Physica A 581 (2021) 126171

Contents lists available at ScienceDirect

Physica A

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

Bayesian space–time modeling of bicycle and pedestrian crash risk by injury severity levels to explore the long-term spatiotemporal effects

Peijie Wu [a], Xianghai Meng [a,*], Li Song [b]

[a] School of Transportation Science and Engineering, Harbin Institute of Technology, 150090, China
[b] Department of Civil and Environmental Engineering, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

ARTICLE INFO

Article history:
Received 14 January 2021
Received in revised form 26 May 2021
Available online 19 June 2021

Keywords:
Vulnerable road users
Zonal modeling
Crash frequency
Multivariate Bayesian space–time models
Spatiotemporal effects

ABSTRACT

Vulnerable road users (VRUs)-related crashes are recognized as an important public safety problem. However, few macro-level studies of VRUs-involved crashes have considered the long-term spatial, temporal, or spatiotemporal effects in the crash risk. This study analyzes the bicycle and pedestrian crash risk in different injury severities by using three multivariate Bayesian space–time models. These models address different spatiotemporal effects to account for possible correlations across injury severities over space and time. Various explanatory variables are used to examine the contributory risk factors, including socio-demographic features, roadway structures, and weather characteristics. Spatio-temporal conditional autoregression with an ANOVA style (ST-CARanova) models outperform other two space–time models in most circumstances. The long-term spatiotemporal effects, such as relatively high temporal autocorrelations, significant spatial heterogeneity, and weak spatiotemporal interactions, are found in this study. The increase of female ratios, young people ratios, unemployment rates, and annual average high temperatures could increase the county-level crash risk of cyclists and pedestrians. The findings provide useful insights for policy makers to improve the safety of cyclists and pedestrians.

1. Introduction

Cycling and walking are active transportation modes that are widely recognized as sustainable, environmentally friendly, and resource-saving for many commuters. Although there are environmental and health benefits of cycling and walking, users of these modes are frequently exposed to severe injury in road crashes [1–3]. In the U.S., there was a 3% increase in the number of pedestrians killed while there was a 10% increase in bicycle deaths in crashes in 2018 when it is compared with 2017 [4]. Safety of vulnerable road users (VRUs) has been paid increasing attention in recent years [5] and is recognized as an important public issue. Thus, it is necessary to make considerable efforts to enhance the safety of VRUs. An efficient approach is the macroscopic crash modeling of VRUs, which can examine the effects of zonal factors on bicycle and pedestrian crashes and analyze the spatiotemporal trends of the crash risk [3,6–8]. Crash risk in this study refers to crash rates, and is calculated as the number of bicycle or pedestrian crashes divided by the total number of cycling or walking trips. By further understanding the contributory zonal factors of the VRUs-involved crash risk, both

* Corresponding author.
E-mail address: mengxianghai100@126.com (X. Meng).

https://doi.org/10.1016/j.physa.2021.126171
0378-4371/

Table 1
Summary of the methods used in crash frequency analysis for VRUs in the macroscopic scale.

Model	Research object	Spatial unit	Data and locations	Explanatory variables	If consider spatiotemporal effects?	Literature
Bayesian joint/ hierarchical model	Pedestrian and bicycle crashes	TAZs	2010–2012 Florida	Traffic and roadway & socio-demographic & land use & commuting variables	No	[16]
	Non-motorist crashes	TAZs	2010–2012 Florida	Traffic and roadway & socio-demographic & commuting variables	No	[3]
	Pedestrian, bicycle, two-wheelers users crashes	Wards	2010–2012 Delhi	Socio-demographic & traffic infrastructure & commuting variables	Only spatial correlation	[17]
	Pedestrian crashes	ZIP areas	2009–2011 Florida	Traffic and roadway & socio-demographic & land use & transit variables	Only spatial heterogeneity and correlation	[18]
	Pedestrian and bicycle crashes	TAZs	2005–2006 Florida	Traffic and roadway & socio-demographic & neighborhood-related variables	Only spatial correlation	[6]
Random parameter negative binomial model	Pedestrian–vehicle crashes	Census tracts	2002–2006 New York	Traffic and roadway & socio-demographic & land use & transit variables	No	[19]
	Bicycle crashes	Super output areas	2012–2013 Great London	Traffic and roadway & socio-demographic & land use variables	No	[20]
Generalized additive model	Vehicle–bicycle, vehicle–pedestrian crashes	Census tracts	2005–2012 Chicago	Traffic and roadway & socio-demographic & land use & commuting variables	Only spatial correlation	[21]
Random coefficients multivariate model	Pedestrian crashes	Census tracts	2009 Manhattan	Traffic and roadway & socio-demographic & land use & commuting & activity intensity & transit variables	Only spatial correlation	[22]
Geographically weighted regression model	Non-motorist crashes	Census tracts	2002–2006 California	Traffic and roadway & socio-demographic & land use & traffic infrastructure & structural measures	Only spatial correlation	[23]
Dual state count model	Pedestrian and bicycle crashes	TAZs	2010–2012 Florida	Traffic and roadway & socio-demographic & land use & commuting variables	Only spatial spillover effects	[2]
Decision tree model	Pedestrian and bicycle crashes	TAZs	2010–2012 Florida	Traffic and roadway & socio-demographic & land use & commuting variables	No	[24]

planning-stage and management-stage strategies could be adopted to proactively improve regional safety performance towards cyclists and pedestrians.

When it comes to crash frequency modeling, unobserved heterogeneity and correlations in the crash data are often regarded as a big challenge, and neglecting these characteristics may result in biased model results [9]. Underlying correlations between different crash severities may produce heterogeneity across observations [9,10]. Additionally, crash numbers are aggregated over space and time, which produces unobserved heterogeneity and correlations as well [9,11]. During the past decades, many contributions have been made to address the unobserved heterogeneity and correlations in vehicle–vehicle crashes from the aspect of space and time [12–15]. However, limited studies focused on the spatiotemporal effects of the crash risk in VRUs–vehicles. Especially, studies of the long-term (over 10 years) spatial trends, temporal trends, and spatiotemporal interactions for different severity levels of VRUs crashes seem to be inadequate so far.

In this study, we utilize three multivariate Bayesian space–time models to investigate the long-term spatiotemporal effects in bicycle and pedestrian crash risk by different injury severities in North Carolina from 2009 to 2018. Multiple spatiotemporal components are proposed and compared in order to choose the most appropriate model for the VRUs-involved crashes at the county-level. Zonal risk factors including socio-demographic features, roadway structures, and weather characteristics are also investigated to explore their effects on the crash risk of cyclists and pedestrians, respectively.

2. Previous work

Many studies have conducted the frequency analysis of VRUs-involved crashes in recent years. Table 1 shows these studies that have been conducted for different types of VRUs (e.g., pedestrian, cyclists, and two-wheeler users), spatial units (e.g., traffic analysis zones (e.g., TAZs, ZIP areas, and census tracts), temporal periods (1–8 years), and regions (e.g., U.S., U.K., Colombia) at the macroscopic scale.

Statistical approaches were widely used in the macro-level crash frequency modeling of VRUs, such as Bayesian joint/hierarchical models [3,6,16–18], random parameter negative binomial models [19,20], and generalized additive models [21,25]. Except for the above methods, random coefficients multivariate models [22], geographically weighted regression models [23], dual state count models [2], and machine learning algorithms [26] also show great potential in exploring zonal contributing factors of the VRUs-involved crashes. Ma et al. [14] suggested that the flexible structure of Bayesian hierarchical models can incorporate spatial correlations and be readily extended to address more complicated models. They asserted that a major drawback of some models is the lack of a correlation structure which can explicitly define the spatial or temporal correlations. Based on the Bayesian framework, a number of studies added spatial correlation components to investigate the spatial effects in crash data [6,17,18]. Recently, Bayesian space–time models have been proven powerful in detecting both spatial and temporal trends of crash numbers. Nevertheless, current researches of Bayesian space–time models mostly concentrated on vehicle–vehicle crashes instead of VRUs–vehicle crashes [12–14].

