A Literature Review on GIS Approach to Determine Traffic Accidents

Farzaneh Farshad

CarltonU Student Number: 101213808

uOttawa Student Number: 300124641

Traffic Engineering Term Paper

CIVE 5305

Table of Contents

Abstract
Traffic accidents are complicated data to be analyzed without a proper system. Geographic Information System (GIS) have made this complex task easier to manage. Many agencies, municipalities, organization, etc. have been using GIS in their transportation department to monitor, analyze, map and manage traffic accidents and assist them with decision making to lead to a safer transportation network and less accident rates. Location of accidents, land use of the area where accident happens, geographic information of the accident, environmental characteristic and socioeconomic condition are a chain of information that requires to be analyzed and linked to identify their relationship. ArcGIS is one of the most powerful software that is able to find a meaningful relationship between contributing features and analyze the result from spatial and temporal point of view. In fact, time and space are two important aspect of traffic events. Based on the accident location the contributing factors are different for each accident occurrence. This spatial disaggregation needs to be identified. GIS can investigate and assess any kind of complex spatial information and its relationship with other components such as traffic accidents which has both spatial and temporal aspect. Not to mention that one of the GIS advantages is that it can integrate, join and visualize various and unrelated databases such as road qualifications, land use, population and environmental factors as mentioned before. It is a powerful technology and approach to support researches and operational needs which ends up in safety improvement and proper allocation of the resources to improve transportation network safety.

Keywords
Geographical Information system, Traffic Accident, Spatial Analysis

1. Introduction

The purpose of this study is to review the literature of Geographical Information system application in urban traffic accident (TA) identification. Determining and analyzing TAs can be quite challenging from geospatial aspects as it depends on many parameters. However, it plays a remarkable role in decision making and safety. One of the major issues worth considering is the TA, after other disasters which affects human life negatively. In this context, different public institute and organizations including World Health Organization (WHO) have invested on studying how to decrease traffic accident rate.

Massive numbers of accidents which results in serious injuries is happening around the world which can be reduced by analyzing those events. Geographical Information system (GIS) and its application has been playing an essential role in analyzing TA patterns and structures which will be reviewed in this study. This research invests on the literature of the previous studies that have worked on geospatial aspects of traffic accident and GIS application in detecting TA and give detailed summary of analyses related to this issue which have been done by transportation engineers and researchers. Moreover, detection of locations that have a high frequency of accidents and models for predicting traffic accidents and Bayesian approach will be reviewed.

At the end of this study it is expected that vulnerable areas to traffic accidents become known to readers. In most articles which will be reviewed in this study, researchers tried to discuss about GIS application to locate the traffic accidents first. Then, they identified the black spots through spatial relationship between traffic accidents and road network elements and they’ll find the causes of the accidents to develop road safety programs.

2. GISs and Traffic Accidents

Application of GIS in analysing the traffic accidents is based on the fact that the traffic accidents have 2 perspectives including spatial and temporal aspect, so space and time are two important foundation in GIS analyses of TA. From spatial aspect, many criteria comprise road type (road curves, ramps and bridges), increasing population density, weather, culture, distribution of facilities may affect TAs which are very important to be understood. GIS comprises of various layers of information and digital maps that can be integrated, analyzed and monetarized in different desired combination and scales. It has been verified as one of the most functional, convenient and helpful tools for spatial analysis and mapping of different events in transportation studies. Not to mention that rise in population influenced the increased risk of transportation problems indirectly by arising the human error, failure and mistake (Budiawan & Purwanggono, 2018). However, the frequency of this issue can be taken under control and be decreased by proper spatial analysis. GIS can visualize TA spatial distribution through various kind of maps which is one of the most proper methods of spatial-temporal analysis. This study will review the literature of detecting TA accidents through GIS. It is obvious that TA is affected by many driving criteria ranging from geography to transportation network condition. Moreover, causes and driving factors of TA should be examined from planning, design and road structure aspect and then they should be formulated, since TA is a complicated issue which should be investigated from different aspects toward a better conception of accident patterns, road safety and decision making. TA can be analyzed and illustrated through single layers and multiple layers of spatial information with the help of GIS technology. GIS makes planning and management of transportation network easier and faster with regards to TA. Application of GIS can enhance the decision-making process easier for accident analysis. It is a proper platform to integrate the spatial relationship between events such as traffic accidents and other objects including socio-economic data, geographical information, transportation network quality and urban design features based on temporal scales. Combining spatial and geographic properties with transportation objects such as network intersections straight or curved links and network connections makes an efficient integration for decision making, planning and transportation management (Deloukas et al., 1997). The death number because of traffic accident around the world is stated around 1 million people in developing countries by WHO (Dereli & Erdogan, 2017). This problem can be fixed by identifying the high risk areas in terms of accidents using spatial analysis and statistical data (Erdogan et al., 2008). So, developing road safety programs should be supported for effective enhancement in the cities. Traffic accident studies are nowadays being fallowed up internationally and locally (Dereli & Erdogan, 2017). Dereli & Erdogan (2017), believe that these studies are aiming to identify concentrated areas in terms of accidents based on statistical models which should be defined quite accurate to decrease the accident. So, dangerous areas should first become known which are called black spots then, these spots should be prioritized, implemented, monitored, and assessed from their affects and costs (Dereli & Erdogan, 2017). The objective of using GIS in mentioned studies is to detect accident locations, classify accidents by their different attributes and map the history of accidents in different locations (Faghri & Raman, 1995). Chen (2012) explains that the location of the accident is usually described with text and it cannot be shown on the map unless it be coded by Geo-Coding methods to be placed on the exact positions on the map. Moreover, a lot of information can be gain from an accident which can be categorized into two main groups: Traffic loss information and other key information (Chen, 2012). Traffic loss information includes the number of deaths, injuries, missing persons, economic loss etc. and the key information contain spatial data like the coordinates of the accident and accident number, time of the accident and district number which are very important for safety analysis (Chen, 2012). To go through some examples in that respect GIS can help designers who are trying to understand the "geometric design conditions of a road section" or planners who are trying to find "the coordinates of an accident site in order to plan for future developments such as the addition of a lane" (Faghri & Raman, 1995). GIS presents accident related spatial information for a transportation network and that is why GIS has been widely applied to urban traffic information management.

