Using GIS to Identify Road Segments for Lighting Curfew

CVEN658 Term Project

Hongmin Zhou
Zachry Department of Civil Engineering
Texas A&M University
Instructor: Dr. Francisco Olivera
CVEN 658 Civil Engineering Applications of GIS

12/6/2010
ABSTRACT

More and more traffic agencies are considering adopting lighting curfew policy to reduce roadway lighting energy cost and preserve environment. A typical duration of lighting curfew is to turn off lights from 10 pm to 5 am at roadway segments where dark condition will not result in significant increase in crashes. This study used geographic data of Harris County and crash data to identify high crash locations where lighting curfew for continuous lighting would potentially raise a safety problem. The author did basic data processing for roadway network and crash data using functions of ArcGIS such as split, spatial join, multipart to singlepart, buffer etc. The author also applied “Hot Spot Analysis (Getis-Ord Gi*)” function to examine random effect of crash locations based on measures of crash cost and crash frequency. This study found that: (1) the simple map of crash measures appeared countywide discrete distribution of crash prone segments; (2) the hot spot analysis showed more meaningful information of crash concentrations. Northwest and southwest of central Houston area had concentrations of crashes with both high crash frequency and crash severity. Thus implementation of lighting curfew in Harris County, Texas should avoid locations close to those areas.
# Table of Contents

Abstract ........................................................................................................................................... 1  
Introduction ..................................................................................................................................... 3  
  Background .................................................................................................................................... 3  
  Literature Review .......................................................................................................................... 4  
  Study Purpose and Outline .......................................................................................................... 6  
Data Collection .................................................................................................................................. 7  
  Geographical Data ...................................................................................................................... 7  
  Crash Data .................................................................................................................................... 7  
Methodology ...................................................................................................................................... 9  
  Basic Data Processing ................................................................................................................. 9  
    Geographical Data ................................................................................................................... 9  
    Crash Data ................................................................................................................................ 10  
  Hot Spot Analysis ....................................................................................................................... 11  
    Gi* Statistics and Z-Score .................................................................................................... 11  
    Spatial Analysis of Hot Spots ............................................................................................... 12  
Results and Discussion .................................................................................................................. 13  
Conclusion and Recommendation ................................................................................................ 16  
Reference ....................................................................................................................................... 17  

Table 1 Variables in Crash data Lighting Curfew Hot Spot Analysis ............................................ 8  
Table 2 Methods of Basic Data Processing in ArcGIS 10 for Geographic and Crash data............ 9  
Figure 1 Network Screening Process ............................................................................................ 4  
Figure 2 Functions of Split Line at Point and Split into Equal Parts ............................................ 10  
Figure 3 Address Locator Interface for Address Match ............................................................... 10  
Figure 4 Spatial Join for Crash Summary per Roadway Segment ................................................ 11  
Figure 5 Hot Spot Analysis (Getis-Ord Gi*) Interface ................................................................. 12  
Figure 6 Site Screening using Simple Mapping of MOEs............................................................ 13  
Figure 7 Hot Spot Analysis Outputs of Crash Cost and Crash Frequency ................................... 14  
Figure 8 Hot Spot Analysis Output Using Combined Rank as MOE ........................................... 15
INTRODUCTION

The term lighting curfew is used to describe the concept of reducing or eliminating lighting during part of the nighttime hours. Currently many traffic agencies are considering adopting lighting curfew strategy on highway and street lighting to save energy costs and reduce lighting pollution. This study is trying to identify roadway locations where implementing lighting curfew will not potentially result in safety problems applying functions of Geographic information systems (GIS) software package.

Background

Roadway lighting is provided on selected highways to improve the visibility of the nighttime environment, and thus improve safety, and benefit local security and beautification. Current practice in the United States is normally lighting the roadway throughout the hours of darkness with the same illuminance and/or luminance level justified by warranting conditions and proper design criteria. There are two types of roadway lighting: safety lighting and continuous lighting. Safety lighting is often installed at intersections, interchanges or ramp areas where high traffic conflicts and/or pedestrian activities exist. Continuous lighting is generally for roadway segments between intersections or interchanges where safety is an issue due to high traffic volume, constrained geometry etc (CIE, 2010).

