Annals of GIS

ISSN: 1947-5683 (Print) 1947-5691 (Online) Journal homepage: https://www.tandfonline.com/loi/tagi20

Traffic collisions in space: four decades of advancement in applied GIS

Shenjun Yao, Becky P.Y. Loo & Bruce Zi Yang

ANNALS OF GIS, 2016
VOL. 22, NO. 1, 1–14
http://dx.doi.org/10.1080/19475683.2015.1085440

Taylor & Francis
Taylor & Francis Group

Traffic collisions in space: four decades of advancement in applied GIS

Shenjun Yao, Becky P.Y. Loo and Bruce Zi Yang

Department of Geography, The University of Hong Kong, Hong Kong, China

ABSTRACT
Traffic collision analysis is essential for reducing traffic injuries. While most traditional approaches focus on the time dimension of traffic collisions, the recent past has witnessed a growing awareness of the spatial dimension in a geographical context. In this paper, 70 studies on the application of GIS to the spatial analysis of traffic collisions are reviewed. The purpose of the paper is to provide a systematic analysis of the major advancements in applied GIS for studying traffic collisions in space since the mid-1970s.

ARTICLE HISTORY
Received 6 June 2015
Accepted 10 August 2015

KEYWORDS
Traffic collision; GIS; spatial analysis; road safety

1. Introduction

The development of GIS has provided valuable tools for performing analysis in public health (Goodchild 2015; Kwan 2012; Neutens 2015). Traffic injury has now been widely acknowledged as a serious worldwide public health problem that requires urgent attention. According to the World Health Organization (WHO) (2013), road traffic injuries are the eighth leading cause of death globally. More than one million deaths and billions of dollars of loss were incurred every year. It was also forecasted that road traffic deaths would become the fifth leading cause of death by 2030 (WHO 2013).

Traffic collision analysis is essential for reducing traffic injuries. As there is no consensus on the best guidelines for the analysis of traffic collisions, researchers and practitioners have developed different approaches for analysing collision events. Before-and-after studies, as a major conventional method, have been widely adopted in collision analysis studies (Hauer 1997). With time series data, the method uses the safety performance of sites provided in the 'before' period to predict what they would be like statistically in the 'after' period. The predicted results are then compared with the actual performance in the 'after' period on the sites to see if there are improvements or not. While most traditional approaches, such as the before-and-after method, focus on the time dimension, recent years have witnessed a growing awareness of the spatial dimension of traffic collisions in a geographical context.

The first application of GIS to traffic collision analysis can date back to 1976, when Moellering (1976) used a computer-animated film to produce snapshots of traffic collisions. Since then, various GIS methods have been employed in analysing traffic collisions over the past four decades. They range from simple mapping and visualization functions to more advanced methods like spatial statistical models and the application of big data techniques in a multidisciplinary framework. In this paper, 70 previous studies on the application of GIS to the spatial analysis of traffic collisions will be reviewed. The purpose of the paper is to provide a systematic analysis of the major advancements in applied GIS for studying traffic collisions in space since the mid-1970s.

2. Methodology

Traffic collisions can be visualized and analysed using GIS with diverse geographical information analysis tools. In order to define the scope of this study, the GIS methods are first roughly categorized into four classes based on the contents. Some key words related to the methods and the primary purpose are selected to search for relevant previous studies for further analysis. This section will present the ways in which the publications are selected among the rich literature on road safety research and GIS applications.

2.1. GIS methods in road safety

Table 1 summarizes the methods of GIS widely used in road safety research. Mapping and visualization are two fundamental functions of GIS that allow researchers to locate traffic collisions on a map and visualize

CONTACT: Becky P.Y. Loo bpyloo@hku.hk

Table 1. Methods of GIS in road safety classified by primary purpose.

Primary purpose	Methods in GIS
Visualizing, mapping and identifying topological relationship	Basic functions of GIS (e.g. geo-coding; buffer; overlay visualization)
Identifying clustering patterns of traffic collisions (global or local pattern)	Exploratory spatial statistical (global or local) methods (e.g. K-function; Moran's *I* statistic)
Analysing effects of contributory variables	Spatial statistical regression models (e.g. spatial lag regression; GWR)
Handling large databases in a multidisciplinary framework	Various (e.g. 3D GIS; data mining; tracking; big data)

Note: Key words used for selecting representative studies for further analysis are underlined.

the spatial distribution of collision events. Topological tools, such as buffer and map overlay, are another set of key features of GIS, whereby safety analysts can identify links between collision cases and the neighbourhood environment. In addition to basic functions, exploratory data analysis techniques as well as spatial statistical regression models have also been well developed in GIS. This enables road safety analysts to identify spatial patterns of interest and possible connections to the environment with various advanced statistical approaches. While the former directly addresses an important issue in road safety, that is, the identification of clustering patterns of traffic collisions; the latter allows the safety practitioners and researchers to better understand the effects of potential contributory factors and to recommend appropriate countermeasures. Moreover, dealing with road safety problems in safety improvement programmes requires multidisciplinary efforts in exploring large databases from different sources. The rapid development of GIS and other information technology, such as 3D GIS and big data, has significantly facilitated research collaboration among geographers and people from other professions, such as medical doctors and leaders of non-governmental organizations, with different disciplinary backgrounds.

2.2. Selection of publications in four decades

In addition to the knowledge of the research team on representative studies, some key words related to the primary purpose and methods of GIS in road safety research are chosen to search for relevant publications in ISI Web of Science (these words are underlined in Table 1). As a detailed study of the contents and methods will be made, the number of papers reviewed is limited to 70. Only papers cited most or published in important journals of geography and road safety fields are selected for further analysis. Table 2 lists the numbers of selected publications by primary purpose and time period.

In this section, the major changes over time are summarized. A more detailed analysis and discussion of the contents of these papers will be made in the next few sections. In this way, methodological advances under each primary purpose can be seen more clearly. Since 1976, visualization was employed to display the spatial distribution of traffic collisions. In the 1986–1995 period, many more other basic GIS functions such as geo-coding and topological tools were applied in road safety studies. Most of them, however, were still limited to the visualization, such as displaying analysis outputs and the use of topological tools to delimit the surrounding environment of traffic collisions.

With the fast development of spatial statistical methods, many more collision cases have been analysed with spatial statistical approaches in a GIS environment in the 1995–2005 period with the aim of identifying clustering patterns of traffic collisions. The spatial statistics used for pattern identification can be divided into *global* and *local* types. The former describes whether an overall configuration differs from a random pattern and the latter tells us where the unusual interactions of collisions are. As the latter is closely related with a fundamental problem in road safety, that is, the identification of hazardous road locations (hot spots or hot zones), it has attracted great attention among researchers in both road safety and geography fields during the recent

Table 2. Number of selected publications by primary purpose and time period.

	Visualizing, mapping and identifying topological relationship	Identifying clustering patterns of traffic collisions		Analysing effects of contributory factors	Handling integrated databases in a multidisciplinary framework
		Global	Local		
1976–1985	1	–	–	–	–
1986–1995	3	2	–	1	1
1996–2005	1	3	6	6	2
2006- now	7	–	18	12	7

two decades. As indicated in Table 2, almost all publications in the past 10 years belong to the latter type.

While knowing *where* means the detection of local clusters (Xie and Yan 2008; Yamada and Thill 2010; Loo 2009), knowing *why* is more related with another important issue on road safety – the influence of contributory variables. With spatial modelling techniques, contributory variables can be identified and analysed by taking spatial autocorrelation or spatial heterogeneity into consideration. For instance, as early as 1995, Levine, Kim, and Nitz (1995b) identified population, employment and road characteristics in the surrounding environment as significant explanatory variables by introducing spatial autocorrelation. Thanks to the development of statistical regression models, more and more road safety researchers have been able to analyse the influence of potential contributory factors by accounting for 'spatial' effects in the recent two decades (see Table 2).

Although professionals from different disciplines, such as engineering and medicine, have always realized the importance of collaboration for the improvement of road safety, the number of publications involving the use of applied GIS is limited and most studies before 2006 only focused on the comparison of collision cases recorded in the police and hospital reports (Aptel et al. 1999; Maas and Harris 1984). In the new century, researchers from a broad range of disciplines, such as geography, psychology and other social science disciplines, joined hands in developing and evaluating road safety improvement programmes. In particular, the past 10 years have witnessed a widespread adoption of the Internet, mobile phones and location-aware technologies (Shaw 2010). The change in people's life style, together with advanced information techniques for big data, has pushed the world into a new era. This allows researchers to tackle road safety problems by integrating traffic collision, human activity, transport system and traffic data through GIS in a 3D space–time context (Yao, Loo, and Lam 2015). Hence, a growing number of publications on road safety using multidisciplinary approaches and enhanced GIS databases appeared.