As for the exposure of VRUs–related crashes, there are different methods to estimate the actual exposure. Lee et al. [27] used a multiple linear regression model to calculate walking exposure (i.e., trips, lengths, or duration). Saad et al. [28] compared three adjustment methods of bicycle exposure, and they found that using both population and field observation data adjustment has the best performance of bicycle crash modeling. In terms of explanatory variables, some studies explored socio-demographic, land use, and commuting variables, and other studies also investigated the influence of traffic infrastructure, transit-related factors in VRUs–involved crashes. However, previous researches have neglected to consider the weather variables at the macro-level VRUs' crash frequency models. Several studies constructed zonal models of VRUs-involved crashes based on TAZs [2,3,6,16] and census tracts [19,21–23] for the accessibility of the data. Nevertheless, the long-term (more than ten years) spatiotemporal effects in the crash risk of VRUs were not investigated in previous studies.

For the spatiotemporal effects in the zonal modeling for VRUs–involved crashes, several studies are focusing on spatial correlation [17,21,22], spatial heterogeneity [18], and spatial spillover effects [2]. Liu and Sharma [15] pointed out that unobserved heterogeneity across space and time should be addressed, and it is often non-negligible in crash frequency modeling. Temporal correlations of crash numbers commonly exist because crash-related factors, such as economy, weather, environment, and travel modes exhibit temporal features. Eksler and Lassarre [12] proposed a Bayesian space–time model to assess the spatial distribution and temporal trends of crash risk in a cohesive way. Besides, multivariate Bayesian space–time models not only have the advantages of goodness-of-fit over univariate Bayesian space–time models but also could consider the correlation across different injury severities in the modeling [14,15].

This study applies three multivariate Bayesian space–time models on the crash risk of VRUs (bicycle–vehicle and pedestrian–vehicle in this study) by injury severities to explore the long-term spatiotemporal effects. The purpose of this study is to answer the following two questions: (i) what are the long-term spatiotemporal effects in the crash risk of different injury severities of VRUs? and (ii) how do the socio-demographic, roadway, weather factors have influence on the crash risk of VRUs by injury severities? The rest of the paper is organized as follows. Section 3 presents the Bayesian space–time models. Section 4 comprises the description of the crash data used for this study. Section 5 includes the analyses and discussions of the observed results. Conclusions are provided in Section 6.

3. Methodology

3.1. General framework

Poisson models are usually utilized to model crash frequency due to the nature of count data [29,30]. Extension models of Poisson distribution, such as Poisson–gamma and Poisson–lognormal models, were usually used to accommodate the overdispersion in crash counts [31]. Studies implied that the multivariate Poisson–lognormal model (MVPLN) is flexible as it allows for a more general correlation structure [31]. Besides, MVPLN have been proved to be more powerful than univariate Poisson–lognormal model (UVPLN) in the previous literature [32–34].

In this study, a Bayesian hierarchical architecture is applied in the general modeling framework by taking consideration of both spatial effects, temporal effects, and spatiotemporal interaction effects. In the first level of the hierarchical model, it is assumed that:

$$y_{stk} \sim Poisson(\lambda_{stk}) \tag{1}$$

$$y_{stk} \sim Poisson(E_{st}\theta_{stk}) \tag{2}$$

where y_{stk} is the crash number of injury severity k (i.e., killed and suspected serious injury (KA) crashes, suspected minor (B) crashes, and possible injury and no injury (CO)) for the area s (i.e., counties, $s = 1, 2, \ldots, 100$) in the tth year($t = 1, 2, \ldots, 10$); λ_{stk} is the mean crash number of injury severity k for the area s in the tth year; θ_{stk} is the crash risk of injury severity k for the area s in the tth year (i.e. crash number per 100,000 cycling or walking trips in this study), and E_{st} is the number of cycling or walking trips of area s in order to represent the exposure.

Next, MVPLN is formulated by specifying the crash risk at the second level of the hierarchical model

$$\log \theta_{stk} = \alpha_k + \boldsymbol{X}_{\boldsymbol{stk}}^{\boldsymbol{T}} \times \boldsymbol{\beta_k} + \psi_{stk} \tag{3}$$

where α_k is the intercept term of injury severity k; $\boldsymbol{\beta_k} = (\beta_{k1}, \beta_{k2}, \ldots, \beta_{km})$ is the m-dimensional regression coefficient vector of injury severity k, and m is the number of explanatory variables; $\boldsymbol{X_{stk}} = (X_{stk1}, X_{stk2}, \ldots, X_{stkm})$ is the m-dimensional explanatory variable vector for the area s in the tth year; ψ_{stk} represents spatiotemporal random effects of injury severity k for the area s in the tth year. Three candidate Bayesian space–time models with different spatiotemporal effects are constructed in this study to further investigate the unobserved heterogeneity across space and time in bicycle crash risk and pedestrian crash risk. Three candidate models are the spatio-temporal conditional autoregression with a first order autoregressive process (ST-CARar), spatio-temporal conditional autoregression with an ANOVA style (ST-CARanova), and spatio-temporal conditional autoregression with spatially adaptive smoothing (ST-CARadaptive).

3.2. ST-CARar model

The ST-CARar model proposes the spatiotemporal structure with a multivariate first-order autoregressive process to consider residual spatiotemporal autocorrelation in the data [35]. The model specification is given below.

$$\psi_{st} = \phi_{st} \qquad \mathrm{s} = 1, \ldots, \mathrm{K} \tag{4}$$

$$\boldsymbol{\phi_t} \big| \boldsymbol{\phi_{t-1}} \sim \mathrm{N}(\rho_T \boldsymbol{\phi_{t-1}}, \tau^2 \mathbf{Q}(\boldsymbol{W}, \rho_S)^{-1}) \qquad t = 2, \ldots, N \tag{5}$$

$$\boldsymbol{\phi}_1 \sim \mathrm{N}(0, \tau^2 \mathbf{Q}(\boldsymbol{W}, \rho_S)^{-1}) \tag{6}$$

$$\tau^2 \sim \mathrm{Inverse} - \mathrm{Gamma}(a, b) \tag{7}$$

$$\rho_S, \rho_T \sim \mathrm{Uniform}(0, 1) \tag{8}$$

where $\boldsymbol{\phi_t} = (\boldsymbol{\phi_{1t}}, \ldots, \boldsymbol{\phi_{St}})$ is the vector of random effects for tth year, which evolved over time through a multivariate first-order autoregressive process with temporal autoregressive parameter ρ_T. Thus, temporal autocorrelation is induced via the mean $\rho_T \boldsymbol{\phi_{t-1}}$. The level of temporal autocorrelation is controlled by ρ_T, $\rho_T = 0$ means temporal independence, and $\rho_T = 1$ means high temporal autocorrelation with a first-order random walk model. $\boldsymbol{W} = (\omega_{ik})$ denotes a binary $K \times K$ adjacency matrix, which is based on the contiguity structure of the area S. Element $\omega_{ik} = 1$ means that areal unit i shares a border with areal unit k, otherwise $\omega_{ik} = 0$. Spatial autocorrelation is induced by the precision matrix $\mathbf{Q}(\mathbf{W}, \rho_S)$ [36] and corresponds to the CAR models. The algebraic form of this matrix is given by:

$$\mathbf{Q}(\boldsymbol{W}, \rho_S) = \rho_S [\mathrm{diag}(\boldsymbol{W} 1) - \boldsymbol{W}] + (1 - \rho_S) \boldsymbol{I} \tag{9}$$

where 1 is the $K \times 1$ vector of ones while $\boldsymbol{I}$ is the $K \times K$ identity matrix; Spatial autocorrelation is induced by the variance $\tau^2 \mathbf{Q}(\boldsymbol{W}, \rho_S)^{-1}$. $\rho_S = 1$ represents the intrinsic CAR prior [37] for high autocorrelation; $\rho_S = 0$ corresponds to independent random effects with constant mean and variance. Therefore, this model induced temporal autocorrelation through the conditional expectation while spatial autocorrelation is induced via the precision matrix.