Energy consumption resulted from such constant lighting especially continuous lighting, however, could be costly and unnecessary when traffic volume drops to a low level after late night. Lighting pollution is becoming another issue for people who prefer dark skies at night. It is also reported that excessive artificial lighting along the coastal roadways is the major cause of population decline in sea turtle hatchlings (Washburn et al., 2003). Therefore, many traffic agencies consider applying advanced techniques and developing new policies for more effective roadway lighting, such as lighting curfew or adaptive lighting. Choosing locations where lights are switched off should consider the following factors:

- It should not be locations where turning off lightings would cause a substantial increase in crash rate and severity.
- If choose traffic volume as a major factor of site screening, the underlying assumption is that low crash frequency or crash rate is associated with low traffic volume.
- The time duration when roadway lights are turned off should be consistent along time of day (or week and season) based on natural lighting level and traffic volume. A typical duration of lighting curfew is from 10 pm to 5 am (Rhode Island DOT, 2010).
- Consider site selection on a network level, there should be priority among all candidate sites based on crash history and available resources such as traffic count station.
Thus the first step of identifying candidate locations for lighting curfew is to exclude high crash locations, namely “hot spot” in the realm of traffic safety. Literature has informed the author that ArcGIS has a couple of functions capable of doing hot spot analysis.

Literature Review

Identification of hazardous sites is one type of network screening, the process of which is shown in Figure 1 (Lord, class material, November 2010).

![Figure 1 Network Screening Process](image)

The first task of network screening is to identify purpose and target specific crashes and types of sites or facilities (i.e. segments, intersections, etc.), and identify control group with sites that have similar characteristics. Then choose measure of effectiveness (MOE) in order to use one of multiple site screening methods. MOEs for traffic safety include crash frequency (number of crashes per site), crash rate (number of crashes per unit of traffic exposure), equivalent property damage only, relative severity index and combined criteria etc. (Lord, class material, November 2010). With the chosen MOEs, proper screening methods can be then applied to identify hazardous locations. As of interest, numerous studies on traffic safety have reported the use of GIS function to analyze spatial characteristics of crashes.

Gundogdu (2010) conducted a GIS-based hot spot analysis for preventing crashes in Turkey by applying a Linear Referencing Model (LRM). A LRM is a method of measuring (e.g., mile point, address, station) along linear elements, such as routes, streets or alignments. It uses a location expression to specify event locations which consists of a linear element and a distance expression. In the study, the linear element applied the normalized relative crash rate of roadway segments, i.e. z-score value. Calculated Z-score values were inquired in the GIS analysis to see
if each of them was greater than the 1.645 threshold (i.e. 95% confident level), so that segments of high crash rate (for total number of crashes or crashes with certain level of crash severity) or high crash potential (segments having z-scores close to the threshold) could be identified. The steps analyzing hot spots in the study were summarized below:

- Collected accident data and geographical data. The crash data should contain enough information to locate the crash site with the geographical data;
- Divided the roads in to 1-km segments and used Poisson distribution to estimate crash potential for each segment. The expectation was the mean number of crashes of all segments;
- Calculated crash rate and relative crash rate of each segment and defined a standardized normal variable using the relative crash rate data;
- Compared the normalized value (z-score) with the threshold 1.645, identified hot-spot segments with equal- or higher-than-threshold z-scores. Segments with less-than but close-to-threshold z-scores were potential hot spots.
- Illustrated the result on the map by specifying crash types for hot spots and potential hot spots;
- For hot spots, locations with all type of crashes clustered were further defined as “intersection segment” which should be given first priority for safety improvement; potential hot spots should receive second priority for further safety enhancement (Gundogdu, 2010).