The following four sections will discuss the advancement of applied GIS in traffic collision analysis by primary purpose (Table 2).To reiterate, it is hoped that the methodological advances under each primary purpose can be seen more clearly.

3. From the beginning: visualizing, mapping and identifying topological relationship

Since the first application by Moellering (1976), the basic functions of GIS have been comprehensively employed in road safety research. Table 3 lists some publications related to basic GIS functions for visualizing, mapping and identifying the topological relationship.

Table 3. Selected publications by basic GIS function.

	Publication	Brief description
Visualizing	Moellering 1976	Used a computer-animated film
	Miller 1995	Used a 'sliding scale'
	Plug, Xia, and Caulfield 2011	Used the comap approach for spatial and temporal visualization
	Lam, Loo, and Yao 2013	
Mapping	Austin 1995	Used buffer zones to match the collision records
	Levine, Kim, and Nitz 1995a	Snapped collisions to the nearest intersections
	Kam 2003	Used three ways to geo-code collisions
	Loo 2006	Developed an automatic collision mapping process for geo-coding/ geo-validation
	Zhang et al. 2010	
	Qin et al. 2013	
	Deepika and Saradha 2014	
Topological tools	Austin 1995	Created buffers to identify topological relationship
	Loo and Tsui 2010	Created buffers to identify underlying patterns

Visualization is a fundamental function of GIS, which has been fully applied since 1976. In a study by Miller (1995) in North Carolina, a 'sliding scale' was used, whereby a segment of a specific length along a road was dynamically moved until that segment met a threshold. Focusing on both spatial and temporal dimensions, spatio-temporal visualization approaches are applied in mapping collision patterns (Plug, Xia, and Caulfield 2011; Lam, Loo and Yao 2013). For instance, Lam, Loo and Yao (2013) employed the comap approach to display the spatial and temporal patterns of traffic collisions involving elderly pedestrians.

In the early period from 1986 to 1995, scholars started to snap traffic collisions onto the transport system by abstraction and simplifications. Austin (1995) chose 24-m buffer zones from road centrelines to match the corresponding variable of the collision records, and the validity varied from over 90% to below 80%. Levine, Kim, and Nitz (1995a) believed that road intersections are key features, so they snapped the collisions to the nearest intersections. Kam (2003) concluded that there are three alternatives that are for geo-coding road collisions, and the choice can be based on different degrees of precise location information available. Later, with the development of algorithms and GIS components, researchers started trying to use computer programs to facilitate the geo-coding process. Loo (2006) proposed a methodology that validates precise collision locations within a linknode system and developed a programme not only to identify, but also to correct miscoded spatial variables in the police records. The results show that the program can validate over 99% of the collisions in recent years. Zhang et al. (2010) developed a program that automatically geo-codes traffic records. They used a hybrid approach to work with the segment-type geo-coding in contrast to traditional geo-coding that converts the addresses to *XY* coordinates. Qin et al. (2013) also developed an automatic collision mapping process that associates the collision records to underlying business data. The methods proposed by these studies significantly improve the efficiency and accuracy of geo-coding and validating procedures.

Since the mid-1990s, topological analysis functions of GIS, such as buffer and map overlay, have been widely used in road collision analysis for identifying topological relationships among data. Austin (1995), for example, generated buffer zones along the road network to assign traffic collisions to specific road sections. In addition, topological functions can help identify underlying spatial patterns. In a project related to bicycle collisions in Hong Kong (Loo and Tsui 2010), the researchers identified the bicycle collision pattern by buffering bicycle tracks to some distance and using point-in-polygon operations to determine how many cases of collisions had occurred in the relevant buffer areas.

4. Beyond mapping: identifying clustering patterns of traffic collisions

The identification of clustering patterns of traffic collisions can be traced back to the early 1990s. Since then, substantial research efforts have been made on the application of exploratory spatial statistical technique to pattern detection. Two categories of techniques are commonly used (Yamada and Thill 2007). One is the event-based approach, which analyses the traffic collision events that are recorded as individual points with *x* and *y* coordinates in space. The other is the link-attribute approach, which was developed on the basis of linear basic spatial units (BSUs) and traffic collision counts or risks as attributes attached to the BSUs. While early applications targeted at describing the global clustering pattern of traffic collisions, the studies in the recent two decades focused on identifying the local clusters, that is, hazardous road locations. Table 4 lists major publications related to the identification of spatial clustering patterns of traffic collisions.

Among the different approaches that have been employed in the identification of global clustering tendency of traffic collisions, the nearest neighbour distance was one of most widely used tools in the literature (Nicholson 1999; Levine, Kim, and Nitz 1995a). For instance, Levine, Kim, and Nitz (1995a) calculated the ratio of the observed to the expected nearest neighbour distance (the Nearest Neighbour Index) and found a significant degree of concentration of traffic collisions in Honolulu. Other common methods include the Quadrat methods (Nicholson 1999), Ripley's K-function (Yamada and Thill 2004; Nicholson 1999) and Moran's *I* (Black and Thomas 1998; Loveday 1991). While these global measures help to understand the system-wide variation in

Table 4. Selected publications by type of spatial statistic.

		Problem	Brief description (type of spatial statistic)
Global	Event-based	Levine, Kim, and Nitz 1995a	Nearest Neighbour Index
		Nicholson 1999	Quadrat methods; Ripley's K-function
		Yamada and Thill 2004	Ripley's K-function
	Link-attribute	Loveday 1991 Black and Thomas 1998	Moran's *I*
Local	Event-based (hot spots or hot zones)	Steenberghen, Dufays, and Flahaut 2004 Sabel et al. 2005 Erdogan et al. 2008 Anderson 2009	Planar KDE (hot spots)
		Xie and Yan 2008 Okabe, Satoh, and Sugihara 2009 Loo, Shenjun Yao, and Jianping 2011 Yu et al. 2014	Network-constrained KDE (hot spots)
		Yamada and Thill 2007	Local network-constrained K-function (hot spots)
		Steenberghen, Aerts, and Thomas 2010	Moving-segment approach (hot spots)
		Loo and Yao 2013	Local network-constrained K-function (hot zones)
	Link-attribute (hot zones)	Flahaut and Thomas 2002 Flahaut et al. 2003 Eckhardt, Flahaut, and Thomas 2004 Steenberghen, Dufays, and Flahaut 2004 Moons, Brijs, and Wets 2009a Truong and Somenahalli 2011 Yu et al. 2014 Yu et al. 2014 Young and Park 2014	Local Moran's *I*
		Yamada and Thill 2010	Local Moran's *I* and local Getis and Ord G
		Nie et al. 2015	Local Getis and Ord G
		Loo 2009 Loo and Yao 2013 Moons, Brijs, and Wets 2009b	Spatial indices for identifying 'high-high' spatial autocorrelation

traffic collisions, they fail to identify traffic collision clusters within the distribution. As the detection of local clusters have consistently played a key role in road safety research during the recent decade, this section will focus on the discussion of spatial statistical methods for identifying clusters of traffic collisions with GIS at the local level.

4.1. Event-based approaches

The event-based approach can be divided into two subtypes depending on whether the traffic collisions are treated as a network-constrained phenomenon or not (Yamada and Thill 2004). The standard methods assume that distance between events is measured as the Euclidean distance on a continuous planar space. As traffic collisions always happen on road network, some researchers argued that the standard (named 'planar' hereafter) methods may be problematic due to their planar space assumptions (Miller 1999; Yamada and Thill 2004). Hence, they extended the planar methods to a network space where the distance between events is usually measured as the shortest path along the road network.

