Accident Analysis and Prevention 135 (2020) 105323

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

A review of spatial approaches in road safety

Apostolos Ziakopoulos*, George Yannis

National Technical University of Athens, Department of Transportation Planning and Engineering, 5 Heroon Polytechniou Str., GR-15773, Athens, Greece

ARTICLE INFO

Keywords:
Road safety
spatial analysis
crash analysis
study characteristics
areal units

ABSTRACT

Spatial analyses of crashes have been adopted in road safety for decades in order to determine how crashes are affected by neighboring locations, how the influence of parameters varies spatially and which locations warrant interventions more urgently. The aim of the present research is to critically review the existing literature on different spatial approaches through which researchers handle the dimension of space in its various aspects in their studies and analyses. Specifically, the use of different areal unit levels in spatial road safety studies is investigated, different modelling approaches are discussed, and the corresponding study design characteristics are summarized in respective tables including traffic, road environment and area parameters and spatial aggregation approaches. Developments in famous issues in spatial analysis such as the boundary problem, the modifiable areal unit problem and spatial proximity structures are also discussed. Studies focusing on spatially analyzing vulnerable road users are reviewed as well. Regarding spatial models, the application, advantages and disadvantages of various functional/econometric approaches, Bayesian models and machine learning methods are discussed. Based on the reviewed studies, present challenges and future research directions are determined.

1. Introduction

Road safety has been a major issue in contemporary societies, with road crashes incurring major human and material costs annually worldwide. Traffic and road safety practices have been implemented to save lives by halting the increase of road traffic fatalities against an ever-rising population (WHO, 2015), though it appears that the global target of halving road traffic deaths by 2020 will not be met (WHO, 2018).

The still occurring and plateauing crash casualties suggest a lot of untapped potential and margins for safety improvements that can be exploited if the occurrence of crashes can be predicted more accurately. Road safety scientists have invested considerable efforts in studying the impacts of several risk factors (e.g. Theofilatos and Yannis, 2014; Papadimitriou et al., 2019) and road safety measures (e.g. Elvik et al., 2009) and have developed or adopted a number of mathematical methodologies to approach crash prediction problems (e.g. Lord and Mannering, 2010) or road safety site prioritization problems (e.g. Lee and Abdel-Aty, 2018).

Since road transport involves distances by nature, it stands to reason that spatial analyses would be considered by researchers. Spatial analyses in road safety typically involve the examination of crashes while taking their absolute or relative locations into account. Crashes face the typical issues of all point data: spatial dependence and spatial heterogeneity.

In simple terms, spatial dependence essentially refers to events at a location being highly influenced by events at neighboring locations. It is usually measured via spatial autocorrelation metrics. In turn, autocorrelation refers to the influence of variable values of given points on variable values of adjacent points (spatially or temporally). Spatial heterogeneity occurs in the modelled relationships as the coefficients between random parameters and observed events are not fixed spatially.

Therefore, researchers have discovered several caveats and merits in conducting spatial analysis. Road crashes are subject to both spatial and temporal variations (Loo and Anderson, 2015), intuitively suggesting spatial analyses as informative. By accounting for spatial dependence and heterogeneity in the estimates, spatial analyses describe how regions affect and are affected by the road safety attributes of their neighbors, and how the influence of explanatory parameters varies across space as well.

As a more specific example, when considering spatial correlation in crash models, estimates are effectively "pooling strength" from neighboring locations, thus improving the produced estimations (Aguero-Valverde and Jovanis, 2008). Road crashes are a complex phenomenon, and their analysis requires assumptions and merging of the examined parameters for a feasible approach, which unavoidably leads to some degree of loss of information or even misrepresentation of the actual

* Corresponding author.
E-mail address: apziak@central.ntua.gr (A. Ziakopoulos).

https://doi.org/10.1016/j.aap.2019.105323
Received 8 July 2019; Received in revised form 27 September 2019; Accepted 3 October 2019
Available online 22 October 2019

conditions (Xu and Huang, 2015). Spatial analyses can counterbalance this loss by providing predictions of counts of crashes (and of similar incidents, such as near-misses) that vary across different units of analyses, thus capturing all the unobserved trends and particularities of each area. Thus not only is better theoretical understanding provided for crash occurrence across space, but the identification of high-risk sites (known as hotspots) becomes more accurate (El-Basyouny and Sayed, 2009; Aguero-Valverde, 2014).

After decades of research, the topic of spatial analysis of traffic crashes covers a wide range, including mapping and visualization of crash counts, identifying clustering patterns of traffic collisions, and use of spatial models to investigate the effects of contributory factors and recommend targeted countermeasures. The mathematic particulars of spatial analyses have been examined in several published studies, for instance in Bivand et al. (2009) for Global and Local Moran's I and in Ver Hoef et al. (2018) for conditional autoregressive priors (CAR) models or simultaneous autoregressive priors (SAR) models. The reader is also referred to Yao et al. (2016), for a review of major advancements of spatial crash analysis using applied GIS tools. The examined research there starts from a significantly older period (in 1976) and includes topics that fall out of the scope of the present research, such as visualizing and mapping of events.

The aim of the present paper is to provide a review of the scientific literature regarding spatial approaches and spatial analyses in road safety. The present study is an endeavor to investigate how road safety researchers handle the dimension of space in its various aspects in their studies, whether that regards modelling of spatial events, selecting the scale of areal units or proximity structures, tackling boundary problems or other specific issues (such as vulnerable road users – VRUs). In order to achieve the aim of the current research, published scientific studies (in English) are critically examined. The selected studies were intended to be representative of a wide array of countries and adopted methodologies, in order to provide a well-rounded summary of the state-of-the art in road safety spatial analyses. Emphasis was given to more recent studies, with some seminal endeavors being included as well for completeness.

The main focus of the current study is on study characteristics, modelling approaches and methodological issues. It should be noted that this research only includes studies that conducted explicit and dedicated spatial or spatio-temporal analyses, as opposed to studies that examine different areas for purposes of cross-sectional or case-control studies (and as such do not examine the spatial aspect of road safety incidents). The second category of studies has its own merits and has been extensively implemented in road safety research, but falls out of the scope of this review.

This paper is organized as follows. Section 2 includes an examination of the different spatial units of analyses, together with famous boundary and zonal problems, as well as the issues of proximity structures. Section 3 outlines various modelling approaches, while Section 4 discusses issues in spatial analyses of VRUs. Finally, a discussion of overall findings from the review process and future research directions are provided in Section 5.

2. Examination of spatial units

Spatial analyses in road safety fundamentally involve the examination of road safety indicators (crash counts or rates, injury severity rates etc.) across spatial units of analyses. The manner in which researchers select and define these spatial units directly influences the scope of the study, as well as the interpretability of results, while this can apply to data preparation as well (Imprialou et al., 2016). There is a structural difference, for instance, in examining spatial distribution of road safety indicators in consequent road segments that feed traffic flow seamlessly into each other compared to examining junction clusters with several inflows and outflows for the distributions of the same indicators.

Different spatial units are discussed in the following section, and study characteristics for each spatial unit level are summarized on Tables 1–4. It was decided to include study characteristics initially considered by researchers on the Tables of this review, even if they were not found significant in the respective final models, to better showcase the scope of each research. The examined crash categories are denoted with the following acronyms with respect to the involved road users: Total Crashes (TC), Motorcycle crashes (MC), Single Vehicle crashes (V), Vehicle-vehicle crashes (V-V), Bicycle-vehicle crashes (B-V) and Pedestrian-vehicle crashes (P-V). When crash category details are not given about the examined crashes in a study, they are noted as TC. Additional details, such as the analysis of a specific crash type are noted as well.

2.1. Road segment and intersection approaches

Initial approaches of spatial analyses involved the more intuitive examination of road safety indicators across singular or multiple road sections, such as straight road segments and intersections. Earlier approaches involve the depiction and analysis of spatial distribution of crashes on (state) highways, in an attempt to perceive visual patterns of heightened concentration and possible correlation with touristic areas (Page and Meyer, 1996), albeit with a small sample. Furthermore, examination of the impact of the length of segments on crash counts and density which were found to follow Poisson distribution in the smaller segment scales growing from more intermediate distributions to normal distributions as segments increased, as shown by a study by Thomas (1996) that also first touched on the modifiable areal unit problem in road safety (discussed in section 2.6).

It has been determined that local environment and road infrastructure are critical factors of crash occurrence (Flahaut, 2004; Wang et al., 2016a). A traditional division when examining straight road segments is road type; highways with divided traffic directions display different road safety mechanisms than undivided two-lane expressways and for decades have been analyzed separately, a practice that is continued in segment-based spatial analyses.

The environment of road segments has been traditionally examined separately in the literature, with researchers distinguishing between urban and rural segments and often producing comparative analyses between different types of segments. A spatial analysis by Flahaut (2004) determined 2-lane configurations as the most unsafe configuration for rural roads. For urban roads, it has been found that increases in the number of crosswalks and the densities of unsignalized intersections both increase crash occurrence (Barua et al., 2014). Furthermore, local and non-local drivers are found to cluster along road segments, and segments with adverse safety interactions between these two groups are estimated to transfer these effects spatially to neighboring segments (Wang et al., 2016a).

In spatial analyses, researchers examine intersections either in groups (Guo et al., 2010; El-Basyouny and Sayed, 2011) or in aggregation (Miaou and Lord, 2003; Wang and Abdel-Aty, 2006). Intersection geometry, location and traffic parameters are important within the context of spatial analyses. The size of intersection, the traffic conditions by turning movement, and the coordination of signal phase have significant impacts on the number of crashes at intersections (Guo et al., 2010). Xie et al. (2013) have shown intersections on segments with lower mean speeds were associated with fewer crashes than those with higher speeds, and that intersections on two-way roads, under elevated roads, and in close proximity to each other, tended to have higher crash frequencies as well. A seminal result of a study by Abdel-Aty & Wang (2006) shows that overall, three-legged intersections tend to exhibit lower crash rates than four-legged intersections, and that they exhibit different road safety mechanisms. Furthermore, effectiveness of implemented road safety treatments can vary between locations when considering injury severity levels (El-Basyouny and Sayed, 2011).

When proximal segments are considered, with the layout of a simple

Table 1
Studies with road safety spatial analyses primarily on the individual road segment/intersection level

Study Characteristics		Dependent variables				Independent variables – parameters						
							Traffic				Road user	
Author(s) (Year)	Country of study	Crash type analyzed	Crash count/ frequency	Crash rate	Injury Severity	Casualty rate	Speed	Traffic volume	Vehicle distance traveled	Number of Trips - OD	Road user/ Population age	Modal distinction
Abdel-Aty and Wang (2006)	United States	TC	●				●	●				
Aguero-Valverde (2014)	United States	TC	●					●				
Aguero-Valverde and Jovanis (2010)	United States	TC	●					●				
Aguero-Valverde and Jovanis (2008)	United States	TC	●					●				
Aguero-Valverde et al. (2016)	United States	TC (6 Crash types)	●					●				
Alarifi et al. (2018)	United States	TC	●					●				●
Alarifi et al. (2017)	United States	TC	●					●				●
Barua et al. (2016)	Canada	TC	●		○			●				
Barua et al. (2014)	Canada	TC	●		○			●				
Chiou et al. (2014)	Taiwan	TC	●		●			●				●
Effati et al. (2015)	Iran	TC			●						●	
El-Basyouny and Sayed (2011)	Canada	TC	●		○			●				
El-Basyouny & Sayed (2009)	Canada	TC	●					●				
Guo et al. (2010)	United States	TC	●					○				
Huang et al. (2017)	China	TC | V/V-V P-V | B-V	●					●			●	○
Huang et al. (2016)	United States	TC	●					●	●	●	●	
Flahaut (2004)	Belgium	TC	●		○			●				
Liu et al. (2017)	United States	TC	●					●				●
Ma et al. (2017)	United States	TC	●		○		●		●			
Miaou & Lord (2003)	Canada	TC		●				●				
Miaou & Song (2005)	Canada | United States	TC	●	●	●			●	●			
Mitra (2009)	United States	TC	●		●			●				
Mountrakis & Gunson (2009)	United States	V-A	●									
Page & Meyer (1996)	New Zealand	TC	●		○							
Thomas (1996)	Belgium	TC	●		○			○				
Wang & Abdel-Aty (2006)	United States	V-V (rear-end only)	●					●				
Wang & Huang (2016)	United States	TC	●					●		●		
Wang et al. 2016a	United States	TC	●		●			●				
Wang et al. (2009)	United Kingdom	TC	●		○		●	●				
Wen et al. (2019)	China	TC	●						●			
Xie et al. (2014)	China	TC	●				●	●				
Xie et al. (2013)	China	TC	●				●	●				
Zeng & Huang (2014)	United States	TC	●				●					

Study Characteristics		Road environment							Spatial aggregation approach			Analysis - Modelling approach
Author(s) (Year)	Country of study	Speed Limit	Curvature	Gradient	Lane width	Lane number	Intersection nr./ density	Roadway length	Regional level	Zonal level	Link/ segment/ intersection level	
Abdel-Aty and Wang (2006)	United States	●			●	●	○	●			Intersections	Negative Binomial Regression with and without Generalized

(*continued on next page*)

Table 1 (*continued*)

Study Characteristics												Analysis - Modelling approach
		Road environment							Spatial aggregation approach			
Author(s) (Year)	Country of study	Speed Limit	Curvature	Gradient	Lane width	Lane number	Intersection nr./ density	Roadway length	Regional level	Zonal level	Link/ segment/ intersection level	
												estimating equations | Cluster analysis
Aguero-Valverde (2014)	United States	●			●	○		●			Rural road segments	Full Bayes hierarchical Poisson model (1) with normal priors for spatial random effects | (2) with CAR priors for spatial random effects | (3) with a joint distribution
Aguero-Valverde and Jovanis (2010)	United States	●	●	●	●	○		●			Rural & Urban road segments	Full Bayes hierarchical Poisson model with CAR priors for spatial random effects
Aguero-Valverde and Jovanis (2008)	United States	●			●	○		●			Rural road segments	Bayesian Multivariate Poisson Lognormal Regression | Bayesian random effects models
Aguero-Valverde et al. (2016)	United States							●			Rural road segments	Full Bayes Poisson Regressions (Univariate, Univariate Spatial, Multivariate, Multivariate Spatial)
Alarifi et al. (2018)	United States	●				●	●	●			Intersections | Road segments	13 Bayesian hierarchical Poisson-lognormal joint spatial models with adjacency-based, adjacency-route, distance-order, and distance-based spatial weight features
Alarifi et al. (2017)	United States	●				●	●	●			Intersections | Road segments	Multilevel Poisson-lognormal joint model (1,2) with corridor and sub-corridor random effects (3,4) with corridor and sub-corridor random parameters
Barua et al. (2016)	Canada					●	●	●			Urban road segments	Full Bayesian Poisson lognormal multivariate random parameters models (1) with heterogenous effects (2) with CAR priors for spatial heterogeneity (3) with both
Barua et al. (2014)	Canada					●	●	●			Urban road segments	Full Bayesian Poisson lognormal univariate and multivariate random parameters models (1) with heterogenous effects (2) with CAR priors for spatial heterogeneity (3) with both
Chiou et al. (2014)	Taiwan		●	●		●	○	●			Highway segments	Multinomial-generalized Poisson with error-components (spatial error and spatial exogenous)
Effati et al. (2015)	Iran			●	●		●	●			Highway segments	Support Vector Machine Algorithms (SVMs) | Coactive neuro-fuzzy inference system
El-Basyouny and Sayed (2011)	Canada						○					

(*continued on next page*)

Table 1 (*continued*)

Study Characteristics		Road environment							Spatial aggregation approach			Analysis - Modelling approach
Author(s) (Year)	Country of study	Speed Limit	Curvature	Gradient	Lane width	Lane number	Intersection nr./density	Roadway length	Regional level	Zonal level	Link/segment/intersection level	
											Intersections	Univariate and Multivariate Poisson Lognormal Regressions | Full Bayes estimations
El-Basyouny & Sayed (2009)	Canada					●	●	●			Urban road segments	Full Bayesian Multivariate Poisson Lognormal with and without CAR Prior | Full Bayesian Multiple Membership model | Full Bayesian Extended Multiple Membership model
Guo et al. (2010)	United States	●				○		○			Intersections	Fixed effects Bayesian Poisson Regression | Fixed and Mixed effects Bayesian Negative Binomial Regression | Spatial CAR Prior extended Poisson/Negative Binomial models
Huang et al. (2017)	China	●					○				Intersections	Poisson Regression (Univariate, Multivariate Lognormal & Spatial random effects models)
Huang et al. (2016)	United States	●				●	●	●		TAZ	Intersections | Road segments	Bayesian spatial model with CAR prior (macroscopic) | Bayesian spatial joint models with CAR prior (microscopic)
Flahaut (2004)	Belgium	●		○		●	○	○			Rural & Highway segments	Logistic regression with and without spatial autocorrelation
Liu et al. (2017)	United States	●						●			Highway segments	Geographically Weighted Negative Binomial Regression | Negative Binomial Regression
Ma et al. (2017)	United States			●		●					Highway segments	Hierarchical Bayesian random parameters models (structured and unstructured spatio-temporal effects)
Miaou & Lord (2003)	Canada	○					○				Intersections	Full Bayes | Empirical Bayes
Miaou & Song (2005)	Canada | United States	○				○		●			Intersections | Rural segments	Multivariate spatial Bayesian generalized linear mixed models with and without CAR Prior
Mitra (2009)	United States										Intersections	Hierarchical Full Bayes Jointly specified spatial model | Negative Binomial Regression | Local Moran's I
Mountrakis & Gunson (2009)	United States							○			Rural segments	Spatial, Temporal & Spatiotemporal kernel estimation | Ripley's K-function
Page & Meyer (1996)	New Zealand							○	National Parks		Highway segments	Percentage descriptive statistics
Thomas (1996)	Belgium							●				

