Application of GIS for Urban Traffic Accidents: A Critical Review

Niloofer Haji Mirza Aghasi

Graduate School of Life and Environmental Science, Division of Spatial Information Science, University of Tsukuba, Tsukuba, Japan
Email: nia922@yahoo.com

Abstract

The analysis of huge data is a complex task that cannot be executed without a proper system. Geographic information systems (GISs) have been used by many transportation agencies and police departments to analyze and manage urban traffic accident (UTA) data and for decision making aimed at decreasing accident rates and increasing safety. The exact location of accidents and environmental characteristics must be analyzed as UTAs occur in specific locations with specific characteristics. ArcGIS software is the best choice for obtaining meaningful information and analysis results from UTAs in an observational time span. GIS technology is a fundamental element for investigating and evaluating the complex spatial relationship among different components and urban traffic accident is one of them. Micro or macro analysis of UTAs through the spatial prospective within the geographical environment and urban structure can make a deep micro understanding of UTAs patterns in addition to assisting in decision making. UTAs can be considered complex events that occur in two aspects which are spatial and temporal or space and time in other word. A GIS can integrate more than two different and unrelated databases. The evaluation among different spatial objects in a geographical environment and associated factors in urban structure which are included but not limited to land use category, road transportation network qualification, population density, etc., is one of the GIS specification. Traffic safety organizations and UTA researchers use GISs as a key technology to support their research and operational needs. In particular, GIS-T is an often-used GIS application used for planning and decision-making in transportation.

Keywords

Geographical Information, Road Accident, Spatial Analysis, Urban Traffic Accidents
1. Introduction

The main objective of this paper is a critical review of literature of GIS or geographical information system in the most negative impact of urban transportation, UTA, as geography can be relevant to various phenomena, including traffic accidents. Analyzing of UTAs through the spatial prospective in a geographical environment with associated factors will express the key factor for safety and decision making. Road accidents can be analyzed as discrete incident not only in space but also in time. The huge number of road transportation vehicle and users drivers has resulted in a significant number of accidents, and developed and undeveloped countries currently suffer from the consequences of these accidents. GISs have played a vital role in the field of transportation engineering. The main technique for this article was review of research with application of GIS in UTAs and urban structure. This paper considers the literature related to existing studies on traffic accidents from multiple perspectives. Previous spatial, temporal, and spatiotemporal analyses of traffic accidents are summarized. Then, we provide a detailed review of GIS applications in traffic accident studies. Traffic accident prediction models and the Bayesian approach are reviewed. In addition, the current methods of identifying hazardous roadway sites for traffic accidents are described.

2. GISs and Traffic Accidents

A fundamental premise of using GISs in relation to traffic accidents is that accidents are discrete events that are localized spatially and temporally and every phenomenon has 2 sides which are space and time. By spatial prospective view to UTAs, it’s clear that many different factors are associated with even an UTA, including road type, population density, weather, culture, distribution of facilities, etc. Understanding the spatial relationships among urban accidents and associated factors is important.

GISs now encompass computerized systems that comprise layers of additional information and a digital map background, which can be viewed in any desired combination and at any scale. GISs have been proven to be one of the most useful tools for spatial analysis and mapping of different subjects in transportation studies. UTAs have become more common with the increase in population and vehicles. Human mistakes and human errors are the most common reason of UTAs. However, the frequency of UTAs and their severity can be reduced by proper and efficient analysis. GISs and spatial analysis are considered the most appropriate methods to do spatiotemporal analysis. The spatial distribution of UTA, as the simplest factors of accidents location, can be show and create as a map by GIS. Geography can be important in the context of various phenomena, including UTAs. This subchapter provides a literature review of GIS accident analysis. Driving is a complex task, and UTAs are affected by multiple factors, which are include but not limited to road transportation network qualification, time as the most important factor in urban dynamic pattern, geography and all
associated factors in geography like as weather, etc. UTAs are complex events, and their analysis is not straightforward.