Kernel density estimation (KDE) is a widely used event-based technique for the identification of traffic collision hot spots. It was developed for calculating a magnitude per unit area from point or line features using a kernel function to fit a smoothly tapered surface to each point or line (Silverman 1986). In the early stage, planar KDE was applied to the detection of spatial concentration of traffic collisions (Sabel et al. 2005; Anderson 2009; Erdogan et al. 2008). With planar KDE, the study area is divided into a grid with a user-defined cell size. A kernel function is used to compute the density of traffic collisions within a user-defined search bandwidth (also known as search radius). To overcome the two-dimensional assumption of planar KDE, Xie and Yan (2008) first proposed a network KDE method for identifying the collision hot spots. Okabe, Satoh, and Sugihara (2009) further discussed the methodological issues in the implementation of network KDE. A major limitation of both planar and network KDEs, as noted by Xie and Yan (2008), is that there is no formal statistical test for hot spots. Nonetheless, the KDE, especially the network-constrained version, has still been a commonly used approach for detecting the spatial concentration of traffic collisions. In addition to KDE, there are some other novel event-based methods for hot spot detection, including local network-constrained K-function (Yamada and Thill 2007) and the moving-segment approach (Steenberghen, Aerts, and Thomas 2010).

Although most studies employed the event-based methods to identify collision hot spots, some researchers aimed to detect hot zones, which consider possible spatial association of neighbouring hot spots as well (Flahaut and Thomas 2002; Loo 2009). An example is Loo and Yao (2013), who identified hot zones based on the local K-function method.

4.2. Link-attribute approaches

Flahaut et al. (2003) and Flahaut and Thomas (2002) are among pioneering studies that introduced local spatial autocorrelation techniques in the identification of hazardous road locations. They divided a study area into non-overlapping BSUs of unique length and counted the number of traffic collisions in each BSU. Using the contiguity (0–1) or distance matrix to define the spatial relationship between BSUs, local Moran's I (local indicator of spatial association, LISA) (Anselin 1995) can be computed to identify statistically significant contiguous road segments with high collision counts, which were first referred to as 'black zones' and later as 'hot zones' by Loo (2009), in accordance with 'hot spots' in road safety terminology. Unlike hot spots whose length is always fixed, the length of hot zones can vary depending on the spatial structure of the BSUs and the number of traffic collisions. Eckhardt, Flahaut, and Thomas (2004) examined the stability of the spatial pattern identified by local Moran's I over time in the same study area and confirmed the superior performance of the technique. Flahaut et al. (2003) and Steenberghen, Dufays, and Flahaut (2004) further compared the LISA with Kernel density methods, and the results indicate that both methods have pros and cons. The best choice relies on a good knowledge of collision locations and a proper understanding of the meaning of the selected parameters.

Although there is no consensus on whether the hot spot or the hot zone methodology is more appropriate for identifying hazardous road locations, hot zone identification based on a link-attribute framework has attracted much attention in the recent decade. In the studies by Loo (2009) and Loo and Yao (2013), an index which only measures the 'high-high' type of spatial autocorrelation was presented for the specific purpose of hot zone detection in road safety analysis. As the local autocorrelation requires the division of the road network into BSUs with equal length, they also introduced a dissolving algorithm to ensure more standardized BSUs. Moons, Brijs, and Wets (2009a) improved the local Moran's I using network-distance weights and a Monte Carlo simulation procedure. Yamada and Thill (2010) applied both local Moran's I statistic (Anselin 1995) and the local Getis and Ord G statistics (Getis and Ord 1992) to the hot zone detection of highway traffic collisions in Buffalo. Their experiments indicated the strengths of link-attribute methods in reflecting the effect of linear traffic features, which is in accordance with the previous findings of Steenberghen, Dufays, and Flahaut (2004).

4.3. Important issues in the context of road safety

From the perspective of safety improvement, hazardous road locations are usually defined as those that could benefit most from safety improvement programmes. However, most studies on identifying spatial patterns of traffic collisions only focus on the spatial aspect while neglecting some other important issues in the context of road safety listed below.

4.3.1. Measures of 'safety'

Using different traffic collision measures (e.g. slight versus fatal) may provide different information on road safety. To deal with this, some recent research efforts have been made to differentiate safety performances. In the work by Truong and Somenahalli (2011), the severity-weighted collision value was used as the attribute to compute the local Moran's *I*. Some studies even made use of a spatial statistic of the collision pattern as the attribute for hot zone identification (Moons, Brijs, and Wets 2009a; Xie and Yan 2013; Nie et al. 2015). A typical example is the integration of network KDE and local Moran's *I*, introduced by Xie and Yan (2013). It is known that a spatial process may be characterized by first-order and second-order properties. While the former describe the way in which the expected value of the process varies across space, the latter describe the correlation between values of the process at different regions in space (Gatrell et al. 1996). Xie and Yan (2013) argued that KDE belongs to the methods examining the first-order effects (region-wide trends), while Moran's *I* can be used for examining the second-order effects (correlation structures) of a spatial process. A combination of both methods may lead to more effective hazardous road locations. In addition, as the KDE technique has no formal statistical inference, they believed that applying a local statistical approach to density values could provide a useful mechanism for conducting rigorous statistical tests, but more sensitivity analysis is necessary for robust analysis.

4.3.2. Regression-to-the-mean (RTM)

RTM is an inevitable statistical phenomenon inherent in collision data (Hauer 1997). With RTM, the direct interpretation of collision data may not be due to the systematic deficiencies on the road but because of the randomness and errors hidden in the data. To address the problem, the Empirical Bayes (EB) approach has proved to be one of the best in terms of consistency and reliability (Cheng and Washington 2005; Elvik 2008). The essence of the EB approach is to combine two pieces of information: the historical collision records from a site and the performance of sites that share similar characteristics. The former keeps the major safety performance on the typical site, and the latter mediates the extreme value from randomness and other errors. Young and Park (2014) took the RTM into account using EB estimates as the basis for hot zone identification.

4.3.3. Contributory factors

Research over time has identified plenty of contributory factors to traffic collisions. For example, in the United Kingdom, the police report summaries 77 contributory factors into nine categories: road environment contributed, vehicle defects, injudicious action, driver/rider error or reaction, impairment or distraction, behaviour or inexperience, vision affected by external factors, pedestrian only factors (casualty or uninjured) and special codes (Department for Transport 2011). But the use of these contributory factors needs attention and further exploration. A failure to account for contributory factors such as traffic flow may result in some locations being identified as hazardous only because of high traffic volume. However, only a very limited number of studies (Yamada and Thill 2010; Loo and Yao 2013) have realized this problem in the spatial analysis of traffic collisions. It is believed that greater research efforts in this area can improve the value of hazardous road location identification.

5. Looking for reasons: identifying and analysing effects of contributory variables

For many years, researchers have attempted to investigate the impacts of potential risk factors on traffic safety. Dealing with this task usually requires a collision prediction model (CPM), which is commonly performed at an aggregate level like a road segment or traffic analysis zone (TAZ). Conventional CPMs, such as generalized linear models (GLMs) (Miaou 1994; Lord 2006), have a basic requirement that observations should be random. However, spatial data almost always violate this fundamental requirement, since they usually have spatial dependency. Ignoring spatial dependence can lead to an underestimation of variability (Congdon 2001). The problem is particularly important in collision modelling, where random variability and small sample sizes are common issues (Aguero-Valverde and Jovanis

2008). From the mid-1990s, safety analysts have begun to incorporate spatial dependence, which can result in a more precise estimation of the parameters (Levine, Kim, and Nitz 1995b). While most early research focused on spatial dependence, recent studies have made great attempts to deal with another problem related to the locational dimension of collision data, that is, spatial heterogeneity (LeSage and Pace 2009) or spatial non-stationarity (Fotheringham, Brunsdon, and Charlton 2002) in the relationships that are modelled (Hadayeghi, Shalaby, and Persaud 2010; Erdogan 2009). In addition, regardless of the model that is chosen to analyse collisions, the modifiable areal unit problem (MAUP) cannot be neglected for the analysis at an aggregate level. This issue has begun to attract growing awareness among researchers in the new century (Xu et al. 2014). Table 5 lists major publications on modelling of contributory factors with GIS.

Table 5. Selected publications on modelling contributory factors.