3.3. ST-CARanova model

The ST-CARanova model is a modification of the model proposed by Knorr-Held [38], and its spatiotemporal structure consists of an overall spatial effect, an overall temporal trend, and an independent space–time interaction.

$$\psi_{stk} = \phi_{sk} + \delta_{tk} + \gamma_{stk} \tag{10}$$

$$\boldsymbol{\phi_{sk}} \big| \boldsymbol{\phi_{-sk}}, \mathbf{W} \sim \mathrm{N}\left(\frac{\rho_S \sum_{j=1}^{S} \omega_{sjk} \phi_{jk}}{\rho_S \sum_{j=1}^{S} \omega_{sjk} + 1 - \rho_S}, \frac{\tau_S^2}{\rho_S \sum_{j=1}^{S} \omega_{sjk} + 1 - \rho_S}\right) \tag{11}$$

$$\boldsymbol{\delta_{tk}} \,|\, \boldsymbol{\delta_{-tk}}, \mathbf{D} \sim \mathrm{N}\left(\frac{\rho_T \sum_{j=1}^{N} d_{tjk} \delta_{jk}}{\rho_T \sum_{j=1}^{N} \omega_{tjk} + 1 - \rho_T}, \frac{\tau_T^2}{\rho_T \sum_{j=1}^{N} d_{tjk} + 1 - \rho_T}\right) \tag{12}$$

$$\gamma_{stk} \sim \mathrm{N}(0, \tau_I^2) \tag{13}$$

$$\tau_S^2, \tau_T^2, \tau_I^2 \sim \mathrm{Inverse} - \mathrm{Gamma}(a, b) \tag{14}$$

$$\rho_S, \rho_T \sim \mathrm{Uniform}(0, 1) \tag{15}$$

where $\boldsymbol{\phi_{sk}} = (\phi_{1k}, \ldots, \phi_{Sk})$ and $\boldsymbol{\delta_{tk}} = (\delta_{1k}, \ldots, \delta_{Nk})$ are common sets of spatial random effects and temporal random effects respectively, and both are modeled by the CAR prior [36]. $\boldsymbol{\gamma_{stk}}$ is the independent space–time interaction of injury severity k for the area s in the tth year, which is assumed to follow a zero-mean multivariate normal distribution. $\tau_S^2, \tau_T^2, \tau_I^2$ are spatial random effects variance, temporal random effects variance, and space–time interaction variance, respectively. Fixed uniform (ρ_S, ρ_T) and conjugate $\tau_S^2, \tau_T^2, \tau_I^2$ priors are specified for the remaining parameters, and the specifications for the latter are set as $a = 1$ and$b = 0.01$.

3.4. ST-CARadaptive model

The ST-CARadaptive model is the extension of the ST-CARar model to allow for spatially adaptive smoothing [39]. It considers localized spatial autocorrelation by allowing spatially neighboring random effects to be correlated or conditionally independent. It is achieved by modeling the non-zero elements of the neighborhood matrix $\mathbf{W}$ as unknown parameters rather than assuming they are fixed at 1. These adjacency parameters are collectively denoted by $\boldsymbol{w}^+ = \omega_{sj} \big| k \sim j$, where $k \sim j$ means (ϕ_{st}, ϕ_{jt}) are conditionally independent for all years given the remaining random effects. These adjacency parameters in $\boldsymbol{w}^+$ are modeled on the unit interval by assuming a multivariate Gaussian prior distribution on the transformed scale $\boldsymbol{v}^+ = \log(\boldsymbol{w}^+, (1 - \boldsymbol{w}^+))$. This prior is a shrinkage model with a constant mean μ and a diagonal variance matrix with variance parameters ζ^2, and is given by:

$$\mathrm{f}(\boldsymbol{v}^+ \big| \tau_\omega^2, \mu) \propto \exp[-\frac{1}{2\tau_\omega^2}(\sum_{v_{is} \epsilon v^+} (v_{is} - \mu)^2)] \tag{16}$$

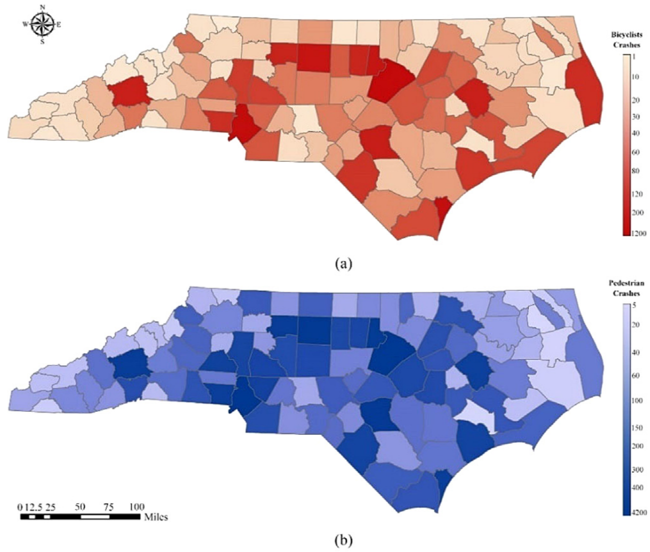

Fig. 1. Spatial distribution of the total number of the bicycle (a) and pedestrian (b) crashes in North Carolina from 2009 to 2018.

$$\tau_{\omega}^{2} \sim \text{Inverse} - \text{Gamma}(a, b) \tag{17}$$

The prior distribution for $\boldsymbol{v}^{+}$ assumes that the degree of smoothing between pairs of adjacent random effects is not spatially dependent [39]. The elements of $\boldsymbol{v}^{+}$ are shrunk to μ under small values of τ_{ω}^{2}. This paper employs $\mu = 15$ because it avoids numerical issues when transforming between $\boldsymbol{v}^{+}$ and $\boldsymbol{w}^{+}$, and implies a prior preference for values of ω_{sj} close to 1.

3.5. Model comparison and checking

Deviance information criteria (DIC) is usually used for evaluating the goodness of fitting for Bayesian hierarchical models in previous literature, and it is defined as [40]:

$$\text{DIC} = \text{D}\left(\overline{\theta}\right) + 2p\text{D} = \overline{D} + p\text{D} \tag{18}$$

where $\overline{\theta}$ is the posterior mean of the parameters; $\text{D}\left(\overline{\theta}\right)$ is the deviance at the posterior mean of the parameters; $p\text{D}$ is the effective number of the model; and $\overline{D}$ is the mean of the sampled deviances from Markov Chain Monte Carlo (MCMC) simulations. Bayesian models with smaller DIC values are preferred. Also, posterior predictive distribution can evaluate the quality of model fitness [41]. The posterior predictive p-value is defined as the following based on the posterior predictive distribution.

$$p = \text{P}[\text{D}(y^{rep}) \geq \text{D(y)}|\,\text{y}] \tag{19}$$

where p represents the probability that the observed measure $\text{D}(y; \theta)$ is more extreme than the replicated discrepancy measure $\text{D}(y^{rep}; \theta)$. p-values around 0.5 indicate that the distribution of the replicated and observed data are close, which means the good fitting of Bayesian models.

4. Data preparation

Bicycle-vehicles and pedestrian-vehicles crashes from counties in North Carolina are used for the analysis. During 2009–2018, 9247 bicycle-involved crashes and 28,787 pedestrian-involved crashes occurred in 100 counties in North Carolina. Bicycle–vehicle crashes and pedestrian–vehicle crashes take the proportion of 24.48% and 75.52% of the total 32,917 crashes, respectively. Fig. 1 presents the spatial distribution of the selected bicycle and pedestrian crashes.

Crash data is collected from the North Carolina Department of Transportation (NCDOT) and the injury severity is recorded as the format of KABCO. Due to the sparseness of injury severity, three severity levels are considered, namely KA crashes, B crashes, and CO crashes. There are 618 KA crashes, 3288 B crashes, and 4151 CO crashes in the selected bicycle crash dataset. There are 3,456 KA crashes, 8917 B crashes, and 12,487 CO crashes in the selected pedestrian crash dataset. The crash risk of bicycles and pedestrians in different crash severities in North Carolina from 2009 to 2018 is shown in Fig. 2. The dependent variables for this study are the crash risk of bicycle–vehicle and pedestrian–vehicle for three injury severities levels.