GIS technology is a fundamental element for investigating and evaluating the complex phenomena. The spatial analysis of UTAs and their relationship to the geographic environment and urban structure can provide deep insight into accident patterns, safety outcomes and decision making process. UTAs are complex events that has two sides, spatial or space and temporal or time. Large number of UTAs in recent years caused by urban expansion and increasing the road transportation network users and vehicles. Currently, both developed and under developing countries face to serious issues of UTAs and their consequences, such as disability, trauma, death, etc. A GIS can integrate two or more previously unrelated databases. Creating spatial relationship between different objects within a geographical environment and urban structure which can be included but not limited to land use category population density, urban dynamic as well as socio-economic data, is one of the most useful application of GIS. Traffic safety organizations and UTA researchers use GISs as a key technology to support their research and operational needs. In particular, GIS-T is an often-used GIS application in transportation planning and decision-making.

A UTA is a negative impact resulting from intercommunication with two or more than two associated factors. Several studies have addressed how GISs help the integration of transportation networks with different elements. For example, Martin described in 2002 how traffic safety and analysis decision programs improved reliability of results by application of GIS in UTAs.

In 1994, Johnson did a study with Demetsky [1] and by this study they confirmed the potential of GIS applications in transportation and accident management. This study highlights the notion of GIS application in UTAs and transportation network.

In addition, other research has shown the application of GIS in UTA and transportation management studies and decision making. A UTAs database which can be used by GIS with attributes was made in 1995 by Faghri and Raman [2] for Kent County, Delaware. This system included attributes of UTAs, such as environmental factors and the frequency of incidents at the accident’s location for different road types. The main fundamentals of GIS, including data acquisition, data preparation pre-processing, data interpretation and analysis, and quantitative and qualitative outcomes and production, were studied by Lewis [3].

3. Spatial Relationships and UTAs

The main objective of analyzing spatial relationship and UTAs is to make a scheme which can generate micro identification of UTAs distribution to identify the causes and consequences of and temporal, spatial or spatiotemporal variation in UTAs.

The next objective of analyzing spatial relationships and UTAs is the investigation of so-called hot spots, or high-activity areas through the urban structure
by application of GIS. This scheme includes the following tasks: identify hot spots using a clustering method, identify the pattern of UTAs for different road types, perform hot spots of UTAs, identify the hot spots pattern through urban structure, identify the spatial distribution of UTAs and hot/cold spots over time, and establish risk areas by population and land use.

**Temporal, Spatial, and Spatiotemporal Pattern Analyses of Traffic Accidents**

Different methodologies and methods have been used for UTAs studies and analysis spatially and temporally. In terms of temporal patterns of UTAs, studies have focused primarily on fluctuations in the quantity or rate of traffic accidents severity which is injuries and fatalities, based on the temporal scales which are 24 hours in micro scale and yearly in macro scale [4] [5] [6]. High levels of temporal aggregation cannot clearly present differences in various variables, while disaggregated analysis requires more temporal information in details and has high variability.

Concerning the spatial patterns of UTAs, various studies have been done to analyze UTAs in different geographical environment and different area characteristics, such as specific roads [7] [8], intersections [9] or corridors [10]. Additionally, many studies have been done at micro and macro scale by different database within different geographical setting of study area ranging from the census tract/traffic analysis zone (TAZ) [11] to city [9], county [5] and state and national levels [12].

Another beneficial of GIS in UTAs is disaggregating data spatially. Before GIS become famous in world, highly aggregated data set used by many researchers. Highly aggregated UTA data which can be but not limited to number and rate of UTAs based on travel distance as an example, or disaggregate by demographic data, such as age, gender, accident type and severity, car category or roadway facility type (urban vs. rural, etc.), rather than spatial aspects [5] [8] [13]. With GIS becoming increasingly popular, it has become easier and less costly to spatially disaggregate data and conduct analyses at finer resolutions.