	Publication	Brief description
Spatial dependence	Levine, Kim, and Nitz 1995b	Developed a spatial lag model
	Miaou, Song, and Mallick 2003	Developed a series of Poisson-based hierarchical FB models
	Flahaut 2004	Developed a logistic model with spatial autocorrelation
	Geurts, Thomas, and Wets 2005	Used frequent item sets
	Aguero-Valverde and Jovanis 2006	Developed a hierarchical FB model (with spatial and temporal effects and space–time interactions)
	Guo, Wang, and Abdel-Aty 2010	Developed a series of Negative binomial and Poisson-based Bayesian models
Spatial heterogeneity	Xiong and Mannering 2013	Developed RPNB models
	Chen and Tarko 2014	
	Hadayeghi, Shalaby, and Persaud 2003	Developed GWR models
	Erdogan 2009	
	Hadayeghi, Shalaby, and Persaud 2010	Developed geographically weighted Poisson regression models (GWPR)
	Li et al. 2013	
	Pirdavani et al. 2014	
	Yao, Loo, and Lam 2015	
	Xu and Huang 2015	Compared RPNB and GWPR
MAUP	Thomas 1996	Investigated the effect of variations of road segment length
	Zhang and Kukadia 2005	Measured urban form through empirical modelling of travel mode choice using data aggregated at different spatial levels
	Abdel-Aty et al. 2013	Developed three different models for TAZs, block groups and census tracts
	Xu et al. 2014	Developed an emerging regionalization method

5.1. Spatial dependence

Levine, Kim, and Nitz (1995b) first developed a spatial lag model to examine the relationship between motor vehicle collisions and population, employment and road characteristics. Controlling for the degree of spatial autocorrelation, the value of the dependent variable Y_i in the model is a function of the value of Y_j at all other locations as well as a function of the independent variables. The spatial weight matrix is defined by an interaction (an inverse distance function) among the centroids of zones. Flahaut (2004) investigated the impact of infrastructure and local environment on 'road unsafety' using a logistic model with spatial autocorrelation. Geurts, Thomas, and Wets (2005) analysed the characteristics of traffic collisions in hot zones compared to those scattered all over the road with frequent item sets (a data mining technique). Although the two studies did not directly employ a spatial statistical model as Levine, Kim, and Nitz (1995b) did, they took into consideration the spatial autocorrelation by categorizing the road locations into hot zones and non-hot zones at the initial stage of the analysis.

A more advanced type of spatial modelling technique that is commonly used by road safety researchers belongs to the Bayesian family (Guo, Wang, and Abdel-Aty 2010; Miaou, Song, and Mallick 2003; Aguero-Valverde and Jovanis 2006). As Full Bayes (FB) models have potential advantages in dealing with and adjusting for the unobserved heterogeneity in space and time, Miaou, Song, and Mallick (2003) used a series of Poisson-based hierarchical FB models to model collision data aggregated at the county level in the state of Texas. A Gaussian conditional autoregressive (CAR) model was used to model spatial correlation and a Markov chain Monte Carlo (MCMC) method was used to sample the posterior probability distribution. Aguero-Valverde and Jovanis (2006) also employed an FB hierarchical model (with spatial and temporal effects and space–time interactions) for modelling injury and fatality data for Pennsylvania, and the results indicated the presence of spatial correlation at an aggregated level.

5.2. Spatial heterogeneity

The estimates of variables of the models introduced in the previous subsection are fixed for all locations. They mainly focus on handling spatial dependence between observations, but do not deal with spatial heterogeneity. In other words, they neglect the fact that the influence of predicting variables on collision counts may be greater in certain spatial units but smaller in others. To address this problem, one approach that is widely applied in safety research at the micro level is the random parameter negative binomial (RPNB) model, in which the parameters are drawn from some random distributions (typically the normal) and are assumed to vary randomly from case to case (Xiong and Mannering 2013; Chen and Tarko 2014). Another method dealing with spatial nonstationarity is the geographically weighted regression (GWR) technique which, when compared with RPNB, has advantages in accounting for spatial correlation existing across adjacent units (Xu and Huang 2015). In the recent decades, GWR has been employed by many researchers in modelling collision data, particularly at the zonal level (Pirdavani et al. 2014; Li et al. 2013; Hadayeghi, Shalaby, and Persaud 2003; Erdogan 2009).

GWR relies on the calibration of multiple regression models for different geographical entities (Fotheringham, Brunsdon, and Charlton 2002). In the road safety literature, Hadayeghi, Shalaby, and Persaud (2003) is among the first studies which developed GWR models to investigate spatial variations in relationships. With an increased R^2, the research indicated that the use of GWR can lead to an improvement in the prediction of the response. In another study by Erdogan (2009), different spatial autocorrelation statistics were calculated to see whether the traffic collisions in Turkey are clustered or not at the provincial level. Since the exploratory analysis indicated the absence of spatial clustering, a GWR model was developed in comparison with the ordinary least square (OLS) model for investigating the risk factors. The results indicated that the former performed better than the latter in terms of an approximate likelihood ratio based on the *F*-test.

The GWR technique can be integrated with GLMs and form geographically weighted generalized linear models (GWGLMs) to handle count data (Fotheringham, Brunsdon, and Charlton 2002). Hadayeghi, Shalaby, and Persaud (2010) combined the two techniques with the Poisson error distribution and applied it to the traffic collision data analysis. The model form was proposed as:

$$\ln(A(\boldsymbol{\mu}_i)) = \ln(\beta_0(\boldsymbol{\mu}_i)) + \beta_1(\boldsymbol{\mu}_i)\ln(Ex) + \beta_2(\boldsymbol{\mu}_i)X_2 + \ldots + \beta_P(\boldsymbol{\mu}_i)X_P \quad (1)$$

where $\ln(A(\boldsymbol{\mu}_i))$ is natural log of collision frequency on location $\boldsymbol{\mu}_i$; *Ex* is the exposure, such as traffic flow or pedestrian flow; X_j is *j*th explanatory variables ($j = 2,\ldots, p$); β_j ($j = 0,1, \ldots, p$) is a function of location $\boldsymbol{\mu}_i$ denoting the coordinates (x_i,y_i) of the *i*th location. The basic idea of GWR technique is that the observed data near location *i* have greater influence on the estimation of the $\beta_j(\boldsymbol{\mu}_i)$ than that located farther from *i*. This scheme can be described by a weighting function that is usually referred to as a kernel. Gaussian and the bisquare functions are two common kernels. Although both kernels and bandwidth influence the GWPR estimates, the choice of bandwidth is more important than the shape of the kernel in affecting the fit of the model (Pirdavani et al. 2014; Fotheringham, Brunsdon, and Charlton 2002). Comprehensive discussion on these issues can be found in Nakaya et al. (2005), Pirdavani et al. (2014), and Fotheringham, Brunsdon, and Charlton (2002). Given the strength of GWGLMs in dealing with count data, GWGLMs have been more commonly adopted than conventional GWR in collision analysis at various spatial levels. Typical examples include Li et al. (2013), who modelled the fatal collisions and countywide factors, including traffic patterns, road network attributes and socio-demographic characteristics; and Yao, Loo, and Lam (2015), who investigated the relationship between pedestrian collisions and exposure at the road segment level. In addition to the choice of kernels and bandwidth, research efforts were also made on the explanation and discussion of the signs and significance of coefficients (Hadayeghi, Shalaby, and Persaud 2010). The findings kept prompting researchers to question the general perception that the influences of collision risk factor are always uniform across the study area.

5.3. Modifiable areal unit problem (MAUP)

Although most previous work relied on a specific segmentation/zoning system, there are some research studies that focus on the assessment of how spatial scale and zoning systems may influence the statistical results of safety models. Focusing on MAUP on a line, Thomas (1996) investigated the effect of variations of road segment length on the statistical description of collision counts using small, medium and large segments, respectively. The experiment demonstrated that the definition of segments could affect the statistical results due to MAUP. In a recent study by Xu et al. (2014), an emerging regionalization method was employed to aggregate 738 TAZs in the county of Hillsborough to 14 zoning schemes at an incremental step size of 50 zones, based on spatial homogeneity of collision risk. Their results revealed that the zoning schemes with higher number of zones tend to have increasing number of significant variables, more stable coefficient estimation, smaller standard error, but worse model performance. Their research, together with other studies such as Zhang and Kukadia (2005) and Abdel-Aty et al. (2013), highlighted MAUP as a pressing issue that is generally neglected by road safety analysts.