(*continued on next page*)

Table 1 (*continued*)

Study Characteristics		Road environment							Spatial aggregation approach			Analysis - Modelling approach
Author(s) (Year)	Country of study	Speed Limit	Curvature	Gradient	Lane width	Lane number	Intersection nr./ density	Roadway length	Regional level	Zonal level	Link/ segment/ intersection level	
											Highway segments	Univariate and bivariate descriptive statistics, chi^2 and W tests
Wang & Abdel-Aty (2006)	United States	●				●	○				Intersections	Generalized Estimating Equations with Negative Binomial link function
Wang & Huang (2016)	United States	●				●	●	●		TAZ	Intersections | Urban segments	Bayesian hierarchical joint Poisson Regression | Bayesian joint Poisson Regression | Negative Binomial Regression
Wang et al. 2016a	United States	●	●	●			●	●			Highway segments	Multivariate Poisson Lognormal regression with CAR Prior
Wang et al. (2009)	United Kingdom		●	●		●		●			Highway segments	Bayesian Multivariate Poisson Lognormal | Negative Binomial Regression | Poisson Models with CAR priors (with first/second order neighbors)
Wen et al. (2019)	China		●	●							Highway segments	(1) Poisson Lognormal regression with CAR Prior | (2) Poisson Lognormal regression with spillover effects | (3) Hybrid of (1) and (2)
Xie et al. (2014)	China					●	○	●			Intersections | Urban segments	Bayesian Negative Binomial regression (basic, random effect, random parameter, hierarchical, hierarchical CAR)
Xie et al. (2013)	China					●	○	●			Intersections | Urban segments	Bayesian Negative Binomial regression (basic, random parameter, hierarchical)
Zeng & Huang (2014)	United States	●				●	●	●			Intersections | Urban segments	Poisson Regression | Negative Binomial Regression | Bayesian spatial model with CAR prior | Bayesian spatial joint models with CAR prior

● Considered in the study design, ○ considered in the study process as filter/defining characteristic

Table 2
Studies with road safety spatial analyses primarily on the zonal level

	Study Characteristics		Dependent variables								Independent variables – parameters		
							Traffic				Road environment		
Author(s) (Year)	Country of study	Crash type analyzed	Crash count/ frequency	Crash rate	Injury Severity	Casualty rate	Speed	Traffic volume	Vehicle distance traveled	Number of Trips - OD	Speed Limit	Curvatu-re	Lane width
Abdel-Aty et al. (2013)	United States	TC	●		○				●	●	●		
Abdel-Aty et al.(2011)	United States	TC	●		○		○			●	●		
Amoh-Gyimah et al. (2017)	Australia	TC	●		○				●		●		
Anderson (2007)	United Kingdom	TC	●		○								
Anderson (2009)	United Kingdom	TC | P-V | B-V	●		○								
Bao et al. (2018)	United States	TC	●		○				●	●	○		
Bao et al. (2017)	United States	TC | V-V | P-V	●					●		●			
Cai et al. (2019a)	United States	TC	●						●				
Cai et al. (2018)	United States	TC	●					●					
Cai et al. (2017b)	United States	TC | P-V | B-V	●						●				
Cai et al. (2016)	United States	P-V | B-V	●						●	●			
Cottrill & Thakuriah (2010)	United States	P-V	●	●	○			●					
Cui et al. (2015)	Canada	TC (on boundary)		●									
Delmelle & Thill (2008)	United States	B-V	●										
Dong et al. (2016)	United States	TC	●						●				
Dong et al. (2015)	United States	TC	●						●	●	○		
Dong et al. (2014)	United States	TC	●						●	●	○		
Erdogan et al. (2008)	Turkey	TC	●		●		○					○	○
Gomes et al. (2017)	Brazil	TC	●		○								
Guo et al. (2017)	Hong Kong	P-V	●		○		●	●		●			
Hadayeghi et al. (2010)	Canada	TC	●				●	●	●				
Hadayeghi et al. (2003)	Canada	TC	●		○		●	●	●				
Jiang et al. (2016)	United States	TC | B-V | P-V	●		○				●				
Ladron de Guevara et al. (2004)	United States	TC	●		○	○			●				
LaScala et al. (2004)	United States	P-V | B-V	●		○			●					
LaScala et al. (2000)	United States	P-V			●			●					
Lee & Abdel-Aty (2018)	United States	B-V	●						●		●		
Lee et al. (2018)b	United States	Crashes of 8 road user types	●	●					●				
Lee et al. (2017)a	United States	TC | P-V | B-V	●		○			●					
Lee et al. (2015)a	United States	V/V-V | P-V | B-V	●						●		●		
Lee et al. (2015b)	United States	P-V	●						●		●		
Lee et al. (2014a)	United States	V/V-V (at-fault)	●										
Lee et al. (2014b)	United States	TC	●		○				●		○		
Levine et al. (1995)	United States	TC	●		○								
Loukaitou-Sideris et al. (2007)	United States	P-V	●		○			●					
Lovegrove & Sayed (2007)	Canada	TC	●		○				●				
Lovegrove & Sayed (2006)	Canada	TC	●		○		●		●	●			
Lovegrove et al. (2009)	Canada	TC	●		○				●		○		
MacNab (2004)	Canada	TC			●								
Naderan & Shahi (2010)	Iran	TC	●		○					●			
Narayanamoorthy et al. (2013)	United States	P-V | B-V	●		●								
Nashad et al. (2016)	United States	P-V | B-V	●						●				

Table 2 (*continued*)

	Study Characteristics		Dependent variables				Traffic				Independent variables – parameters Road environment		
Author(s) (Year)	Country of study	Crash type analyzed	Crash count/ frequency	Crash rate	Injury Severity	Casualty rate	Speed	Traffic volume	Vehicle distance traveled	Number of Trips - OD	Speed Limit	Curvatu-re	Lane width
Ng et al. (2002)	China	TC | P-V	●		○								
Noland & Quddus (2005)	United Kingdom	TC | P-V	●		○								
Noland & Quddus (2004)	United Kingdom	TC	●		○			○					
Pirdavani et al. (2014)a	Belgium	TC	●		○			●	●	●	●		
Pirdavani et al. (2014)b	Belgium	V-V P-V | B-V	●		○				●	●			
Pirdavani et al. (2013)	Belgium	V-V P-V | B-V	●		○				●	●			
Quddus (2008)	United Kingdom	TC	●		○		●	●				●	
Rhee et al. (2016)	South Korea	TC	●		○				●	●	○		
Siddiqui & Abdel-Aty (2012)	United States	P-V (interior & boundary)	●								●		
Siddiqui et al. (2012)	United States	P-V | B-V	●				○				●		
Soltani & Askari (2017)	Iran	V-V	●		●								
Tasic et al. (2017)	United States	TC | V-V | P-V | B-V	●		○				●	●			
Ukkusuri et al. (2012)	United States	P-V	●		○								●
Ukkusuri et al. (2011)	United States	P-V	●										
Wang et al. (2016)b	China	P-V	●		○								
Wang & Kockelman (2013)	United States	P-V	●		○				●				
Wei & Lovegrove (2013)	Canada	B-V	●					●					
Wier et al. (2009)	United States	P-V	●		○			●					
Xu and Huang (2015)	United States	TC	●		●		○		●		●		
Xu et al. (2017)a	United States	TC (interior & boundary)	●		○				●	●	●		
Xu et al. (2017)b	United States	TC	●					●		●			
Yasmin & Eluru (2016)	Canada	B-V	●					●					
Zhai et al. (2019)a	United States	TC (interior & boundary)	●		●					●	●		
Zhai et al. (2018)	United States	TC (interior & boundary)	●						●	●	●		

	Study Characteristics	Independent variables – parameters Road environment			Demographic			Socio-economic		Land Use				
Author(s) (Year)	Country of study	Lane number	Intersecti-on nr./ density	Roadway length	Populatio-n number/ density	Road user/ Populatio-n age	Modal distinc-tion	Househol-d/ Personal income	Employm-ent percen-tage/ density	Land use factor(s)	Regional level	Zonal level	Link/ segment/ intersec-tion level	
Abdel-Aty et al. (2013)	United States		●	●	●	●		●		●		TAZ | CT | BG	Intersecti-ons	Bayesian Multivariate Poisson Lognormal Regression
Abdel-Aty et al.(2011)	United States		●	●			○					TAZ		Negative Binomial Regression
Amoh-Gyimah et al. (2017)	Australia				●	●	●	●		●				Random parameter negative binomial model |

(*continued on next page*)

Table 2 (*continued*)

	Study Characteristics	Independent variables – parameters												
		Road environment			Demographic			Socio-economic		Land Use				
Author(s) (Year)	Country of study	Lane number	Intersection nr./ density	Roadway length	Population number/ density	Road user/ Population age	Modal distinction	Household/ Personal income	Employment percentage/ density	Land use factor(s)	Regional level	Zonal level	Link/ segment/ intersection level	
												SA1 | SA2 | TAZ | SED | ZIP		Semi-parametric Poisson GWR (also on custom grid cells)
Anderson (2007)	United Kingdom											CT	Urban road segments	Kernel density estimation | Network analysis | Census Output Area estimation
Anderson (2009)	United Kingdom		○	●			●			●		Hotspot clusters		Kernel density estimation | K-means clustering
Bao et al. (2018)	United States		●	●	●	●		●	●			ZIP		Poisson GWR | Latent Dirichlet Allocation
Bao et al. (2017)	United States		●	●	●	●	○	●	●			TAZ		Geographically Weighted Regression (GWR)
Cai et al. (2019a)	United States	●	●	●	●	●		●		●		TAD		Bayesian Poisson Lognormal Regression: (1) at macro- level; (2) at micro- level; (3) integrated at macro- and micro- levels
Cai et al. (2018)	United States	●	●	●	●	●		●		○	County	TAD		Poisson-lognormal models: (1) Fixed param. univariate model; (2) Grouped random param. univ. spatial model; (3) Grouped random param. univ. spatial model with zonal factors; (4) Grouped random param. multiv. spatial model with zonal factors
Cai et al. (2017b)	United States		●	●	●	●	●	●	●			TAD		Bayesian Negative Binomial regression | Bayesian Logit regression model | Bayesian Joint model [of the two] | Elasticity analysis
Cai et al. (2016)	United States		●	●	●		●		●	●		TAZ		Negative Binomial spatial and aspatial models (basic, zero-inflated & hurdle)
Cottrill & Thakuriah (2010)	United States			●	●	●	○	●		●		EJ (CT)		Poisson Regression with heterogeneity | Poisson Regression with exogenous underreporting
Cui et al. (2015)	Canada		●	●							2 city areas	Neighborhoods		(1) Entropy-based histogram thresholding

(*continued on next page*)

Table 2 (*continued*)

Author(s) (Year)	Study Characteristics	Independent variables – parameters												
		Road environment			Demographic			Socio-economic		Land Use				
	Country of study	Lane number	Intersection nr./density	Roadway length	Population number/density	Road user/Population age	Modal distinction	Household/Personal income	Employment percentage/density	Land use factor(s)	Regional level	Zonal level	Link/segment/intersection level	
														(2) Collision density probability distribution (3) Collision aggregation through density ratio
Delmelle & Thill (2008)	United States		●	○	●	●	○	●		●		CT		OLS Regression | Kernel density
Dong et al. (2016)	United States					●	●		●			TAZ		Bayesian Multivariate Poisson Lognormal Regression | Bayesian spatial-temporal interaction models
Dong et al. (2015)	United States		●	●				●		●		TAZ		ν-Support Vector Machine with Correlation-based Feature Selector | Bayesian Multivariate Poisson Lognormal with CAR Prior
Dong et al. (2014)	United States		●	●	●			●		●		TAZ		Bayesian Multivariate Poisson Lognormal with CAR Prior Regression for boundary and non-boundary area models
Erdogan et al. (2008)	Turkey			●								Hotspot clusters		Poisson test | Chiˆ2 test | Kernel density analysis
Gomes et al. (2017)	Brazil		●	●		●		●		●		TAZ		Negative binomial regression | Poisson GWR | Negative Binomial GWR
Guo et al. (2017)	Hong Kong		●	●		●	○			●		TAZ		Space Syntax | Poisson Lognormal Regression | Bayesian Poisson Lognormal with CAR Prior Regression with (1) contiguity (2) geometry-centroid distance and (3) road network connectivity
Hadayeghi et al. (2010)	Canada		●	●	●	●			●	●		TAZ		Poisson GWR | Negative Binomial Regression | Poisson regression
Hadayeghi et al. (2003)	Canada		●	●	●	●			●	●		TAZ		GWR | Negative Binomial Regression
Jiang et al. (2016)	United States		●	●	●	●	○	●		●		TAZ		Random Forest Models (CART trees) | Wiloxon Tests
Ladron de Guevara et al. (2004)	United States		●	●	●	●			●	●		TAZ		

(*continued on next page*)

Table 2 (*continued*)

Author(s) (Year)	Study Characteristics: Country of study	Independent variables – parameters: Road environment: Lane number	Intersection nr./ density	Roadway length	Demographic: Population number/ density	Road user/ Population age	Modal distinction	Socio-economic: Household/ Personal income	Employment percentage/ density	Land Use: Land use factor(s)	Regional level	Zonal level	Link/ segment/ intersection level	
														Negative Binomial Regression | Simultaneous equation estimation
LaScala et al. (2004)	United States			●	●	●	○	●	●	●	Communities	Geographic units		Linear regression models
LaScala et al. (2000)	United States		○	●	●	●	○	●	●	●		CT		Spatial autocorrelation regression log-linear model
Lee & Abdel-Aty (2018)	United States		●		●	●	●	●	●	●		ZIP		Bayesian Poisson lognormal CAR models
Lee et al. (2018)b	United States		●	●	●	●	●	●	●	●		TAZ		Fractional Split Multinomial Model
Lee et al. (2017)a	United States		○		●	●	○	●		●	County | County Division	TAD | ZIP | TAZ | CT | BG | CB	Intersections	Mixed effects Negative Binomial models with: (1) micro-level variables, (2) micro- and macro-level variables and (3) micro- and macro-level variables with random-effects
Lee et al. (2015)a	United States				●	●	○		●	●		TAZ		Univariate and Multivariate Bayesian Poisson Lognormal with CAR Prior Regression
Lee et al. (2015b)	United States		●	●	●	●	●	●	●	●		ZIP		Bayesian Poisson lognormal simultaneous equations spatial error model
Lee et al. (2014a)	United States				●	●	●	●	●			ZIP		Bayesian Poisson-lognormal model
Lee et al. (2014b)	United States			●	●	●				●		TSAZ | TAZ		Brown-Forsythe test | Bayesian Multivariate Poisson Lognormal Regression
Levine et al. (1995)	United States		○	●	●				●	●		BG		Spatial lag regression model
Loukaitou-Sideris et al. (2007)	United States		○	○	●	●	○	●	●	●		CT		OLS regression
Lovegrove & Sayed (2007)	Canada		●	●	●				●	●		Neighborhood - TAZ		Groups of Macrolevel Crash Prediction Models using GLMs
Lovegrove & Sayed (2006)	Canada		●	●	●			●	●	●		Neighborhood - TAZ		Groups of Macrolevel Crash Prediction Models using GLMs
Lovegrove et al. (2009)	Canada		●	●	●				●	●		TAZ		Groups of Collision Prediction GLMs | Modified T-tests
MacNab (2004)	Canada				●	●		●	●	●				

(*continued on next page*)

Table 2 (*continued*)

