This is an Accepted Manuscript of an article published by Taylor & Francis in Transportmetrica A: Transport Science on 10 May 2021 (Published online), available at: http://www.tandfonline.com/10.1080/23249935.2021.1922536.

Revisiting spatial correlation in crash injury severity: A Bayesian generalized ordered probit model with Leroux conditional autoregressive prior

Qiang Zeng

[a] School of Civil Engineering and Transportation, South China University of Technology

Guangzhou 510641, P.R. China

[b] Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies

Southeast University Road #2, Nanjing 211189, P.R. China

Email: zengqiang@scut.edu.cn

Qianfang Wang

[a] School of Civil Engineering and Transportation, South China University of Technology

Guangzhou 510641, P.R. China

Email: qianfangwang@163.com

Fangzhou Wang

[a] School of Civil Engineering and Transportation, South China University of Technology

Guangzhou 510641, P.R. China

Email: wangfangzhou2019@163.com

N.N. Sze (Corresponding author)

[c] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University

Hung Hom, Kowloon, Hong Kong

Tel: +852 2766-6062; Email: tony.nn.sze@polyu.edu.hk

Revisiting spatial correlation in crash injury severity: A Bayesian generalized ordered probit model with Leroux conditional autoregressive prior

ABSTRACT

To properly account for the spatial correlation of the crashes that are in close proximity, this study proposes a Bayesian spatial generalized ordered probit (SGOP) model with Leroux conditional autoregressive (CAR) prior for crash severity analysis, using the comprehensive crash data of Kaiyang Freeway in Guangdong Province of China in 2014. The proposed model can accommodate the ordinal nature of injury severity and relax the assumption of monotonic effects of explanatory factors on the crash injury severity. More importantly, strength of spatial correlation is considered in the proposed model. Results indicate that the proposed SGOP model with Leroux CAR prior outperforms the conventional generalized ordered probit (GOP) model and SGOP model with intrinsic CAR prior in terms of model fit and classification accuracy. Also, there is moderate spatial correlation for the crashes that are in close proximity. Results of parameter estimation indicate that factors including vehicle type, horizontal curvature, vertical grade, precipitation, visibility, traffic composition, day of the week, crash type, and response time of emergency medical service all affect the crash injury severity. Additionally, the marginal effects of explanatory variables on the probabilities of slight injury and fatal and severe injury (FSI) crashes are estimated. Findings of this study can advance the understanding on the relationship between crash injury severity and possible factors, and indicate the effective engineering countermeasures that can mitigate the risk of more severe crashes on the freeways.

Keywords: Traffic crash; injury severity; generalized ordered probit model; spatial correlation; Leroux conditional autoregressive prior.

1. INTRODUCTION

Road crash is one of the leading causes of deaths resulting from injuries round the world. More than 1.3 million peoples die on the roads every year. It costs up to 3% of GDP in some developing countries (World Health Organization, 2018). In China, there were more than 240,000 road crashes in 2018. It resulted in about 63,000 deaths, 260,000 personal injuries, and property loss of 1.4 billion CNY (i.e. 0.2 billion USD), according to the *Annual Statistical Report on Roadway Traffic Accidents* published by the Ministry of Public Security of China (2019). Therefore, it is of paramount importance to understand the effects of contributory factors on the crash and injury risk. This can then facilitate the development of effective policy initiatives and engineering countermeasures of government agencies, transport planners and engineers, and vehicle manufacturers that can mitigate the potential road safety hazards and minimize the toll of road death and casualty.

Discrete outcome models based on logit and probit regression approaches have been extensively adopted to examine the relationship between possible contributory factors and crash injury severity. To address the problems including endogeneity, unobserved heterogeneity, and multivariate correlation, advanced methodological approaches like generalized ordered model (Eluru et al., 2008; Xie et al., 2009), bivariate and multivariate models (Dong et al., 2016; Winston et al., 2006), random parameter model (Milton et al., 2008), latent class model (Yasmin et al., 2014), Markov switching model (Malyshkina and Mannering, 2009), and Bayesian hierarchical model (Huang et al., 2011) are proposed. As revealed in two comprehensive reviews on the methodological developments of crash data analytics, it is crucial to accommodate the effect of spatial correlation (also known as "spatial dependence") when modelling the prevalence of crashes that are in close proximity (Savolainen et al., 2011; Mannering and Bhat, 2014). This is sensible because crashes and related attributes are location specific and are often mapped to some spatial units of analysis (e.g. intersections, roadway segments, and areal units, etc.). There are often unmeasured (or 'unobserved') factors that affect the severity levels of crashes that are in close proximity in the same way (Castro et al., 2013).