6. The new trend: multidisciplinary investigation of road safety in an enhanced GIS database

As early as the mid-1980s, professionals from engineering and medical sciences had realized the importance of linking police collision databases with hospital injury databases to ensure data completeness, representativeness and reliability (Maas and Harris 1984). As data quality is crucial for any collision analysis, this issue has always been investigated over the past four decades (Aptel et al. 1999; Rosman 2001; Amoros et al. 2008). However, more recent studies not only aim to find discrepancies and complementarity among multidata sources but also attempt to bring multidisciplinary teams to work more closely using advanced approaches. For example, a research team consisting of members with engineering and epidemiology backgrounds developed an algorithm to assess the road traffic injuries and quantified the underreporting cases in the Greek Island of Corfu (Petridou et al. 2009). In another recent study, delegates from the police and insurance industry have collaborated with a group of researchers from a hospital to investigate the factors and reporting issue of bicycle collisions in Germany (Juhra et al. 2012). Recently, a group of experts from geography, engineering and medical disciplines have worked together with community leaders in addressing road safety problems in Hong Kong (Loo et al. 2013). These efforts enlighten us about the great potential of using GIS to support multidisciplinary approaches in road safety analysis.

Furthermore, the new information era has allowed users to handle a huge amount of dynamic information (e.g. movement of individual objects) in an enhanced GIS database, which can help better understand human perceptions and behaviours in health (Goodchild 2015). As this dynamic GIS has proven to be very useful in tracking human exposures to the environment, collision analysts have attempted to employ it for measuring exposures to the collision risks. In the specific context of road safety, movements (including vehicles and pedestrians) are treated from a people-based perspective, which focuses directly on individuals' activities in space and time (Miller 2007; Kwan 2012; Shaw 2010). The foundation can be traced to time geography (Hägerstraand 1970). In traditional time geography, space–time path (SP) and potential path area (PPA) are computed in Euclidean space. However, for vehicles travelling or pedestrian walking along road networks, a network-based framework is more appropriate. Following this rationale, SP is calculated as the shortest path along the network between any two consecutive points and all possible locations reachable during the specified time interval form a potential path tree (PPT) (Miller 1991; Neutens et al. 2008). Network-based formulations of time geography have been used in a variety of transport settings (Downs and Horner 2012; Neutens et al. 2008; Kim and Kwan 2003), but the application in measuring traffic collision risk is limited. Till now, the only exceptions are those conducted by Loo et al. (Lam, Loo, and Yao 2013; Lam, Yao, and Loo 2014; Yao, Loo, and Lam 2015).

Lam, Loo, and Yao (2013) measured pedestrian exposure in a network-constrained time geography framework. Based on the Travel Characteristics Survey (TCS) database, which stores the detailed information on trips made by Hong Kong residents, the authors applied the concept of SP to model the movements of elderly pedestrians and calculated collision risks at the segment level. Lam, Yao, and Loo (2014) further proposed a PPT method, which was only applicable to home-destination-home trips. In a more recent work by Yao, Loo, and Lam (2015), the PPT method was extended to all types of trips and a weighted scheme (Figure 1) was developed by considering the fact that most pedestrians would like to choose the shortest path (Guo and Loo 2013). The empirical results of the three studies indicated that methods developed in the context of network-constrained time geography are useful tools in measuring pedestrian exposure. However, as the authors concluded in Yao, Loo, and Lam (2015), there are still a number of issues requiring further research efforts, such as applying proposed models to different transport modes and differentiation of trips by activity/trip purpose.

7. Conclusions

Most general collision prediction methods are not capable of answering one simple but important question in road safety: where are the hazardous road locations? To answer this question, the spatial pattern of collisions must be explored in a scientific manner. Advanced GIS-based methods do not only provide a simple visualization of dangerous road sections but, more importantly, also offer the capability of analysing spatial dependency and spatial interactions with other geographical factors. The methodological advancement of applying GIS in road safety research over the last four decades has helped us to gain a more comprehensive understanding

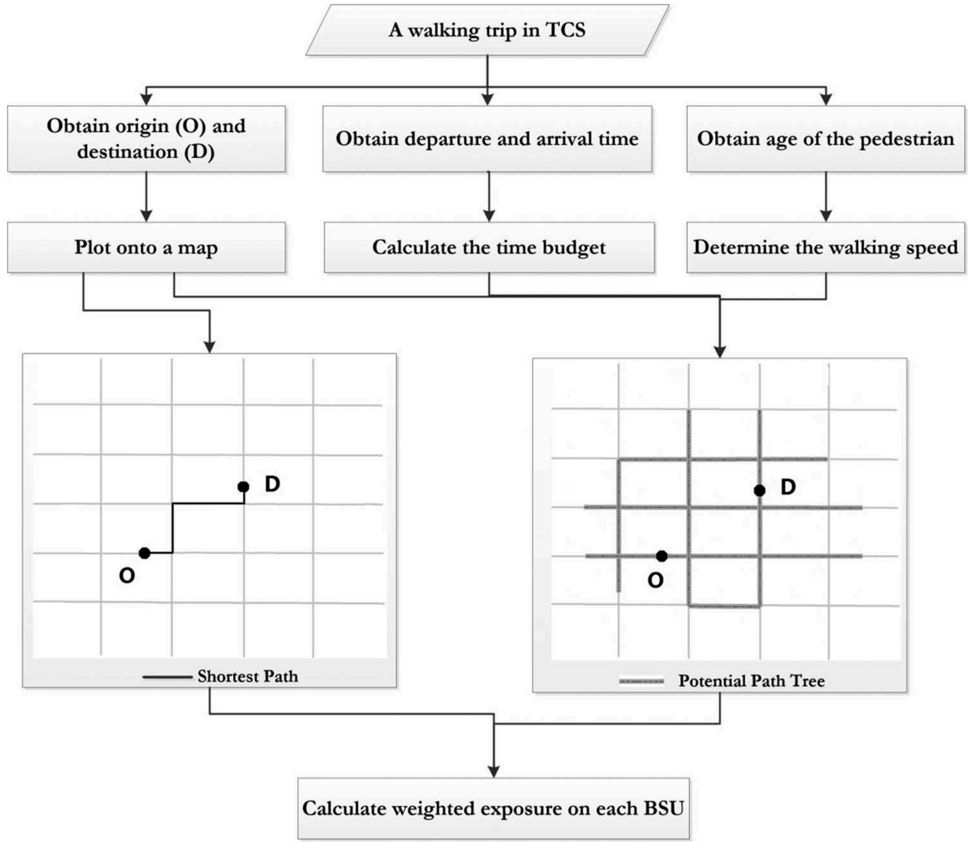

Figure 1. Flow chart for deriving weighted pedestrian exposures (Yao, Loo, and Lam 2015).

of traffic collisions to answer another important question: in what ways can road safety be improved ?An emerging research direction is the multidisciplinary approach to integrate various types of data (including mobile data, smart data and big data) and analyse them (meaningfully) with a set of advanced analytical tools in an enhanced GIS database. In particular, information is generated at an unprecedented speed and volume nowadays. Big data, which describe the exponential increase in data volume, velocity, variety and value, allow us to manage large amount of structured, semi-structured and unstructured data. However, a lot of 'big' data comprises 'noise' (information or metadata having low or no real value). Filtering out the noise and holding the 'meaningful' data (also called 'smart data') is a big challenge for road safety researchers. To tackle this, a multidisciplinary approach is needed to integrate various types of data, such as mobile data that describe the transmission of data with a smartphone, tablet or other portable device, and analyse them (meaningfully) with a set of advanced analytical tools to further improve road safety and benefit the human race.

Seshan and Maitra (2014), for example, suggest that autonomous or self-driving vehicles will join the web as the biggest revolution of the early twenty-first century. They have the potential to eliminate virtually all driver errors and most traffic collisions (Lutin, Kornhauser, and Lerner-Lam 2013). A self-driving vehicle is capable of sensing its environment and navigating without human input (Seshan and Maitra 2014). Its elements can be divided into four basic component categories: sensors, mapping, perception and communication (Lutin, Kornhauser, and Lerner-Lam 2013). Hardware sensors are used for sensing real-time environment conditions; mapping can be regarded as a GIS database with geographical information ranging from single points representing traffic lights to 3D terrain models created with LIDAR; perception includes a set of software processes used for determining the reaction of the vehicle to various inputs; and communication technology allows a vehicle to exchange information with other vehicles about position and movement intentions (vehicle to vehicle, V2V) or to communicate with traffic control devices (vehicle to infrastructure, V2I). Although it is estimated that self-driving or autonomous cars could plausibly be present on the roads in significant numbers within a decade (Belzowski and McManus 2010), accommodating them requires changes in the other parts of the transport system such as parking, lane widths design and speed-limit enforcement (Lutin, Kornhauser, and Lerner-Lam 2013). The LIDAR sensors on the vehicles, for example, can only detect the rough shape of the objects on the road, which means that they cannot detect the hand signals of a cyclist or follow the instructions of a policeman conducting traffic. Furthermore, there is the additional risk of equipment failure, computer bugs and data acquisition inaccuracy in any of the four basic component categories of an autonomous vehicle. Currently, much effort about autonomous vehicles has been spent on finding optimal routes and the avoidance of traffic collisions (Seshan and Maitra 2014). However, there is a huge potential of using and integrating the data collected by autonomous vehicles to analyse road safety problems too. For instance, the Google Driverless Car Project has collected comprehensive data over 700,000 autonomous driving miles. The ways of using these data fruitfully to identify, analyse and tackle road safety problems, especially with an enhanced GIS database and a multidisciplinary team, remain highly relevant and important.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Becky P.Y. Loo http://orcid.org/0000-0003-0822-5354

References

Abdel-Aty, M., J. Lee, C. Siddiqui, and K. Choi. 2013. "Geographical Unit based Analysis in the Context of Transportation Safety Planning." *Transportation Research Part A: Policy and Practice* 49: 62–75. doi:10.1016/j.tra.2013.01.030.