Author(s) (Year)	Study Characteristics	Independent variables – parameters												
		Road environment			Demographic			Socio-economic		Land Use				
	Country of study	Lane number	Intersection nr./density	Roadway length	Population number/density	Road user/Population age	Modal distinction	Household/Personal income	Employment percentage/density	Land use factor(s)	Regional level	Zonal level	Link/segment/intersection level	
												Local health area		Bayesian spatial model with spatial autocorrelation
Naderan & Shahi (2010)	Iran				●							TAZ		Negative Binomial regression
Narayanamoorthy et al. (2013)	United States			○	●	●	●	●		●		CT		Customized generalized ordered-response spatial multivariate count model
Nashad et al. (2016)	United States		●	●	●		●		●	●		sTAZ		Negative binomial regression (copula-based)
Ng et al. (2002)	China				●		○			●		TAZ		Negative Binomial Regression with Empirical Bayes approach | Cluster Analysis
Noland & Quddus (2005)	United Kingdom		●	●	●		●	●	●	●		Enumeration District		Negative Binomial Regression | ANOVA
Noland & Quddus (2004)	United Kingdom		●	●	●	●		●	●	●		Ward		Negative Binomial Regression
Pirdavani et al. (2014)a	Belgium		●	●	●			●	●			TAZ		Geographically Weighted GLM | Negative Binomial Regression
Pirdavani et al. (2014)b	Belgium		●	○			●	●				TAZ		Geographically Weighted Regression (GWR)
Pirdavani et al. (2013)	Belgium		●				○	●				TAZ		Negative Binomial regression Zonal Crash Prediction Models
Quddus (2008)	United Kingdom		●	●	●	●	○		●			Ward		Negative Binomial Regression | Spatial autoregressive model | Spatial error model | Bayesian hierarchical models for spatial units
Rhee et al. (2016)	South Korea	○	●	●	●	●		●	●	●		TAZ		OLS regression | Spatial lag regression | Spatial error regression | Poisson GWR
Siddiqui & Abdel-Aty (2012)	United States		●	●	●		○		●	●		TAZ		Multivariate Negative Binomial regression | Multivariate Bayesian Negative Binomial regression for boundary and non-boundary area models
Siddiqui et al. (2012)	United States		●	●	●		○	●	●	●		TAZ		Bayesian Multivariate Poisson Lognormal |

(*continued on next page*)

Table 2 (*continued*)

Author(s) (Year)	Study Characteristics	Independent variables – parameters												
		Road environment			Demographic			Socio-economic		Land Use				
	Country of study	Lane number	Intersection nr./density	Roadway length	Population number/density	Road user/Population age	Modal distinction	Household/Personal income	Employment percentage/density	Land use factor(s)	Regional level	Zonal level	Link/segment/intersection level	
														Negative Binomial Regression
Soltani & Askari (2017)	Iran				●		○			●		TAZ		Moran's I | Getis-Ord Gi* index
Tasic et al. (2017)	United States		●	●	●		●	●	●	●		CT		Generalized Additive Models
Ukkusuri et al. (2012)	United States	●	●	●	●	●				●		CT | ZIP		Negative binomial regression | Negative binomial regression with heterogeneity in dispersion parameter | Zero-inflated negative binomial regression
Ukkusuri et al. (2011)	United States	○	●	●	●	●	○			●		CT		Negative Binomial Regression with random parameters
Wang et al. (2016)b	China		●	●	●		○			●		TAZ		Bayesian Conditional Autoregressive (CAR) models with seven different spatial weight features
Wang & Kockelman (2013)	United States			●	●		○		●	●		CT		Multivariate Poisson Lognormal Regression with and without CAR Priors
Wei & Lovegrove (2013)	Canada		●	●	●	●	●	●	●	●		TAZ		Negative Binomial Macrolevel Crash Prediction Models
Wier et al. (2009)	United States		●	●	●	●	○		●	●		CT		Log-linear multivariate OLS regression model
Xu and Huang (2015)	United States		●	●	●			●				TAZ		Negative Binomial regression | Bayesian negative binomial model with CAR prior | Random parameter negative binomial model | Semi-parametric Poisson GWR
Xu et al. (2017)a	United States		●	●	●	●		●	●			TAZ		Bayesian spatially varying coefficients model
Xu et al. (2017)b	United States		●	●	●	●	●	●	●	●		TAZ		Semi-parametric Poisson GWR | One-way ANOVA tests
Yasmin & Eluru (2016)	Canada		●	●	●		●	●	●	●		TAZ		Poisson Regression | Negative Binomial regression (basic and Latent Segmentation)

Table 2 (*continued*)

Author(s) (Year)	Study Characteristics	Independent variables – parameters												
		Road environment			Demographic			Socio-economic		Land Use				
	Country of study	Lane number	Intersection nr./density	Roadway length	Population number/density	Road user/Population age	Modal distinction	Household/Personal income	Employment percentage/density	Land use factor(s)	Regional level	Zonal level	Link/segment/intersection level	
Zhai et al. (2019)a	United States		●	●	●	●		●				BG | TAZ | CT | ZIP		Bayesian Poisson-lognormal models with Multivariate CAR priors
Zhai et al. (2018) Bayesian Poisson-lognormal model with CAR prior	United States		●	●	●			●					TAZ	

● Considered in the study design, ○ considered in the study process as filter/defining characteristic

road network, it is important to note that there are spatial correlations between intersections and their adjacent segments, which have been found to be significant in the literature (Abdel-Aty and Wang, 2006; Quddus, 2008; Aguero-Valverde and Jovanis, 2010; Dong et al., 2014; Dong et al., 2015; Wang and Huang, 2016). Spatial correlation is also found in crashes of intersections along the same corridor, due to similar traffic flow patterns, presence of traffic signals and geographic characteristics (Guo et al., 2010), an issue which ought to be properly addressed with proper modelling tools (Xie et al., 2014). Additionally, several studies have integrated corridor-level characteristics into segment-level or intersection-level analysis in an effort to capture factors explaining heterogeneity (Abdel-Aty and Wang, 2006; Guo et al., 2010; Xie et al., 2014).

A different effort was made by Zeng and Huang (2014), who endeavored to model crash counts on road segments and intersections simultaneously. They used Bayesian spatial joint models to account for spatial correlations between adjacent road segments and intersections that were found to be more accurate than simple Poisson and negative binomial models. The joint model integrated junctions and segments to the basic link function. An indicator variable which denoted whether a segment or intersection was examined was utilized. The authors highlight that the spatial correlations between intersections and their connected segments were more significant than those found between intersections or between segments only, presumably due to common unobserved parameters such as speed. The approach of joint simultaneous modelling of intersections and segments was further advanced by Alarifi et al. (2017) who developed four multi-level Bayesian joint models for that purpose. Specifically, the reasoning was to complement the intersection/segment examination by including corridor-level characteristics in the models. Because corridor characteristics vary along their length, random forest models were used to divide corridors into-sub corridors of fixed-value characteristics. Ultimately there were statistically significant variables at the segment level, at the intersection level and at the corridor/sub-corridor level; the importance of median opening density for crash occurrence was underlined from the results. However, spatial autocorrelation of adjacent road entities was not examined in that study. Moreover, Alarifi et al. (2018) (discussed in Section 2.7) also conducted analyses including intersection-, road segment- and corridor-level parameters, in an attempt to explore that research question.

Reviewed studies that primarily focus on spatial analyses at the individual road segment/intersection level are shown on Table 1.

2.2. Zonal approaches

A number of zonal units have been adopted by researchers, from smaller to larger ones. Their boundaries can be census-based, administrative-based or traffic-based, and are dependent on the country or environment of study. Studies in the UK might utilize enumeration districts, namely areas averaging circa 200 households (Noland and Quddus, 2005) or census wards, which include about 2000 households (Noland and Quddus, 2004; Quddus, 2008). Similarly, studies from other countries have used locally available spatial units, such as the Australian ABS structure units (Statistical areas 1,2 (SA1,2), state electoral divisions (SED)) used by Amoh-Gyimah et al. (2017).

Many studies originate from the US and have utilized units that are used there: Census Blocks (CBs) are the smallest unit, averaging 85 people and are expanded to Census Block Groups (CBGs), averaging 39 blocks with about 1500 people (Lee et al., 2017a). CBGs have been utilized by road safety researchers to some extent (Levine et al., 1995; Abdel-Aty et al., 2013).

Traffic Analysis Zones (TAZs) are created primarily in the US with the explicit purpose of collecting trip and traffic statistics and data, though they have been implemented in other countries as well (Ng et al., 2002; Gomes et al., 2017). From traditional zonal approaches, TAZs are the only traffic-related zone system (Lee et al., 2017a), which

Table 3
Studies with road safety spatial analyses primarily on the regional level

Study Characteristics			Dependent variables				Independent variables – parameters						
							Traffic				Road environ-ment		
Author(s) (Year)	Country of study	Crash type analyzed	Crash count/ frequency	Crash rate	Injury Severity	Casualty rate	Speed	Traffic volume	Vehicle distance travelled	Number of Trips - OD	Speed Limit	Curvature	Gradient
Aguero-Valverde (2013)	Costa Rica	TC	●		●				●				
Aguero-Valverde & Jovanis (2006)	United States	TC	●		○				●				
Atubi (2012)	Nigeria	TC	●		○								
Bu et al. (2018)	United States	TC	●		●			●			●		
Erdogan (2009)	Turkey	TC		●	●	●							
Flask & Schneider (2013)	United States	MC	●		○							●	●
Han et al. (2018)	United States	TC	●					●					
Huang et al. (2010)	United States	TC	●		●			○	●				
LaScala et al. (2001)	United States	P-V			●	●		●					
Lee et al. (2019)a	United States	P-V		○	○	●				●			
Lee et al. (2019)b	Italy, United States	TC | P-V | B-V	●										
Lee et al. (2018)a	United States	TC	●		○								
Lee et al. (2018)c	United States	P-V | B-V	●		○					●			
Lee et al. (2017)b	United States	MC	●		○								
Li et al. (2019)	United States	TC	●		○				●				
Li et al. (2013)	United States	TC	●		○				●	●			
Liu and Sharma (2018)	United States	TC	●		●				●				
Moeinaddini et al. (2014)	20 Cities Worldwide	TC	●		○								
Noland & Oh (2004)	United States	TC	●		○			●				●	
Song et al. (2006)	United States	TC	●		○				○			●	
Zhai et al. (2019)b	Hong Kong	P-V			●								

Study Characteristics		Independent variables – parameters										Spatial aggrega-tion approach	Analysis - Modelling approach
					Demogra-phic			Socio-economic			Land Use		
Author(s) (Year)	Country of study	Lane width	Lane number	Intersecti-on nr./ density	Roadway length	Populatio-n number/ density	Road user/ Populatio-n age	Modal distinction	Household/ Personal income	Employme-nt percentage/ density	Land use factor(s)	Regional level	
Aguero-Valverde (2013)	Costa Rica				●	●	●		●			Canton	Full Bayes hierarchical approach Poisson multivariate CAR model for spatial random effects.
	United States				●	●	●					County	

(*continued on next page*)

Table 3 (*continued*)

Study Characteristics		Independent variables – parameters										Spatial aggregation approach	Analysis - Modelling approach
					Demographic			Socio-economic			Land Use		
Author(s) (Year)	Country of study	Lane width	Lane number	Intersection nr./ density	Roadway length	Population number/ density	Road user/ Population age	Modal distinction	Household/ Personal income	Employment percentage/ density	Land use factor(s)	Regional level	
Aguero-Valverde & Jovanis (2006)													Negative Binomial Regression | Full Bayesian hierarchical models
Atubi (2012)	Nigeria				●	●						State	Multivariate linear regression
Bu et al. (2018)	United States		●			●						Metropolitan areas	Simple Density distribution analysis
Erdogan (2009)	Turkey				●	●		●		●		County	Moran's I and Geary's c values, Z and G statistics
Flask & Schneider (2013)	United States		○		●	●	●		●			County | Township	Bayesian Negative Binomial Regression with mixed effects
Han et al. (2018)	United States		○	●	●							County (spec. road type)	Bayesian hierarchical random parameter model | Bayesian hierarchical random intercept model | Bayesian Poisson lognormal model
Huang et al. (2010)	United States			●	●	●	●	○	●	●	●	County	Bayesian Spatial CAR Priors regression
LaScala et al. (2001)	United States			●	○	●	●	○	●	●	●	Communities	Spatial autocorrelation regression log-linear model
Lee et al. (2019)a	United States					●	●	●	●		●	Metropolitan areas	Multiple linear regression model integrated in a Poisson Lognormal Model
Lee et al. (2019)b	Italy, United States					●	●	●	●			County | Provincia	Negative Binomial Regression | Calibration factors | Transferability Indexes
Lee et al. (2018)a	United States											State	Crash Modification Factors
Lee et al. (2018)c	United States					●	●	●	●			Metropolitan areas	Bayesian integrated and non-integrated Bivariate Models
Lee et al. (2017)b	United States					●			●	●	●	County | Parish	Before-and-After Study (1) with Comparison Group | (2) With Empirical Bayes | Safety Performance Functions | Crash Modification Factors
Li et al. (2019)	United States		●			●	○		●	●	●	County	Hierarchical Bayesian random parameters

(*continued on next page*)

Table 3 (*continued*)

Study Characteristics		Independent variables – parameters										Spatial aggregation approach	Analysis - Modelling approach
					Demographic			Socio-economic			Land Use		
Author(s) (Year)	Country of study	Lane width	Lane number	Intersection nr./density	Roadway length	Population number/density	Road user/Population age	Modal distinction	Household/Personal income	Employment percentage/density	Land use factor(s)	Regional level	
													models (structured and unstructured spatio-temporal effects)
Li et al. (2013)	United States				●	●	●		●	●		County	Negative Binomial Regression | Poisson GWR
Liu and Sharma (2018)	United States								●	●	●	County	Hierarchical Bayesian random parameters models (structured and unstructured spatio-temporal effects)
Moeinaddini et al. (2014)	20 Cities Worldwide			●	●							City	Gamma-distributed GLM
Noland & Oh (2004)	United States	●	●		●	●			●			County	Negative Binomial Panel Regression
Song et al. (2006)	United States			○								County	Bayesian Multivariate Poisson Lognormal Regression with and without CAR Prior
Zhai et al. (2019)b	Hong Kong			●	●		●	●				City	Binary & Mixed logit models with and without variable interaction terms

● Considered in the study design, ○ considered in the study process as filter/defining characteristic

Table 4
Studies with road safety spatial analyses primarily by conditional approaches

Study Characteristics		Dependent variables				Independent variables – parameters							
						Traffic					Road environment		
Author(s) (Year)	Country of study	Crash type analyzed	Crash count/ frequency	Crash rate	Injury Severity	Speed	Traffic volume	Vehicle distance traveled	Number of Trips - OD	Speed Limit	Curvature	Gradient	Lane width
Bao et al. (2019)	United States	TC	●	●	●			●					
Bíl et al. (2013)	Czech Republic	TC	●										
Cai et al. (2019b)	United States	TC	●					●		●			
Cai et al. (2017)a	United States	TC | P-V | B-V	●		○			●		●			
Chung et al. (2018)	United States	TC	●		○			●					
Imprialou et al. (2016)	United Kingdom	TC	●		○	●	●				●	●	
Kim et al. (2006)	United States	TC | V-V | P-V | B-V	●										
Loo et al. (2011)	China	V-V | P-V	●										○
Mohaymany et al. (2013)	Iran	TC	●										○
Ossenbruggen et al. (2010)	United States	TC	●		○		●	●	●				
Xie et al. (2017)	United States	P-V	○		●			●	●				
Xie and Yan (2008)	United States	TC	●										○

Study Characteristics		Independent variables – parameters												
		Road environment			Demographic				Socio-economic	Land Use		Spatial aggregation approach		Analysis - Modelling approach
Author(s) (Year)	Country of study	Lane number	Intersection nr./ density	Roadway length	Population number/ density	Road user/ Population age	Modal distinction	Household/ Personal income	Employment percentage/ density	Land use factor(s)	Zonal level	Link/ segment/ intersection level	Condition-based level	
Bao et al. (2019)	United States		●	●	●		○			●			Multiple grids (approx. to ZIP areas)	Convolutional Neural Network augmented with a Long Short-term Memory Network
Bíl et al. (2013)	Czech Republic		○	●								Rural segments	Rural road network split into fundamental segments	Network Kernel Density Estimation with significance verification
Cai et al. (2019b)	United States	●	●	●			●			●			9-mi^2 grid structure divided to smaller cells	Convolutional Neural Networks (GLM and Artificial Neural Networks for benchmarking purposes)
Cai et al. (2017)a	United States		●	●		●	●				TAD | TAZ | CT		Multiple grids from 1 to 100 mi^2	Multivariate Poisson Lognormal Regression with and without

(*continued on next page*)

Table 4 (*continued*)

Study Characteristics		Independent variables – parameters												
		Road environment			Demographic				Socio-economic	Land Use		Spatial aggregation approach		Analysis - Modelling approach
Author(s) (Year)	Country of study	Lane number	Intersection nr./density	Roadway length	Population number/density	Road user/Population age	Modal distinction	Household/Personal income	Employment percentage/density	Land use factor(s)	Zonal level	Link/segment/intersection level	Condition-based level	
														spatial autocorrelation
Chung et al. (2018)	United States			○									Areas within 20 mi of 2271 weather stations	Categorical analysis (sensitivity, positive predictive value, Cohen's Kappa) | Negative Binomial Regression
Imprialou et al. (2016)	United Kingdom	●		●								Rural & Highway segments	Pre-crash conditions	Bayesian Multivariate Poisson Lognormal Regression
Kim et al. (2006)	United States				●		○		●	●			0.1-mi^2 grid structure	Negative Binomial Regression | OLS Regression
Loo et al. (2011)	China			○			○					Urban & suburban segments	Urban and suburban network split into fundamental segments	Network Kernel Density Estimation
Mohaymany et al. (2013)	Iran			○	●							Rural segments	Rural road split into fundamental segments	Network Kernel Density Estimation
Ossenbruggen et al. (2010)	United States												1-mi^2 grid structure	Homogeneous Poisson process spatial testing
Xie et al. (2017)	United States				●	●	●	●	●	●			300 × 300 $feet^2$ grid structure	Linear Regression Model | Tobit Model | Potential for Safety Improvement
Xie and Yan (2008)	United States			○									Urban network split into fundamental lixels	Network Kernel Density Estimation

● Considered in the study design, ○ considered in the study process as filter/defining characteristic

might explain their popularity for utilization in spatial analyses (e.g. Ng et al., 2002; Hadayeghi et al., 2003; Ladron de Guevara et al., 2004; Lovegrove and Sayed, 2006; Lovegrove and Sayed, 2007; Hadayeghi et al., 2010; Naderan and Shahi, 2010; Abdel-Aty et al., 2011; Abdel-Aty et al., 2013; Dong et al., 2014; Lee et al., 2014b; Dong et al., 2015; Lee et al., 2015a; Xu and Huang, 2015; Dong et al., 2016; Nashad et al., 2016; Xu et al., 2017a, 2017b; Bao et al., 2017; Gomes et al., 2017). TAZs can be also expanded for road safety assessment purposes by aggregating TAZs groups with similar crash rates, thus creating Traffic Safety Analysis Zones (TSAZs), (Lee et al., 2014b; Abdel-Aty et al., 2016).