3. Spatiotemporal Analysis of Traffic Accidents

Causes of traffic accidents are being identified by researchers analyzing of where accidents happen (spatial identification) and date and time of the crash (temporal information). These spatiotemporal analyses of the accidents usually show non-random patterns which aim hot-spots detection and ends up in a better prevention of accident occurrence by characterisation of the causes and consequences of events. Also, spatiotemporal analysis helps testing the implemented safety improvements and traffic safety (Kaygisiz et al., 2015). Available databases of the accidents usually include the exact position and time of the crash and these information make spatiotemporal analysis of accidents available (Bíl et al., 2019). These analyses are progressed easily with the help of Geographic Information system (GIS). In other word GIS plays a vital role in traffic accident analysis (Steenberghen et al., 2004). Spatial and temporal aspects of accidents are usually considered separately and only a few studies has been investigated on considering both aspects together at the same time (Bíl et al., 2019).

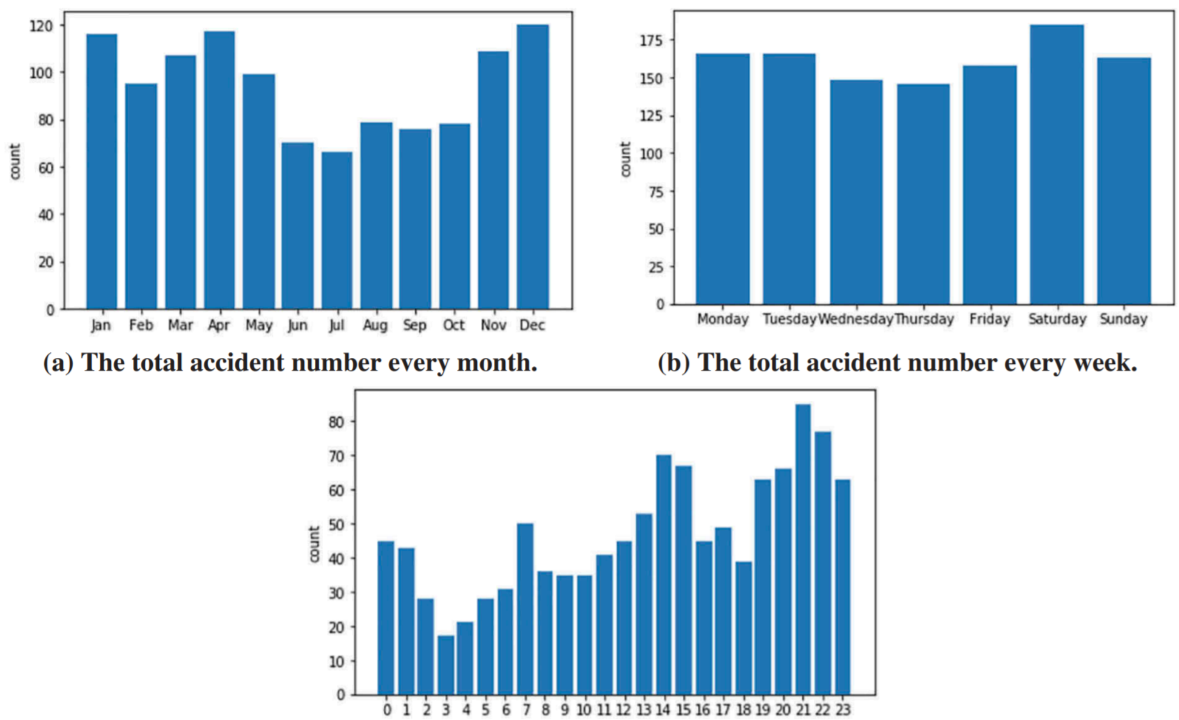
3.1. Spatial Analysis

"Spatial distribution" of the events are defined from traffic databases which include exact (GPS) location of the accidents (Gundogdu, 2010). Regarding Spatial distribution of accidents, number of studies has been developed which point to the possible methods of analysis including Kernel Density Estimation (KDE) and Nearest Neighbourhood Distance (NND) which can be done through ArcGIS software.

Using spatial data GIS can provide spatial integration, spatial analysis, spatial visualization and many other spatial functions (Liu & Zhu, 2004) specially in traffic analysis and urban planning. The location of the traffic accidents can be visualized on digital maps in ArcGIS. There are various ways of traffic accident visualization on maps which depends on the collected location data through GPS including: showing by the latitude and longitude coordinates of the accident occurrence, geocoding the addresses that were collected from the accident reports on a reference layer, analyzing characteristics of a specific location on the transportation network over time and then make a query from the attribute table which includes the names of roads and other measurements (Aghasi, 2019). These features made GIS the most practical tool in traffic accidents analysis. Moreover, GIS can integrate datasets based on their geospatial source (coordinates) and make it easier to solve the problem and give a better understanding of what is under consideration which may otherwise not be easily achievable by other methods. Macroscopic parameters in traffic accident analysis such as traffic speed, traffic volume, density, spot speed, space mean speed can be linked with each other and environmental data and socio-economic elements through GIS to illustrate the relationship of traffic accidents and mentioned features (Ng et al., 2002). Furthermore, the queries' results can be shown in spatial forms and models which is one the GIS capabilities to enable users observe the spatial patterns of a query (specific traffic variable). For example, a certain type of traffic cluster can be shown on a map by requesting the traffic accidents that happened on a certain days and time such as weekend mornings from 12 to 1 AM for young drivers on a specific highway. Also, traffic accidents which has common features can be selected by queries.