Researchers have also explored the disaggregation of traffic accident data at both time and spatial scales [14] [15] to identify temporal and spatial changes in traffic accident patterns, distributions or risks. GIS facilitates both spatial disaggregation and temporal disaggregation; its applications in traffic accident analysis are discussed in the next section.

**4. GIS Applications in Traffic Accident Analysis**

GISs can provide many more analytical capabilities than other mapping tools. For example, while a GIS is a database that contains geographic information, it is also an intelligent mapping system that can link features with other features and an information transformation tool that can create new geo-databases based on existing ones by applying analytic functions [16].
With stored spatial (geographic) data, GISs enable spatial display, spatial integration, spatial query, spatial analysis and processing, etc., almost all of which have applications in transportation studies, especially in traffic accident analysis [17].

4.1. Spatial Display

Traffic accidents can be shown at their locations of occurrence on digital maps. At least three methods are used to accurately locate traffic accidents on maps. The selection of the method depends on the recorded location data.

- The exact location of accident is important as an accident can display spatially only if the XY coordinator known.
- Geocoding addresses can come from the exact addresses only. This approach normally applies to the two format types of traffic accident addresses that were recorded on traffic accident reports. One is the exact address. The other format is an intersection of two streets. Both of these address formats can be recognized and added to the correct locations if the correct address locator manager is selected. However, this method requires the input of a reference layer, which stores all the geographic information and attributes of each road section.
- Analysis overtime can observe the exact location of UTA on road transportation network with qualification and characteristics of that specific location. This method requires a route layer, which includes linear features that store unique identifiers and measurement systems. The traffic accident data should be stored with an attribute indicating the route name and an attribute indicating the measurement. With the development of computer science and GIS techniques, numerous researchers have used GISs for displaying UTAs on a digital map and doing spatial analysis, GIS is the most appropriate technology.

4.2. Spatial Integration

GISs enhance the integration of data from different sources based on geographic locations. In traffic accident analyses, researchers have used GISs to link traffic accident data with traffic data (e.g., traffic volume and speed limit) [18], roadway inventory data (e.g., pavement, geometry, road condition and number of lanes) [18] environmental data (e.g., land use) [11], demographic data (e.g., employment and population) [19], socio-economic and other potential contributing factors, and particular locations of interest such as schools [20] to evaluate and investigate relationships between the occurrence of traffic accidents and contributing factors. It has also been argued that GIS applications are fundamental element in any spatial as well as temporal even if the main focus will not GIS [21].

4.3. Spatial Queries

The major advantages of spatial queries in GISs are that the results of a query
from the database can be viewed in a spatial format [22] and the results of a query from digital maps can be viewed in a tabular format. The linkage between a database format and a spatial format is one of the main characteristics of GIS data.

This capability facilitates contributing factor analysis. One can select traffic accidents based on variables and visually observe the spatial patterns of selected traffic accidents to discern geographic relationships. For example, by limiting traffic accidents which happened on weekdays and weekends from night to early morning (between 21:00 and 6:00) for drivers younger than 24 years old, certain types of spatial clusters can be found [23]. Alternatively, one can spatially select traffic accidents to determine whether these traffic accidents share common attributes. For example, Steenberghen, Dufays, Thomas and Flahaut [24] found in their regional study that UTAs in schools zones always occurred in a 1 km buffer zone from schools. The spatial query and tabular query can also be used simultaneously. For example, one can determine whether a traffic accident was caused by an earlier traffic accident by examining a space criterion (e.g., within 1600 m of an earlier traffic accident) by using a spatial query within the time (e.g., [25]).

4.4. Spatial Analysis

The spatial analysis of traffic accidents is a highly quantitative statistical process. In traffic accident analysis, cluster analysis is commonly conducted to find hot spots for traffic accidents by either a two-dimensional approach or a linear approach [14] [24] [26] [27] [28].

One essential problem with traffic accident cluster analysis is spatial autocorrelation, which has already been discussed by Black [14] and Black and Thomas [29]. Grid-based traffic accident cluster analysis was suggested and implemented by Choi and Park [30], and Steenberghen et al. [24].