Census Tracts (CTs, or census output areas) are larger units containing about 4000 people of comparable socio-economic statuses in the US (or about 2500 people in the UK). They too have been adequately explored in road safety spatial analyses in the literature (e.g. LaScala et al., 2000; Loukaitou-Sideris et al., 2007; Delmelle and Thill, 2008; Wier et al., 2009; Cottrill and Thakuriah, 2010; Ukkusuri et al., 2011; Narayanamoorthy et al., 2013).

Similar to TAZs, Traffic Analysis Districts (TADs) are newly created, larger geographic traffic related units used for transport analyses. A few recent studies have utilized TADs as basis for analysis (e.g. Abdel-Aty et al., 2016, Cai et al., 2017b; Lee et al., 2017a). Other zonal areas have been used as well by exploiting existing utility systems, such as postal-ZIP codes (e.g. Lee et al., 2014a; Bao et al., 2018) and urban/rural areas defined by healthcare authorities (e.g. MacNab, 2004; Bu et al., 2018).

Reviewed studies that primarily focus on spatial analyses at zonal levels are shown on Table 2.

TAZ approaches can conceptually include elements of segment approaches nested in them. An example is the study of Yasmin and Eluru (2016) that employed latent segmentation count models where TAZs are allocated probabilistically to different segments. This was in order to limit external factor impact and to classify segments within a TAZ to high- and low- risk based on empirical expected crash means. Studies have also developed models on several zonal systems for comparison purposes between them. Abdel-Aty et al. (2013) claimed that while TAZs and CBGs are equally desirable for spatial analysis, TAZs allow the examination of more transport-related factors, and thus are easier to integrate in transport contexts. Furthermore, the aggregation of TAZs into TSAZs with a rate of about 1:2 was found to be preferable for macroscopic safety modeling (Lee et al., 2014b). Cai et al. (2017a) conducted comparative Poisson lognormal models for three crash types with and without considering spatial autocorrelation effects, and recommended that CTs are better used for socio-demographic data collection, TAZs are used for transportation demand forecasting and TADs are used for transportation safety planning. Different zonal levels have also been used in conjunction for simultaneous aggregate and disaggregate modelling; it has been shown that aggregate models using ZIP codes were more volatile in parameter values and significance levels, while disaggregate CT models provided more consistent results (Ukkusuri et al., 2012). Lastly, it has been determined that separate considerations for crashes near TAZ boundaries revealed unique predictor variables (Siddiqui and Abdel-Aty, 2012), a finding worthy of examination in all spatial units.

2.3. Regional approaches

Regional areas (counties, cities, metropolitan areas, states) that are larger than the zonal ones examined above have also been implemented in the literature. Regional areas are administrative units, with often different governance laws and frameworks than their neighboring areas, as is often the case in US states. In the US, entire Metropolitan Statistical Areas (MSAs) have been used for the National Household Travel Survey, which has provided data for pedestrian trips (Lee et al., 2019a). The benefit of using regional units can lie in the interpretation of model results and possible evaluation of risk factors or road safety interventions, such as legislation changes. For instance, a study by Song et al. (2006) applied Bayesian multivariate spatial models in county-level data in Texas, and results indicated that eastern Texas counties had higher crash risks than western Texas counties, with less safe sites being near large city conglomerations. Studies have examined road safety indicators at the level of geographic units formed from communities (LaScala et al., 2001, 2004), at the city level (Moeinaddini et al., 2014), at the metropolitan area level (Bu et al., 2018), at the county level (Noland and Oh, 2004; Song et al., 2006; Erdogan, 2009; Huang et al., 2010; Li et al., 2013) or similarly at the state level (Atubi, 2012).

Regional-wide crash modification factors (CMFs) have also been developed for a single change affecting the traffic environment uniformly, e.g. for legal changes in some U.S. States or across the entire country (Lee et al., 2017b, 2018a), however this approach does not take spatial effects explicitly into account. As the area size increases, it is important to remember that unobserved heterogeneity is more difficult to capture, due to multiple unobserved parameters being introduced in the occurrence of events; as Wang et al. (2016b) state, it becomes more difficult to capture spatial trends and problems in a larger area. If differences in comparable units between remote areas such as different countries are taken into account, it is reasonable to assume that transferability of results for macroscopic spatial analysis is far from seamless. In a study seeking to examine transferability of results across regions of different countries (from US counties to Italian provincias) Lee et al. (2019b) employed negative binomial models using data from both countries and calculated the respective transferability indexes and calibration factors. Models for total crashes and bicycle crashes were transferable from Italy to the US; the opposite, however, was found to be untrue for most study areas. In addition, no model for pedestrian crashes was found to be transferrable between the two countries. It is important to note that this statistical disagreement emerged even while several significant variables were common across the two countries, and without accounting for spatial effects in the models of the study.

Reviewed studies that primarily focus on spatial analyses at the zonal level are shown on Table 3.

2.4. Conditional approaches

Apart from defined zones, conditional approaches have been adopted. As conditional is hereby defined any approach that does not utilize any of the previous segment, zonal or regional approaches but a more rigid ruleset set by researchers. An example is fix-distance grid structures, such as 0.1 square mile grids (Kim et al., 2006), 1 square mile grids (Ossenbruggen et al., 2009) and multiple grid sizes from 1 to 100 square miles (Cai et al., 2017a). While the impacts of grid-based characteristics on crash counts have been proven to be statistically significant, a grid of a particular size might be improper for certain areas, depending on spatial distributions of safety-related parameters (Kim et al., 2006).

An example of approaches that are conditional not by area, but by crash circumstance, are link-based approaches that utilize crash-mapping algorithms and assign crashes to each road segment, and assuming that the crashes happening on the same link have the same underlying conditions, which might not always be the case. Link-based approaches can be problematic in providing interpretable results, however. Conversely, crashes can also be grouped by pre-crash conditions, regardless of their actual location, for the purposes of spatial analyses. Pre-crash conditional approaches have appeared to be more transferable overall (Imprialou et al., 2016).

Reviewed studies that primarily focus on conditional spatial analyses are shown on Table 4.

2.5. Integration of different areal units

The aforementioned integration of characteristics of the corridor level to road segment or intersection level analysis by several studies (Zeng and Huang, 2014; Alarifi et al., 2017, 2018) is a considerable achievement in road safety. In these studies, the levels of analysis can be considered to be close in geographical characteristics (i.e. a segment is similar to a corridor). There have been other endeavors, however, to integrate factors from units of more different scales in spatial analyses, such as zonal-level characteristics to segment-level analysis.

As stated before, the zonal level has become a promising medium during the more recent years for the exploration of new approaches of spatial analyses. Zonal factors, such as Vehicle Miles Traveled (VMT), are considered to be shared by segments of both segments and intersections of the same zone. It has been hypothesized that both observed and unobserved heterogeneity at the zonal level would influence crash frequency at both segments and intersections inside these zones. Cai et al. (2018) investigated crashes at the TAD level across three counties to determine the influence of any observed and unobserved zonal factors. Results indicate that including zonal factors improve model performance for both segment and intersection crash frequency prediction.

Another concept is incorporating macro-level variables into micro-level safety analysis. This has been attempted by Lee et al. (2017a) across seven areal units of varying sizes for intersection crashes. They determined that accounting for macro-level variables and introducing macro-level random-effects leads to models of better performance than the baseline, though performance varies when using data of different areal unit size. Additionally, there have been endeavors to link crash counts of micro- and macro-levels through their spatial interaction (Cai et al., 2019a). A spatial interaction matrix was created based on whether a road segment (micro-level) was inside a zone (macro level), and an adjustment factor was introduced to bridge the different estimates of expected crashes that would occur for the two levels. Once again, following an integrated approach increased model performance; moreover, the determination of both macro- and micro-level risk factors that influenced crashes were possible, as well as crash hotspots on both levels.

Conversely, road-level factors have been shown to influence safety by varying effects across regions, and can be considered to be correlated with unobserved heterogeneity, to an extent. To demonstrate this, a dedicated study examined specifically urban two-lane roadway segments in 34 counties in Florida, US. Regression coefficients of Poisson lognormal models and hierarchical models were found to fluctuate considerably for crash counts across the examined counties (Han et al., 2018). However, neither factors at the regional level nor spatial correlations at the microscopic level were taken into account in that particular study.

Huang et al. (2016) investigated a possible bridging of the macro- and micro-level approaches for an integrated crash prediction and hotspot identification approach. Crashes were analyzed both jointly at the micro-level (road segment/intersection level) and at the macro-level (TAZ level). The authors developed both a micro-level Bayesian spatial joint model and a macro-level Bayesian spatial model; as expected, the models included different statistically significant variables. Results reaffirmed the known model merits: micro-level modelling provided more informative and precise insights for directly improving road safety, while macro-level modelling allows for incorporating safety improvements in long term transportation planning. The authors acknowledge that TAZs may have unobserved scale and zonal effects and further, the boundary issue – explained in the following – needs to be accounted for.

2.6. Boundary problem and Modifiable areal unit problem

Apart from conducting studies across many different areal levels and bridging aspects and attributes of different spatial levels, researchers have also shown interest on how to define areas and areal units and how to treat events on their boundaries. The boundary problem, or boundary effect, refers to the manner in which crashes recorded on (or very close to) the borders of neighboring study areas are allocated and treated in statistical analyses. Fotheringham and Wegener (1999) claimed that neighboring zones influence crashes close to the borders of areal units. Since then, several studies have explored the problem, each proposing a solution. Delmelle and Thill (2008) mention simple solutions such as (1) assigning the locations as they were assigned by police records, (2) double-counting boundary crashes or (3) apportioning crashes, dividing the counts per neighboring zones.

Separate predictor sets have been prepared for boundary and interior pedestrian crashes per TAZ, introducing buffer zones around 2-D borders. This mutually exclusive separation and modelling within a hierarchical Bayesian framework has led to increased model fit. However, this approach was adopted due to the limited distance travelled by pedestrians, and accounting for additional road user types might differ due to higher amounts of areal units that are typically crossed (Siddiqui and Abdel-Aty, 2012). Instead of using a fixed buffer zone, Cui et al. (2015) introduced an entropy-based method applied on histogram thresholding, to obtain a variable buffer zone size. The crash density probability distribution was then calculated, and boundary crashes were aggregated into neighborhoods. The case study resulted in 6 m and 9 m buffer zones for central areas and south areas in Edmonton, Canada, respectively. The authors concluded that the entropy-based method was precise when compared to ground truth data, though more variables are required to verify this finding; especially traffic-related variables such as speed and traffic volume.

An alternative was proposed by Zhai et al. (2018), who adopted an iterative data aggregation approach to compensate for the boundary effect. The reasoning behind this method was the division of each zone into boundary and interior, the development of a crash prediction model for each zone based on interior crashes only, the aggregation of crashes based on crash model predictions, the assignment of boundary crashes to each zone based on the proportions of expected interior crashes, and, as a last step, re-run the prediction model until convergence. The crash assignment based using the CAR Poisson Lognormal Bayesian Spatial Model. It is notable that the impact of several independent variables were found to be influenced by the boundary effect in the case study in Florida, US. Both Cui et al. (2015) and Zhai et al. (2018) demonstrated that certain analytical approaches outperform conventional rules such as the various ratio methods that split boundary crashes based on numerical rules or exposure parameters). It is also worth noting that certain Bayesian statistical models can express the interaction of neighboring zones on crashes close to zone boundaries via the utilization of corresponding spatial weights (e.g. Wang et al., 2016b).

The modifiable areal unit problem (MAUP) occurs when boundaries are changed inside the study areas, causing possible influences on the statistical models and resulting inferences (Openshaw, 1984). The issue is particularly present in road safety when area boundaries are arbitrary or malleable, without any hard geographical borders, such as administrative areas or grids. Two studies did experiment with the discrepancies caused by MAUP on different aggregation levels (Ukkusuri et al., 2012; Abdel-Aty et al., 2013). While the areas which provided more accurate predictions were determined, no uniform solutions were proposed. When outlining MAUP, Xu et al. (2018) outlined four potential solutions. These were: (1) using disaggregate data as possible (2) capturing the spatial non-stationarity, which refers to capturing local space variation for each explanatory variable, (3) designing optimal zoning systems, an approach which presents its own limitations and (4) conduct sensitivity analysis for MAUP effects specifically.

A recent study has empirically highlighted the important effects of MAUP on four different zonal configurations using an identical dataset (Zhai et al., 2019a). It was determined that the impact of MAUP was significant on parameter estimates, model assessment and hotspot identification. Larger zones, such as CTs and ZIP codes led to models of higher predictive accuracy in that study. It has also been considered that the zonal systems may have inherent limitations by Lee et al. (2014b), who developed ten new zonal systems to tackle both the boundary and the MAUP problems. The Brown-Forsythe homogeneity of variance test was implemented to obtain the optimal zonal scale, which was found to be at the custom TSAZ level, as zones cannot be scaled up indefinitely to reduce boundary crash percentages. However, the authors state that the boundary issue still needs to be accounted for in TSAZs, and that further research on additional crash types such as non-motorized (VRU) crashes is needed.

2.7. Examination of spatial proximity structures

A critical point that attracts researcher interest is the creation of different spatial proximity structures and the examination of the effects these structures have on model performance and fit. Various spatial proximity structures have been formulated both at the microscopic and macroscopic levels. Regarding the microscopic level, Aguero-Valverde and Jovanis (2010) concluded that by including route information in the neighboring structure, especially in a simple neighboring structure (direct adjacency), model performance is improved.

Regarding the macroscopic level, Dong et al. (2014) evaluated crash prediction models at the TAZ level using four different types of spatial proximity structures (0–1 first-order adjacency, common-boundary length, geometry-centroid distance, and crash-weighted centroid distance). The best model fit was provided when weighting the common-boundary length of neighboring TAZs, though cross-zonal spatial correlations was identified as present in crash occurrence for all four different configurations. The authors comment that the inclusion of all possible spatial correlations increases model complexity, thus resulting in decreased prediction performance.

Moreover, Alarifi et al. (2018) sought to investigate spatial weights configuration for a hierarchical spatial proximity structure, including intersection-, road segment- and corridor-level parameters. The authors examined four different types of conceptualization of spatial relationships and calibrated 13 Bayesian hierarchical Poisson-lognormal joint model with spatial effects. The adjacency-based first-order model (where directly adjacent road entities and feeding road entities are considered for each segment) was among the best performing models and once again significant variables were found in all configurations for all unit levels. The authors suggest that the sensitivity of AADT in the models is a matter for further investigation.