3.2. Temporal Analysis

Temporal analysis of accidents is very beneficial in giving a long-term evaluation of total accidents' trends in a region(Bíl et al., 2019). It is very important to figure out the precise accident patterns. A frequency distribution of the accidents would link the time of the accident and other effecting factors such as the traffic flow, weather condition, etc. Accident time series, graphs and spider plots are some of the visualization approaches of temporal analysis including year, month, date and hours of accidents. Temporal analysis can be done based on various scales like daily, weekly, monthly or even seasonal based on the purpose of the study. For example, Rodríguez-Morales et al., (2013) conducted seasonal analysis because the collision of vehicles with animals were being discussed in that research and some animals have different life cycle in different seasons.



(a) The total accident number every month. **(b) The total accident number every week.**

(c) The total accident number every hour.

Figure 1. Example of temporal analyses (Le et al., 2020)

3.3. Hotspots

One of the main results of spatiotemporal analysis is to identify hot spots or vulnerable areas in urban structures to the accidents which can be done through GIS which ends up in safety improvement and proper allocation of the resources to improve network safety. A hot-spots' spatiotemporal analysis will visualize the hazardous segment of traffic network which can affect decision making of the drivers and behavioral elements of their driving (Kaygisiz et al., 2015). These hot-spots can differ from time to time due to the taken safety prevention or their "physical environment" which can be detected by "Hotspot Detection Methods (HDM) in a spatio-temporal context" (Kaygisiz et al., 2015). Hot spots can be extracted by clustering methods and also through urban structure, land use and population. They can also be identified through measurments including accident rates, accident frequencies, and accident severity methods (Dereli & Erdogan, 2017).

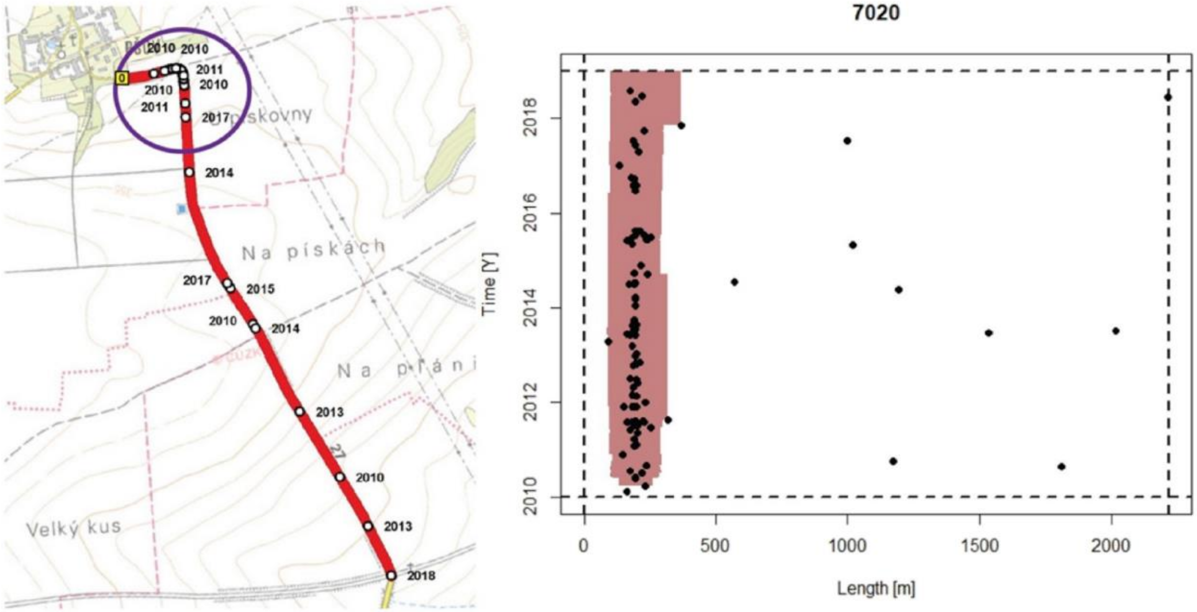


Figure 2. An example of a dangerous segment which presents a stable hotspot on a road (Bíl et al., 2019)

4. Traffic Accident Analysis with GIS

Geographic Information System (GIS) has made it easier to detect the traffic accident hotspots in transportation networks. Integration of the GIS with the Global Positioning System (GPS) added more capabilities and advances to traffic accident analyses (ZahranEl-Said et al., 2019). This combination has more potential than other mapping tools. GIS is an intelligent analysis tool that is usually applied to deal with multicriteria issues and combine multiple maps and layers to create an integrated output. Using GIS which has always been an integrated tool that could analyze the accessibility plans, enabled the researcher to take advantage of analyzing collected data concerning their importance and allowed carrying out complex assessment of the situation and provide network analysis and hotspots and creates a basis for applying more exact and scientifically influencing features, tools and methods for future application and studies.

5. Traffic Accident Hotspots Analysis methods

Various statistical methods have been developed to extract the traffic accident hotspots in different segments of a traffic network including Cluster Analysis which group similar items or dissimal items together "to improve the accuracy of the estimates from mathematical events" (Ng et al., 2002), Regression Analysis which shows the relationship between numbers of accidents and the causes features, General Linear model, Kernel Density, Bayesian Approach which can assess the risk of accidents, etc. Multiple linear regression, Poisson regression and negative binomial regression models can be applied to develop accident event models that shows the relationship of accident and contributing elements (Ng et al., 2002). Because the accident events are random by nature, the estimation of accident events can be modeled by mentioned distributions (Ng et al., 2002). Poisson and negative binomial regression models perform better than the regression models (Abdel-Aty & Radwan, 2000). Generally, these algorithms help researchers within their main purposes of establishing an accident event estimation model, identifying factors that have significant effect on accidents among the potential contributing factors, and identifying high accident risk areas. Other statistical techniques are commonly adopted for secondary objectives such as improving the accuracy of the estimations and estimating the confidence of the calculations.

