815</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>TAZ-Level crashes in 2006</td>
<td>TAZ-Level crashes in 2009</td>
<td>0.1422</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>TAZ-Level crashes in 2007</td>
<td>TAZ-Level crashes in 2009</td>
<td>0.1493</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>TAZ-Level crashes in 2007</td>
<td>TAZ-Level crashes in 2008</td>
<td>0.2080</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>TAZ-Level crashes in 2008</td>
<td>TAZ-Level crashes in 2009</td>
<td>0.1163</td>
<td>&lt; 0.05</td>
</tr>
</tbody>
</table>

RESULTS OF LISA

The global analyses have suggested non-randomness in the overall spatial-temporal pattern of crashes. More information on what kinds of clustering may be present is provided by an analysis of LISA. The bivariate LISA cluster map (for different permutations and p-values) for crashes in year 2006 as the original variable and crashes in years 2007, 2008 and 2009 as the spatial lag have been indicated in [Fig. 2]. It is evident that most of the TAZs are highly statistically significant (p-value<0.05). Similar analysis by inferring the bivariate LISA and the related significant maps for crashes in year 2007 and
Bivariate Moran’s I and LISA to explore the crash risky locations in urban areas
Ali Akbar Matkan, Matin Shahri, Mojgan Mirzaie

N-AERUS XIV
Enschede 12 - 14th September 2013

[fig. 2] the LISA map for crashes in comparing years a) 2006 and 2007 b) 2006 and 2008 c) 2007 and 2009 d,e,f) the corresponding significant maps
Bivariate Moran’s I and LISA to explore the crash risky locations in urban areas
Ali Akbar Matkan, Matin Shahri, Mojgan Mirzaie

N-AERUS XIV
Enschede 12 - 14th September 2013

[fig. 3] The LISA map for crashes in comparing years a) 2007 and 2008 b) 2007 and 2009 c) 2008 and 2009 d,e,f) the corresponding significant maps
2008 as the original variable and crashes in the year 2008 and 2009 as the spatial lag has also been illustrated in [Fig. 3]. As can be seen, this shows local patterns of spatial correlation between number of crashes per TAZ in one year and the average of the number of crashes in the second year for its neighbor TAZs. The results of the LISA inferred by local bivariate Moran’s I enable dividing the study area into four sub regions corresponding to the clusters and outliers. These four categories match to the four quadrants in the Moran’s I scatter plot as shown in [Fig. 1]. The high-high and low-low locations (positive local spatial autocorrelation) represent spatial clusters, while the high-low and low-high locations (negative local spatial autocorrelation) represent spatial outliers. The high-high sub region is related to the areas having high (low) number of crashes in the first year of study which surrounded by the areas with high (low) weighted average of crashes for the second year of study (e.g. number of crashes in year 2006 and weighted average of number of crashes in year 2007 in [Fig. 2a]). In total, the results reflect a distinct pattern of crash occurrences spatially and temporally. It can be inferred from maps that the high-high sub regions corresponding to the crash hotspots have been concentrated mainly in the center of the Mashhad expanding from northwest to southeast. The stability of spatial pattern indicates the likelihood of occurring crashes in the same regions. Such areas as so-called hotspots demand more safety attention. The significant maps also reveal that for most of the TAZs located in high-high sub region, the results are highly significant (p-value<0.01). Another interesting result that can be inferred is that a majority of the units of these measures which fall into the “Not-Significant” category, remains the same for all three years as well.

CONCLUSION

In this paper we investigated whether the patterning of crash occurrences follows an organized spatial and temporal pattern or not. As presented, the consistent answer obtained in this paper is that the crash occurrences over TAZs in Mashhad are spatially-temporally dependent in the four successive years of study which was yield through bivariate global and local Moran’s I indicator. Such results indicate that the actual spatial patterns of crashes are strongly associated spatially and temporally. Stability of spatial patterns of crashes over the time leads the safety specialists to infer that the crash occurrences might be affected by the same spatial or non-spatial factors. Such a focus is consistent with the strategies of Transportation and Traffic Organization, which aims to reduce and ultimately eliminate crash occurrence. Therefore, since most of the data needed for crash prediction models are not available for the years of study and the data preparation depends directly on the study area, the results of spatial and temporal analysis can help the traffic engineers develop their safety models for the particular year and generalize the results to three years with high spatial and temporal dependency. A geographic analysis helps to identify TAZs or regions of the study area that have relatively high crash occurrences and may need to be particularly targeted with safety-attention programs. Such results also contributes the safety specialists to assign the resources such as budgets and time by including the high-risk TAZs in detailed engineering study sites. The results of our analysis highlighted the overall pattern of hotspots in the study area, although investigating for the factors affecting the crashes in such detected regions needs further research.
REFERENCES


Anselin, L. (2005). GeoDa version 0.9.5-i


Bivariate Moran’s I and LISA to explore the crash risky locations in urban areas

Ali Akbar Matkan, Matin Shahri, Mojgan Mirzaie

N-AERUS XIV
Enschede 12 - 14th September 2013