Spatial correlation does not only exist in crash severity models, but also in crash frequency models (Wen et al., 2019a). Several advanced spatial modeling approaches including simultaneous autoregressive model (Quddus, 2008), conditional autoregressive (CAR) model (Quddus, 2008), and generalized estimating equations (Abdel-Aty and Wang, 2006), have been advocated to account for the effect of spatial correlation in crash frequency models (Ziakopoulos and Yannis, 2020). Results indicate that problem of model misspecification can be resolved, and prediction performance can be improved when spatial correlation among neighboring units are accounted (Aguero-Valverde and Jovanis, 2008). Otherwise, variances of parameters would be underestimated, and some (statistically) insignificant factors can be misinterpreted as 'significant' (Zeng et al., 2019b).

Spatial analyses of crash frequency, whether or not accounting for spatial correlation, are prevalent (Cai et al., 2019; Huang et al., 2016; Soroori et al., 2019; Wen et al., 2019b; Zeng et al., 2020b; Zhai et al., 2019). However, spatial correlation is often ignored in crash severity models (i.e. individual crashes are often assumed to be spatially 'independent'). Nevertheless, Xu et al. (2016) first developed a spatial binary logit model with CAR prior to account for the spatial correlation when modeling the injury severity of pedestrians involving in the crashes at signalized intersections. Later on, Meng et al. (2017) extended the spatial model to account for spatio-temporal correlation when modeling the injury severity of occupants of taxi-related crashes. Furthermore, Prato et al. (2018) proposed a linear spatial binary logit model, with which the spatial correlation was accounted for using a spatial lag structure, to analyze the pedestrian injury severity in Denmark. For multi-level crash outcomes (i.e. more than two), Castro et al. (2013) developed a spatial generalized ordered probit (SGOP) model, with which the spatial dependency was accounted using a spatial lag in the error terms. It was adopted in the truck-related crash severity model in New York (Zou et al., 2017). Recently, Zeng et al. (2019a) proposed a Bayesian spatial generalized ordered logit model with CAR prior to analyze the freeway crash severity in China.

Previous studies indicate that overall model fit can be improved and precision of parameter estimation can be enhanced after accounting for the spatial correlation in the crash severity model. In particular, Bayesian spatial model with CAR prior is more commonly used, considering the computation efficiency, compared with the models based on spatial lag and spatial error structures (Quddus, 2008). Nonetheless, the CAR prior adopted in these studies is intrinsic. It is a conditional specification of the Gaussian Markov random field. Intrinsic CAR prior specification indicates the presence of spatial correlation, regardless of the degree of strength, across the entire study area. This may hinder the identifiability and convergence of Bayesian estimation (Eberly and Carlin, 2000). To this end, alternate CAR prior specifications have been proposed (Cressie 1993; Leroux et al., 1999). In particular, the CAR formulation proposed by Leroux et al. (1999) (known as "Leroux CAR" in the rest of the paper) specifies the joint distribution of independent and spatially structured random effects. It is capable of representing varying degree of spatial correlation (i.e., strong, medium and weak, etc.), compared with the intrinsic CAR prior. Lee (2011) compared the performances among the models using different CAR prior formulations. Results indicate that the Leroux CAR prior is superior. In addition, it is suitable to model the spatial distribution of crash frequency (Xu et al., 2017). Despite that there are possible improvements in modeling efficiency and model fit, it is rare that the Leroux CAR prior is applied to model the spatial correlation in the crash severity models.

Therefore, in this study, we aim to: (1) propose a SGOP model with Leroux CAR prior to model the crash severity, with which the spatial correlation in the injury severities of crashes that are in close proximity are controlled; and (2) assess the performances of the SGOP model with Leroux CAR prior, a SGOP model with intrinsic CAR prior, and a conventional generalized ordered probit (GOP) model, using the comprehensive crash data of a freeway in Guangdong Province of China. In particular, the models are estimated using the Bayesian approach, given its superior performance as compared to the maximum likelihood estimation approach (Xie et al., 2009).