Pulugurtha et al. (2007) applied GIS to rank high pedestrian crash zones. Identification of high pedestrian crash zones involved three steps: 1) geocoded pedestrian crash data, 2) created crash concentration map, and 3) identified zones, their shapes and sizes. For geocoding process, the “address match” feature of Geographical Information Systems (GIS) software was used to locate crash site to the map, by either the names of intersecting streets for crashes happened at intersections or street addresses for crashes happened at mid-block locations. For crash concentration map, two methods were used. In the Simple Method, a circular search area was drawn around each cell, and the cell density was the ratio of number of crashes that fell within the search area to the size of the search area. In the Kernel Method, the circular search area was drawn around each crash, and the surface value was highest at the location of the crash and diminished with increasing distance from the crash, reaching 0 at the radius distance from the crash. The Kernel Method created a smoother looking density surface than the Simple Method and thus was more suitable for identifying locations with severe crash problems. To identify zones, two types were considered, i.e. linear zones and circular zones. Clusters of crashes closely spaced along a road were called linear zones and elsewhere are called circular zones. Using crash concentration map to identify zones was less subjective than using simple crash distribution map.
The study also introduced two types of methods to evaluate high pedestrian crash zones. Individual method was to rank crash risk based on single criteria while composite method was to rank risky zones based on combination of individual criteria. The single criteria used include crash frequency, crash density and crash rate, each of which considered absolute crash number, crashes per unit of zone size or crashes per unit of traffic exposure, respectively. In the composite method, each zone were ranked based on these single criteria and the average rank of three criteria, or the sum of standardized criteria was the total score of each zone (Pulugurtha et al., 2007).

There was also a study specifically focusing on roadway lighting and crashes (Saraiji et al., 2009). It mainly dealt with geo-mapping and demonstrating streetlights information and nighttime crash at intersections. In order to find the percentage of a particular street being lit, the authors assigned a circle to each streetlight indicating the illuminance contours of the light based on lighting installation spacing on particular road classification specified by the ISEN lighting guide (ISEN 2000). For intersection crashes, the study considered the following factors:

- Impact area of an intersection on crashes was 250 ft from the center of the intersection.
- Discrimination between daytime and nighttime crashes based on actual time of sunset and sunrise with daylight saving time adjusted.
- Night-to-day crash ratio was the MOE for intersection safety and was used for simple mapping and information demonstration (Saraiji et al., 2009).

**Study Purpose and Outline**

None of previous studies, according to the author’s knowledge, focused on the GIS-based identification of continuous lighting locations for lighting curfew. In the state of Texas, Houston area of Harris County has a roadway network with a number of continuous lighting facilities which consume a substantial amount of energy and costs every year. Thus, this study focuses on hot spot analysis of nighttime crashes in Harris County, Texas to identify candidate roadway segments with continuous lighting for lighting curfew implementation.

- Obtain geographical data and crash data in Harris County, Texas for major highways.
- Extract nighttime crashes to consider lighting condition related segments.
- Use crash frequency and crash severity (revealed by crash cost) as the MOEs for each roadway segment between interchanges. Perform simple geo-mapping for high crash locations.
- Use the functions of spatial analysis in ArcGIS 10 to conduct hot spot analysis for combined MOE from crash frequency and crash cost.
- Exclude high crash locations and recommend candidate roadway segments for further site study of lighting curfew.
DATA COLLECTION

For the purpose of interest, data needed of this study include two types: geographical data of Harris County, Texas, such as major highways with high percentage of continuous lighting, other roadways intersecting those major highways at interchanges; crash data with location, time and light condition variables for extracting nighttime crashes.

Geographical Data

Selection of target study area considers locations of continuous lighting and availability of hourly traffic data. Approximate lighting location information is obtained via email surveys sent out to district traffic agencies in Texas. According to the responses received by the time of this study, Houston area has a roadway network with the densest continuous lighting among all districts replied. Also, eight major highways in Houston area have been equipped with Automatic Traffic Recorders (ATRs) at many roadway cross-sections: US-290, US-59, IH-10, IH-45, IH-610, ShH-35, SH-146 and SH-249. Thus geographic data of Harris County is accessed via TIGER (Topologically Integrated Geographic Encoding and Referencing) system of U.S. Census Bureau. The latest shapefiles on TIGER website for Harris County record the boundaries for legal entities as of January 1, 2009, though a few states also have their shapefiles updated as of January 1, 2010 (U.S. Census Bureau, 2009). The roadway polyline shapefile of Harris County contains 327571 records of all types of roadway using a geographic coordinate system of GCS_North_American_1983. Each of those eight highways consists of hundreds of individual roadways that need basic data processing to be further analyzed. Other roadways of interest are secondary roadways intersecting the major highways at interchanges.