Aguero-Valverde, J., and P. P. Jovanis. 2006. "Spatial Analysis of Fatal and Injury Crashes in Pennsylvania." *Accident Analysis & Prevention* 38 (3): 618–625. doi:10.1016/j.aap.2005.12.006.

Aguero-Valverde, J., and P. P. Jovanis. 2008. "Analysis of Road Crash Frequency with Spatial Models." *Transportation Research Record: Journal of the Transportation Research Board* 2061 (1): 55–63. doi:10.3141/2061-07.

Amoros, E., J.-L. Martin, S. Lafont, and B. Laumon. 2008. "Actual Incidences of Road Casualties, and Their Injury Severity, Modelled from Police and Hospital Data, France." *The European Journal of Public Health* 18 (4): 360–365. doi:10.1093/eurpub/ckn018.

Anderson, T. K. 2009. "Kernel Density Estimation and K-means Clustering to Profile Road Accident Hotspots." *Accident Analysis & Prevention* 41 (3): 359–364. doi:10.1016/j.aap.2008.12.014.

Anselin, L. 1995. "Local Indicators of Spatial Association-LISA." *Geographical Analysis* 27 (2): 93–115. doi:10.1111/j.1538-4632.1995.tb00338.x.

Aptel, I., L. Rachid Salmi, F. Masson, A. Bourdé, G. Henrion, and P. Erny. 1999. "Road Accident Statistics: Discrepancies between Police and Hospital Data in a French Island." *Accident Analysis & Prevention* 31 (1–2): 101–108. doi:10.1016/S0001-4575(98)00051-7.

Austin, K. 1995. "The Identification of Mistakes in Road Accident Records: Part 1, Locational Variables." *Accident Analysis & Prevention* 27 (2): 261–276. doi:10.1016/0001-4575(94)00065-T.

Belzowski, B. M., and W. McManus. 2010. *Alternative Powertrain Strategies and Fleet Turnover in the 21st Century*. Ann Arbor: University of Michigan Transportation Research Institute. http://deepblue.lib.umich.edu/bitstream/handle/2027.42/78001/102673.pdf.

Black, W. R., and I. Thomas. 1998. "Accidents on Belgium's Motorways: A Network Autocorrelation Analysis." *Journal of Transport Geography* 6 (1): 23–31. doi:10.1016/S0966-6923(97)00037-9.

Chen, E., and A. P. Tarko. 2014. "Modeling Safety of Highway Work Zones with Random Parameters and Random Effects Models." *Analytic Methods in Accident Research* 1: 86–95. doi:10.1016/j.amar.2013.10.003.

Cheng, W., and S. P. Washington. 2005. "Experimental Evaluation of Hotspot Identification Methods." *Accident Analysis & Prevention* 37 (5): 870–881. doi:10.1016/j.aap.2005.04.015.

Congdon, P. 2001. *Bayesian Statistical Modeling*. Chichester: John Wiley & Sons.

Deepika, S., and R. Saradha. 2014. "Clustering Crash Hotspots to Organize Police Dispatch Routes Using GIS." *International Journal of Science and Research* 3 (2): 44–50. http://www.ijsr.net/archive/v3i2/MDIwMTM4Njk=.pdf.

Department for Transport. 2011. "Reported Road Casualties Great Britain: Annual Report 2011." https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/9277/rrcgb2011-04.pdf

Downs, J. A., and M. W. Horner. 2012. "Probabilistic Potential Path Trees for Visualizing and Analyzing Vehicle Tracking Data." *Journal of Transport Geography* 23: 72–80. doi:10.1016/j.jtrangeo.2012.03.017.

Eckhardt, N., B. Flahaut, and I. Thomas. 2004. "The Spatio-Temporal Distribution of Road Traffic Accidents in a Periurban Environment: The Case of Brussels." *Recherche, Transports, Sécurité* 82: 35–46. doi:10.3166/rts.82.35-46.

Elvik, R. 2008. "The Predictive Validity of Empirical Bayes Estimates of Road Safety." *Accident Analysis & Prevention* 40 (6): 1964–1969. doi:10.1016/j.aap.2008.07.007.

Erdogan, S. 2009. "Explorative Spatial Analysis of Traffic Accident Statistics and Road Mortality among the Provinces of Turkey." *Journal of Safety Research* 40 (5): 341–351. doi:10.1016/j.jsr.2009.07.006.

Erdogan, S., I. Yilmaz, T. Baybura, and M. Gullu. 2008. "Geographical Information Systems aided Traffic Accident Analysis System Case Study: City of Afyonkarahisar." *Accident Analysis & Prevention* 40 (1): 174–181. doi:10.1016/j.aap.2007.05.004.

Flahaut, B. 2004. "Impact of Infrastructure and Local Environment on Road Unsafety: Logistic Modeling with Spatial Autocorrelation." *Accident Analysis & Prevention* 36 (6): 1055–1066. doi:10.1016/j.aap.2003.12.003.

Flahaut, B., M. Mouchart, E. S. Martin, and I. Thomas. 2003. "The Local Spatial Autocorrelation and the Kernel Method for Identifying Black Zones: A Comparative Approach." *Accident Analysis & Prevention* 35 (6): 991–1004. doi:10.1016/S0001-4575(02)00107-0.

Flahaut, B., and I. Thomas. 2002. "Identifier Les Zones Noires D'un Reseau Routier Par L'autocorrelation Spatiale Locale. Analyses De Sensibilite Et Aspects Operationnels." *Revue Internationale De Géomatique* 12 (2): 245–261. doi:10.3166/rig.12.245-261.

Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: Wiley.

Gatrell, A. C., T. C. Bailey, P. J. Diggle, and B. S. Rowlingson. 1996. "Spatial Point Pattern Analysis and Its Application in Geographical Epidemiology." *Transactions of the Institute of British Geographers* 21 (1): 256–274. doi:10.2307/622936.

Getis, A., and J. Keith Ord. 1992. "The Analysis of Spatial Association by Use of Distance Statistics." *Geographical Analysis* 24 (3): 189–206. doi:10.1111/j.1538-4632.1992.tb00261.x.

Geurts, K., I. Thomas, and G. Wets. 2005. "Understanding Spatial Concentrations of Road Accidents Using Frequent Item Sets." *Accident Analysis & Prevention* 37 (4): 787–799. doi:10.1016/j.aap.2005.03.023.

Goodchild, M. F. 2015. "Space, Place and Health." *Annals of GIS* 21 (2): 97–100. doi:10.1080/19475683.2015.1007895.

Guo, F., X. Wang, and M. A. Abdel-Aty. 2010. "Modeling Signalized Intersection Safety with Corridor-Level Spatial Correlations." *Accident Analysis & Prevention* 42 (1): 84–92. doi:10.1016/j.aap.2009.07.005.

Guo, Z., and B. P. Y. Loo. 2013. "Pedestrian Environment and Route Choice: Evidence from New York City and Hong Kong." *Journal of Transport Geography* 28: 124–136. doi:10.1016/j.jtrangeo.2012.11.013.

Hadayeghi, A., A. S. Shalaby, and B. N. Persaud. 2003. "Macrolevel Accident Prediction Models for Evaluating Safety of Urban Transportation Systems." *Transportation Research Record: Journal of the Transportation Research Board* 1840 (1): 87–95. doi:10.3141/1840-10.