The rest of this paper is structured as follows. Crash data used are described in Section 2. Model formulation and estimation method are specified in Section 3. Results are presented and possible implications are suggested in Section 4. Finally, concluding remarks and recommendations for future research are provided in Section 5.

2. DATA

In this study, comprehensive dataset of the Kaiyang Freeway in Guangdong Province of China in 2014 is used. **Figure 1** illustrates the location of Kaiyang Freeway.

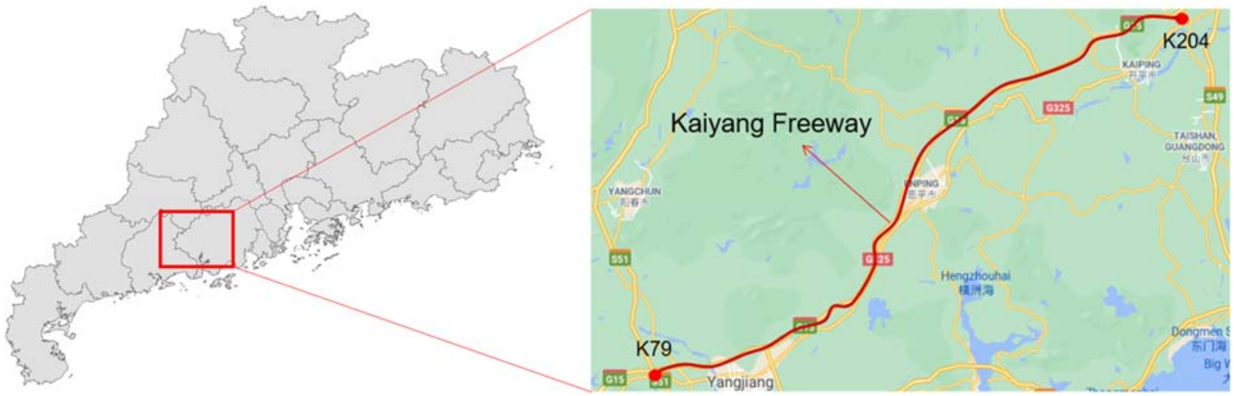

Figure 1 Illustration of the Study Sites

As shown in Figure 1, Kaiyang Freeway is a part of the national level expressway, G15 Shenyang–Haikou Expressway, of China, starting from the kilometer marker K79 and ending at K204. Total length of Kaiyang Freeway is 125 kilometers. The dataset consists of four parts: (i) crash data, (ii) roadway inventory, (iii) traffic flow characteristics, and (iv) weather data.

2.1. Crash data

Crash data are extracted from the Highway Maintenance and Administration Management System, which is administered by the Guangdong Transportation Group. For each crash, injury severity level is defined based on that of the most severely injured person. Crash injury severity is divided into four classes: (i) no injury (i.e., property damage only), (ii) slight injury, (iii) severe injury, and (iv) fatality. In China, fatality refers to the one who dies within 7 days after crash. There were 691 crashes in total at the Kaiyang freeway in 2014. In particular, there were 556 (80.5%) no injury crashes, 95 (13.7%) slight injury crashes, 21 (3.0%) severe injury crashes, and 19 (2.8%) fatal crashes. Consider the modeling deficiency attributed to small sample sizes, severe injury crashes and fatal crashes are combined into one class as “fatal and severe injury (FSI) crashes”.

Other than crash severity, data including response time of emergency medical service (EMS), vehicle class (i.e., passenger car, bus, truck, and other vehicles), license type (i.e. whether the vehicle is registered in the Guangdong Province), crash type (i.e., single vehicle crash, rear-end crash, and angled crash), day of the week (i.e., weekday and weekend), crash time (i.e., before dawn, morning, afternoon, and evening), and crash location (in term of mileage) of every crash are also available in the dataset.