Crash Data

Texas Department of Transportation (TxDOT) provide online crash data request system. TxDOT maintain an automated database of all reported motor vehicle traffic crashes in Texas for recent six calendar years. This crash database is required by federal highway safety laws for TxDOT to obtain federal safety improvement funds. But not all reported motor vehicle crash information is releasable according Texas Transportation Code for tort liability concern (Texas Transportation Department, 2010).

Among all data filter choices, crash data by geographical area is applied. Also available are crashes by severity, crashes by types of involvement such as alcohol, speed, pedestrian and bicyclist, or by driver age group and vehicle types. Crash data by geographical area requested contain all types of crashes in Harris County in 2003, and it takes about one week to receive the crash data after submitting the request. The data set includes variables of interest such as light condition, crash time and crash severity. Though the crash data set does not contain crash cost.
information, TxDOT provide their estimation of crash cost by severity. Thus this study simply assigns a cost to each crash according to crash severity variable. Table 1 shows the brief summary of these variables.

**Table 1 Variables in Crash data Lighting Curfew Hot Spot Analysis**

<table>
<thead>
<tr>
<th>Crash Time</th>
<th>Light Condition</th>
<th>Crash Severity and Crash Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time of Day</td>
<td>Dawn</td>
<td>Incapacitating Injury</td>
</tr>
<tr>
<td>Day of Week</td>
<td>Dark</td>
<td>Kiln</td>
</tr>
<tr>
<td>Month</td>
<td>Unlighted</td>
<td>Not Injured</td>
</tr>
<tr>
<td>Year</td>
<td>Daylight</td>
<td>$1,000</td>
</tr>
<tr>
<td></td>
<td>Dark Lighted</td>
<td>$180,000</td>
</tr>
<tr>
<td></td>
<td>Dark Unlighted</td>
<td>$36,000</td>
</tr>
<tr>
<td></td>
<td>Dawn</td>
<td>Possible Injury</td>
</tr>
<tr>
<td></td>
<td>Dusk</td>
<td>$19,000</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>Killed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$2,600,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not Injured</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$2,000</td>
</tr>
</tbody>
</table>
METHODOLOGY

Data sets obtained are quite coarse thus need basic data processing using ArcGIS before proceeding to spatial hot spot analysis. Table 2 Methods of Basic Data Processing in ArcGIS 10 for Geographic and Crash data

Table 2 Methods of Basic Data Processing in ArcGIS 10 for Geographic and Crash data

<table>
<thead>
<tr>
<th>Geographic Data</th>
<th>Crash Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection by attribute, Dissolve, Clip, Intersect, Append, Multipart To Singlepart, Buffer, Split Line At Points, Selection by location, Split</td>
<td>Selection by Location</td>
</tr>
</tbody>
</table>

Basic Data Processing

In traffic safety study, MOEs are often based on a site or a unit of roadway segment. Thus the purpose of basic data processing here is to get the MOEs of crash frequency and crash cost per segment with a relative uniform length, say number of crashes per mile and crash cost per mile.

Geographical Data

The major roadways of Harris County are all freeways or highways that are accessed through interchanges or ramps from other roads. The roadway geographic data contained in the polyline shapefile, however, originally consists of separated segments for both traffic directions between any roadway facilities, regardless of the real existence of the intersections at those facilities. Thus in order to extract crash locations of segments between interchanges, all individual segments for one traffic direction are selected and dissolved to one single roadway for each of the eight major highways. Then secondary highways intersecting the major highways at interchanges are identified by browsing Google map and extracted from the shapefile.

The impact area of an interchange is about 0.5 mile from the center of interchange to each direction of the mainline roadway, thus a 0.5-mile buffer zone should be excluded from every interchange on the major highways. One difficulty is that ArcGIS 10 package does not have a “subtract” or a “minus” function. Alternative solvation is to intersect major highways with buffer boundaries and use the “split at point” function under the Editor menu to break major highways at those intersecting points. When the point features do not locate exactly on the line features, a “search distance” is available to get the approximate location of the points and the corresponding perpendicular points on the line become the breaking points. Then select roadway segments outside the buffers, and split each of those segments into approximate 1-mile sub-
segments for further analysis. Note that the number of parts in Split function interface does not allow decimal input. See Figure 2 for the two interfaces.