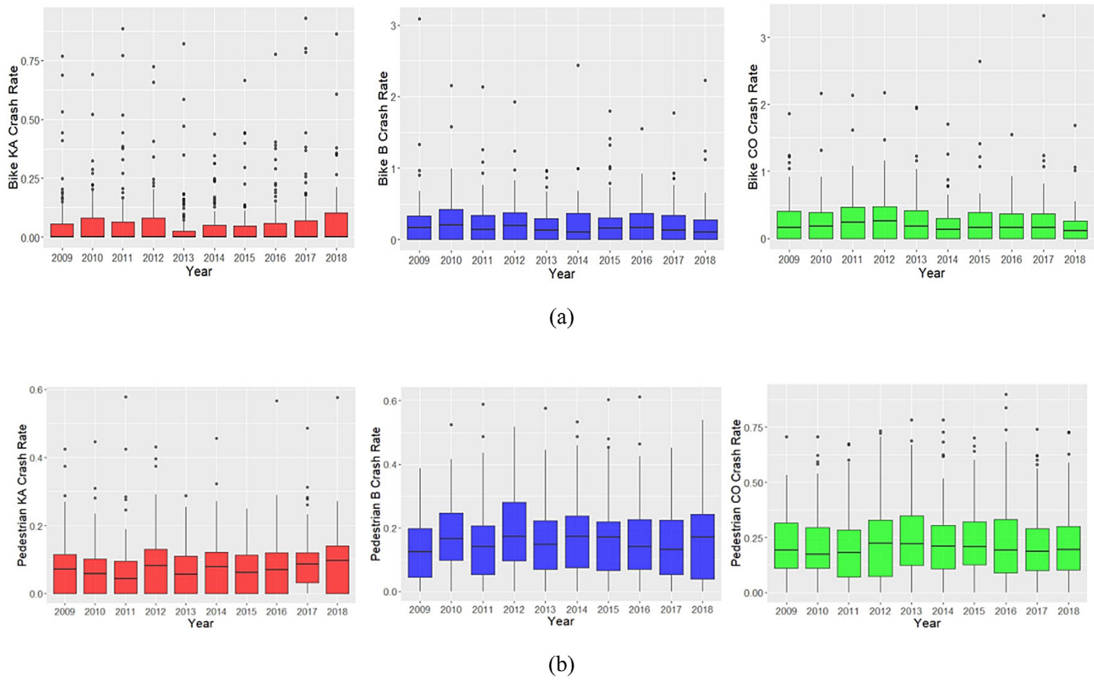

Fig. 2. Crash risk of different severities (KA, B and CO) in bicycle and pedestrian crashes per year, (a) crash risk of bicycle crashes, (b) crash risk of pedestrian crashes.

Table 2
Summary statistics of explanatory variables.

Variables	Description	Mean	StandardDeviation	Maximum	Minimum
Exposure					
ExposureBike	Annual cycling trips per county (*100,000)	0.032	0.003	0.397	0.001
ExposurePed	Annual walking trips per county (*100,000)	0.106	0.027	1.193	0.004
Socio-demography					
PopuDens	Population density per county (number of people per mile^2)	202.542	278.607	2077.640	8.480
MaleRate	Ratios of male people per county	0.491	0.017	0.556	0.464
FemaleRate	Ratios of females per county	0.509	0.017	0.344	0.086
YoungRate	Ratios of people aged between 15 and 24 per county	0.129	0.032	0.344	0.086
OldRate	Ratios of people aged 65 or over per county	0.175	0.045	0.313	0.073
Unemployment	Unemployment rates per county	0.082	0.032	0.181	0.030
Median-Income	Median household income per county (dollars)	44.314	9.072	81.123	27.622
Traffic and roadways					
RoadDensity	Road density per county (mile per mile^2)	5.764	0.668	7.423	4.216
FreewayRate	Proportions of length of freeways	0.068	0.048	0.185	0.010
ArterialRate	Proportions of length of arterials	0.101	0.058	0.252	0.009
CollectorRate	Proportions of length of collectors	0.188	0.078	0.352	0.001
LocalRate	Proportions of length of local roads	0.644	0.117	0.957	0.361
Weather					
Hightemp	Annual average high temperatures per county (°F)	70.662	3.069	79.275	58.617
Lowtemp	Annual average low temperatures per county (°F)	48.734	3.751	56.725	37.300
Precipitation	The total annual amount of precipitation per county (inch)	51.126	12.558	112.140	0
Snowfall	The total annual amount of snowfall per county (inch)	3.805	7.739	94.900	0

According to 2009 National Household Travel Survey, the national walking and bike mode share are 2.8% and 0.8%, respectively [42]. According to 2017 National Household Travel Survey, the national walking and bike mode share are 3.3% and 1.0% [43]. In this study, we use the product of populations, daily trips per person, walking or bike mode share, and time period (365 days) as the exposure for pedestrian or bicycle crashes in each county per year. A host of explanatory variables is considered for the analysis and grouped into three categories: socio-demographic characteristics (i.e., population density, male ratios, female ratios, young people ratios, old people ratios, unemployment rates, median household incomes), roadway structure information (i.e., roadway density, the proportion of freeways/arterial roads/collectors/local roads length), and weather features (i.e., average high temperature, average low temperature, precipitation, and snowfall). Socio-demographic data for each county per year is obtained from the U.S. Census Bureau and the U.S. Bureau of Labor Statistics. The length of total roadways, freeways, arterial roads, collectors, and local roads is obtained from the Roadway Characteristics Inventory (RCI) of NCDOT. To investigate the influence of annual weather changes on bicycle and pedestrian crashes, an array of weather-related explanatory variables in each county are obtained from US Climate Data, which recorded historical daily, monthly and annual weather data since 2009. A statistical summary of explanatory variables is presented in Table 2.

Table 3
Model comparisons of three Bayesian space–time models for cyclist and pedestrian crashes.

Models	Injury severity	$\overline{D}$	$D(\overline{\theta})$	pD	DIC
Bicycles crashes					
ST-CARar	KA	1608.1	1540.9	67.3	1675.4
	B	2851.6	2667.8	183.8	3035.5
	CO	2891.1	2690.4	200.7	3091.9
ST-CARanova	KA	1626.5	1577.5	49.0	1675.5
	B	**2836.7**	**2668.0**	**168.7**	**3005.5**
	CO	**2878.6**	**2713.2**	**165.4**	**3044.1**
ST-CARadaptive	**KA**	**1610.0**	**1544.8**	**65.2**	**1675.2**
	B	2850.6	2666.0	184.6	3035.2
	CO	2890.6	2690.0	200.7	3091.3
Pedestrian crashes					
ST-CARar	**KA**	**3192.1**	**3097.6**	**94.5**	**3286.6**
	B	3885.9	3718.7	167.2	4053.1
	CO	4177.7	3910.1	267.6	4445.2
ST-CARanova	KA	3126.0	3126.0	82.6	3291.6
	B	**3883.9**	**3726.1**	**157.9**	**4041.8**
	CO	**4177.6**	**3929.7**	**247.9**	**4425.6**
ST-CARadaptive	KA	3192.5	3095.6	96.9	3289.5
	B	3886.0	3719.5	166.5	4052.5
	CO	4180.3	3913.3	267.1	4447.4

Note: bold numbers mean that the lowest DIC among three Bayesian space–time models at the same injury severity datasets.

Before incorporating variables into Bayesian space–time models, the preliminary work is to examine spatial and temporal correlations in crash data [13,15]. Moran's I statistic is frequently utilized to test spatial correlations in the traffic safety field [6,44,45]. With statistical significance p-value, (i) if Moran's I is positive and close to 1, it represents the incremental spatial autocorrelation; (ii) if Moran's I is equal to 0, it indicates a random pattern; and (iii) if Moran's I is less than 0, it represents a dispersed pattern [46]. The global Moran's I statistics of different injury severities of bicycle and pedestrian crash risk in each year from 2009 to 2018 are calculated using the toolbox of ArcGIS with inverse-distance. Results show that the existence of some spatial autocorrelations of crash risk in every injury severity [14,15].