2.2. Roadway inventory

For the road geometry, the freeway profile is obtained from the Guangdong Province Communication Planning and Design Institute. As the characteristics including number of lanes, lane width, pavement type, central median, shoulder type and posted speed limit are fixed across the whole freeway sketch in accordance with Design Specification for Highway Alignment (2006), we only include the factors that have considerable variations, e.g., horizontal curvature, vertical grade, presence of bridge, and presence of entrance and exit ramps, etc., among the individual segments in the proposed crash severity models. The freeway under investigation is divided into 154 consecutive segments, each of which have homogeneous horizontal curvature and vertical grade (Ahemed et al., 2011; Wen et al., 2019a). **Figure 2** illustrates the distribution of crashes across different segments.

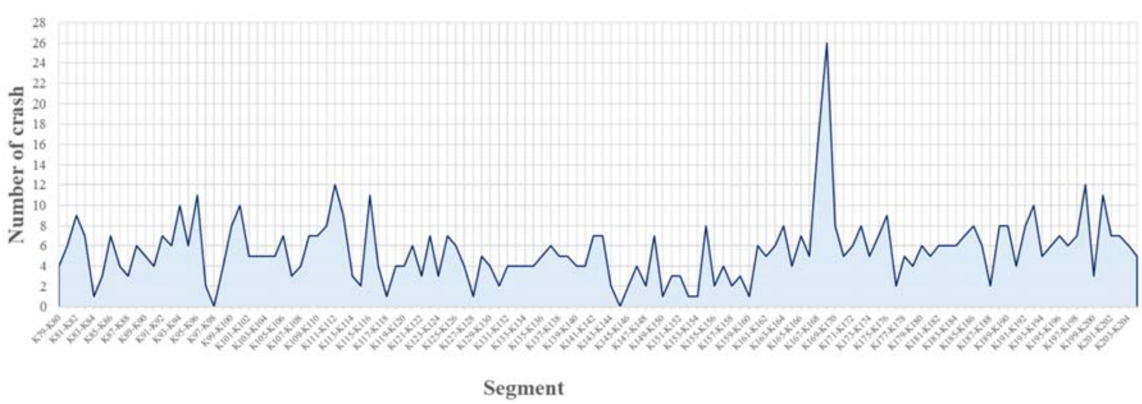

Figure 2 Distribution of crashes across different segments

To indicate whether the crashes are spatially correlated, an indicator that is commonly adopted in geo-informatics - Moran's I is determined as:

$$Moran'I = \frac{N \sum_m \sum_n \omega_{m,n} (C_m - \overline{C})(C_n - \overline{C})}{(\sum_{m \neq n} \omega_{m,n}) \sum_n (C_n - \overline{C})^2} \tag{1}$$

where N is the sample size, C_m and C_n are the numbers of crashes of segment m and segment n, $\overline{C}$ is the overall mean, and $\omega_{m,n}$ is the spatial adjacency weight between m and n.

The spatial adjacency weight is defined using the first-order neighboring structure (Wen et al., 2019b, Xu et al., 2016; Zeng et al., 2019a). Hence, $\omega_{m,n} = 1$ when m and n are neighbors (i.e. sharing a common end) and $\omega_{m,n} = 0$ otherwise. Overall, the value of Moran's I is 0.203 (*Z*-score = 2.605). This implies that the crashes are spatially clustered at the 5% level of significance.

2.3. Traffic flow characteristics

Comprehensive traffic data are obtained from the database of Guangdong Freeway Network Toll System. In accordance to the standard of the Guangdong Transportation Department, there are five vehicle classes: (i) Vehicle Class I - with two axles, two to four wheels, wheelbase less than 3.2 m, and height less than 1.3 m, i.e., passenger cars; (ii) Vehicle Class II - with two axles, four wheels, wheelbase greater than 3.2 m, and height greater than 1.3 m, i.e., light trucks; (3) Vehicle Class III - with two axles, six wheels, wheelbase greater than 3.2 m, and height greater than 1.3 m, i.e., medium trucks and light buses; (4) Vehicle Class IV - with three axles, six to ten wheels, wheelbase greater than 3.2 m, and height greater than 1.3 m, i.e., large trucks and buses; and (5) Vehicle Class V - with more than three axles, more than ten wheels, wheelbase greater than 3.2 m, and height greater than 1.3 m, i.e., super large trucks and trailers.