Another sophisticated approach was the utilization of the space syntax technique for modelling street patterns. Space syntax acknowledges the configuration of the urban grid itself is responsible for generation of movement patterns (Hillier et al., 1993), though its exact use for deriving certain route choices has been challenged in the past (Ratti, 2004). Guo et al. (2017) considered simple geographical proximity as inadequate to properly describe spatial relationships of crashes. Rather, they sought to integrate road network characteristics in a zonal level examination. They used space syntax to quantify road network structures in Hong Kong through three main parameters on the TAZ level: (1) connectivity, (2) local integration and (3) global integration. After calculating global integration for three road network patterns (grid, deformed grid and irregular), it was determined that global integration was positively related with increased pedestrian-vehicle crashes. Furthermore, the more structured patterns featured the highest global integration values, thus irregular patterns were found to be the safest, followed by deformed grids and lastly (regular) grids.

2.8. Further topics of areal unit analysis

In spatial analysis, study designs sometimes appear to be data-driven, conducted where there is availability of information instead of intuition or previous experience. Availability of data does not necessarily imply its fitness for use in studies. As an indication, weather data measured from stations may or may not describe the situation at crash sites accurately. A study was conducted to evaluate the effectiveness of coverage of weather stations for use in spatially analyzing traffic crashes (Chung et al., 2018). Hourly data which are observed from land-based stations was contrasted with data from fatal crash databases. Through categorical analysis, sensitivity, positive predictive value, and Cohen's Kappa were examined, and it was determined that there were agreements of data in rain and snow weather conditions but not in fog, which displayed a 91% rate of false alarm. The authors suggest that fog may present higher spatio-temporal sensitivity as a parameter. While the weather station data was found adequate overall for use in crash analyses, the finding regarding the fog parameter ought to make researchers carefully consider possible data sources for their studies.

Furthermore, instead of analyzing crashes collectively in each areal unit, or treating them as separate variables, different crash categories can be examined while taking their interactions into account. A study by Lee et al. (2018b) analyzed the proportions of crashes of each vehicle type at the TAZ level, using a fractional split multinomial model. The fractional approach ensures the summation of crash proportions of all categories to 100%, thus forcing interactions between each category. Findings showed considerable differences as to which variables were statistically significant for each vehicle type. Moreover, the spatial distribution of hot zones varied considerably per vehicle type considered. On that matter, hotspots have also been found to vary temporally. Soltani and Askari (2017) conducted a spatial autocorrelation analysis of crashes and hotspots at the TAZ-level in Iran. Moran's I and Getis-Ord Gi* methods were used, and were found to provide significant clustering. The authors examined crashes based on location, time of day and injury severity, which is a very rare combination of parameters. This time, hotspots were found to vary considerably across the various times of day. Another important finding is that zones located at intersections connecting other zones were identified as clusters with high crash rates. Despite the hotspot identification, however, no other explanatory characteristics were introduced in the analysis. It appears thus reasonable to assume that the identified hotspots may vary considerably if certain elements are introduced to a study or omitted from it.

3. Modelling approaches

This section provides a brief overview of the various modelling approaches implemented so far in the literature of spatial analysis in road safety. A multitude of tools have been developed that endeavor to predict road safety indicators (Lord and Mannering, 2010; Mannering and Bhat, 2014) and explain spatial correlation and unobserved heterogeneity and to incorporate the effects of various spatial characteristics that are difficult to be represented individually. Several studies have been testing various advanced models against simpler ones for performance assessment (e.g. Miaou and Song, 2005; Chiou et al., 2014; Dong et al., 2016; Aguero-Valverde et al., 2016; Cai et al., 2019b).

Multivariate models are found to have better goodness-of-fit and precision due to correlation between dependent variables, such as crashes of different severity levels while accounting for spatial correlation (Barua et al., 2014) or simultaneous crash frequency and severity examination (Chiou et al., 2014). The benefits of multi-level data have been discussed in spatial analyses, for instance the multilevel structural hierarchy proposed by Huang and Abdel-Aty (2010) combining driver-level and site-level data with geographic region characteristics.

Spatial analyses often test for spatial autocorrelation or heterogeneity of events, and also consider size and structure for the various research areas and spatial units of analysis in the adopted approaches. For the precise examination of autocorrelation phenomena, various geo-spatial statistics have been adopted by scientists for decades, such as Moran's I, Local Moran's I, and Getis-Ord-Gi* statistics.

Generalized Linear Models (GLMs) have been used extensively in the road safety literature for decades, since they assume crashes are independent, random and sporadic countable events (Hauer et al., 1988; El-Basyouny and Sayed, 2009). Their intricacies and limitations have been covered in past studies (e.g. Lord and Mannering, 2010).

While GLMs in their basic form are aspatial, they can be extended to incorporate spatial effects in their structure, eventually becoming quite advanced. An example is the EMGP model by Chiou and Fu (2013), further advanced by Chiou et al. (2014), which originated as an extension of the multinomial-Poisson regression model with added error components, to which spatial correlation effects were also added. Better predictions have been obtained from GLMs including random effects rather from fixed effects, and from GLMs including zonal factors as opposed to those not including them (Cai et al., 2018).

3.1. Geographically Weighted Regression

A method that accounts for spatial variation is the simultaneous development of several localized models using Geographically Weighted Regression (GWR). First proposed by Fotheringham et al. (2002), these models extend the traditional regression framework to allow for a continuous surface of parameter values, with measurements at points that indicate the spatial variability of such a surface. A number of road safety GWR analyses have been published (Hadayeghi et al., 2003, 2010; Pirdavani et al., 2014a, 2014b; Rhee et al., 2016; Gomes et al., 2017; Liu et al., 2017). As Pirdavani et al. (2014b) note, GWR models offer explanatory and descriptive power and provide intuitive results that enable researchers and stakeholders to investigate varying effects of explanatory variables on crash occurrence throughout the study areas.

Gomes et al. (2017) compared the performance of GWR extended in a GLM context and highlight that Geographically Weighted Negative Binomial Regression (GWNBR) is appropriate for spatially analyzing crash data while accounting for their over-dispersion. Additionally, GWNBR models significantly reduced the spatial dependence of model residuals. GWNBR models were also utilized by Liu et al. (2017) to produce localized models at the roadway segment level, without restrictions by jurisdiction boundaries. The variation of three calculated parameters (intercept, AADT and segment length) was found to be substantial in highway segments across Virginia, US, though the effects of several factors remain to be examined. Additionally, the introduced parameter of segment length is present in spatial structures, which might introduce bias to GWNBR estimations. The authors comment that GWNBR models are highly localized, thus the transferability of their predictions is limited and need to be reapplied to each area.

Xu and Huang (2015) extended GWR to semiparametric GWR (S-GWR), which combines geographically varying parameters with geographically constant parameters. Although their composite approach outperformed a random parameter negative binomial (RPNB) model, the authors claimed that S-GWR models are not transferable spatially, and that each region would need to develop separate S-GWR models (a common conclusion with the GWNBR method). S-GWR was compared again with RPNB by a study conducting crash analysis across six spatial units and three injury severity levels (Amoh-Gyimah et al., 2017). Again, results indicated that S-GWR performed better than the RPNB overall, based on mean absolute deviation (MAD) and Akaike information criterion (AIC) metrics, and had increased prediction accuracy. On the other hand, RPNB displayed increased sensitivity when examining the effect of variation of spatial units on unobserved heterogeneity compared to S-GWR. It should be noted that the latter study did not examine any geometrical characteristics such as segment length or intersection density.

S-GWR has also been employed to investigate possible correlations between jobs-housing balance and road safety, since disruptions in that balance have been found to lead to reduced road network efficiency (Xu et al., 2017b). The authors converted jobs-housing ratio to a categorical variable and then applied S-GWR models at the TAZ level. Considerable spatial variations were discovered for different jobs-housing ratio categories, through elasticity analysis of the model results for each jobs-housing ratio category. However, the study did not compare the S-GWR results with those of another baseline model.

3.2. Autoregressive prior models

A common problem in geographical studies with spatial dataset can be the selection of the appropriate size and scale units for analyses. This has a direct impact on results, as experience suggests that increasing granularity (i.e. spatial resolution) can weaken correlations between output areas and introduce spatial autocorrelation issues (Loo and Anderson, 2015). To counter this, studies have introduced spatial autocorrelation effects (e.g. Aguero-Valverde and Jovanis, 2006, 2008; Guo et al., 2010; Flask and Schneider, 2013; Chiou et al., 2014) or temporal autocorrelation effects in crash count models (e.g. Wang and Abdel-Aty, 2006). The respective models often use CAR or SAR

with the former being more frequently implemented in road safety spatial analyses. A seminal study by Besag et al. (1991) presented a normal distribution for spatial autocorrelation effects using a CAR prior, which has been implemented in many studies since (e.g. Huang et al., 2016; Cai et al., 2018; Zhai et al., 2018; Wen et al., 2019).

CAR models have been found to perform better than Poisson models and Multiple Membership models (where higher level units are formed by each unit and its adjacent neighbors), by explaining a high degree of spatial heterogeneity and by being more lenient in spatial variable omission (El-Basyouny and Sayed, 2009). However, Yasmin and Eluru (2016) note that considering spatial autocorrelation effects and latent segmentation simultaneously can be analytically challenging. Autoregressive models can also be developed within a Bayesian Framework as shown in Aguero-Valverde et al. (2016); CAR models have been found to be convenient to compute while using a Gibbs sampler in the Bayesian inference (Huang et al., 2010). Bayesian CAR models have been shown as capable to function with a variety of customizable spatial weights (Aguero-Valverde and Jovanis, 2010; Alarifi et al., 2018). These weights can be calculated based on several different bases (e.g. by geometric distance of zone centroids or by land use type). Of these weight sets, it is natural that some will outperform others for a specific study configuration, though not always in the expected manner, as shown by Wang et al. (2016b), where a simple 0-1 configuration based on proximity outperformed land use type- and intensity-based weights for pedestrian crash prediction (population was used as exposure parameter for pedestrians only, without a corresponding parameter for vehicles).

3.3. Bayesian modelling

The process of Bayesian inference has led to the development of several interesting methodologies during more recent years. Bayesian hierarchical joint models have been developed in various complexities using regression and regression methods for parameter estimation, possibly with regression splines, as shown in an early Bayesian approach by MacNab (2004). Moreover, multivariate Bayesian models are capable of estimating excess crash frequencies at different severity levels in the same spatial analysis unit (Aguero-Valverde, 2013). Bayesian hierarchical joint models have been shown to highlight significant variables at both micro and macro levels while accounting for spatial correlations between entities (e.g. in Cai et al., 2019a). Such an application by Wang and Huang (2016) determined higher AADT, more lanes and accesses for segments on the micro level, signal control, more intersection legs, and higher speed limit for segments for intersections on the micro level and higher road network and trip generation densities as significant risk factors, among others.

As studies often report, models with Bayesian approaches have been found to perform consistently better than their non-Bayesian counterparts (e.g. Miaou and Song, 2005; Siddiqui et al., 2012; Wang and Huang, 2016). Bayesian models with CAR effects have been shown to simultaneously account the spatial correlation and uncorrelated heterogeneity present in aggregated crash count data, and to reveal more significant variables with the same signs as frequentist modelling (Quddus, 2008). However, Bayesian models are not without drawbacks, as a main strength of their applications is reduced in cases without any solid basis of prior knowledge (uninformed priors). Furthermore, they require a considerable amount of calibration cases (sometimes mentioned as burn-outs) which leads to some loss of information and might require considerable computational time and power to obtain.

A noteworthy development is the recent investigation of spatio-temporal heterogeneity using multivariate hierarchical Bayesian models across injury severity categories. Relevant studies have endeavored to capture data heterogeneity with spatial and temporal effects, with the hierarchical framework serving to predict crash counts of different severities simultaneously. Spatial and temporal components are specified with several structured and unstructured components, and random effects can be inserted in the models to address the underlying data structure. Specifically, Ma et al. (2017) aggregated crash counts from 100 homogenous US highway segments into injury/no injury crash categories using high temporal resolution (daily intervals). They identified vehicle-distance travelled and some geometric characteristics as significant crash predictors, as well as variables that are more sensitive temporally, such as wet pavement and average speed.

In a recent study by Liu and Sharma (2018) examining injury crashes, both spatial and temporal effects were bound to be important in approximately the same magnitude across spatial, temporal and spatio-temporal structures. Crash frequencies showed significant spatial, but not temporal, autocorrelations. Similarly, Li et al. (2019) mentioned the issues of spatio-temporal instability in crash data, apart from the typical unobserved heterogeneity that is inherent to data collection. They calibrated Bayesian random parameters models (with both structured and unstructured spatio-temporal effects) which show that daily VMT, proportion of males, unemployment rate and education are found to positively increase crash frequency and are normally distributed across crash severities for crashes related to substance consumption.

3.4. Empirical Bayes and Full Bayes analyses

Since several decades, Empirical Bayes (EB) methods have been implemented in road safety by contrasting crash counts of a road segment with sites with comparable true crash risk, which are the reference population. EB estimations have displayed better predicting capabilities and eliminate regression to the mean issues than Naive before-after comparisons (Hauer, 1997; Geurts and Wets, 2003). EB methods have been also used in a before-after study in complementarity to a before-after study with a comparison group in order to obtain more reliable CMFs (Lee et al., 2017b).

Further to that direction, Full Bayes (FB) extended models can be used to account for heterogeneity due to unobserved road geometric characteristics, traffic characteristics, environmental factors and driver behavior (El-Basyouny and Sayed, 2011; Ma et al., 2017). The FB approach has also been shown to be more reliable empirically in hotspot identification compared to EB (Huang et al., 2009). The advantage of FB over EB is that it takes into account that model parameter estimates include an amount of uncertainty and can provide a quantitative measure of said uncertainty (Miaou and Lord, 2003). The FB approach is the basis of several recent developments discussed in the following.

3.5. Spatial spillover effects

An emerging aspect of spatial analyses is the examination of spatial spillover effects. Spatial spillover effects are the effects that exogenous observed variables have on the dependent variable at both the target and the neighboring locations. Spatial spillover effects differ from spatial autocorrelation (or error correlation) effects, which entail unobserved exogenous variables at one location affecting dependent variables at the targeted and neighboring locations (Narayanamoorthy et al., 2013; Cai et al., 2016; Lee et al., 2018b).

Past studies have utilized spatial lag regression models in an effort to capture spillover effects. LaScala et al. (2000) and Quddus (2008) converted count variables into continuous approximations for their analyses. They then used an explanatory variable in the expression of a spatially lagged dependent variable to form a spatial autoregressive (SAR or spatial lag) model.

Cai et al. (2016) included spatial spillover effects in the examination of pedestrian and bicyclist crashes. Via the application of dual-state GLMs, it was determined that taking observed spatial spillover effects into consideration results to models with better performance consistently. The zero-inflated negative binomial models were found to have the best fit for pedestrian and bicycle crashes, though unobserved spatial autocorrelation effects were not simultaneously examined in the study. To evaluate the impacts of significant factors, marginal effects were calculated as well.

In addition, Wen et al. (2019) aimed to capture both spatial autocorrelation and spillover effects using a hybrid model. The hybrid model featured the traditional Poisson-lognormal basis. The authors expressed spatial autocorrelation effects as the CAR prior and spillover effects as exogenous variables of neighboring road segments. Homogeneous highway segments were used for the analysis. Both of spatial autocorrelation and spatial spillover effects were found to be significantly correlated with the respective crash data. This hybrid approach yielded better estimates than both of its individual components, with coefficients that showed lower standard deviations. The authors suggest that accounting for spatial heterogeneity may further refine the model, but a much more complex structure would be required.

3.6. Alternative Prior Distributions

Apart from the widely used CAR model, other approaches can be implemented to account for spatial effects in models through different prior distributions. Mitra (2009) adopted a hierarchical Full Bayes spatial model to investigate the presence of possible influences of spatially structured factors on injury crashes at intersections. The reasoning behind such an approach is an attempt to capture both heterogeneity from spatial effects (implying a common global structure) and excess heterogeneity (originating from spatially unstructured effects). The first level of the hierarchy is a Poisson-lognormal specification. The Poisson rate then included the typical intercept and covariates, and also two separate effect terms, spatially structured and unstructured, to capture spatial and excess heterogeneity respectively. The spatially structured effects used a multivariate normal joint prior. Results indicated considerable spatial autocorrelation effects at the intersection level, while a comparison with aspatial Negative Binomial regression revealed similar coefficient estimates but increased model precision.