All candidate variables are regarded as model inputs at first, then several variables are selected by the convergence results of each model. Meanwhile, variance inflation factors (VIFs) and correlation tests are used to diagnose multi-collinearity in each model. Seven explanatory variables are remained at last, including female ratios, young people ratios, unemployment rates, median household income, roadway density, average high temperatures, and snowfall.

5. Results and discussion

Model fitting is performed by using three separate MCMC chains in package 'CARBayesST' in the R platform [47]. The potential scale reduction factor (PSRF) is used to assess the convergence of multiple chains [48]. Convergence is assumed to occur when PSRF is less than 1.2. Three simulation chains were run with 25,000 iterations for each chain, and 50,000 samples are discarded as burn-ins. The remaining 25,000 samples are retained to obtain the posterior distributions of parameters with a thinning interval of 5. Three Bayesian space–time models, ST-CARar model, ST-CARanova model, and ST-CARadaptive model, are estimated and compared for three injury severities of bicycle and pedestrian crashes to explore the long-term spatiotemporal effects in crash risk.

5.1. Model comparisons

The goodness-of-fit of candidate models is judged based on the posterior mean of deviance ($\overline{D}$), deviance evaluated at the posterior mean ($D(\overline{\theta})$), the effective number of parameters (pD), and deviance information criterion (DIC). Besides, the posterior predictive performance is also tested, and the results show that p-values of all the estimated Bayesian space–time models are around 0.5. PSRFs of all space–time models are less than 1.2, which indicates the convergence of each chain in Bayesian models. Model comparison results are presented in Table 3.

A comparison between three candidate space–time models highlights the importance of including spatial main effects, temporal main effects, and spatiotemporal interaction term in the modeling (the DIC value of the ST-CARanova model is lowest for B and CO severities of bicycle crashes among three different space–time models). It is shown in Table 3 that the DIC values do not show big differences among three Bayesian models for KA severity of bicycles crashes. A possible reason for this is that a large number of zeros (more than 60% of crashes risk is zero) in the KA injury severity of cyclist crashes displayed similar spatiotemporal effects, and separable or inseparable spatial or temporal effects seem to be almost the same. However, Liu and Sharma [13] found that fatal crash frequencies of vehicle–vehicle exhibited some linear trend.

The benefit of introducing the independent spatiotemporal effects is revealed by the significant lowest DIC values of ST-CARanova models for B and CO injury severities of pedestrian crashes. This finding indicates that there are significant overall spatial effects, temporal trends, and an independent space–time interaction in the pedestrian crash risk during ten years, which need to be incorporated in the modeling of minor and no injury severity levels. Additionally, both spatial and temporal effects play important roles in unobserved heterogeneity cross injury severities and thus need to be considered in the Bayesian modeling. These conclusions are supported by previous literature [6,13,14]. Although the ST-CARadaptive model is the improved version of the ST-CARar model by adding a localized spatially adaptive smoothing structure, the

Table 4a
ST-CARanova model for different severities of bicycle crashes in North Carolina from 2009 to 2018.

Variables	KA		B		CO	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
Intercept	**−7.10**	**(−13.57, −0.77)**	**−8.61**	**(−15.03, −2.02)**	**−8.45**	**(−15.12, −1.37)**
Ratios of females	**11.49**	**(0.59, 22.71)**	**12.15**	**(1.91, 21.74)**	**13.06**	**(2.31, 23.14)**
Ratios of people aged between 15 and 24	**3.90**	**(0.09, 7.62)**	**5.47**	**(1.80, 8.83)**	**5.00**	**(0.71, 9.25)**
Unemployment rates	2.19	(−1.97, 6.44)	**5.15**	**(1.35, 8.58)**	**3.96**	**(−0.04, 7.41)**
Median household income	0.01	(0.00, 0.003)	**0.02**	**(0.01, 0.04)**	0.02	(0.00, 0.04)
Road density	**−0.41**	**(−0.61, −0.21)**	−0.16	(−0.44, 0.09)	−0.06	(−0.42, 0.04)
Annual average high temperature	**0.07**	**(0.03, 0.12)**	**0.08**	**(0.04, 0.12)**	**0.07**	**(0.02, 0.11)**
τ_S^2	**0.19**	**(0.08, 0.44)**	**0.35**	**(0.21, 0.69)**	**0.55**	**(0.34, 1.05)**
τ_T^2	**0.01**	**(0.00, 0.03)**	**0.01**	**(0.00, 0.05)**	**0.01**	**(0.00, 0.04)**
τ_I^2	**0.01**	**(0.00, 0.07)**	**0.04**	**(0.02, 0.08)**	**0.03**	**(0.02, 0.05)**
ρ_S	**0.14**	**(0.01, 0.92)**	**0.05**	**(0.00, 0.28)**	**0.06**	**(0.00, 0.30)**
ρ_T	**0.41**	**(0.02, 0.92)**	**0.22**	**(0.01, 0.81)**	**0.49**	**(0.03, 0.94)**

Note: bold numbers mean that the estimated coefficients are significant.

Table 4b
ST-CARadaptive model for different severities of bicycle crashes in North Carolina from 2009 to 2018.

Variables	KA		B		CO	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
Intercept	**−8.33**	**(−14.29, −2.76)**	**−8.21**	**(−12.34, −4.01)**	−11.92	(−16.33, −7.36)
Ratios of females	**12.84**	**(3.06, 22.61)**	**13.59**	**(5.69, 21.24)**	**19.36**	**(11.16, 27.31)**
Ratios of people aged between 15–24	**3.75**	**(0.34, 7.16)**	**5.92**	**(3.14, 8.39)**	**7.56**	**(4.83, 10.41)**
Unemployment rates	2.71	(−0.94, 6.40)	**6.08**	**(3.55, 8.50)**	**6.83**	**(4.02, 9.62)**
Median household income	**0.01**	**(0.00, 0.02)**	**0.03**	**(0.02, 0.04)**	**0.02**	**(0.01, 0.03)**
Road density	**−0.40**	**(−0.59, −0.21)**	−0.14	(−0.29, 0.02)	−0.07	(−0.23, 0.08)
Annual average high temperature	**0.08**	**(0.04, 0.13)**	**0.05**	**(0.03, 0.08)**	**0.06**	**(0.03, 0.09)**
τ^2	**0.07**	**(0.02, 0.19)**	**0.09**	**(0.06, 0.15)**	**0.11**	**(0.07, 0.17)**
ρ_S	**0.13**	**(0.00, 0.57)**	**0.07**	**(0.00, 0.32)**	**0.12**	**(0.01, 0.44)**
ρ_T	**0.89**	**(0.67, 0.99)**	**0.94**	**(0.87, 0.99)**	**0.96**	**(0.90, 1.00)**
τ_ω^2	**134.41**	**(99.25, 178.81)**	**135.21**	**(97.81, 179.47)**	**135.83**	**(99.47, 180.80)**

Note: bold numbers mean that the estimated coefficients are significant.

Table 5a
ST-CARanova model for different severities of pedestrian crashes in North Carolina from 2009 to 2018.