Hadayeghi, A., A. S. Shalaby, and B. N. Persaud. 2010. "Development of Planning Level Transportation Safety Tools using Geographically Weighted Poisson Regression." *Accident Analysis & Prevention* 42 (2): 676–688. doi:10.1016/j.aap.2009.10.016.

Hägerstraand, T. 1970. "What About People in Regional Science?" *Papers in Regional Science* 24 (1): 7–24. doi:10.1111/j.1435-5597.1970.tb01464.x.

Hauer, E. 1997. *Observational Before-After Studies in Road Safety*. Bradford: Emerald Group Publishing Limited.

Juhra, C., B. Wieskötter, K. Chu, L. Trost, U. Weiss, M. Messerschmidt, A. Malczyk, M. Heckwolf, and M. Raschke. 2012. "Bicycle Accidents – Do We Only See the Tip of the Iceberg?: A Prospective Multi-Centre Study in a Large German City Combining Medical and Police data." *Injury* 43 (12): 2026–2034. doi:10.1016/j.injury.2011.10.016.

Kam, B. H. 2003. "A Disaggregate Approach to Crash Rate Analysis." *Accident Analysis & Prevention* 35 (5): 693–709. doi:10.1016/S0001-4575(02)00048-9.

Kim, H.-M., and M.-P. Kwan. 2003. "Space-Time Accessibility Measures: A Geocomputational Algorithm with a Focus on the Feasible Opportunity Set and Possible Activity Duration." *Journal of Geographical Systems* 5 (1): 71–91. doi:10.1007/s101090300104.

Kwan, M.-P. 2012. "How GIS Can Help Address the Uncertain Geographic Context Problem in Social Science Research." *Annals of GIS* 18 (4): 245–255. doi:10.1080/19475683.2012.727867.

Lam, W. W. Y., B. P. Y. Loo, and S. Yao. 2013. "Towards Exposure-based Time-Space Pedestrian Crash Analysis in Facing the Challenges of Ageing Societies in Asia." *Asian Geographer* 30 (2): 105–125. doi:10.1080/10225706.2012.735436.

Lam, W. W. Y., S. Yao, and B. P. Y. Loo. 2014. "Pedestrian Exposure Measures: A Time-Space Framework." *Travel Behaviour and Society* 1 (1): 22–30. doi:10.1016/j.tbs.2013.10.004.

LeSage, J., and R. K. Pace. 2009. *Introduction to Spatial Econometrics*. London: CRC press.

Levine, N., K. E. Kim, and L. H. Nitz. 1995a. "Spatial Analysis of Honolulu Motor Vehicle Crashes: I. Spatial Patterns." *Accident Analysis & Prevention* 27 (5): 663–674. doi:10.1016/0001-4575(95)00017-T.

Levine, N., K. E. Kim, and L. H. Nitz. 1995b. "Spatial Analysis of Honolulu Motor Vehicle Crashes: II. Zonal Generators." *Accident Analysis & Prevention* 27 (5): 675–685. doi:10.1016/0001-4575(95)00018-U.

Li, Z., W. Wang, P. Liu, J. M. Bigham, and D. R. Ragland. 2013. "Using Geographically Weighted Poisson Regression for County-Level Crash Modeling in California." *Safety Science* 58: 89–97. doi:10.1016/j.ssci.2013.04.005.

Loo, B. P. 2006. "Validating Crash Locations for Quantitative Spatial Analysis: A GIS-based Approach." *Accident Analysis & Prevention* 38 (5): 879–886. doi:10.1016/j.aap.2006.02.012.

Loo, B. P. 2009. "The Identification of Hazardous Road Locations: A Comparison of the Blacksite and Hot Zone Methodologies in Hong Kong." *International Journal of Sustainable Transportation* 3 (3): 187–202. doi:10.1080/15568310801915583.

Loo, B. P., Y. Shenjun Yao, and W. Jianping. 2011. "Spatial Point Analysis of Road Crashes in Shanghai: A GIS-based Network Kernel Density Method." *Geoinformatics* 2011. doi:10.1109/GeoInformatics.2011.5980938.

Loo, B. P. Y., C. B. Chow, M. Leung, T. H. J. Kwong, S. F. A. Lai, and Y. H. Chau. 2013. "Multidisciplinary Efforts toward Sustained Road Safety Benefits: Integrating Place-based and People-based Safety Analyses." *Injury Prevention* 19 (1): 58–63. doi:10.1136/injuryprev-2012-040400.

Loo, B. P. Y., and K. L. Tsui. 2010. "Bicycle Crash Casualties in a Highly Motorized City." *Accident Analysis & Prevention* 42 (6): 1902–1907. doi:10.1016/j.aap.2010.05.011.

Loo, B. P. Y., and S. Yao. 2013. "The Identification of Traffic Crash Hot Zones under the Link-Attribute and Event-based Approaches in a Network-Constrained Environment." *Computers, Environment and Urban Systems* 41: 249–261. doi:10.1016/j.compenvurbsys.2013.07.001.

Lord, D. 2006. "Modeling Motor Vehicle Crashes using Poisson-Gamma Models: Examining the Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter." *Accident

Analysis & Prevention 38 (4): 751–766. doi:10.1016/j.aap.2006.02.001.

Loveday, J. 1991. "Spatial Autocorrelation and Road Accident Migration." Paper presented at the 23rd Annual Conference of Universities Transport Studies Group, University of Nottingham, January.

Lutin, J. M., A. L. Kornhauser, and E. Lerner-Lam. 2013. "The Revolutionary Development of Self-Driving Vehicles and Implications for the Transportation Engineering Profession." *ITE Journal*. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=5C3DA095A071DEBA0779C8694379F5D3?doi=10.1.1.377.6954&rep=rep1&type=pdf.

Maas, M. W., and S. Harris. 1984. "Police Recording of Road Accident In-patients: Investigation into the Completeness, Representativity and Reliability of Police Records of Hospitalized Traffic Victims." *Accident Analysis & Prevention* 16 (3): 167–184. doi:10.1016/0001-4575(84)90011-3.

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. doi:10.1016/0001-4575(94)90038-8.

Miaou, S.-P., J. J. Song, and B. K. Mallick. 2003. "Roadway Traffic Crash Mapping: A Space-Time Modeling Approach." *Journal of Transportation and Statistics* 6: 33–58. http://apps.bts.gov/publications/journal_of_transportation_and_statistics/volume_06_number_01/html/paper_03/.

Miller, H. 2007. "Place-based versus People-based Geographic Information Science." *Geography Compass* 1 (3): 503–535. doi:10.1111/j.1749-8198.2007.00025.x.

Miller, H. J. 1991. "Modelling Accessibility using Space-Time Prism Concepts within Geographical Information Systems." *International Journal of Geographical Information System* 5 (3): 287–301. doi:10.1080/02693799108927856.

Miller, H. J. 1999. "Potential Contributions of Spatial Analysis to Geographic Information Systems for Transportation (GIS-T)." *Geographical Analysis* 31 (4): 373–399. doi:10.1111/j.1538-4632.1999.tb00991.x.

Miller, S. 1995. "Design and Development of a Crash Referencing and Analysis System." In *Proceedings of the* 1995 *Geographic Information Systems for Transportation (GIS-T) Symposium*. Sparks, NV: American Association of State Highway and Transportation Officials (AASHTO).

Moellering, H. 1976. "The Potential Uses of a Computer Animated Film in the Analysis of Geographical Patterns of Traffic Crashes." *Accident Analysis & Prevention* 8 (4): 215–227. doi:10.1016/0001-4575(76)90007-5.

Moons, E., T. Brijs, and G. Wets. 2009a. "Identifying Hazardous Road Locations: Hot Spots versus Hot Zones." In *Transactions on Computational Science VI*, edited by C. J. K. Tan, 288–300. Heidelberg: Springer. doi:10.1007/978-3-642-10649-1_17.

Moons, E., T. Brijs, and G. Wets. 2009b. "Improving Moran's Index to Identify Hot Spots in Traffic Safety." In *Geocomputation and Urban Planning*, edited by B. Murgante, G. Borruso, and A. Lapucci, 117–132. Heidelberg: Springer. doi:10.1007/978-3-540-89930-3_7.

Nakaya, T., A. S. Fotheringham, C. Brunsdon, and M. Charlton. 2005. "Geographically Weighted Poisson Regression for Disease Association Mapping." *Statistics in Medicine* 24 (17): 2695–2717. doi:10.1002/sim.2129.