A similar jointly-specified approach was adopted by Aguero-Valverde (2014), to determine the effective range after which no lingering correlation is found at the road segment level. The Poisson rate function featured one parameter for heterogeneity among segments, using a normal distribution, and one for spatially correlated random effects per segment, using a jointly specified prior. Additionally, a temporal indicator for the evolution of crashes in years in covariate values and predicted crash counts was included. Ultimately, the joint prior model outperformed a random-effects model and a CAR prior model and the effective range was determined (at about 168 m). The author states that the manner in which distance is measured (e.g. Euclidean distance, ground route distance or any other way) also has an impact on model predictions.

A different form is the Full Bayes Multiple Membership (MM) spatial model proposed by El-Basyouny and Sayed (2009). The approach includes similar spatially structured and unstructured effects as the previous studies. In addition, MM models consider each site as a member of a higher-level unit that contains its nearest neighbors. They also include a parameter measuring the strength of association between structured and unstructured spatial effects. The authors further extended MM models by adding an additional component to allow for variance in the values of crash risks and characteristics between mutually exclusive corridors. When tested, the extended MM model slightly outperformed a CAR model, which in turn outperformed a basic MM model, though the overall DIC metrics showed quite close values.

Xu et al. (2017a) introduced another methodological alternative in the form of a very detailed Bayesian spatially varying coefficients approach, based on the hierarchy proposed by Huang and Abdel-Aty (2010). The process again started with a Poisson function in a Full Bayesian framework, and the parameters were modelled using a CAR prior. The innovation of the study lied in the utilization of a single set of random effects ranging from purely unstructured to purely spatially structured effects; this simultaneous process is considered superior by the authors, however it features a mathematical structure that is quite complicated.

3.7. Machine learning & Deep learning approaches

Given their popularity as a powerful, data-driven family of prediction tools, machine learning (ML) methods have been implemented for spatial and spatio-temporal road safety analyses. Indicative methods used in road safety spatial analyses are outlined below. ML methods can operate with increased degrees of freedom without requiring traditional assumptions as regression models do, and are more resilient to data outliers. They are methods typically used in conjunction with big data in transport and road safety.

Random forest (RF) models are collections of numerous superimposed decision trees that emerge from a selection and validation process, as described in Chang and Wang (2006). RF models have been used in road safety studies by researchers. For instance in Jiang et al. (2016) the feasibility of RF models for ranking hot-zones on a TAZ level and identifying critical parameters for crash occurrence when utilizing big data was investigated. Road network distribution (density) and socio-economic features such as school enrollment and car ownership percentages were found as the most statistically significant variables for crash occurrence. The study concludes that RF models provide classification with about 80% accuracy in hotspot identification.

Support Vector Algorithms (SVM) have been successfully implemented as alternatives to traditional statistical-regression modelling. In a relevant study, SVMs were employed together with a coactive neuro-fuzzy inference system (CANFIS) algorithm (Effati et al., 2015). SVMs were found to be considerably better performing when examining crash injury severity, especially when utilizing a radial basis kernel function (RBF). The researchers propose the enhancement of spatial analyses with machine learning algorithms as the key to unveiling significant factors affecting crash injury severity while accounting for spatial correlation and heterogeneity effects. The study of Dong et al. (2015) implemented SVMs as a tool for handling big and complex data structures. They examined zone-level crash prediction while taking spatial autocorrelation into account, and SVMs were found to perform better when including a spatial weight feature with an RBF kernel as opposed to SVM models. SVMs have been also used in conjunction with Bayesian methods, though, to the authors' knowledge, not yet in a spatial analysis framework; for instance, Wang et al. (2019) used Bayesian logistic regression to detect factors contributing to highway ramp crashes.

Latest technological progressions make neural network implementation much more feasible than past years. Bao et al. (2019) utilized a deep learning approach for short-term crash risk prediction for crash risk on an urban level. They augmented a convolutional neural network (CNNs) with a long short-term memory network in order to examine variables that varied spatially, temporally or spatio-temporally, proposed by earlier research for traffic speed and congestion prediction (Ma et al., 2015a, 2015b). Weekly, daily and hourly prediction models with varying spatial grids were produced as a result. The authors mention that prediction performance of the proposed model decreases as the spatiotemporal prediction outcome resolution increases towards the hourly level. It is noteworthy that machine learning models exhibited better performance on the daily level, while benchmark econometric models generally performed better on the weekly level, suggesting that neither approach is clearly superior. Another interesting application is described in Zhu et al. (2018); the CNNs developed in the study take into account spatio-temporal network and traffic structure. However, they are used for traffic incident detection/identification, and not road safety prediction or causation analysis.

Cai et al. (2019b) explored that research direction by applying CNNs for road safety prediction by collecting and utilizing high-resolution data: 3mile x 3mile grids with crash counts and data, each grid containing 100×100 cells with width and height of 158.4 feet, examined in 17 layers of data matrices. By feeding data of a higher resolution into a CNN, the authors allowed variables to fluctuate across locations more freely, thus increasing the model accuracy. It was stated that the hierarchical structure enables better understanding of the circumstances of crash occurrence. While the authors demonstrated a viable approach for crash prediction, it is obvious that extra effort is required for the creation of this high-resolution grid and the complementing database. Some variables might be readily available for calculation in high-resolution or inferred via the existing road geometry (such as segment lengths), while others may be harder to obtain in case of missing data (such as land uses). Approaches such as CNNs might require custom, tailor-made data collection frameworks in order to provide their full potential, as the authors suggest. Furthermore, no specific framework is established for assigning the values of required hyperparameters during the CNN training phase.

3.8. Kernel Density Estimation

Another crash and hotspot analysis method is kernel density estimation (KDE), which allows the generalization of incident locations to an entire area. It should be noted that this is not a direct analytical method, but rather an interpolation technique (Anderson, 2007) mainly used for the identification of clustering patterns of traffic collisions. KDE can be advantageous in predicting the spread of crash risks, though the kernel radius has been a matter of debate in several scientific fields (e.g. Raykar and Duraiswami, 2006; Hart and Zandbergen, 2014). It appears that bandwidth determination influences the outcome of the hotspots (Fotheringham et al., 2000; Anderson, 2009; Loo and Anderson, 2015). Furthermore, the fact that KDE treats discrete events as a continuous area effect can be presented as a limitation (Anderson, 2009). Erdogan et al. (2008) conducted an analysis of hotspot clusters in a province of Turkey and utilized KDE together with a repeatability analysis of hotspot crashes for a decade. The authors reported considerable overlap of the outcomes, though KDE determined less hotspot locations overall. An interesting approach by Mountrakis and Gunson (2009) investigated the development of KDE spatially (determining varying density peaks among roads) and temporally (determining an exponentially increasing trend with annual periodicity and a seasonal cyclic component) for animal-related crash hotspots in Vermont, US.

Kernels are projected over 2-D spaces, while road crashes usually occur in a 1-D linear area, which most road environments approach, as Xie and Yan (2008) note. In order to overcome this discrepancy, KDE has been expanded to network KDE approaches, in which the network is represented as fundamental units of equal network length (termed lixels). Xie and Yan (2008) investigated this method and how fundamental lengths and regular kernel bandwidth affect its performance for road crash prediction. They conclude that network KDE describes crash densities and network borders more precisely than regular KDE, and that lixel length appears more important than Kernel function selection. However, Loo et al. (2011) implemented network KDE in areas of varying land use and found that kernel bandwidth critically affects the spatial distribution of resulting density estimates. Furthermore, wider bandwidths appeared to be more appropriate for non-urban areas where crash density is lower.

Similarly, Mohaymany et al. (2013) applied network KDE to a rural road in order to determine hazardous segments; apart from static spatial autocorrelation of crashes they also investigated its temporal evolution through a three-year period. Bíl et al. (2013) also used KDE in a 1-D area by separating the network into sections. They explored an alternative venue for better refining KDE results by providing a method to test their statistical significance. The proposed method utilized relative spatial positions of crashes and roadway length to calculate kernel strength, which allows detection and prioritization of the most hazardous locations, which included classifying clusters with values above the 95th percentile of the kernel density function as hazardous.

4. Vulnerable Road Users

In road safety, vulnerable road users (VRUs) include pedestrians, bicyclists and other road users who are often children, elderly, people with impairments and disabilities. Due to their vulnerability to injuries or fatalities compared to vehicle users, VRUs have increased safety needs. The use of spatial analyses, or approaches in a spatial context, to examine aspects of road safety concerning VRUs warrants specific examination. A notable example is the study of Tasic et al. (2017) which investigated crashes involving vehicles and VRUs by using models that accounted for spatial correlation effects. Data was aggregated on a CT level for a large array of about a hundred variables for vehicle-only, pedestrian and bicycle crashes. The data were analyzed using an extension of GLMs, Generalized Additive Models (GAMs), which included a two-dimensional smooth function to account for spatial correlation. A remarkable finding was that the expected pedestrian or bicyclist crashes increased less than proportionally with the exposure variables of vehicle, pedestrian or bicyclist trips, confirming the safety-in-numbers effect on a macroscopic level while accounting for spatial correlation effects.

Analyzing pedestrians' walking exposure and crashes in an integrated manner was proposed in a dedicated study on the MSA level (Lee et al., 2019a). For estimating exposure, multiple linear regression models were calibrated, followed by a Poisson-lognormal regression model for fatality estimation using the estimated exposure as input. Walking hours was determined as the best performing exposure variable. The proposed integrated model outperforming the non-integrated ones. Spatial correlation of trips was not investigated in the study, however, and pedestrian safety features were not examined either. VRU exposure, in the form of trips, has also been estimated at a macroscopic level in an integrated manner. These trip numbers were used to calibrate VRU crash prediction models in a study across 23 Metropolitan areas, and it was found that estimated exposure (VRU trips) led to models with calibrated performance compared to observed exposure for both pedestrians and cyclists (Lee et al., 2018c).

Pedestrian crash hotspots have been examined through spatial processing of their respective costs using big data from multiple sources such as taxi trips and social media (Xie et al., 2017) by employing a grid structure divided in higher resolution cells, similar to Cai et al. (2019b). Crash costs were assigned to cells using a kernel density estimation function, and sites were identified using tobit models with potential safety improvements (PSIs) and ranked as potential hotspots based on the potential of pedestrian crash cost reduction. The authors claim that their method can be transferred to less populated regions by adjusting kernel bandwidths.

Pedestrian crashes do not necessarily occur in the zone of residence of the pedestrians involved; Lee et al. (2015b) sought to identify zones where pedestrian crashes occur, and zones where pedestrian crashes originated from. Using different exposure variables, a variation of a Bayesian lognormal model with Poisson structure was applied. The occurrence of crashes with pedestrian involvement was revealed to be significantly affected by more location-related factors, while pedestrian origin was revealed to be significantly affected by more demographic-related factors. A similar concept of investigating both ZIP codes of crash locations for bicyclists and the number of crash-involved bicyclists in their ZIP of residence was explored in a study by Lee and Abdel-Aty (2018). Bayesian Poisson lognormal CAR models were used to examine bicycle crashes, and the contributing factors were not identical in each case. For instance, increases in the number of schools per mi^2 were only found to lead to increases in bicycle crashes in the crash location ZIP. Conversely, lower income areas were found to be a contributing factor overall through the significance of many related variables. Again, PSI was used to identify VRU crash hotspots in both studies.

A noteworthy finding is that of Siddiqui et al. (2012), who produced Bayesian models for pedestrian and bicyclist crashes at the TAZ level, noting the necessity of accounting for spatial correlation while examining VRU crashes at the macroscopic level, which is also corroborated by Guo et al. (2017). In addition, spatial spillover effects have also been examined in a VRU context, as mentioned before (Cai et al., 2016).

Apart from methodological and modelling approaches, the influence of parameters for pedestrian crashes have also been examined in high resolution. Specifically, the effects of weather conditions have been investigated using GIS within a spatial context (Zhai et al., 2019b). Binary and mixed logit models were used in the study, in a basic form and in a more advanced form including terms of interaction between weather conditions and risk factor variables. Both high temperatures and precipitation were found to be associated with pedestrian crashes of increased severity. Hotter weather and the presence of rain were also found to exacerbate the effect of risk factors, such as jaywalking or unsafe driver behavior.

5. Discussion

5.1. Findings from reviewed studies

The examination of the studies that was carried out in this research has led to some noteworthy conclusions for spatial analyses in road safety. It appears that a multitude of different approaches and modelling methodologies has been adopted in the literature, with a trend towards advanced Bayesian models and methods in the past decade. This has led to the development of powerful tools that provide accurate predictions for crash counts per area with increasingly complex model configurations. However these approaches also lead to a lack of a common established methodology or framework to compare results of spatial analyses. Additionally, this finding does not imply that more traditional functional/econometrics methods, such as GLM models or GWR are not found useful still, at least for benchmarking purposes. Functional models appear to be more straightforward in their interpretation and assessment of results. In both cases, results of spatial studies have also been reported to have limited transferability as well.

Recently, machine learning approaches have come to challenge the dominance of Bayesian models by being implemented alongside or instead of them. It should be noted that these are mostly data-driven approaches, which have also been reported as containing inherently biased samples, especially when examining big data (e.g. Bao et al., 2017, 2019). While the aforementioned transferability issues are mostly solved with machine learning methods, there are often difficulties in the interpretation of results: A commonly cited example is the hidden layers of neural networks and the meaning of each contributing factor. Approaches such as SVM are subpar in determining the significance of revealed patterns in the data they examine or the utility each variable offers in prediction tasks.

Further on the results of spatial studies, another important finding is the revelation of sensitivity of hotspot locations. Researchers have shown that hotspots are radically different across users of different vehicles and ages, and that hotspots display significant variation throughout the time of day. It can be reasonably surmised that many elements that are introduced to an analysis radically change the hotspot map. Naturally, the employed methodologies also affect the final outcome of spatial studies. Researchers should be vigilant and try to convert unobserved factors into observed ones, in order to receive more substantial and precise hotspot maps.

Though studies have been published internationally, spatial analyses have been more common in more modernized and developed countries (especially USA), while developing countries are considerably less represented. The use of different sizes of spatial units as basis for spatial analyses has been examined extensively, and it appears that apart from information and data availability, spatial areas of each size have different advantages and disadvantages. Several studies include exposure parameters in order to establish a common baseline for crash risk comparisons between models (Imprialou et al., 2016). When exposure parameters such as road length, AADT and vehicle distance travelled are examined, they are found to increase crash risk overall, as expected, however there are particular cases where these results might not apply or even be reversed (e.g. Dong et al., 2014).

It has been demonstrated that the parametrization of the spatial correlation term, namely, its inclusion as a variable in models, can aid in situations where data is scarce or difficult to obtain. Its use can be further expanded, however, as a complementary feature to even variable-rich models, in order to explain parts of variation in the data.

That being said, data availability remains a critical issue, and lack of consistent data across a respectable duration of time can be a critical obstacle in conducting spatial and spatio-temporal analysis. Spatial analyses in road safety appear data-driven most of the time, stemming from the drive of researchers to prove or test a concept. There are variables that have not been extensively tested due to lack of data, for instance pavement condition. Similarly, there are study areas that merit more attention, such as extensive urban network environments formed by roads of lower categories.

Traffic speed does not appear to be as frequently used as in past decades, though speed limits are taken into account as network characteristics, rather than traffic characteristics. Moreover, it can be observed that certain geometrical features seem to be used less frequently, such as road gradient, curvature and lane width. As an indication, the 'gradient' column on Table 2 was blank at the end of the reviewing process and was thus removed. This decline in use can be attributed to missing data for many study areas, or to difficulty in data acquisition. Another reason may be the lower prioritization of geometrical features from researchers: studies often seek to include crash data, traffic data, socio-economic data, demographic data and land-use data. Therefore traditional road geometry data is receiving less attention in comparison to past decades.

5.2. Future research directions

This section outlines research directions that do not appear to be adequately investigated from the present literature of road safety spatial analyses and can constitute meaningful future research endeavors. An important aspect that was does not appear to be adequately investigated is that of micro-level road safety and event analysis with spatial modelling considerations. A small number of studies has been found to explore concepts such as automated conflict extraction via trajectory analyses using automated data (Saunier and Sayed, 2007; St-Aubin et al., 2015). The inclusion of spatial effects in such design concepts would be very interesting for the determination of the influence of spatial effects at a small-unit level.

While crash counts have been examined extensively, their distributions over several categories have received less focus within a spatial context. The recent fractional approach by Lee et al. (2018b) that examines crash distribution across vehicle types is an example towards that direction, as is the examination per crash type proposed by Aguero-Valverde et al. (2016). Nonetheless, more research is needed on the manner in which various categories of crashes occur across study areas. The distribution of exact crash proportions and the factors that affect them needs to be researched within a spatial context. For instance, injury severity distributions have not been investigated as frequently as crash counts; rather, they have mostly been used as a categorization mechanism. By jointly examining crash severities and occurrence while taking spatial effects into account, more informative results can be reached for practitioners. Similar potential exists for studies aiming to examine casualty rates. In addition to the previous, it would be interesting to spatially analyze other road safety indicators, such as those related to driver behavior: conflicts, near-misses, harsh events and traffic law violations. These can aid in determining high crash concentrations and locations of poor road safety performance (hotspots).