5.1. Establishing Estimation Model

There are many approaches available to develop models and identify black spots including Multiple Linear Regression, Poisson regression and Negative Binomial method. However, Abdel-Aty & Radwan, (2000) stated in his research from Jovanis and Chang (1986) that multiple linear regression is not appropriate to be applied on all accident analysis since it gives undesirable statistical properties. Dereli & Erdogan, (2017) applied the "Poisson regression, Negative Binomial regression, and Empirical Bayes" methods together to identify the relationship between features that has influence on accidents and the number of events occurrences and Maximum likelihood was used to estimate the coefficients of the models. Then, they "developed an integrated GIS supported model" based on the obtained results. Poisson regression and Negative Binomial regression are more accurate when it comes to predict an accident and develop an accident pattern for shorter road segments (Dereli & Erdogan, 2017). It is advised in many researches that it would be the best to use Poisson regression model and Negative Binomial model together rather than pure Poisson because a certain amount of over dispersion must always be expected in traffic accident analysis (Abdel-Aty & Radwan, 2000). Erdogan et al., (2008) applied the Poisson method with kernel density analysis to find the accident clusters. Lord & Mannering, (2010) summarized existing models for analyzing accident-frequency data and their benefits and drawbacks as below:

Model type	Advantages	Disadvantages
Poisson	Most basic model, easy to estimate	Cannot handle over- and under-dispersion; Negatively influenced by the low sample-mean and small sample size bias
Negative binomial/ Poisson-gamma	Easy to estimate can account for over-dispersion	Cannot handle under dispersion; can be adversely influenced by the low sample-mean and small sample size bias
Poisson-lognormal	More flexible than the Poisson-gamma to handle overdispersion	Cannot handle under dispersion; can be adversely influenced by the low sample-mean and small sample size bias (less than the Poisson-gamma), cannot estimate a varying dispersion parameter

Zero-inflated Poisson and negative binomial	Handles datasets that have a large number of zero-crash observations	Can create theoretical inconsistencies; zero-inflated negative binomial can be adversely influenced by the low sample-mean and small sample size bias
Conway–Maxwell–Poisson	Can handle under- and over-dispersion or combination of both using a variable dispersion (scaling) parameter	Could be negatively influenced by the low sample-mean and small sample size bias; no multivariate extensions available to date
Gamma	Can handle under-dispersed data	Dual-state model with one state having a long-term mean equal to zero
Generalized estimating equation	Can handle temporal correlation	May need to determine or evaluate the type of temporal correlation a priori; results sensitive to missing values
Generalized additive	More flexible than the traditional generalized estimating equation models; allows non-linear variable interactions	Relatively complex to implement; may not be easily transferable to other datasets
Random effects	Handles temporal and spatial correlation	May not be easily transferable to other datasets; Negative multinomial; Can account for over-dispersion and serial correlation; panel count data Cannot handle under-dispersion; can be adversely influenced by the low sample-mean and small sample size bias
Random-parameters	More flexible than the traditional fixed parameter models in accounting for unobserved heterogeneity	Complex estimation process; may not be easily transferable to other datasets
Bivariate/multivariate	Can model different crash types simultaneously; more flexible functional form than the generalized estimating equation models (can use non-linear functions)	Complex estimation process; requires formulation of correlation matrix

Finite mixture/Markov switching	Can be used for analyzing sources of dispersion in the data	Complex estimation process; may not be easily transferable to other datasets
Duration	By considering the time between crashes (as opposed to crash frequency directly), allows for a very in-depth analysis of data and duration effects	Requires more detailed data than traditional crash-frequency models; time-varying explanatory variables are difficult to handle
Hierarchical/multilevel	Can handle temporal, spatial and other correlations among groups of observations	May not be easily transferable to other datasets; correlation results can be difficult to interpret
Neural network, Bayesian neural network, and support vector machine	Non-parametric approach does not require an assumption about distribution of data; flexible functional form; usually provides better statistical fit than traditional parametric models	Complex estimation process; may not be transferable to other datasets; work as black boxes; may not have interpretable parameters

Table 1. Existing models to analyze accident-frequency (Lord & Mannering, 2010)

Lord & Mannering, (2010) also summarized studies based on the methodology that they applied on accident frequency as it is shown in appendix A. based on the literature review Poisson regression and Negative Binomial regression are more appropriate approaches for accident modeling which are going to be discussed in this review.

5.1.1. Poisson regression

The basic assessment of the accident data and its distribution is preferred to be identified by Poisson regression. Poisson is suitable to be applied when the mean and variance of the accident frequencies are approximately equal to model the relationship between events and affecting features (Miaou, 1994). Dereli & Erdogan, (2017) explained that "If the number of accidents occurring in a certain period of time in the i^{th}point is expressed as Yi, then Poisson distribution is equal to":

$$Y_i \mid \mu_i \sim poisson(\mu_i)$$

If there are n regions of studies, the $Y_i = (Y_1, Y_2, \ldots\ldots, Y_n)$ shows the number of them and $\mu_i = (\mu_1, \mu_2, \ldots\ldots, \mu_n)$ shows the average of the accidents, assuming $E(Y_i \mid \mu_i) = VAR\ (Y_i \mid \mu_i) = \mu$ (Dereli & Erdogan, 2017).

In the same article the "likelihood of i objects (segment, intersection, etc.) of Y_i accidents in a given time period is expressed as in Equation (1)" in Poisson regression model:

$$P_{(y_i)} = \frac{EXP\ (-\mu_i)\mu_i{}^{y_i}}{Yi!} \quad (1)$$

By evaluating μ_i , the Poisson estimated model was determined. The likelihood function is as equation (2):

$$L_{(\mu_i)} = \prod \frac{{\mu_i}^{y_i}\, e^{-\mu_i}}{Yi!} \quad (2)$$

The model parameters are evaluated based on maximizing the likelihood function in equation (3).

$$ln\, L_{(\mu_i)} = \sum_{i=1}^{n} [y_i \ln(\mu_i) - \mu_i \ln(y_i!)] \quad (3)$$

Lord & Mannering, (2010) believe that the Poisson models are not appropriate under-dispersion and over-dispersion conditions. It is one of the most suitable methods to analyse the numerical values such as traffic accidents when the variance in the data is equal to the mean and linear regression method is not appropriate to analyse traffic accidents (Dereli & Erdogan, 2017). GRETL and STATA are software that are able to perform Poisson regression through Generalized Linear Model solutions (Dereli & Erdogan, 2017).