Traffic flow data are aggregated at daily level. Considering the differences in the velocity and acceleration performances, normalized daily traffic volume is calculated based on the weighted sum of traffic by vehicle classes (with weight equal to 1, 1.5, 2, 3 and 3.5 for Vehicle Class I, II, III, IV and V respectively) as,

$$NDTV_{i,t} = TV\_1_{i,t} + 1.5TV\_2_{i,t} + 2TV\_3_{i,t} + 3TV\_4_{i,t} + 3.5TV\_5_{i,t} \quad (2)$$

where $NDTV_{i,t}$ denotes the normalized daily traffic volume at Segment i on Day t and $TV\_1_{i,t}$, $TV\_2_{i,t}$, $TV\_3_{i,t}$, $TV\_4_{i,t}$, and $TV\_5_{i,t}$ are the observed daily traffic at Segment i on Day t for Vehicle Class I, II, III, IV and V respectively.

To reflect the effect of traffic composition on crash injury severity, proportions of different vehicle classes in the traffic flow mix are also estimated and considered in the proposed analysis as:

$$PVC\_1_{i,t} = \frac{TV\_1_{i,t}}{TV\_1_{i,t}+1.5TV\_2_{i,t}+2TV\_3_{i,t}+3TV\_4_{i,t}+3.5TV\_5_{i,t}} \quad (3)$$

$$PVC\_2_{i,t} = \frac{1.5TV\_2_{i,t}}{TV\_1_{i,t}+1.5TV\_2_{i,t}+2TV\_3_{i,t}+3TV\_4_{i,t}+3.5TV\_5_{i,t}} \quad (4)$$

$$PVC\_3_{i,t} = \frac{2TV\_3_{i,t}}{TV\_1_{i,t}+1.5TV\_2_{i,t}+2TV\_3_{i,t}+3TV\_4_{i,t}+3.5TV\_5_{i,t}} \quad (5)$$

$$PVC\_4_{i,t} = \frac{3TV\_4_{i,t}}{TV\_1_{i,t}+1.5TV\_2_{i,t}+2TV\_3_{i,t}+3TV\_4_{i,t}+3.5TV\_5_{i,t}} \quad (6)$$

$$PVC\_5_{i,t} = \frac{3.5TV\_5_{i,t}}{TV\_1_{i,t}+1.5TV\_2_{i,t}+2TV\_3_{i,t}+3TV\_4_{i,t}+3.5TV\_5_{i,t}} \quad (7)$$

where $PVC\_1_{i,t}$, $PVC\_2_{i,t}$, $PVC\_3_{i,t}$, $PVC\_4_{i,t}$, and $PCV\_5_{i,t}$ are the proportions of Vehicle Class I, II, III, IV and V respectively at Segment i on Day t.

2.4. Weather data

The meteorological observation and measurement data at three county-level weather stations along the Kaiyang Freeway are obtained from the Meteorological Information Management System, which is maintained by the Guangdong Climate Center. For every crash, the weather data is matched in accordance to the distance from the nearby weather station (Naik et al., 2016; Yu and Abdel-Aty, 2014). Information on the wind speed, precipitation, temperature, humidity, and visibility in the hour of crash are available.

Table 1 summarizes the distribution of the sample (i.e. 691 crashes). We have conducted the multi-collinearity test for the candidate variables. In particular, variables that have high variance inflation factor (VIF) values (i.e. % of Vehicle Class V) would not be considered in subsequent analyses.