![Figure 2 Functions of Split Line at Point and Split into Equal Parts](image)

**Crash Data**

Filtering nighttime crashes is based on two crash variables: light condition and time of day. For light condition, crashes under dark condition and dark lighted condition are extracted. For time of day, crashes occurred from 10 pm to 5 am with daylight time saving adjusted are selected. The two sets of crash data are parallel to each other and are analyzed to examine the consistency of the crash records about lighting location. This comparison is necessary to decide whether further data collection is needed.

The dataset has three sets of variables recording crash locations: latitude and longitude, street name and number, two intersecting roadways. Though latitude and longitude information is preferred for locating crashes on the map, not all crashes have complete information of latitude and longitude. Thus the process of geocoding crash locations consists of two parts: display and export XY data for crashes with latitude and longitude information, address match for crashes without latitude and longitude records. When create address locator for the address match, choose “US Address – One Range” since the side of street in the output data is not essential for crash location. See Figure 3 for the interface of address locator.

![Figure 3 Address Locator Interface for Address Match](image)
Once all crashes are located on the map, use “Selection by Location” to extract crashes on roadway segments, i.e. exclude interchange crashes. Then for the MOEs chosen, number of crashes per segment and crash cost per segment require crash information on each segment be summarized. For this summary purpose, use “Spatial Join” function shown in Figure 4.

Till now, each roadway segment has two fields of MOEs: number of crashes per mile and crash cost per mile. The spatial distribution of crashes can be simply demonstrated by setting the symbology of both fields as “Graduated Colors”. Segments with MOEs greater than the countywide average values (defined as MOE thresholds) should be high crash locations.

**Hot Spot Analysis**

When screening sites from a large scale roadway network based on limited information, random effect would be one of the important issues that need to consider. In ArcGIS, there is a function in spatial analysis called “Hot Spot Analysis (Getis-Ord Gi*)”. This function can examine if there is a random effect in the spatial distribution of the target MOEs. Note that “hot spot” is a term describing high crash location in traffic engineering, while in ArcGIS it is used to name a point or area with spatial concentration of any subject value of interest, such as pollution, 911 emergency calls etc.

**Gi* Statistics and Z-Score**

“Hot Spot Analysis (Getis-Ord Gi*)” in ArcGIS spatial analysis has two important concepts: Gi* Statistics and Z-Score. The Gi* statistic uses weighted data points to identify point values higher than that could be found in random distribution. The calculation of Gi* statistic follows:
Where: $x_j$ is the data point, i.e. MOE in this case; $\bar{X}$ and $S$ are the mean value and standard deviation of the data points respectively. $\omega_{i,j}$ is the weight value for each data point. ArcGIS provides several choices for the weight determination.

Z-score is the major output of this Gi* function. It represents the statistical significance of clustering for a specified distance. This function tests the null hypothesis that there is complete spatial randomness in the Gi* statistic using normal distribution. Large z-scores and corresponding small p-values indicate a tendency of rejecting the null hypothesis. This study adopts a z-score threshold of ±1.65 (i.e. confidence interval at 90 percent) to screen high or low crash locations, namely “hot spots” or “cold spots”. Cold spots are candidate locations for lighting curfew (ArcGIS Desktop 9.3 Help, 2010).

Spatial Analysis of Hot Spots

Figure 5 shows the interface of “Hot Spots Analysis”. The “Input Feature Class” is the roadway network containing crash summary information (i.e. the output after spatial join of roadway polyline feature and crash point feature). The “Input Field” is the field of MOE. The “Conceptualization of Spatial Relationships” provides methods of determining weight information to calculate Gi* statistic. This study chooses “ZONE_OF_INFERENCE”, i.e. the weight is set as 1 within the zone distance (one mile in this study), and diminish with distance outside the zone. The “Distance Method” has two ways of calculating distance from each feature to neighboring features. This study use simple straight line, i.e. Euclidean Distance.

![Figure 5 Hot Spot Analysis (Getis-Ord Gi*) Interface](image-url)
Results and Discussion

Figure 6 shows the result of this basic site screening using nighttime crash data extracted based on Time of Day variable. Note that segments in Orange and Red might be hot spots according to Figure 6.