5.1.2. Negative Binomial Regression

Unlike Poisson model the negative binomial regression model is particularly useful when data is overdispersion. When the predicted variance is smaller than the observed variance in the data, the overdispersion happens and if the overdispersion is moderate or high, the negative binomial regression is a proper choice in estimating models (Aghasi, 2019). It is a particular modified version of Poisson regression to deal with overdispersion. When the mean of data is equal to the variance in the data negative binomial regression model is used (Dereli & Erdogan, 2017). Dereli & Erdogan, (2017) discussed that the variance is defined as Variance= $\mu_i + k * {\mu_i}^2$. K is overdispersion factor here. Equation (4) shows the likelihood of i objects in negative binomial regression model where y_i is the number of road accidents:

$$P_{(y_i)} = \frac{\Gamma\left(y_i + \frac{1}{\alpha}\right)}{Yi!\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \left(\frac{1}{1+\alpha\mu_i}\right)^{1/\alpha} \quad y_i = 1, 2, \dots, n \quad (4)$$

Where $\mu_i = E(y_i) = V_i \exp\,(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_n X_{in}) \quad i = 1, 2, 3, \dots, n$

Variance of $(y_i$ is defined as in equation (5):

$$Var_{(y_i)} = \mu_i + \alpha{\mu_i}^2 \quad (5)$$

α is expressed as the over-dispersion parameter and "coefficients of$\beta_i = \beta_0, \beta_1, \beta_2, \dots, \beta_n$ is determined according to the most likelihood function in the Negative Binomial regression" (Dereli & Erdogan, 2017). Maximum likelihood is generally used to estimate the coefficients of the models. One of the most important "advantage of maximum likelihood estimation is that closed-form functions often exist for the most common distributions used" (Lord & Mannering, 2010). However, if the likelihood function is difficult to be identified, maximum likelihood estimation is impossible to be applied (Lord & Mannering, 2010). Likelihood function of negative binomial is shown in equation (6) below:

$$L_{(\mu_i,\alpha)} = \prod_{i=1}^{n} \frac{\Gamma\left(y_i + \frac{1}{\alpha}\right)}{Yi!\,\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \left(\frac{1}{1+\alpha\mu_i}\right)^{1/\alpha} \quad (6)$$

5.2. Identifying contributor factors to the event

Matching and linking accident data to the geospatial information, travel information, land use of the crash location and socioeconomic information through GIS helps researchers to identify the relationship between affecting factors and accident occurrence (Le et al., 2020). Various factors such as road design, weather condition, lighting factor affect the safety performance and crash occurrence, “either qualitatively or quantitatively” (Khan et al., 2008). Accident contributing factors differ from each other by changing the location of its occurrence which needs spatial disaggregation to be identified (Le et al., 2020). Affecting factors can be comprised of the dark environment, lack of good lighting, land use, etc. Studies show that land use is one of the main factors which have significant effects on the total number of crashes (Ng et al., 2002). To produce 2 dimensional visualization the width and color of the roadways can be adjusted in GIS (Le et al., 2020). To identify high risk locations, vehicle accidents spatial patterns should be stablished (Jones et al., 1996). These spatial patterns are shaped in light of various demographic characteristics. Without Geographical Information System, the relationships between spatial and non spatial phenomena in accident information system cannot be found (Erdogan et al., 2008). Researchers conducted several studies to stablish spatial patterns in vehicle accidents (Jones et al., 1996). Abdel-Aty & Radwan, (2000) applied an elasticity method to detect the most affecting features that influence the accident occurrence and evaluate their “relative significance”. This method can measure the “relative effect of the variable on accident frequency” (Abdel-Aty & Radwan, 2000). They used equation (7) to calculate the elasticity “where λ is the mean number of accidents and x are the explanatory variables”.

$$E_{(y)} = \frac{\partial \lambda}{\partial x} \frac{x}{\lambda} \quad (7)$$

5.3. Identify the high-risk areas

Various approaches have been applied to identify hazardous areas in terms of traffic accidents. Erdogan et al., (2008) used Poisson regression with kernel density analysis to find the hotspots and then GIS was used to identify high risk areas. The world Road Association used Frequency method to rank the high risk areas based on the high frequency of accidents to measure the safety of locations and finding hot-spots (Aghasi, 2019). The disadvantage of this method is that it has regression-to-the-mean bias and it cannot be applied in “vehicle exposure counting” which is one the traffic accident ratio dependencies (Aghasi, 2019). The other method worth mentioning is to rank accident locations based on the correlation between the number of traffic accidents and vehicle exposure (Aghasi, 2019). In this method accident locations with “higher rate than expected” are known as high risk sites (Aghasi, 2019). Where the traffic accident rate is higher than normal rate (average), those location are known as hot spots in quantitative analysis of traffic accidents (Aghasi, 2019).

5.3.1. Bayesian Approach

Several research and operational project has been done by applying the Bayesian approach recently (Li et al., 2007). Bayesian approach gives researchers the opportunity to predict risk of events even “for sparse data or rare events” (Withers, 2002). In many areas such as health, disease mapping, risk assessment and prediction Bayesian method is being used (Besag, 1974). Incorporating basic knowledge without limitation of “classical distribution assumption” is ne of the advantages of this method in various fields (Withers, 2002). Bayes method, Empirical Bayes method, Full Bayes method and Hierarchical Bayesian model are all being used to estimate accident risk ad predict accident frequency in many studies (Brüde & Larsson, 1988). Equation (7) shows Bayes method to model and analyze data where y is the observed traffic accident frequency:

$$P_{(\theta|y)} = \frac{P_\theta P_{(y|\theta)}}{P_y} = \frac{P_\theta P_{(y|\theta)}}{\int P_\theta \, P_{(y|\theta)} d\theta} \qquad (7)$$