Variables	KA		B		CO	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
Intercept	−3.21	(−6.42, 0.03)	**−5.98**	**(−9.93, −2.28)**	−3.41	(−8.45, 1.26)
Ratios of females	**6.25**	**(1.32, 11.34)**	**11.01**	**(4.97, 17.23)**	7.00	(−0.08, 14.48)
Ratios of people aged between 15–24	1.40	(−0.5, 3.28)	**3.28**	**(0.97, 5.43)**	2.97	(−0.18, 5.67)
Median household income	**−0.01**	**(−0.01, 0.00)**	0.00	(−0.01, 0.01)	0.00	(−0.01, 0.01)
Road density	0.01	(−0.1, 0.11)	**0.15**	**(0, 0.33)**	**0.31**	**(0.09, 0.52)**
Annual average high temperature	**0.05**	**(0.02, 0.08)**	**0.04**	**(0.02, 0.07)**	**0.03**	**(0, 0.06)**
Annual total amount of snowfall	0.00	(−0.01, 0.01)	**0.01**	**(0, 0.01)**	0.00	(−0.01, 0.01)
τ_S^2	**0.06**	**(0.02, 0.16)**	**0.12**	**(0.07, 0.22)**	**0.24**	**(0.15, 0.46)**
τ_T^2	**0.02**	**(0.01, 0.05)**	**0.01**	**(0, 0.02)**	**0.01**	**(0, 0.02)**
τ_I^2	**0.01**	**(0, 0.03)**	**0.01**	**(0.01, 0.03)**	**0.03**	**(0.02, 0.04)**
ρ_S	**0.17**	**(0.01, 0.69)**	**0.05**	**(0, 0.27)**	**0.07**	**(0, 0.33)**
ρ_T	**0.50**	**(0.03, 0.94)**	**0.30**	**(0.01, 0.85)**	**0.37**	**(0.02, 0.89)**

Note: bold numbers mean that the estimated coefficients are significant.

fitting performance of these two models shows great proximity in terms of four evaluation indexes. This implies that there is no significant localized spatial autocorrelation that need to be emphasized in the crash risk for different injury severity levels in bicycle and pedestrian crashes.

5.2. Long-term spatiotemporal random effects

Three Bayesian space–time models are utilized to explore the long-term spatiotemporal random effects of bicycle and pedestrian crash datasets by different serveries in the present study. The ST-CARanova and ST-CARadaptive model results of bicycle crashes are shown in Table 4a and Table 4b. The ST-CARanova and ST-CARar model results of pedestrian crashes are shown in Tables 5a and 5b.

For the ST-CARanova models, the temporal autoregressive parameters for KA, B, and CO crashes of cyclists are 0.41, 0.22, and 0.49, respectively, while these for KA, B, and CO crashes of pedestrian are 0.50, 0.30, and 0.37, respectively. This result indicates that there are relatively high temporal autocorrelation in crash risk of bicycle crashes and pedestrian crashes in the long time period. Besides, small spatiotemporal interaction variances of all severities in bicycle and pedestrian crashes are found. In terms of the DIC values displayed in Table 3, the goodness-of-fit of the ST-CARanova model indicates that the independent spatiotemporal interaction structure and spatial and temporal main effects played an important role in the crash modeling. It is necessary to incorporate them when we estimating Bayesian space–time models of VRU-related crashes.

The spatial autocorrelation parameters for KA, B, and CO crashes of cyclists in the ST-CARadaptive model are 0.07, 0.09, and 0.11, respectively. Those for KA, B, and CO crashes of pedestrians in the ST-CARar model are 0.89, 0.03, and

Table 5b
ST-CARar model for different severities of pedestrian crashes in North Carolina from 2009 to 2018.

Variables	KA		B		CO	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
Intercept	−2.57	(−5.4, 0.23)	**−5.56**	**(−8.16, −2.86)**	**−3.35**	**(−6.63, −0.38)**
Ratios of females	**5.89**	**(1.22, 10.78)**	**11.85**	**(6.89, 16.37)**	**9.53**	**(3.91, 15.59)**
Ratios of people aged between 15–24	1.31	(−0.39, 2.96)	**4.34**	**(2.77, 5.92)**	**3.84**	**(1.81, 5.67)**
Median household income	0.00	(−0.02, 0.00)	0.00	(−0.01, 0.00)	−0.01	(−0.01, 0.00)
Road density	0.01	(−0.08, 0.11)	**0.19**	**(0.08, 0.28)**	**0.27**	**(0.16, 0.39)**
Annual average high temperature	**0.04**	**(0.02, 0.07)**	**0.03**	**(0.01, 0.05)**	**0.01**	**(0.00, 0.03)**
Annual total amount of snowfall	0.00	(−0.01, 0.01)	**0.01**	**(0.00, 0.01)**	0.00	(−0.01, 0.00)
τ^2	**0.05**	**(0.02, 0.09)**	**0.03**	**(0.02, 0.06)**	**0.07**	**(0.05, 0.11)**
ρ_S	**0.89**	**(0.53, 0.98)**	**0.07**	**(0.00, 0.34)**	**0.11**	**(0.01, 0.32)**
ρ_T	**0.92**	**(0.73, 0.99)**	**0.96**	**(0.89, 1.00)**	**0.93**	**(0.87, 0.98)**

Note: bold numbers mean that the estimated coefficients are significant.

0.07, respectively. Unobserved spatial heterogeneity in different counties of bicycle and pedestrian crashes is further confirmed to exist across injury severities as their estimated spatial autocorrelation parameters are very small (close to 0). Temporal autoregressive parameters for three injury severities of both cyclists and pedestrians are much larger than spatial autocorrelation parameters (all close to 1) in the ST-CARaptive and ST-CARar models. This finding indicates that high temporal autocorrelations are found to exist in the bicycle and pedestrian crash risk of different injury severities at the county-level, which is consistent with other similar researches of Liu and Sharma [13].

Overall, the long-term spatiotemporal effects of bicycle and pedestrian crash risk in different injury severities are different from each other which confirms some heterogeneity of spatial and temporal effects across three crash severities [13,14]. However, both bicycle and pedestrian crash risk have some similar patterns, such as relatively high temporal autocorrelations, significant spatial heterogeneity, and weak spatiotemporal interactions. This important finding provides valuable suggestions for the future work in Bayesian space–time modeling for crash frequency modeling for VRUs in a long time period.

5.3. Model parameter interpretations

5.3.1. Bicycles crashes

The final ST-CARanova model of bicycle–vehicle crashes (Table 4a) retains five, six, and five explanatory variables which significantly different from zero at 95% Bayesian credible interval (BCI) in KA, B, and CO crashes, respectively. These variables are ratios of females, ratios of young people (aged between 15 to 24), median household incomes, unemployment rates, roadway density, and annual average high temperatures. In the comparison of three injury severities in bicycle crashes, it is found that socio-demographic characteristics (i.e., female ratios, young people ratios, and unemployment rates) have the largest impact on the crash risk of bicycles–vehicles crashes, followed by the roadway features (i.e., roadway density) and weather (i.e., annual average high temperature).

Ratios of females and young people show relatively huge effects on three injury severities of bicycle crash risk. The estimated coefficients of female ratios and young people ratios variables are positive and the largest among all the fitted coefficients in the ST-CARanova model. These outcomes imply that high ratios of females in Durham, Edgecombe, Vance counties would result in more bicycle crashes of different severities. This result is interesting and in accord with the study of Carvajal et al. [25]. Ratios of females in counties of North Carolina do not show much fluctuation and they only vary at a small scale from 2009 to 2018, which is similar to the temporal trend of crash risk for cyclists. However, ratios of males are discovered to improve the bicycle crash frequency by some scholars [7,20]. High ratios of young people also bring a great threat to bicycle safety performance which significantly contributed to minor and no injury bicycle crashes. Studies suggested that younger people are more likely to be associated with some unsafe behaviors when they are cycling, such as using cellphones, not wearing helmets, and cycling against the traffic flow [49,50].

Unemployment rates produce significant positive impacts on B and CO crashes but insignificant positive impact on KA crashes of cyclists. This study conducts the marginal effects analysis of the ST-CARanova model for bicycle crashes. The estimated covariate effects are the crash risk for one unit increase in each variable and are obtained through the analysis of the marginal effects. Results showed that the crash risk of one dollar increase in median household income is 1.025, suggesting that such an increase corresponds to a 2.5% additional crash risk of minor injury bicycle crashes. Ding et al. [20] also believed that the increase in median household income by 100% is correlated to the increase in bicycle crashes by 19%. One possible reason could be that cycling popularity will be increased as the median household income increased. Increased road density is found to significantly decrease the bicycle crash risk of KA in this study. One possible reason is that denser roads in the urban regions usually have more traffic, which slow down the speed of bicycles and vehicles. However, previous studies found that, ratios of local streets greatly increase crash frequency rather than the total road density, because cyclists frequently commune in the urban areas where have many local streets [2,16].