Yamada and Till did a study on network autocorrelation analysis which approached to doing spatial analysis through the point data within a transportation network [31]. Statistical approaches for cluster analysis are also widely available in multiple software packages, including CrimStat III [32], SAS [33], Splus [34], and ArcGIS [35].

In the linear method for UTAs spatial analysis, in order to address safety, dividing roadways into micro or macro units based on the analysis methodology is a common method. There are different ways to define segments. One highway safety manual (HSM) attempts to make each segment “homogeneous with considering road qualification which include but not limited to road’s width, area land use, population, safety island” for two-lane highways [35] [36]. Furthermore, a new segment begins at intersections, 250 ft before and after the center of each intersection [36]. The sliding window method is another useful technique for grid sliding and cup specific data by the cells in to several and different segments [30].
5. Traffic Accident Prediction Models

Researchers over the past two decades have developed a variety of statistical methods to predict traffic accidents in different roadway sections and establish relationships between traffic accident between UTA qualifications which are severity for UTA, accident type, traffic condition, weather, road qualification, dominant land use category and driver behavior.

Models that have been widely used since the early days of traffic accident research include multivariate linear regression models and log-linear regression models. These two models assume that the random error term in the function is normally independent; therefore statistical results and outcomes for coefficients as well as variables can be easily obtained. Unfortunately, many practical situations arise in which the assumption of a normally distributed error term cannot hold. Traffic accident count data, binary responses and other continuous variables with positive and highly skewed distributions cannot be modeled with a normally distributed error term [37].

In order to allow fitting regression model characterized by or depending on only one random variable data, GLM model (Generalized Linear Model) was developed to allow fitting regression models for univariate response data that follow general distributions. This is referred to as the exponential family, which includes normal, binomial, Poisson, negative binomial, geometric, and gamma distributions, etc. [37]. GLMs typically consist of three different components: randomly, functional link and systematically [38]. GLM model is more beneficial comparing to other linear model as it provides flexibility in choosing a link function and a distribution of the error term [39]. For example, a log-linear link function and negative binomial error distribution can be chosen for traffic accident count data analysis. The formal structure of GLMs was summarized by Myers et al. [40]. GLM function shows in Equations (2.1) and (2.2) for statistical models of safety. It is convenient to use a software program to fit the data and estimate coefficients.

\[
y_i = \beta_0 \times F_i \exp \left( \sum_{j=1}^{k} \beta_j x_{ij} \right)
\]

\[
\ln y_i = \ln \beta_0 + \beta_1 \ln F_i + \sum_{j=1}^{k} \beta_j x_{ij}
\]

where \( y_i \) = the outcome or response variable, traffic accident count per unit of time;
\( F_i \) = traffic flow;
\( \beta_0 \) = intercept;
\( \beta_1 \) = coefficient for traffic flow;
\( \beta_j \) = unknown coefficients;
\( x_{ij} \) = covariates or explanatory variables.

In 2003, Miaou [13] did a study and expressed that some regression models are not appropriate for not only UTA analysis but also other events on road transportation network. When the mean and variance of the traffic accident fre-
quences were approximately equal, Poisson regression was found to be a more appropriate model for examining the relationships between traffic accidents [13]. Overdispersion occurs when the observed variance in the data is larger than the predicted variance. When over dispersion moderate or high, the use of negative binomial regression has been found to be more appropriate [13].

Finding out the meaningful relation between the rate of UTA and number of urban facilities has been of interest to researchers. A linear relationship \( (β = 1) \) between the number of traffic accidents and AADT was supported by Chipman [41], Gårder [42], Hauer [43], Janssens [44] and Persaud and Džbik [45]. This finding indicates that the individual probability of being involved in a traffic accident increases linearly as exposure increases. A result found by Lord in 2002 [46] expressed that \( β = 1.206 > 1 \) for accident’s severity in highways with more than 3 lanes. However, most researchers agree that the pattern of UTAs regarding the traffic volume does not follow linear relationship \( (β < 1) \). This means by increasing traffic volume, driving speed decreases [47] [48]. The results expressed that UTAs risk factors and vulnerability decrease by increasing traffic volume as drivers have more alert, drive in low speed.