Table 1. Descriptive statistics of the sample

Scope of work	Variable	Description	Mean	S.D.	Min.	Max.	Proportion
Vehicle involved	Car^	All vehicles involved are cars = 1; else = 0	—	—	—	—	0.57
	Bus	Not less than one bus involved = 1; else = 0	—	—	—	—	0.07
	Truck	Not less than one truck involved = 1; else = 0	—	—	—	—	0.32
	Others	Not less than one other vehicle involved = 1; else = 0	—	—	—	—	0.08
	Non-local vehicle	No less than one non-local (e.g. Guangdong Province) registered vehicle involved = 1; else = 0	—	—	—	—	0.27
Road geometry	Horizontal curvature	Horizontal curvature of crash location (0.1 km^{-1})	1.9	1.2	0.0	4.4	—
	Vertical grade	Vertical grade of crash location (%)	0.8	0.7	0.0	2.9	—
	Bridge	Crash located on a bridge = 1; else = 0	—	—	—	—	0.57
	Ramp	Crash located near the freeway ramp = 1; else = 0	—	—	—	—	0.24
Traffic composition	% of Vehicle Class I^	Percentage of Class I vehicle	42.2	12.2	24.9	79.6	—
	% of Vehicle Class II	Percentage of Class II vehicle	2.5	0.7	1.4	5.7	—
	% of Vehicle Class III	Percentage of Class III vehicle	21.3	3.3	13.5	33.7	—
	% of Vehicle Class IV	Percentage of Class IV vehicle	6.1	2.1	1.5	9.4	—
	% of Vehicle Class V	Percentage of Class V vehicle	27.9	9.7	3.6	43.1	—
Weather condition	Wind speed	Average wind speed in the crash hour (m/s)	2.9	2.0	0.0	16.7	—
	Precipitation	Total precipitation in the crash hour (mm)	0.8	3.6	0.0	54.8	—
	Temperature	Average temperature in the crash hour (℃)	23.2	6.5	4.8	36.8	—
	Humidity	Average humidity in the crash hour (%)	78.6	16.1	21.0	100.0	—
	Visibility	Average visibility in the crash hour (km)	17.6	18.7	0.1	80.0	—
Time of the day	Before dawn^	Crash time is within [12:00 a.m., 5:59 a.m.]= 1; else = 0	—	—	—	—	0.22
	Morning	Crash time is within [6:00 a.m., 11:59 a.m.] = 1; else = 0	—	—	—	—	0.39
	Afternoon	Crash time is within [12:00 p.m., 5:59 p.m.]= 1; else = 0	—	—	—	—	0.21

	Evening	Crash time is within [6:00 p.m., 11:59 p.m.] = 1; else = 0	—	—	—	—	18
Crash type	Single-vehicle crash^	Crash involved one vehicle only = 1; crash involved multiple vehicles = 0	—	—	—	—	44
	Rear-end crash	A rear-end crash = 1; else = 0	—	—	—	—	26
	Angled crash	An angled crash = 1; else = 0	—	—	—	—	30
Day of the week	Weekend	Crash occurred on weekend = 1; crash occurred on weekday = 0	—	—	—	—	33
Traffic flow	Traffic volume	Normalized daily traffic volume at the crash location (10^3 passenger car unit, PCU)	5.66	1.07	2.25	10.17	—
EMS response	EMS response time	Time interval between crash reporting and EMS arrival on the scene (min)	20.7	18.0	2	260	—

^ *Reference category*

3. METHOD

Crash injury severity is often modeled using the ordered response models (e.g. ordered probit/logit model) given its ordinal nature (Savolainen et al., 2011). In this part, the formulations of conventional GOP model and two proposed SGOP models (i.e. SGOP with intrinsic CAR prior and SGOP with Leroux CAR prior) are specified. Also, the formulations of Bayesian estimation process, assessment criteria, and compution of marginal effect would be given.

3.1. Model formulation

3.1.1. Conventional GOP model

In the standard ordered response model, effects of explanatory variables on the outcomes are restricted to be monotonic. In the conventional GOP model, such restriction is relaxed by allowing the thresholds to vary with explanatory variables (Yasmin et al., 2014). Specifically, a latent injury severity propensity z_i is denoted as the base for the rank ordering (i.e. crash injury severity) of ith observation. Propensity is formulated as a linear function of the explanatory variables $\mathbf{X}_i$ as,

$$z_i = \boldsymbol{\beta}\mathbf{X}_i + \varepsilon_i. \tag{8}$$

where $\boldsymbol{\beta}$ is a vector of estimable coefficients (including a constant) associated with the explanatory variables $\mathbf{X}_i$; and ε_i is a residual term following a standard normal distribution.

The latent propensity z_i of ith crash is mapped to the observed injury severity level y_i using the thresholds $\mu_{i,k}(k = 1,2,\cdots,J-1)$ as,

$$y_i = \begin{cases} 1, & z_i \le \mu_{i,1} \\ 2, & \mu_{i,1} < z_i \le \mu_{i,2} \\ \vdots & \vdots \\ j & \mu_{i,j-1} < z_i \le \mu_{i,j} \\ \vdots & \vdots \\ J & z_i > \mu_{i,J-1} \end{cases}, \tag{9}$$

where $j \in \{1, 2, \dots, J\}$ represents the ordinal injury severity outcome in an ascending order (i.e. 1

refers to ‘no injury crash’, 2 refers to ‘slight injury crash’ and 3 refers to ‘FSI crash’ respectively).