Average high temperature shows significant positive effects on three injury severities of cyclists. The results of the marginal effects analysis show that for one unit increase of annual high temperature, an increase in the crash risk of three severities 7.8%%, 7.9%%, and 6.7%% would be expected. This result indicates that annual average high temperature is a very informative covariate of bicycle crashes, as higher temperatures and warm seasons will attract more trips of cycling, which in line with the study of Ding et al. [20].

5.3.2. Pedestrian crashes

Table 5a shows significant explanatory variables for KA, B, and CO injury for pedestrian–vehicle crashes in the ST-CARanova model. The significant variables include ratios of females, ratios of young people, median household income, road density, annual average temperature, and annual average snowfall amount. It is found that socio-demographic characteristics have the biggest effects on the crash risk of pedestrian crashes, which is the same with the results of bicycle–vehicle crash modeling.

In terms of female ratios, it is shown that this variable displayed significant positive effects on the crash risk of KA and B in pedestrian crashes, especially on the crash risk of B crashes. There is no firm evidence shows that the female ratio is closely correlated with the crash frequency of pedestrians in the current studies. Thus, maybe females do not drive as much as male people, and they are likely to be exposed to the traffic environment. Another possible reason is that females are physically more vulnerable than males. Ratios of young people aged between 15 to 25 show significant positive effects on minor injury severity of pedestrian crashes. It has been proven that young people are more likely to be involved in aggressive driving [51]. Besides, median household income is found to be weakly associated with the crash risk of KA crashes of pedestrians in the present study. It is possible that people from lower-income families may walk or use public transportation frequently and have more walking exposure [18].

Snowfall do not show significant effects on KA and CO injury severities of pedestrians. Liu and Sharma [13] thought that adverse weather probably resulted in more crashes in the short term but may also reduce people's traveling, leading to lower crash numbers. However, the annual high temperature is strongly associated with the increased crash risk of three injury severities of pedestrian crashes. After conducting the analysis of the marginal effects for the ST-CARanova model in pedestrian crashes, an increase in the high temperature of 1 Fahrenheit degree associates with a 5.1%, 4.3%, and 2.9% increase in the crash risk of KA, B, and CO. It was also found that pedestrians tended to have a higher likelihood of KA crashes when the air temperature is higher than 86 °F [52]. A reasonable explanation for this is that people like to walk on the outside at a warm weather. However, a very high temperature (over 90 °F) is found to decrease the pedestrian trips according to the study of Lee et al. [53].

To draw a conclusion, socio-demographic indicators have significant impacts on the crash risk of different injury severities levels for cyclists and pedestrians in the long term. When we make traffic safety planning or improvement countermeasures in the county-level, the demographic data (e.g., sex ratios, different age gap ratios, median household income) should be taken full consideration of as they have non-negligible effects on the overall safety performance of VRUs. In contrast with the modeling results of pedestrian crashes, bicycle crashes have a higher possibility of being easily influenced by the changes of socio-demographic features according to the results of the marginal effects, which need special attention from the transportation authorities in U.S.

6. Conclusions

Focusing on the bicycle crashes and pedestrian crash risk in three injury severity levels in the counties of North Carolina, three multivariate Bayesian space–time models with different spatiotemporal effects are developed. This study examines the effects of socio-demographic features (e.g., ratios of female, ratios of young people, and unemployment rates), roadway characteristics (e.g., roadway density), and weather conditions (e.g., snowfall and average high temperature). Then, spatiotemporal interactions of the bicycle and pedestrian crash risk by different injury severities in the ten years are analyzed.

Results of Bayesian space–time models indicate that the ST-CARanova model outperforms the ST-CARar model and the ST-CARadaptive model in most cases of bicycle and pedestrian crashes. ST-CARanova models are recommended in the modeling of VRU-related crashes because they can take account of the overall trends of space, time, and spatiotemporal interaction i. Meanwhile, the fitting performance of the ST-CARar model and the ST-CARadaptive model is almost the same in terms of evaluation indexes. Furthermore, the long-term spatiotemporal effects in the bicycle and pedestrian crash risk show similar patterns across three injury severities.

A key contribution of this study is the finding of the temporal correlation, spatial heterogeneity, and spatiotemporal interaction in the crash risk of cyclists and pedestrians in the long-term, which should be addressed in future studies of VRUs. Another contribution of the paper is the inclusion of yearly changed socio-demographic, roadway structure, and weather variables in the estimated models. Socio-demographic variables like female ratios, young people (aged 15–24) ratios, and weather variables like annual average high temperatures are found to have a significant and positive correlation with the crash risk of bicycles and pedestrians. We also find that the effects of contributory factors of both bicycle and pedestrian crashes on the crash risk varied between different injury severity levels. Given the context of the Highway Safety Plan of North Carolina 2020, we hope that our findings would help planners and policy makers make more informed decisions towards VRUs to reduces related crashes, such as pay more attention to counties with relatively high ratios of females and young people, provide appropriate education and enforcement measures to these people, and conduct the Road Awareness Program for vehicle drivers.

In future researches, it would be worthy to analyze the spatiotemporal patterns of smaller space and time scales for VRUs-involved crashes rather than county-level. Besides, there is a need to include more explanatory variables at the macroscopic modeling of VRUs-involved crashes, such as traffic infrastructure, commuting features and cycling and walking facilities. Due to the limitation of collecting the yearly county-level exposure data for cyclists and pedestrians, the estimated exposure of walking and cycling may be inaccurate when compared with the actual exposure, and we suggest that field observation of vehicle–pedestrian and Vehicle–bicycle exposure is needed to conduct in order to adjust the estimated exposure data in the future. .