![Figure 6 Site Screening using Simple Mapping of MOEs](image_url)
Hot spots screened from simple mapping are spatially discrete. Locations with crash cost higher than the countywide average crash cost $546,000 per mile or with crash frequency higher than countywide average crash frequency 9.285 crashes per mile are considered as hot spots. Lighting curfew implementation based on this simple screening might be problematic because nighttime drivers might have difficulty in visual adaptation when driving on a roadway with frequently changed lighting conditions. Therefore, further look into the outputs of “Hot Spot Analysis (Getis-Ord Gi*)”. Figure 7 shows the maps of crash concentration of crash cost and crash frequency. Locations with z-scores higher than 1.65 are hot spots, with z-scores lower than -1.65 are cold spots.

![Hot Spot Analysis Outputs of Crash Cost and Crash Frequency](image)

However, there is discrepancy in the results of hot spot analysis using both MOEs. Locations with high concentration of crash severity may be of low crash frequency and vice versa (see the two sets of ellipses in Figure 7). Thus this study further averages the rank of z-scores for both MOEs and uses it as a new MOE to redo Hot Spot analysis. Figure 8 is the output of crash concentration using this rank-combined MOE. This figure shows that northwest and southwest of central Houston area have concentrations of crashes with both high crash frequency and crash severity. Segments in green are candidate locations for lighting curfew with confidence interval 90 percent or higher.
Hot spot analysis for crash data extracted using Light Condition variable is found less reliable. When compare the concentration map of crashes under “Dark Lighted” and “Dark Unlighted” conditions, there are some segments with high z-scores in both maps. This is problematic since lighting conditions at the same locations should not probably vary very much in the same year. Thus the crash concentration of both data sets should be complementary to each other with the exception of hot spot resulted by factors other than lighting condition. Another doubt lies in the crash reports for the light condition variable, based on which the crash database is drawn. There is probably difference between police officers recorded natural or artificial lighting conditions and the actual lighting conditions. Due to this worry, crash variable of Light Condition is not recommended for lighting curfew hot spot analysis.
CONCLUSION AND RECOMMENDATION

Hot spot analysis in this study for lighting curfew consists of the following process:

- Obtain geographic data and crash data and process the data using basic functions of ArcGIS. Extract crash data based on proper crash variables.
- Choose proper measures of effectiveness (MOEs) to evaluate roadway segments with crash summary information. Simply map the MOEs on the network for basic site screening.
- Use the “Hot Spot Analysis (Getis-Ord Gi*)” under spatial analysis tool to examine the random effect that might lie in the result of simple mapping. Catch the discrepancy among results based on different MOEs, and use the combined rank of different MOEs redo the hot spot analysis to reach a single comprehensive output.

Based on the process of hot spot analysis, the following conclusion and recommendation can be drawn for tips of software usage, study results and further research:

- Spatial Join function is capable of combining roadway polyline feature and crash point feature and providing summary of crashes for each roadway segment.
- Operation of Intersecting features is necessarily followed by a Multipart to Singlepart operation in order to obtain the right amount of features for further data analysis.
- Split Line at Point and Split line Into Equal Parts are two useful Editor functions to create a roadway network for hot analysis.
- Results from simple mapping of crash measures appeared countywide discrete distribution of crash prone segments, while the hot spot analysis showed more meaningful information of crash concentrations.
- There is discrepancy in the results using different MOEs it is necessary to use a combined MOEs for comprehensive analysis.
- Thus implementation of lighting curfew in Harris County, Texas should avoid locations close to the northwest and southwest of central Houston areas due to high concentration of crash cost and crash frequency in those areas.
- Since it is suspecting to crash variable of Light Condition, data collection of lighting condition is necessary for further site study toward feasibility of lighting curfew.
- It might be more reasonable to use crash rate, the number of crashes per veh-mile, as the MOE for hot spot analysis, thus further study would be inference of complete traffic flow information based limited traffic data and redo the analysis.
- Crash information used in this study contains only one year data. It is also necessary to process data from multiple years for examination of long term concentration of crash locations.
REFERENCE


U.S. Census Bureau (October 1, 2009), TIGER, TIGER/Line® and TIGER-Related Products, U.S. Census Bureau, http://www.census.gov/geo/www/tiger/ (November 5, 2010)