However, the result of a power coefficient of traffic flow less than 1 can cause problems in the optimization of network safety. Maher et al. [49] try to optimize safety and vehicular delay simultaneously on digital road network and found that the traffic flow tends to concentrate on a few roads rather than dispersing to many roads when the network is optimized solely for safety. Therefore, Lord [46] argued that is not possible to get suitable methodology for measuring traffic volume and accident by literature. It was suggested that the gamma function might be more appropriate for describing the relationship between the traffic accident count and traffic flow and that density might be a suitable measure to represent exposure in a traffic accident function.

5.1. Bayes Method

One of the famous scientific disciplines which widely has been used for UTA studies is the Bayesian Method. The Bayesian approach has been widely used in statistics and scientific disciplines over the past decade. One of the major advantages of the Bayesian approach is that it enables predictive forecasting of risks even for sparse data or rare events [50]. The Bayes Method is powerful and effective predicting method which can be used in many different fields.

The Bayes method for modeling and data analysis is shown in the following equation. When \( y \) represents the observed traffic accident frequency, what is the probability of occurrence of \( θ \) traffic accidents?

\[
p(θ \mid y) \propto \frac{p(θ)p(y \mid θ)}{p(y)} = \frac{p(θ)p(y \mid θ)}{\int p(θ)p(y \mid θ)dθ}
\]

where

\[ p(θ \mid y) = \text{posterior probability conditional on } y; \]
\[ p(\theta) \] = prior distribution (can be informative or non-informative);  
\[ p(y|\theta) \] = likelihood function;  
\[ p(y) \] = prior distribution of projective.

**EB Method or Empirical Bayes Method**

Empirical Bayes Method is a method with the ability of estimation of prior distribution form big and real data.

This approach challenges the fundamentals of the full Bayes method because the EB method uses the data twice. Different weights are assigned to standard estimation. This estimation is based on the each segment assumption which follow gamma distribution.

\[
\hat{\mu}_{\text{EB}_i} = \gamma_{ai} \hat{\mu}_{ai} + (1 - \gamma_{ai}) \hat{y}_i
\]  
where  
\[ \mu_{\text{EB}_i} = \text{EB posterior estimate for segment } i \text{ at time } t; \]  
\[ \hat{\mu}_{ai} = \text{expected traffic accident count of segment } i \text{ at time } t \text{ (the maximum likelihood estimate, the mean of } a, \text{ can be estimated from a reference site).} \]

**5.2. Full Bayes Method**

The full Bayes method is a computationally intensive process that is commonly implemented using computers. The common method for posterior distribution estimation now is simulation methods by Markov Chain Monte Carlo (MCMC).

This approach commonly uses multiple levels of analysis in an iterative way, so it is called the hierarchical Bayes model. The hierarchical model allows the modeler to make a meaningful relationship between study’s parameters with a logical prospective. The hierarchical procedures, whether from a full Bayes or an EB model, typically result in a smoothing of estimates for each unit towards the average outcome rate and have generally been shown to improve precision and predictive performance [51]. This thesis uses the hierarchical Bayesian approach to model relative traffic accident risks. A detailed modeling process is shown in Section 3.5.

In public health research, the main consideration of Hierarchical Bayed Model is creating new methodologies for disease mapping and disease prediction [52] [53] [54] [55] [56].

Some studies have been shown and discussed about the advantages and disadvantages of Hierarchical Bayed Model in comparison of previous and classical methods. Studies have shown that performing a risk estimation using a hierarchical Bayes model has several advantages over classical methods. The occurrence of disease is typically rare for an analysis sites, therefore, large variability exists across analysis units, especially for analysis units with small population sizes, making it challenging to differentiate chance variability and genuine difference in the estimates.