$P_{(\theta|y)} =$ 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

The Empirical Bayes Method is a method which is able to estimate prior distribution form big and real data which uses the data twice and "Different weights are assigned to standard estimation"(Aghasi, 2019). Each segment assumption has its own gamma distribution which is shown in equation (8) (Aghasi, 2019).

$$\mu_{EBT} = \gamma_{it}\mu_{it} + (1 - \gamma_{it})y_{it} \qquad (8)$$

$\mu_{EBT} =$ EB posterior estimate for segment i at time t

$\mu_{it} =$ expected traffic accident count of segment i at time t

The full Bayes method is using computers because it is very intensive. The hierarchical model is an approach in which multiple levels of analysis in an iterative way is used and a meaningful relationship is make between factors (Aghasi, 2019). All of mentioned methods from Bayes method to Hierarchical Bayesian model lead to smooth estimation to improve performance. To estimate and analyze the accident and its risk GIS, regression analysis and Bayesian approaches were integrated recently (Ng et al., 2002). The Bayesian method is able to smooth the spatial relative risks in all segments (Li et al., 2007). Li et al., (2007) shows the relative risk distribution in Harris County between 2:00 and 3:00 a.m. on Saturdays in the figure 3,4 and 5. In the pictures below it is illustrated that relative risks were much higher than at other times. The same pattern can be seen on Sundays.

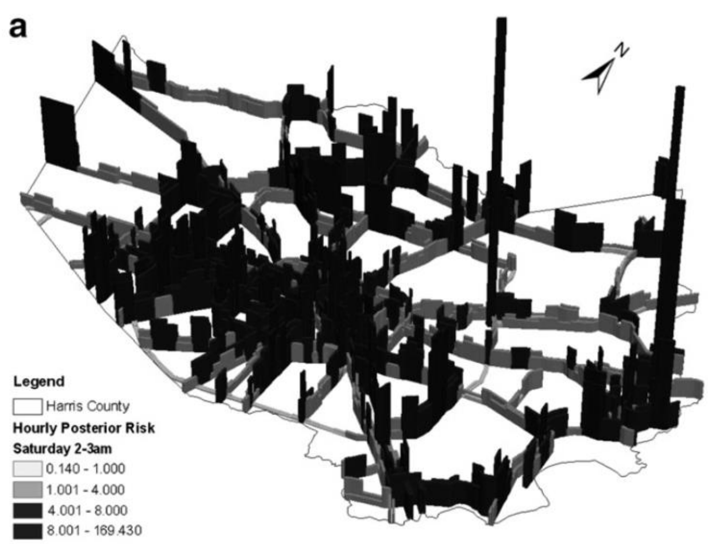


Figure 3. Posterior relative crash risk, 2:00–3:00 a.m., Saturdays (Li et al., 2007)

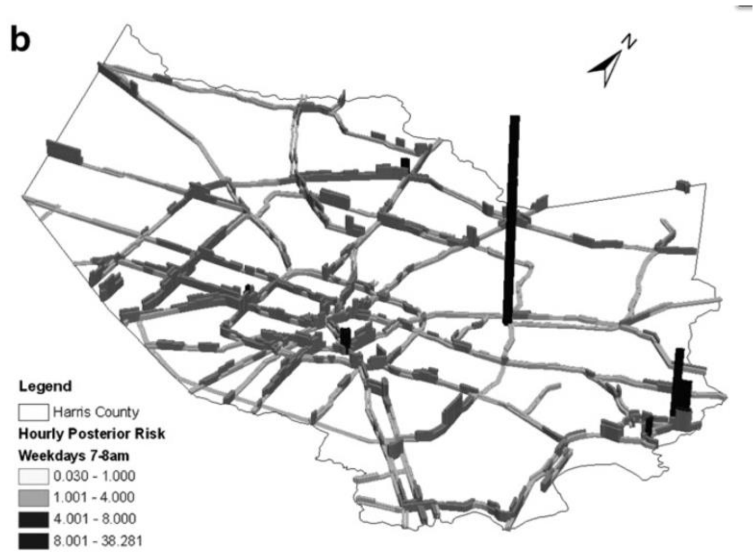


Figure 4.posteriorrelative crash risk, 7:00–8:00 a.m., weekdays (Li et al., 2007)

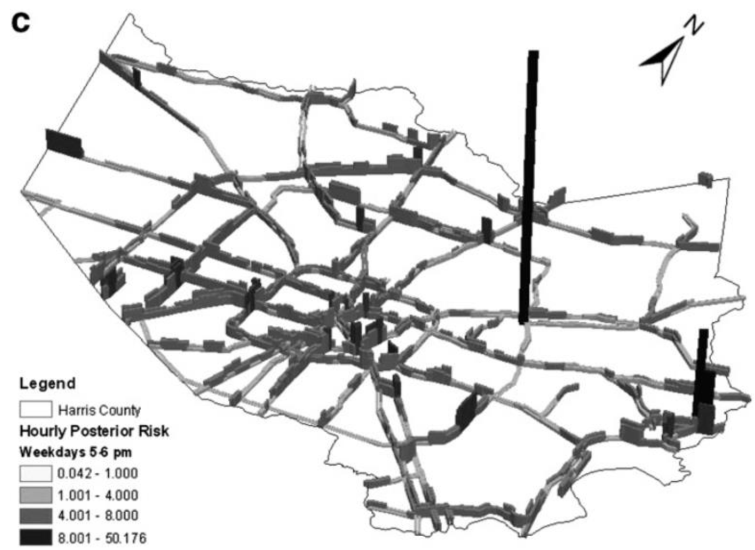


Figure 5. posterior relative crash risk, 5:00–6:00 p.m., weekdays (Li et al., 2007)