Hotspot detection, or problematic region identification in greater scales, is a crucial advantage typically provided by spatial analyses for locating problems. Therefore, the determination of the spatial impacts of implemented road safety measures would also be very beneficial. Before-after studies within a spatial context (or even a spatiotemporal context, if a dedicated data collection scheme can be set) would allow observation of crash reductions due to targeted observations from the initial analyses. Such study designs would also allow the examination of the variation of spatial autocorrelation of events (and whether any exists) before and after interventions, and would offer interesting insights in any possible crash mitigation phenomena. Another promising research direction is the transfer and application of more focused spatial analysis methods for the examination of segments of a contiguous road network, similar to network KDE approaches, so that segments are assessed instead of areal units, but in the form of an extended and complex road network, as an expansion of the segment analysis approaches mentioned in section 2.1.

Some spatial issues, while proven to exist, need to be further analyzed to increase comprehensiveness. The specific effective range of spatial correlation among analysis units, as studied by Aguero-Valverde (2014) and Wang et al. (2016b) needs to be expanded upon. Again, there is a need for results for different road environments, road users, crash types and injury severities in order to obtain measures of the extent that spatial dependency needs to be accounted for. In addition, different countries are expected to produce varying results, possibly due to differences in driving culture or other unobserved factors.

Another direction that would increase the low transferability of results of spatial analysis is the creation of common frameworks for the two famous problems (boundary and MAUP), preferably on the international scale. The establishment of an acceptable boundary value in order to address boundary issues under different conditions, as suggested by Zhai et al. (2018b), is such an example. More effort is needed to be devoted to understanding the impacts of both the boundary issue and MAUP across areal unit sizes as well, especially if different contributor variables are found in boundaries. Similarly, methods to obtain more homogeneous road segments or areal units need to be developed, in an effort to reduce heterogeneity. They would have to be comprehensible and straightforward in order to be more widely accepted and applied by practitioners worldwide.

Yet another finding from the reviewed studies is that built environment is not very strictly defined in the sense that every study selects some of its characteristics to examine. In a dedicated study, Ukkusuri et al. (2012) include in the term built environment factors such as land use patterns, population characteristics such as age profiles and professional driver percentages, road infrastructure and transit characteristics. This review has not exhausted all built environment parameters, and the investigation of more specific variables such as the presence of refuge islands or crosswalks or proximity to health or education buildings merit additional investigation, and can be a future direction of targeted road safety spatial analyses.

These endeavors can all be further augmented by new technological developments, such as transport applications of big data, cloud computing and connected & autonomous vehicle technologies that can be used to provide a more connected spatial environment (e.g. as in Bao et al., 2018). For instance, it has been found that smartphone technology sampling can provide a vast amount of driving data in real conditions, including risk factors such as distraction and speeding (Papadimitriou et al., 2018), while achieving a seamless transition from data collection to data analysis (Yannis et al., 2017). This framework could enable not only a collection of a wealth of real-time information across several spatial unit levels, but also allow for easier calibration of spatial models without the doubt of transferability that is often present in spatial analyses.

Acknowledgements

This research is co-financed by Greece and the European Union (European Social Fund - ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project "Strengthening Human Resources Research Potential via Doctorate Research" (MIS-5000432), implemented by the State Scholarships Foundation (IKY).

The authors would also like to thank three anonymous reviewers for providing valuable suggestions to increase the completeness of this study.

References

Abdel-Aty, M.A., Lee, J., Eluru, N., Cai, Q., Al Amili, S., Alarifi, S., 2016. Enhancing and generalizing the two-level screening approach incorporating the highway safety manual (HSM) methods, Phase 2. University of Central Florida, Department of Civil, Environmental and Construction Engineering.
Abdel-Aty, M., Lee, J., Siddiqui, C., Choi, K., 2013. Geographical unit based analysis in the context of transportation safety planning. Transportation Research Part A: Policy and Practice 49, 62–75.
Abdel-Aty, M., Siddiqui, C., Huang, H., Wang, X., 2011. Integrating trip and roadway characteristics to manage safety in traffic analysis zones. Transportation Research Record 2213 (1), 20–28.
Abdel-Aty, M., Wang, X., 2006. Crash estimation at signalized intersections along corridors: analyzing spatial effect and identifying significant factors. Transportation Research Record: Journal of the Transportation Research Board 1953, 98–111.
Aguero-Valverde, J., Wu, K.F., Donnell, E.T., 2016. A multivariate spatial crash frequency model for identifying sites with promise based on crash types. Accident Analysis and Prevention 87, 8–16.
Aguero-Valverde, J., 2014. Direct spatial correlation in crash frequency models: estimation of the effective range. Journal of Transportation Safety & Security 6 (1), 21–33.
Aguero-Valverde, J., 2013. Multivariate spatial models of excess crash frequency at area level: Case of Costa Rica. Accident Analysis & Prevention 59, 365–373.
Aguero-Valverde, J., Jovanis, P.P., 2010. Spatial correlation in multilevel crash frequency models: Effects of different neighboring structures. Transportation Research Record 2165 (1), 21–32.
Aguero-Valverde, J., Jovanis, P.P., 2008. Analysis of road crash frequency with spatial models. Transportation Research Record 2061 (1), 55–63.
Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis & Prevention 38 (3), 618–625.
Alarifi, S.A., Abdel-Aty, M.A., Lee, J., Wang, X., 2018. Exploring the effect of different neighboring structures on spatial hierarchical joint crash frequency models. Transportation research record 2672 (38), 210–222.
Alarifi, S.A., Abdel-Aty, M.A., Lee, J., Park, J., 2017. Crash modeling for intersections and segments along corridors: a Bayesian multilevel joint model with random parameters. Analytic methods in accident research 16, 48–59.
Amoh-Gyimah, R., Saberi, M., Sarvi, M., 2017. The effect of variations in spatial units on unobserved heterogeneity in macroscopic crash models. Analytic methods in accident research 13, 28–51.
Anderson, T.K., 2009. Kernel density estimation and K-means clustering to profile road accident hotspots. Accident Analysis & Prevention 41 (3), 359–364.
Anderson, T., 2007. Comparison of spatial methods for measuring road accident 'hotspots': a case study of London. Journal of Maps 3 (1), 55–63.
Atubi, A.O., 2012. Determinants of road traffic accident occurrences in Lagos State: Some lessons for Nigeria. International Journal of Humanities and Social Science 2 (6), 252–259.
Bao, J., Liu, P., Ukkusuri, S.V., 2019. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. Accident Analysis & Prevention 122, 239–254.
Bao, J., Liu, P., Qin, X., Zhou, H., 2018. Understanding the effects of trip patterns on spatially aggregated crashes with large-scale taxi GPS data. Accident Analysis & Prevention 120, 281–294.
Bao, J., Liu, P., Yu, H., Xu, C., 2017. Incorporating twitter-based human activity information in spatial analysis of crashes in urban areas. Accident Analysis & Prevention 106, 358–369.
Barua, S., El-Basyouny, K., Islam, M.T., 2016. Multivariate random parameters collision count data models with spatial heterogeneity. Analytic methods in accident research 9, 1–15.
Barua, S., El-Basyouny, K., Islam, M.T., 2014. A full Bayesian multivariate count data model of collision severity with spatial correlation. Analytic Methods in Accident Research 3, 28–43.
Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Annals of the institute of statistical mathematics 43 (1), 1–20.
Bíl, M., Andrášik, R., Janoška, Z., 2013. Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. Accident Analysis & Prevention 55, 265–273.
Bivand, R., Müller, W.G., Reder, M., 2009. Power calculations for global and local Moran's I. Computational Statistics & Data Analysis 53 (8), 2859–2872.
Bu, L., Wang, F., Gong, H., 2018. Spatial and factor analysis of vehicle crashes in Mississippi state. Natural Hazards 1–22.
Cai, Q., Abdel-Aty, M., Sun, Y., Lee, J., Yuan, J., 2019a. Applying a deep learning approach for transportation safety planning by using high-resolution transportation and land use data. Transportation Research Part A: Policy and Practice 127, 71–85.
Cai, Q., Abdel-Aty, M., Lee, J., Huang, H., 2019b. Integrating macro-and micro-level safety analyses: a Bayesian approach incorporating spatial interaction. Transportmetrica A: Transport Science 15 (2), 285–306. https://doi.org/10.1080/23249935.2018.1471752.
Cai, Q., Abdel-Aty, M., Lee, J., Wang, L., Wang, X., 2018. Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling. Analytic methods in accident research 19, 1–15.
Cai, Q., Abdel-Aty, M., Lee, J., Eluru, N., 2017a. Comparative analysis of zonal systems for macro-level crash modeling. Journal of safety research 61, 157–166.
Cai, Q., Abdel-Aty, M., Lee, J., 2017b. Macro-level vulnerable road users crash analysis: a Bayesian joint modeling approach of frequency and proportion. Accident Analysis & Prevention 107, 11–19.
Cai, Q., Lee, J., Eluru, N., Abdel-Aty, M., 2016. Macro-level pedestrian and bicycle crash analysis: Incorporating spatial spillover effects in dual state count models. Accident Analysis & Prevention 93, 14–22.
Chang, L.Y., Wang, H.W., 2006. Analysis of traffic injury severity: An application of non-parametric classification tree techniques. Accident Analysis & Prevention 38 (5), 1019–1027.
Chiou, Y.C., Fu, C., Chih-Wei, H., 2014. Incorporating spatial dependence in simultaneously modeling crash frequency and severity. Analytic methods in accident research 2, 1–11.
Chiou, Y.C., Fu, C., 2013. Modeling crash frequency and severity using multinomial-generalized Poisson model with error components. Accident Analysis & Prevention 50, 73–82.
Chung, W., Abdel-Aty, M., Lee, J., 2018. Spatial analysis of the effective coverage of land-based weather stations for traffic crashes. Applied geography 90, 17–27.
Cottrill, C.D., Thakuriah, P.V., 2010. Evaluating pedestrian crashes in areas with high low-income or minority populations. Accident Analysis & Prevention 42 (6), 1718–1728.
Cui, G., Wang, X., Kwon, D.W., 2015. A framework of boundary collision data aggregation into neighbourhoods. Accident Analysis & Prevention 83, 1–17.
Delmelle, E., Thill, J.C., 2008. Urban bicyclists: spatial analysis of adult and youth traffic hazard intensity. Transportation Research Record: Journal of the Transportation Research Board 2074, 31–39.
Dong, N., Huang, H., Lee, J., Gao, M., Abdel-Aty, M., 2016. Macroscopic hotspots identification: a Bayesian spatio-temporal interaction approach. Accident Analysis & Prevention 92, 256–264.
Dong, N., Huang, H., Zheng, L., 2015. Support vector machine in crash prediction at the level of traffic analysis zones: assessing the spatial proximity effects. Accident Analysis & Prevention 82, 192–198.
Dong, N., Huang, H., Xu, P., Ding, Z., Wang, D., 2014. Evaluating spatial-proximity structures in crash prediction models at the level of traffic analysis zones. Transportation Research Record: Journal of the Transportation Research Board 2432, 46–52.
Effati, M., Thill, J.C., Shabani, S., 2015. Geospatial and machine learning techniques for wicked social science problems: analysis of crash severity on a regional highway corridor. Journal of Geographical Systems 17 (2), 107–135.
El-Basyouny, K., Sayed, T., 2011. A full Bayes multivariate intervention model with random parameters among matched pairs for before–after safety evaluation. Accident Analysis & Prevention 43 (1), 87–94.
El-Basyouny, K., Sayed, T., 2009. Urban arterial accident prediction models with spatial effects. Transportation Research Record: Journal of the Transportation Research Board 2102, 27–33.
Elvik, R., Vaa, T., Hoye, A., Sorensen, M. (Eds.), 2009. The handbook of road safety measures. Emerald Group Publishing.
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.
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.
Flahaut, B., 2004. Impact of infrastructure and local environment on road unsafety: Logistic modeling with spatial autocorrelation. Accident Analysis & Prevention 36 (6), 1055–1066.
Flask, T., Schneider IV, W., 2013. A Bayesian analysis of multi-level spatial correlation in single vehicle motorcycle crashes in Ohio. Safety science 53, 1–10.
Fotheringham, A.S., Brunsdon, C., Charlton, M., 2002. Geographically weighted regression. John Wiley & Sons, Limited, West Atrium, pp. 159–183.
Fotheringham, S., Brunsdon, C., Charlton, M., 2000. Quantitative Geography: Perspectives on Spatial Data Analysis. Sage, London.
Fotheringham, S., Wegener, M., 1999. Spatial models and GIS: New and potential models Vol. 7 CRC press.
Gomes, M.J.T.L., Cunto, F., da Silva, A.R., 2017. Geographically weighted negative binomial regression applied to zonal level safety performance models. Accident