Hierarchical Bayes methods, in which proper spatial-correlated random effects are modeled, can account for the high-variance estimate in low population...
areas and retain overall spatial trends [55] [57].

Traffic accidents are similar to disease incidents in their rare occurrence and large variability across analysis units, so it is appropriate to adopt hierarchical Bayes models from disease mapping for modeling traffic accident analysis, risk assessment and traffic accident mapping. Many studies have been used the hierarchical Bayesian approach in order to estimate traffic accident performance in terms of safety [15] [58]. The urban traffic accidents hazard location map was also explored by Sun and Miaou [15] [58].

6. Identification of Hazardous Locations

Different and various methodologies have been used to investigation and identify hazard specific location or spatial location in other word.

A variety of methods have been used to identify hazardous locations, and this is a prerequisite for engineering studies intended to examine countermeasures.

6.1. Frequency Method

The World Road Association summarized the number of traffic accidents for each site and ranked them in descending order. Sites with high prearranged number of traffic accidents decided as areas with high-frequency classification. This method is commonly used to measure the safety for a spot location (hot spot identification). However, this method cannot apply in vehicle exposure counting (e.g., traffic volumes), which is directly related with traffic accident ratio. This method also suffers from a regression-to-the-mean bias in which an abnormally high count is likely to subsequently decrease even if no improvements are implemented.

6.2. Traffic Accident Rate Method

The traffic accident rate method ranks sites according to the ratio between the number of traffic accidents and vehicle exposure. Rates are given as traffic accidents/million, entering vehicles (EVs) for a specific area or location and traffic accidents by vehicle miles travelled (VMT) for segments. Sites which have higher rates than expected rate are classified as high-rate sites. The advantage of this method is that it includes vehicle exposure as the denominator. The importance of using traffic volume for normalization was emphasized by Affum and Taylor [59].

Another traffic accident rate method is the sliding window-based ranking approach. The traffic accident frequency, VMT and other explanatory covariate values are combined and calculated for each cell. Then, the traffic accident rate is calculated independently for each cell, and the cells are ranked by the rates [30]. Another problem with this method is that the observed rates have a large amount of uncertainty in low vehicle exposure (low EV, short length or low VMT) sites. Because of this high uncertainty, the rates in low exposure segments tend not to be useful since they can be extremely high or extremely low.
6.3. Critical Traffic Accident Rate Method, Combined Criteria Method

The quantitative analysis of traffic accidents and rate method identifies hot spots where the traffic accident rate is higher than the average traffic accident rate for same locations or areas as well as regions in national scale.

Another application of this method is significance determination of each site’s traffic accident rate relative to the mean traffic accident rate of same areas. This method suffers from the same small area estimation problem as the traffic accident rate method.

Geography and UTAs are unquestionably linked. Whether the problems of concern are related to low quality of transportation network and safety, environmental characteristics, or resource accessibility, geography can apply to all traffic safety issues.

Sites are first ranked by one method, and the sites that receive high rankings are then investigated using another method. For investigation, different weights can also be assigned to different methods to select priority locations. This approach can avoid the limitations of using a single method.

7. Conclusions

A traffic accident is the event with multiple associated factors which include but are not limited to human act and behavior, urban structure, culture, etc.

The associate factors underlying traffic accidents are usually different and interdependent. GISs represent unique spatial methodologies which provide answers to questions about the complex causation of traffic accidents. Such systems can be effective for integrating and analyzing physical, social, and cultural environments. GIS technology offers many advantages related to data integration, interactive querying of databases and design, and presentation of findings in the form of maps. Both the visual impact and the data analysis provided by GISs are advantages that support their use. The ability to overlay data layers allows interpretation beyond that available when using traditional research and statistical methods.

Acknowledgements

Publication fee was provided by Tohoku Kozosha Co., Ltd.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References


Corporation, Seattle.