6. Summary and Conclusions

Traffic accidents are very complex events from analysis perspective because there are typically many contributing factors such as road design, urban structure, culture, environment, socio-economic factors, etc., affecting them spatially. These factors can be completely independent, and their relation should be identified by spatial and statistical methodologies which were discussed in this literature review. GIS offers integration, spatial analysis models and visualization to support researches and operational tasks on this issue by enabling users to overlay different spatial layers. GIS can provide spatial integration, spatial analysis, spatial visualization and many other spatial functions by using spatial data. In other words, GIS can integrate datasets based on their geospatial source and it make it easier to give a better understanding of what is happening by its mapping power which may otherwise not be easily achievable by other methods. One of the main gains of spatiotemporal analysis is to detect hot spots or risky areas in urban structures to the accidents which can be done through GIS which ends up in safety improvement and proper allocation of the resources to improve network safety. A hot-spots' spatiotemporal analysis will visualize the vulnerable segment of traffic network. GIS provides services to traffic accident reporters to be more precise about the causes, effects and statistical analysis about the events. Various statistical methods and models has been developed in the last decades including Poisson, Negative binomial/Poisson-gamma, Poisson-lognormal, Zero-inflated Poisson and negative binomial, Conway–Maxwell–Poisson, Gamma, Generalized estimating equation, Generalized additive, Random effects, Random-parameters, Bivariate/multivariate, Finite mixture/Markov switching, Duration, Hierarchical/multilevel, Neural network, Bayesian neural network, and support vector machine to analyse traffic accidents and predict the risk of crash occurrence in different segments of the road network and establish a meaningful relationship between the causing features. Among mentioned statistical approaches, Poisson and Negative binomial were discussed in detail in this study. Then identifying contributing factors to crashes basics were discussed including an elasticity method to detect the most affecting features that influence the accident occurrence and evaluate their relative significance which can measure the relative effect of the variable on accident frequency. The last part of this study was dedicated to review how researchers can identify the high-risk areas through various methodologies including Bayes method, Empirical Bayes method, Full Bayes method and Hierarchical Bayesian model. Geographical information system has functions to illustrate spatial evets in forms of maps. The technology and statistical models which were discussed in this literature review are available today in real world. A comprehensive method is required to equip the knowledge and information which were discussed in this study with an integrated efficient information system to meet the needs of agencies, organizations, municipalities and other stakeholders.

7. References

Abdel-Aty, M. A., & Radwan, A. E. (2000). Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention*, *32*(5), 633–642. https://doi.org/10.1016/S0001-4575(99)00094-9

Aghasi, N. H. M. (2019). Application of GIS for Urban Traffic Accidents: A Critical Review. *Journal of Geographic Information System*, *11*(01), 82–96. https://doi.org/10.4236/jgis.2019.111007

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*(2), 192–225. https://doi.org/10.1111/j.2517-6161.1974.tb00999.x

Bíl, M., Andrášik, R., & Sedoník, J. (2019). A detailed spatiotemporal analysis of traffic crash hotspots. *Applied Geography*, *107*, 82–90. https://doi.org/10.1016/j.apgeog.2019.04.008

Brüde, U., & Larsson, J. (1988). The use of prediction models for eliminating effects due to regression-to-the-mean in road accident data. *Accident Analysis & Prevention*, *20*(4), 299–310. https://doi.org/10.1016/0001-4575(88)90057-7

Budiawan, W., & Purwanggono, B. (2018). Clustering Analysis of Traffic Accident in Semarang City. *E3S Web of Conferences*, *73*, 12001. https://doi.org/10.1051/e3sconf/20187312001

Chen, H. (2012). Black Spot Determination of Traffic Accident Locations and Its Spatial Association Characteristic Analysis Based on GIS. *Journal of Geographic Information System*, *04*(06), 608–617. https://doi.org/10.4236/jgis.2012.46063

Deloukas, A., Kokkinos, I., Kiousis, G., & Zannou, D. (1997). GIS-Based Transportation Planning and Analysis: A Practical Implementation. *IFAC Proceedings Volumes*, *30*(8), 417–426. https://doi.org/10.1016/S1474-6670(17)43859-6

Dereli, M. A., & Erdogan, S. (2017). A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods. *Transportation Research Part A: Policy and Practice*, *103*, 106–117. https://doi.org/10.1016/j.tra.2017.05.031

Erdogan, S., Yilmaz, I., Baybura, T., & Gullu, M. (2008). Geographical information systems aided traffic accident analysis system case study: City of Afyonkarahisar. *Accident Analysis & Prevention*, *40*(1), 174–181. https://doi.org/10.1016/j.aap.2007.05.004

Faghri, A., & Raman, N. (1995). A GIS-Based traffic accident information system. *Journal of Advanced Transportation*, *29*(3), 321–334. https://doi.org/10.1002/atr.5670290305

Gundogdu, I. B. (2010). Applying linear analysis methods to GIS-supported procedures for preventing traffic accidents: Case study of Konya. *Safety Science*, *48*(6), 763–769. https://doi.org/10.1016/j.ssci.2010.02.016

Jones, A. P., Langford, I. H., & Bentham, G. (1996). The application of K-function analysis to the geographical distribution of road traffic accident outcomes in Norfolk, England. *Social Science & Medicine*, *42*(6), 879–885. https://doi.org/10.1016/0277-9536(95)00186-7

Kaygisiz, Ö., Düzgün, Ş., Yildiz, A., & Senbil, M. (2015). Spatio-temporal accident analysis for accident prevention in relation to behavioral factors in driving: The case of South Anatolian Motorway. *Transportation Research Part F: Traffic Psychology and Behaviour*, *33*, 128–140. https://doi.org/10.1016/j.trf.2015.07.002

Khan, G., Qin, X., & Noyce, D. A. (2008). Spatial Analysis of Weather Crash Patterns. *Journal of Transportation Engineering*, *134*(5), 191–202. https://doi.org/10.1061/(ASCE)0733-947X(2008)134:5(191)

Le, K. G., Liu, P., & Lin, L.-T. (2020). Determining the road traffic accident hotspots using GIS-based temporal-spatial statistical analytic techniques in Hanoi, Vietnam. *Geo-Spatial Information Science*, *23*(2), 153–164. https://doi.org/10.1080/10095020.2019.1683437