Neutens, T. 2015. "Accessibility, Equity and Health Care: Review and Research Directions for Transport Geographers." *Journal of Transport Geography* 43: 14–27. doi:10.1016/j.jtrangeo.2014.12.006.

Neutens, T., N. Van De Weghe, F. Witlox, and D. M. Philippe. 2008. "A Three-Dimensional Network-based Space-Time Prism." *Journal of Geographical Systems* 10 (1): 89–107. doi:10.1007/s10109-007-0057-x.

Nicholson, A. 1999. "Analysis of Spatial Distributions of Accidents." *Safety Science* 31 (1): 71–91. doi:10.1016/S0925-7535(98)00056-3.

Nie, K., Z. Wang, D. Qingyun, F. Ren, and Q. Tian. 2015. "A Network-Constrained Integrated Method for Detecting Spatial Cluster and Risk Location of Traffic Crash: A Case Study from Wuhan, China." *Sustainability* 7 (3): 2662–2677. doi:10.3390/su7032662.

Okabe, A., T. Satoh, and K. Sugihara. 2009. "A Kernel Density Estimation Method for Networks, its Computational Method and a GIS-based Tool." *International Journal of Geographical Information Science* 23 (1): 7–32. doi:10.1080/13658810802475491.

Xu, P., and H. Huang. 2015. "Modeling Crash Spatial Heterogeneity: Random Parameter versus Geographically Weighting." *Accident Analysis & Prevention* 75: 16–25. doi:10.1016/j.aap.2014.10.020.

Xu, P., H. Huang, N. Dong, and A.-A. Mohamed. 2014. "Sensitivity Analysis in the Context of Regional Safety Modeling: Identifying and Assessing the Modifiable Areal Unit Problem." *Accident Analysis & Prevention* 70: 110–120. doi:10.1016/j.aap.2014.02.012.

Petridou, E. T., G. Yannis, A. Terzidis, N. Dessypris, E. Germeni, P. Evgenikos, N. Tselenti, A. Chaziris, and I. Skalkidis. 2009. "Linking Emergency Medical Department and Road Traffic Police Casualty Data: A Tool in Assessing the Burden of Injuries in Less Resourced Countries." *Traffic Injury Prevention* 10 (1): 37–43. doi:10.1080/15389580802526400.

Pirdavani, A., T. Bellemans, T. Brijs, and G. Wets. 2014. "Application of Geographically Weighted Regression Technique in Spatial Analysis of Fatal and Injury Crashes." *Journal of Transportation Engineering* 140 (8): 04014032. doi:10.1061/(ASCE)TE.1943-5436.0000680.

Plug, C., J. C. Xia, and C. Caulfield. 2011. "Spatial and Temporal Visualisation Techniques for Crash Analysis." *Accident Analysis & Prevention* 43 (6): 1937–1946. doi:10.1016/j.aap.2011.05.007.

Qin, X., S. Parker, Y. Liu, A. J. Graettinger, and S. Forde. 2013. "Intelligent Geocoding System to Locate Traffic Crashes." *Accident Analysis & Prevention* 50: 1034–1041. doi:10.1016/j.aap.2012.08.007.

Rosman, D. L. 2001. "The Western Australian Road Injury Database (1987-1996): Ten Years of Linked Police, Hospital and Death Records of Road Crashes and Injuries." *Accident Analysis & Prevention* 33 (1): 81–88. doi:10.1016/S0001-4575(00)00018-X.

Sabel, C. E., S. Kingham, A. Nicholson, and P. Bartie. 2005. "Road Traffic Accident Simulation Modelling - A Kernel Estimation Approach." Paper presented at SIRC 2005 - The 17th Annual Colloquium of the Spatial Information Research Centre University of Otago, Dunedin, New Zealand. http://www.researchgate.net/publication/

252156589_Road_Traffic_Accident_Simulation_Modelling_-_A_Kernel_Estimation_Approach.

Seshan, J., and S. Maitra. 2014. "Efficient Route Finding and Sensors for Collision Detection in Google's Driverless Car." *International Journal of Computer Science and Mobile Computing* 3 (12): 70–78. http://www.ijcsmc.com/docs/papers/December2014/V3I12201418.pdf.

Shaw, S.-L. 2010. "Geographic Information Systems for Transportation: From a Static Past to a Dynamic Future." *Annals of GIS* 16 (3): 129–140. doi:10.1080/19475683.2010.513155.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. London: CRC Press.

Steenberghen, T., K. Aerts, and I. Thomas. 2010. "Spatial Clustering of Events on a Network." *Journal of Transport Geography* 18 (3): 411–418. doi:10.1016/j.jtrangeo.2009.08.005.

Steenberghen, T. T., I. T. Dufays, and B. Flahaut. 2004. "Intra-Urban Location and Clustering of Road Accidents Using GIS: A Belgian Example." *International Journal of Geographical Information Science* 18 (2): 169–181. doi:10.1080/13658810310001629619.

Thomas, I. 1996. "Spatial Data Aggregation: Exploratory Analysis of Road Accidents." *Accident Analysis & Prevention* 28 (2): 251–264. doi:10.1016/0001-4575(95)00067-4.

Truong, L. T., and S. V. C. Somenahalli. 2011. "Using GIS to Identify Pedestrian-Vehicle Crash Hot Spots and Unsafe Bus Stops." *Journal of Public Transportation* 14 (1): 99–114. doi:10.5038/2375-0901.

World Health Organization. 2013. *Global Status Report on Road Safety 2013: Supporting a Decade of Action*. Geneva: WHO Press.

Xie, Z., and J. Yan. 2008. "Kernel Density Estimation of Traffic Accidents in a Network Space." *Computers, Environment and Urban Systems* 32 (5): 396–406. doi:10.1016/j.compenvurbsys.2008.05.001.

Xie, Z., and J. Yan. 2013. "Detecting Traffic Accident Clusters with Network Kernel Density Estimation and Local Spatial Statistics: An Integrated Approach." *Journal of Transport Geography* 31: 64–71. doi:10.1016/j.jtrangeo.2013.05.009.

Xiong, Y., and F. L. Mannering. 2013. "The Heterogeneous Effects of Guardian Supervision on Adolescent Driver-Injury Severities: A Finite-Mixture Random-Parameters Approach." *Transportation Research Part B: Methodological* 49: 39–54. doi:10.1016/j.trb.2013.01.002.

Yamada, I., and J.-C. Thill. 2004. "Comparison of Planar and Network K-functions in Traffic Accident Analysis." *Journal of Transport Geography* 12 (2): 149–158. doi:10.1016/j.jtrangeo.2003.10.006.

Yamada, I., and J.-C. Thill. 2010. "Local Indicators of Network-Constrained Clusters in Spatial Patterns Represented by a Link Attribute." *Annals of the Association of American Geographers* 100 (2): 269–285. doi:10.1080/00045600903550337.

Yamada, I., and J. Thill. 2007. "Local Indicators of Network-Constrained Clusters in Spatial Point Patterns." *Geographical Analysis* 39 (3): 268–292. doi:10.1111/j.1538-4632.2007.00704.x.

Yao, S., B. P. Y. Loo, and W. W. Y. Lam. 2015. "Measures of Activity-based Pedestrian Exposure to the Risk of Vehicle-Pedestrian Collisions: Space-Time Path vs. Potential Path Tree Methods." *Accident Analysis & Prevention* 75: 320–332. doi:10.1016/j.aap.2014.12.005.

Young, J., and P. Y. Park. 2014. "Hotzone Identification with GIS-based Post-Network Screening Analysis." *Journal of Transport Geography* 34: 106–120. doi:10.1016/j.jtrangeo.2013.11.007.

Yu, Hao, P. Liu, J. Chen, and H. Wang. 2014. "Comparative Analysis of the Spatial Analysis Methods for Hotspot Identification." *Accident Analysis & Prevention* 66: 80–88. doi:10.1016/j.aap.2014.01.017.

Zhang, J., S. You, L. Chen, and C. Chen. 2010. "A Hybrid Approach to Segment-Type Geocoding of New York City Traffic Data." In *Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research & Application*. doi:10.1145/1823854.1823871.

Zhang, M., and N. Kukadia. 2005. "Metrics of Urban Form and the Modifiable Areal Unit Problem." *Transportation Research Record: Journal of the Transportation Research Board* 1902: 71–79. doi:10.3141/1902-09.