Analysis & Prevention 106, 254–261.
Geurts, K., Wets, G., 2003. Black spot analysis methods: Literature review. Flemish Research Center for Traffic Safety, Diepenbeek, Belgium.
Guo, Q., Xu, P., Pei, X., Wong, S.C., Yao, D., 2017. The effect of road network patterns on pedestrian safety: A zone-based Bayesian spatial modeling approach. Accident Analysis & Prevention 99, 114–124.
Guo, F., Wang, X., Abdel-Aty, M.A., 2010. Modeling signalized intersection safety with corridor-level spatial correlations. Accident Analysis & Prevention 42 (1), 84–92.
Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2010. Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. Accident Analysis & Prevention 42 (2), 676–688.
Hadayeghi, A., Shalaby, A., Persaud, B., 2003. Macrolevel accident prediction models for evaluating safety of urban transportation systems. Transportation Research Record: Journal of the Transportation Research Board 1840, 87–95.
Han, C., Huang, H., Lee, J., Wang, J., 2018. Investigating varying effect of road-level factors on crash frequency across regions: a Bayesian hierarchical random parameter modeling approach. Analytic methods in accident research 20, 81–91.
Hart, T., Zandbergen, P., 2014. Kernel density estimation and hotspot mapping: Examining the influence of interpolation method, grid cell size, and bandwidth on crime forecasting. Policing: An International Journal of Police Strategies & Management 37 (2), 305–323.
Hauer, E., 1997. Observational before/after studies in road safety. Estimating the effect of highway and traffic engineering measures on road safety.
Hauer, E., Ng, J.C., Lovell, J., 1988. Estimation of safety at signalized intersections (with discussion and closure) (No. 1185).
Huang, H., Zhou, H., Wang, J., Chang, F., Ma, M., 2017. A multivariate spatial model of crash frequency by transportation modes for urban intersections. Analytic methods in accident research 14, 10–21.
Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., Abdel-Aty, M., 2016. Macro and micro models for zonal crash prediction with application in hot zones identification. Journal of Transport Geography 54, 248–256.
Huang, H., Abdel-Aty, M., 2010. Multilevel data and Bayesian analysis in traffic safety. Accident Analysis & Prevention 42 (6), 1556–1565.
Huang, H., Abdel-Aty, M., Darwiche, A., 2010. County-level crash risk analysis in Florida: Bayesian spatial modeling. Transportation Research Record: Journal of the Transportation Research Board 2148, 27–37.
Huang, H., Chin, H., Haque, M., 2009. Empirical evaluation of alternative approaches in identifying crash hot spots: naive ranking, empirical Bayes, and full Bayes methods. Transportation Research Record: Journal of the Transportation Research Board 2103, 32–41.
Imprialou, M.I.M., Quddus, M., Pitfield, D.E., Lord, D., 2016. Re-visiting crash–speed relationships: A new perspective in crash modelling. Accident Analysis & Prevention 86, 173–185.
Jiang, X., Abdel-Aty, M., Hu, J., Lee, J., 2016. Investigating macro-level hotzone identification and variable importance using big data: A random forest models approach. Neurocomputing 181, 53–63.
Kim, K., Brunner, I.M., Yamashita, E.Y., 2006. Influence of land use, population, employment, and economic activity on accidents. Transportation research record 1953 (1), 56–64.
Ladron de Guevara, F., Washington, S., Oh, J., 2004. Forecasting crashes at the planning level: simultaneous negative binomial crash model applied in Tucson. Arizona. Transportation Research Record: Journal of the Transportation Research Board 1897, 191–199.
LaScala, E.A., Gruenewald, P.J., Johnson, F.W., 2004. An ecological study of the locations of schools and child pedestrian injury collisions. Accident Analysis & Prevention 36 (4), 569–576.
LaScala, E.A., Johnson, F.W., Gruenewald, P.J., 2001. Neighborhood characteristics of alcohol-related pedestrian injury collisions: a geostatistical analysis. Prevention Science 2 (2), 123–134.
LaScala, E.A., Gerber, D., Gruenewald, P.J., 2000. Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. Accident Analysis & Prevention 32 (5), 651–658.
Lee, J., Abdel-Aty, M., 2018. Macro-level analysis of bicycle safety: Focusing on the characteristics of both crash location and residence. International journal of sustainable transportation 12 (8), 553–560.
Lee, J., Abdel-Aty, M., Huang, H., Cai, Q., 2019a. Transportation Safety Planning Approach for Pedestrians: An Integrated Framework of Modeling Walking Duration and Pedestrian Fatalities. Transportation Research Record 2673 (4), 898–906.
Lee, J., Abdel-Aty, M., De Blasiis, M.R., Wang, X., Mattei, I., 2019b. International transferability of macro-level safety performance functions: a case study of the United States and Italy. Transportation Safety and Environment.
Lee, J., Abdel-Aty, A., Park, J., 2018a. Investigation of associations between marijuana law changes and marijuana-involved fatal traffic crashes: A state-level analysis. Journal of Transport & Health 10, 194–202.
Lee, J., Yasmin, S., Eluru, N., Abdel-Aty, M., Cai, Q., 2018b. Analysis of crash proportion by vehicle type at traffic analysis zone level: A mixed fractional split multinomial logit modeling approach with spatial effects. Accident Analysis & Prevention 111, 12–22.
Lee, J., Abdel-Aty, M., Cai, Q., Wang, L., Huang, H., 2018c. Integrated modeling approach for non-motorized mode trips and fatal crashes in the framework of transportation safety planning. Transportation research record 2672 (32), 49–60.
Lee, J., Abdel-Aty, M., Cai, Q., 2017a. Intersection crash prediction modeling with macro-level data from various geographic units. Accident Analysis & Prevention 102, 213–226.
Lee, J., Abdel-Aty, M., Wang, J.H., Lee, C., 2017b. Long-term effect of universal helmet law changes on motorcyclist fatal crashes: comparison group and empirical Bayes approaches. Transportation Research Record 2637 (1), 27–37.
Lee, J., Abdel-Aty, M., Jiang, X., 2015a. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. Accident Analysis & Prevention 78, 146–154.
Lee, J., Abdel-Aty, M., Choi, K., Huang, H., 2015b. Multi-level hot zone identification for pedestrian safety. Accident Analysis & Prevention 76, 64–73.
Lee, J., Abdel-Aty, M., Choi, K., 2014a. Analysis of residence characteristics of at-fault drivers in traffic crashes. Safety science 68, 6–13.
Lee, J., Abdel-Aty, M., Jiang, X., 2014b. Development of zone system for macro-level traffic safety analysis. Journal of transport geography 38, 13–21.
Levine, N., Kim, K.E., Nitz, L.H., 1995. Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. Accident Analysis & Prevention 27 (5), 675–685.
Li, Z., Chen, X., Ci, Y., Chen, C., Zhang, G., 2019. A hierarchical Bayesian spatiotemporal random parameters approach for alcohol/drug impaired-driving crash frequency analysis. Analytic Methods in Accident Research.
Li, Z., Wang, W., Liu, P., Bigham, J.M., Ragland, D.R., 2013. Using geographically weighted Poisson regression for county-level crash modeling in California. Safety science 58, 89–97.
Liu, C., Sharma, A., 2018. Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity. Analytic methods in accident research 17, 14–31.
Liu, J., Khattak, A.J., Wali, B., 2017. Do safety performance functions used for predicting crash frequency vary across space? Applying geographically weighted regressions to account for spatial heterogeneity. Accident Analysis & Prevention 109, 132–142.
Loo, B.P., Anderson, T.K., 2015. Spatial Analysis Methods of Road Traffic Collisions. CRC Press.
Loo, B.P., Yao, S., Wu, J., 2011. Spatial point analysis of road crashes in Shanghai: A GIS-based network kernel density method. June In: In 2011 19th international conference on geoinformatics. IEEE. pp. 1–6.
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.
Loukaitou-Sideris, A., Liggett, R., Sung, H.G., 2007. Death on the crosswalk: A study of pedestrian-automobile collisions in Los Angeles. Journal of Planning Education and Research 26 (3), 338–351.
Lovegrove, G., Lim, C., Sayed, T., 2009. Community-based, macrolevel collision prediction model use with a regional transportation plan. Journal of transportation engineering 136 (2), 120–128.
Lovegrove, G., Sayed, T., 2007. Macrolevel collision prediction models to enhance traditional reactive road safety improvement programs. Transportation Research Record: Journal of the Transportation Research Board 2019, 65–73.
Lovegrove, G.R., Sayed, T., 2006. Macro-level collision prediction models for evaluating neighbourhood traffic safety. Canadian Journal of Civil Engineering 33 (5), 609–621.
Ma, X., Chen, S., Chen, F., 2017. Multivariate space-time modeling of crash frequencies by injury severity levels. Analytic Methods in Accident Research 15, 29–40.
Ma, X., Tao, Z., Wang, Y., Yu, H., Wang, Y., 2015a. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research Part C: Emerging Technologies 54, 187–197.
Ma, X., Yu, H., Wang, Y., Wang, Y., 2015b. Large-scale transportation network congestion evolution prediction using deep learning theory. PloS one 10 (3), e0119044.
MacNab, Y.C., 2004. Bayesian spatial and ecological models for small-area accident and injury analysis. Accident Analysis & Prevention 36 (6), 1019–1028.
Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. Analytic methods in accident research 1, 1–22.
Mitra, S., 2009. Spatial autocorrelation and Bayesian spatial statistical method for analyzing intersections prone to injury crashes. Transportation research record 2136 (1), 92–100.
Miaou, S.P., Lord, D., 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes methods. Transportation Research Record: Journal of the Transportation Research Board 1840, 31–40.
Miaou, S.P., Song, J.J., 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence. Accident Analysis & Prevention 37 (4), 699–720.
Moeinaddini, M., Asadi-Shekari, Z., Shah, M.Z., 2014. The relationship between urban street networks and the number of transport fatalities at the city level. Safety science 62, 114–120.
Mohaymany, A.S., Shahri, M., Mirbagheri, B., 2013. GIS-based method for detecting high-crash-risk road segments using network kernel density estimation. Geo-spatial Information Science 16 (2), 113–119.
Mountrakis, G., Gunson, K., 2009. Multi-scale spatiotemporal analyses of moose–vehicle collisions: a case study in northern Vermont. International Journal of Geographical Information Science 23 (11), 1389–1412.
Naderan, A., Shahi, J., 2010. Aggregate crash prediction models: Introducing crash generation concept. Accident Analysis & Prevention 42 (1), 339–346.
Narayanamoorthy, S., Paleti, R., Bhat, C.R., 2013. On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. Transportation research part B: methodological 55, 245–264.
Nashad, T., Yasmin, S., Eluru, N., Lee, J., Abdel-Aty, M.A., 2016. Joint modeling of pedestrian and bicycle crashes: copula-based approach. Transportation Research Record: Journal of the Transportation Research Board 2601, 119–127.
Ng, K.S., Hung, W.T., Wong, W.G., 2002. An algorithm for assessing the risk of traffic accident. Journal of safety research 33 (3), 387–410.
Noland, R.B., Quddus, M.A., 2005. Congestion and safety: A spatial analysis of London. Transportation Research Part A: Policy and Practice 39 (7-9), 737–754.
Noland, R.B., Oh, L., 2004. The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data.

Accident Analysis & Prevention 36 (4), 525–532.
Noland, R.B., Quddus, M.A., 2004. A spatially disaggregate analysis of road casualties in England. Accident Analysis & Prevention 36 (6), 973–984.
Openshaw, S., 1984. The modifiable areal unit problem. Concepts and techniques in modern geography. Geo Books, Norwich.
Ossenbruggen, P.J., Linder, E., Nguyen, B., 2009. Detecting unsafe roadways with spatial statistics: point patterns and geostatistical models. Journal of Transportation Engineering 136 (5), 457–464.
Page, S.J., Meyer, D., 1996. Tourist accidents: an exploratory analysis. Annals of Tourism Research 23 (3), 666–690.
Papadimitriou, E., Filtness, A., Theofilatos, A., Ziakopoulos, A., Quigley, C., Yannis, G., 2019. Review and ranking of crash risk factors related to the road infrastructure. Accident Analysis & Prevention 125, 85–97.
Papadimitriou, E., Tselentis, D.I., Yannis, G., 2018. Analysis of Driving Behaviour Characteristics Based on Smartphone Data. In: Proceedings of 7th Transport Research Arena TRA 2018. Vienna, Austria. pp. 2018 April 16-19.
Pirdavani, A., Bellemans, T., Brijs, T., Wets, G., 2014a. Application of geographically weighted regression technique in spatial analysis of fatal and injury crashes. Journal of Transportation Engineering 140 (8), 04014032.
Pirdavani, A., Bellemans, T., Brijs, T., Kochan, B., Wets, G., 2014b. Assessing the road safety impacts of a teleworking policy by means of geographically weighted regression method. Journal of transport geography 39, 96–110.
Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., Wets, G., 2013. Evaluating the road safety effects of a fuel cost increase measure by means of zonal crash prediction modeling. Accident Analysis & Prevention 50, 186–195.
Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accident Analysis & Prevention 40 (4), 1486–1497.
Ratti, C., 2004. Space syntax: some inconsistencies. Environment and Planning B: Planning and Design 31 (4), 487–499.
Raykar, V.C., Duraiswami, R., 2006. Fast optimal bandwidth selection for kernel density estimation. In: Proceedings of the 2006 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics. pp. 524–528.
Rhee, K.A., Kim, J.K., Lee, Y.I., Ulfarsson, G.F., 2016. Spatial regression analysis of traffic crashes in Seoul. Accident Analysis & Prevention 91, 190–199.
Saunier, N., Sayed, T., 2007. Automated analysis of road safety with video data. Transportation Research Record: Journal of the Transportation Research Board 2019, 57–64.
Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accident Analysis & Prevention 45, 382–391.
Siddiqui, C., Abdel-Aty, M., 2012. Nature of modeling boundary pedestrian crashes at zones. Transportation Research Record 2299 (1), 31–40.
Soltani, A., Askari, S., 2017. Exploring spatial autocorrelation of traffic crashes based on severity. Injury 48 (3), 637–647.
Song, J.J., Ghosh, M., Miaou, S., Mallick, B., 2006. Bayesian multivariate spatial models for roadway traffic crash mapping. Journal of multivariate analysis 97 (1), 246–273.
St-Aubin, P., Saunier, N., Miranda-Moreno, L., 2015. Large-scale automated proactive road safety analysis using video data. Transportation Research Part C: Emerging Technologies 58, 363–379.
Tasic, I., Elvik, R., Brewer, S., 2017. Exploring the safety in numbers effect for vulnerable road users on a macroscopic scale. Accident Analysis & Prevention 109, 36–46.
Theofilatos, A., Yannis, G., 2014. A review of the effect of traffic and weather characteristics on road safety. Accident Analysis & Prevention 72, 244–256.
Thomas, I., 1996. Spatial data aggregation: exploratory analysis of road accidents. Accident Analysis & Prevention 28 (2), 251–264.
Ukkusuri, S., Miranda-Moreno, L.F., Ramadurai, G., Isa-Tavarez, J., 2012. The role of built environment on pedestrian crash frequency. Safety science 50 (4), 1141–1151.
Ukkusuri, S., Hasan, S., Aziz, H., 2011. Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency. Transportation Research Record: Journal of the Transportation Research Board 2237, 98–106.
Ver Hoef, J.M., Peterson, E.E., Hooten, M.B., Hanks, E.M., Fortin, M.J., 2018. Spatial autoregressive models for statistical inference from ecological data. Ecological Monographs 88 (1), 36–59.
Wang, L., Abdel-Aty, M., Lee, J., Shi, Q., 2019. Analysis of real-time crash risk for expressway ramps using traffic, geometric, trip generation, and socio-demographic predictors. Accident Analysis & Prevention 122, 378–384.
Wang, Y., Veneziano, D., Russell, S., Al-Kaisy, A., 2016a. Traffic Safety Along Tourist Routes in Rural Areas. Transportation Research Record: Journal of the Transportation Research Board 2568, 55–63.
Wang, X., Yang, J., Lee, C., Ji, Z., You, S., 2016b. Macro-level safety analysis of pedestrian crashes in Shanghai. China. Accident Analysis & Prevention 96, 12–21.
Wang, J., Huang, H., 2016. Road network safety evaluation using Bayesian hierarchical joint model. Accident Analysis & Prevention 90, 152–158.
Wang, Y., Kockelman, K.M., 2013. A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. Accident Analysis & Prevention 60, 71–84.
Wang, C., Quddus, M.A., Ison, S.G., 2009. Impact of traffic congestion on road accidents: A spatial analysis of the M25 motorway in England. Accident Analysis & Prevention 41 (4), 798–808.
Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis & Prevention 38 (6), 1137–1150.
Wei, F., Lovegrove, G., 2013. An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression. Accident Analysis & Prevention 61, 129–137.
Wen, H., Zhang, X., Zeng, Q., Lee, J., Yuan, Q., 2019. Investigating spatial autocorrelation and spillover effects in freeway crash-frequency data. International journal of environmental research and public health 16 (2), 219.
Wier, M., Weintraub, J., Humphreys, E.H., Seto, E., Bhatia, R., 2009. An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. Accident Analysis & Prevention 41 (1), 137–145.
World Health Organization – WHO, 2018. Global status report on road safety 2018. Available from:. . https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/.
World Health Organization – WHO, 2015. Global status report on road safety 2015. Available from:. . http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/.
Xie, K., Ozbay, K., Kurkcu, A., Yang, H., 2017. Analysis of traffic crashes involving pedestrians using big data: Investigation of contributing factors and identification of hotspots. Risk analysis 37 (8), 1459–1476.
Xie, K., Wang, X., Ozbay, K., Yang, H., 2014. Crash frequency modeling for signalized intersections in a high-density urban road network. Analytic methods in accident research 2, 39–51.
Xie, K., Wang, X., Huang, H., Chen, X., 2013. Corridor-level signalized intersection safety analysis in Shanghai, China using Bayesian hierarchical models. Accident Analysis & Prevention 50, 25–33.
Xie, Z., Yan, J., 2008. Kernel density estimation of traffic accidents in a network space. Computers, environment and urban systems 32 (5), 396–406.
Xu, P., Huang, H., Dong, N., 2018. The modifiable areal unit problem in traffic safety: basic issue, potential solutions and future research. Journal of traffic and transportation engineering (English edition) 5 (1), 73–82.
Xu, P., Huang, H., Dong, N., Wong, S.C., 2017a. Revisiting crash spatial heterogeneity: a Bayesian spatially varying coefficients approach. Accident Analysis & Prevention 98, 330–337.
Xu, C., Li, H., Zhao, J., Chen, J., Wang, W., 2017b. Investigating the relationship between jobs-housing balance and traffic safety. Accident Analysis & Prevention 107, 126–136.
Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: random parameter versus geographically weighting. Accident Analysis & Prevention 75, 16–25.
Yannis, G., Tselentis, D.I., Vlahogianni, E.I., Argyropoulou, A., 2017. Monitoring distraction through smartphone naturalistic driving experiment. In: 6th International Naturalistic Driving Research Symposium. The Hague, Netherlands. 7-9 June 2017.
Yasmin, S., Eluru, N., 2016. Latent segmentation based count models: analysis of bicycle safety in Montreal and Toronto. Accident Analysis & Prevention 95, 157–171.
Zeng, Q., Huang, H., 2014. Bayesian spatial joint modeling of traffic crashes on an urban road network. Accident Analysis & Prevention 67, 105–112.
Zhai, X., Huang, H., Xu, P., Sze, N.N., 2019a. The influence of zonal configurations on macro-level crash modeling. Transportmetrica A: transport science 15 (2), 417–434.
Zhai, X., Huang, H., Sze, N.N., Song, Z., Hon, K.K., 2019b. Diagnostic analysis of the effects of weather condition on pedestrian crash severity. Accident Analysis & Prevention 122, 318–324.
Zhai, X., Huang, H., Gao, M., Dong, N., Sze, N.N., 2018. Boundary crash data assignment in zonal safety analysis: an iterative approach based on data augmentation and Bayesian spatial model. Accident Analysis & Prevention 121, 231–237.
Zhu, L., Guo, F., Krishnan, R., Polak, J.W., 2018. The Use of Convolutional Neural Networks for Traffic Incident Detection at a Network Level (No. 18-00321).