Li, L., Zhu, L., & Sui, D. Z. (2007). A GIS-based Bayesian approach for analyzing spatial–temporal patterns of intra-city motor vehicle crashes. *Journal of Transport Geography*, *15*(4), 274–285. https://doi.org/10.1016/j.jtrangeo.2006.08.005

Liu, S., & Zhu, X. (2004). Accessibility Analyst: An Integrated GIS Tool for Accessibility Analysis in Urban Transportation Planning. *Environment and Planning B: Planning and Design*, *31*(1), 105–124. https://doi.org/10.1068/b305

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, *44*(5), 291–305. https://doi.org/10.1016/j.tra.2010.02.001

Miaou, S.-P. (1994). The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis & Prevention*, *26*(4), 471–482. https://doi.org/10.1016/0001-4575(94)90038-8

Ng, K., Hung, W., & Wong, W. (2002). An algorithm for assessing the risk of traffic accident. *Journal of Safety Research*, *33*(3), 387–410. https://doi.org/10.1016/S0022-4375(02)00033-6

Rodríguez-Morales, B., Díaz-Varela, E. R., & Marey-Pérez, M. F. (2013). Spatiotemporal analysis of vehicle collisions involving wild boar and roe deer in NW Spain. *Accident Analysis & Prevention*, *60*, 121–133. https://doi.org/10.1016/j.aap.2013.07.032

Steenberghen, T., Dufays, T., Thomas, I., & Flahaut, B. (2004). Intra-urban location and clustering of road accidents using GIS: A Belgian example. *International Journal of Geographical Information Science*, *18*(2), 169–181. https://doi.org/10.1080/13658810310001629619

Withers, S. D. (2002). Quantitative methods: Bayesian inference, Bayesian thinking. *Progress in Human Geography*, *26*(4), 553–566. https://doi.org/10.1191/0309132502ph386pr

ZahranEl-Said, M. M., Soon Jiann, T., 'Asri Putra, N. A. 'Atiqah B. M., Tan, E. H. A., Yap, Y. H., & Rahman, E. K. A. (2019). Evaluation of various GIS-based methods for the analysis of road traffic accident hotspot. *MATEC Web of Conferences*, *258*, 03008. https://doi.org/10.1051/matecconf/201925803008

Appendix A

Summary of previous research analyzing crash-frequency data (Lord & Mannering, 2010)

Model Type	Previous Research
Poisson	Jovanis and Chang (1986), Joshua and Garber (1990), Jones et al. (1991), Miaou and Lum (1993), and Miaou (1994)
Negative binomial/ Poisson-gamma	Maycock and Hall (1984), Hauer et al. (1988), Brüde and Larsson (1993), Bonneson and McCoy (1993), Miaou (1994), Persaud (1994), Kumala (1995), Shankar et al. (1995), Poch and Mannering (1996), Maher and Summersgill (1996), Mountain et al. (1996), Milton and Mannering (1998), Brüde et al. (1998), Mountain et al. (1998), Karlaftis and Tarko (1998), Persaud and Nguyen, 1998, Turner and Nicholson (1998), Heydecker and Wu (2001), Carson and Mannering (2001), Miaou and Lord (2003), Amoros et al. (2003), Hirst et al. (2004), Abbas (2004), Lord et al. (2005a), El-Basyouny and Sayed (2006), Lord (2006), Kim and Washington (2006), Lord and Bonneson (2007), Lord et al. (in press), Malyshkina and Mannering (2010b), Daniels et al. (2010), and Cafiso et al. (in press)
Poisson-lognormal	Miaou et al. (2005), Lord and Miranda-Moreno (2008), and Aguero-Valverde and Jovanis (2008)
Zero-inflated Poisson and negative binomial	Miaou (1994), Shankar et al. (1997), Carson and Mannering (2001), Lee and Mannering (2002), Kumara and Chin (2003), Shankar et al. (2003), Qin et al., 2004, Lord et al. (2005b), Lord et al. (2007), and Malyshkina and Mannering (2010a)
Conway–Maxwell–Poisson	Lord et al. (2008), Sellers and Shmueli (in press) and Lord et al. (2010)
Gamma	Oh et al. (2006) and Daniels et al. (2010)
Generalized estimating equation	Lord and Persaud (2000), Lord et al. (2005a), Halekoh et al. (2006), Wang and Abdel-Aty (2006), and Lord and Mahlawat (2009)
Generalized additive	Xie and Zhang (2008) and Li et al. (2009)
Random effects	Johansson (1996), Shankar et al. (1998), Miaou and Lord (2003), Flahaut et al. (2003), MacNab (2004), Noland and Quddus (2004), Miaou et al. (2003), Miaou et al. (2005), Aguero-Valverde and Jovanis (2009), Li et al. (2008), Quddus (2008), Sittikariya and Shankar (2009), Wang et al. (2009) and Guo et al. (2010)
Random parameters	Ulfarsson and Shankar (2003), Hauer (2004), and Caliendo et al. (2007)
Bivariate/multivariate	Anastasopoulos and Mannering (2009) and El-Basyouny and Sayed (2009b)
Finite mixture/Markov switching	Miaou and Lord (2003), Miaou and Song (2005), N'Guessan and Langrand (2005a), N'Guessan and Langrand (2005b), Bijleveld (2005), Song et al. (2006), Ma and Kockelman

	(2006), Park and Lord (2007), N'Guessan et al. (2006), Bonneson and Pratt (2008), Geedipally and Lord (in press), Ma et al. (2008), Depaire et al. (2008), Ye et al. (2009), Aguero-Valverde and Jovanis (2009), El-Basyouny and Sayed (2009a), N'Guessan (2010), and Park et al. (in press)
Duration	Jovanis and Chang (1989), Chang and Jovanis (1990), Mannering (1993), and Chung (2010)
Hierarchical/multilevel	Jones and Jorgensen (2003) and Kim et al. (2007)
Neural network, Bayesian neural network, and support vector machine	Abdelwahab and Abdel-Aty (2002), Chang (2005), Riviere et al. (2006), Xie et al. (2007), and Li et al. (2008)