CRediT authorship contribution statement

Peijie Wu: conceptualization, Methodology, software, Writing - original draft. **Xianghai Meng:** Investigation, Supervision. **Li Song:** Software, Methodology, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Amoh-Gyimah, M. Saberi, M. Sarvi, Macroscopic modeling of pedestrian and bicycle crashes: a cross-comparison of estimation methods, Accid. Anal. Prev. 93 (2016) 147–159.
[2] Q. Cai, J. Lee, N. Eluru, M. Abdel-Aty, Macro-level pedestrian and bicycle crash analysis: incorporating spatial spillover effects in dual state count models, Accid. Anal. Prev. 93 (2016) 14–22.
[3] Q. Cai, M. Abdel-Aty, J. Lee, Macro-level vulnerable road users crash analysis: a Bayesian joint modeling approach of frequency and proportion, Accid. Anal. Prev. 107 (2017) 11–19.
[4] National Highway Traffic Safety Administration (NHSTA), Highway safety plan 2020 north carolina, 2020a, https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/nc_fy20_hsp.pdf, (accessed 2020).
[5] ATC (Australian Transport Council), National road safety strategy 2011–2020 Australia. Austroads, 2011.
[6] C. Siddiqui, M. Abdel-Aty, K. Choi, Macroscopic spatial analysis of pedestrian and bicycle crashes, Accid. Anal. Prev. 45 (3) (2012) 382–391.
[7] D. Saha, P. Alluri, A. Gan, W. Wu, Spatial analysis of macro-level bicycle crashes using the class of conditional autoregressive models, Accid. Anal. Prev. 118 (2018) 166–177.
[8] C. Lizarazo, V. Valencia, Macroscopic spatial analysis of pedestrian crashes in Medellin, Colombia, Transp. Res. Rec.: J. Transp. Res. Board (2018) http://dx.doi.org/10.1177/0361198118758639.
[9] F.L. Mannering, C.R. Bhat, Analytic methods in accident research: methodological frontier and future directions, Anal. Methods Accid. Res. 1 (2014) 1–22.
[10] F.L. Mannering, V. Shankar, C.R. Bhat, Unobserved heterogeneity and the statistical analysis of highway accident data, Anal. Methods Accid. Res. 11 (2016) 1–16.
[11] D. Lord, S.P. Washington, J.N. Ivan, Poisson, Poisson-Gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory, Accid. Anal. Prev. 37 (1) (2005) 35–46.
[12] V. Eksler, S. Lassarre, Evolution of road risk disparities at small-scale level: example of Belgium, J. Saf. Res. 39 (4) (2008) 417–427.
[13] C. Liu, A. Sharma, Exploring spatio-temporal effects in traffic crash trend analysis, Anal. Methods Accid. Res. 16 (2017) 104–116.
[14] X. Ma, S. Chen, F. Chen, Multivariate space–time modeling of crash frequencies by injury severity levels, Anal. Methods Accid. Res. 15 (2017) 29–40.
[15] C. Liu, A. Sharma, Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity, Anal. Methods Accid. Res. 17 (2018) 14–31.
[16] T. Nashad, S. Yasmin, N. Eluru, J. Lee, M. Abdel-Aty, Joint modeling of pedestrian and bicycle crashes: copula-based approach, Transp. Res. Rec.: J. Transp. Res. Board 2601 (2016) 119–127.
[17] R. Goel, P. Jain, G. Tiwari, Correlates of fatality risk of vulnerable road users in delhi, Accid. Anal. Prev. 111 (2018) 86–93.
[18] J. Lee, M. Abdel-Aty, K. Choi, H. Huang, Multi-level hot zone identification for pedestrian safety, Accid. Anal. Prev. 76 (2015) 64–73.
[19] S. Ukkusuri, S. Hasan, H. Aziz, Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency, Transp. Res. Rec.: J. Transp. Res. Board 2237 (2011) 98–106.
[20] H. Ding, N.N. Sze, H. Li, Y. Guo, Roles of infrastructure and land use in bicycle crash exposure and frequency: A case study using greater London bike sharing data, Accid. Anal. Prev. 144 (2020) 105652.
[21] I. Tasic, R. Elvik, S. Brewer, Exploring the safety in numbers effect for vulnerable road users on a macroscopic scale, Accid. Anal. Prev. 109 (2017) 36–46.
[22] C.R. Bhat, S. Astroza, P.S. Lavieri, A new spatial and flexible multivariate random-coefficients model for the analysis of pedestrian injury counts by severity level, Anal. Methods Accid. Res. 16 (2017) 1–22.
[23] Y. Zhang, J. Bigham, D. Ragland, X. Chen, Investigating the associations between road network structure and non-motorist accidents, J. Transp. Geogr. 42 (2015) 34–47.
[24] M.S. Rahman, M. Abdel-Aty, S. Hasan, Q. Cai, Applying machine learning approaches to analyze the vulnerable road-users' crashes at statewide traffic analysis zones, J. Saf. Res. 70 (2019a) 275–288.
[25] G.A. Carvajal, O.L. Sarmiento, A.L. Medaglia, S. Cabrales, D.A. Rodriguez, D.A. Quistberg, S. Lopez, Bicycle safety in Bogotá: a seven-year analysis of bicycles' collisions and fatalities, Accid. Anal. Prev. 144 (2020) 105596.
[26] M.S. Rahman, M. Abdel-Aty, S. Cai, Applying machine learning approaches to analyze the vulnerable roadusers' crashes at statewide traffic analysis zones, J. Saf. Res. (2019b) http://dx.doi.org/10.1016/j.jsr.2019.04.008.
[27] J. Lee, Abdel-Aty, H. Huang, Q. Cai, Transportation safety planning approach for pedestrians: An integrated framework of modeling walking duration and pedestrian fatalities, J. Transp. Res. Rec. 2673 (4) (2019a) 898–906.
[28] M. Saad, M. Abdel-Aty, J. Lee, Bicycle safety analysis at intersections from crowdsourced data, J. Transp. Res. Record 2673 (4) (2019) 1–14.
[29] P. Jovanis, H.L. Chang, Modeling the relationship of accidents to miles travelled, Transp. Res. Rec.: J. Transp. Res. Board 1068 (1986) 42–51.
[30] V. Shankar, F. Mannering, W. Barfield, Effect of roadway geometrics and environmental factors on rural freeway accident frequencies, Accid. Anal. Prev. 27 (1995) 371–389.
[31] J. Ma, K.M. Kockelman, Damien, A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods, Accid. Anal. Prev. 40 (3) (2008) 964–975.
[32] J.J. Song, M. Ghosh, S. Miaou, B. Mallick, BayesIan multivariate spatial models for roadway traffic crash mapping, J. Multivariate Anal. 97 (1) (2006) 246–273.
[33] Y. Wang, K.M. Kockelman, A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods, Accid. Anal. Prev. 60 (2013) 71–84.
[34] J. Aguero-Valverde, K.F. Wu, E.T. Donnell, A multivariate spatial crash frequency model for identifying sites with promise based on crash types, Accid. Anal. Prev. 87 (2016) 8–16.

[35] A. Rushworth, D. Lee, R. Mitchell, A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London, Spat Spatiotemp. Epidemiol 10 (2014) 29–38.
[36] B.G. Leroux, X. Lei, N. Breslow, Statistical Models in Epidemiology, the Environment, and Clinical Trials, Chapter Estimation of Disease Rates in Small Areas: A New Mixed Model for Spatial Dependence, Springer-Verlag, New York, 2000, pp. 179–191.
[37] J. Besag, J. York, A. Mollié, Bayesian Image restoration, with two applications in spatial statistics, Ann. Inst. Statist. Math. 43 (1) (1991) 1–20.
[38] L. Knorr-Held, BayesIan modeling of inseparable space-time variation in disease risk, Stat. Med. 19 (2000) 2555–2567.
[39] A. Rushworth, D. Lee, C. Sarran, An adaptive spatiotemporal smoothing model for estimating trends and step changes in disease risk, J. R. Stat. Soc. Ser. C. Appl. Stat. 66 (1) (2017).
[40] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, A.V.D. Linde, BayesIan measures of model complexity and fit, J. R. Stat. Soc. Ser. B Stat. Methodol. (2002).
[41] A. Gelman, X.-L. Meng, H. Stern, Posterior predictive assessment of model fitness via realized discrepancies, Statist. Sinica 6 (4) (1996) 733–807.
[42] Federal Highway Administration (FHWA), Summary of travel trends: 2009 national household travel survey, 2011.
[43] Federal Highway Administration (FHWA), Summary of travel trends: 2017 national household travel survey, 2018.
[44] F. Guo, X. Wang, M.A. Abdel-Aty, Modeling signalized intersection safety with corridor-level spatial correlations, Accid. Anal. Prev. 42 (1) (2010) 84–92.
[45] K. Xie, X. Wang, K. Ozbay, H. Yang, Crash frequency modeling for signalized intersections in a high-density urban road network, Anal. Methods Accid. Res. 2 (2014) 39–51.
[46] L. Anselin, Spatial Econometrics: Methods and Models, Springer, Netherlands, 1988.
[47] R. Core Team , R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2016.
[48] S.P. Brooks, A. Gelman, General methods for monitoring convergence of iterative simulations, J. Comput. Graph. Statist. 7 (4) (1998) 434–455.
[49] J. Vanparijs, L.I. Panis, R. Meeusen, B. Geus, Characteristics of bicycle crashes in an adolescent population in flanders (Belgium), Accid. Anal. Prev. 97 (2016) 103–110.
[50] X. Wu, W. Xiao, C. Deng, D.C. Schwebel, G. Hu, Unsafe riding behaviors of shared-bicycle riders in urban china: a retrospective survey, Accid. Anal. Prev. 131 (2019) 1–7.
[51] E. Constantinou, G. Panayiotou, N. Konstantinou, A. Loutsiou-Ladd, A. Kapardis, Risky and aggressive driving in young adults: personality matters, Accid. Anal. Prev. 43 (4) (2011) 1323–1331.
[52] X. Zhai, H. Huang, N.N. Sze, Z. Song, K.K. Hon, Diagnostic analysis of the effects of weather condition on pedestrian injury severity, Accid. Anal. Prev. 122 (2018) 318–324.
[53] J. Lee, M. Abdel-Aty, I. Sha, Evaluation of surrogate measures for pedestrian trips at intersections and crash modeling, Accid. Anal. Prev. 130 (2019b) 91–98.