To increase the model flexibility, the thresholds are parameterized as follow (Eluru et al., 2008),

$$\mu_{i,k} = \mu_{i,k-1} + \exp\left(\boldsymbol{\alpha}_{k-1}\mathbf{Z}_{i,k-1}\right), \forall k \in \{2, \ldots, J-2\}, \tag{10}$$

where $\mathbf{Z}_{i,k-1}$ is the vector of covariates associated with the kth threshold and $\boldsymbol{\alpha}_{k-1}$ is the vector of corresponding parameters (also include constant).

For the uniqueness of identification, value of $\mu_{i,1}$ is fixed at zero. Therefore, the model generalization process will not be affected (Yasmin et al., 2014). In other word, it is necessary to estimate parameter $\boldsymbol{\alpha}_1$ only when establishing the threshold between slight injury and FSI crashes, $\mu_{i,1}$ in the empirical analysis.

As the residual term ε_i is standard normally distributed, for some crash i, the cumulative probability of having a crash with injury severity level of j or below, $P_{i,j}$, can be calculated as,

$$P_{i,1} = \Phi(-\boldsymbol{\beta}\mathbf{X}_i), \tag{11}$$

$$P_{i,j} = \Phi\left(\mu_{i,j} - \boldsymbol{\beta}\mathbf{X}_i\right) = \Phi\left(\sum_{k=1}^{j-1} \exp\left(\boldsymbol{\alpha}_k\mathbf{Z}_{i,k}\right) - \boldsymbol{\beta}\mathbf{X}_i\right), \forall j \in \{2, \ldots, J-1\}, \tag{12}$$

$$P_{i,J} = 1. \tag{13}$$

In Eqs. (4) and (5), $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution.

Therefore, the probability of crash i of injury severity level j, $p_{i,j}$, is calculated as:

$$p_{i,1} = P_{i,1} = \Phi(-\boldsymbol{\beta}\mathbf{X}_i), \tag{14}$$

$$p_{i,j} = P_{i,j} - P_{i,j-1} = \Phi\left(\mu_{i,j} - \boldsymbol{\beta}\mathbf{X}_i\right) - \Phi\left(\mu_{i,j-1} - \boldsymbol{\beta}\mathbf{X}_i\right), \forall j \in \{2, \ldots, J-1\}, \tag{15}$$

$$p_{i,J} = 1 - P_{i,J-1} = 1 - \Phi\left(\mu_{i,J-1} - \boldsymbol{\beta}\mathbf{X}_i\right). \tag{16}$$

3.1.2. SGOP model with intrinsic CAR prior

In the GOP model framework, spatial correlation of the crashes that are in close proximity can be accounted in the latent propensity formulation, by incorporating a residual term with intrinsic CAR prior (Zeng et al., 2019a). This reflects the effects of common (unobserved) factors in the neighboring spatial units on crash injury severity. Specifically, the latent injury propensity z_i for crash i at road segment m, is formulated as:

$$z_i = \boldsymbol{\beta}\mathbf{X}_i + \phi_m + \varepsilon_i, \tag{17}$$

in which the residual term ϕ_m denotes the spatial effects of crashes at roadway segment m, and is specified using the intrinsic CAR prior advocated by Besag et al. (1991) as,

$$\phi_m \sim N\left(\frac{\sum_{n\neq m}\phi_n\omega_{m,n}}{\sum_{n\neq m}\omega_{m,n}}, \frac{\sigma^2}{\sum_{n\neq m}\omega_{m,n}}\right), \tag{18}$$

where $\sigma(>0)$ denotes the standard deviation parameter of the spatial term.

Therefore, probability for crash i having the outcome of injury severity level j is specified as,

$$p_{i,1} = \Phi(-\boldsymbol{\beta}\mathbf{X}_i - \phi_m), \tag{19}$$

$$p_{i,j} = \Phi(\mu_{i,j} - \boldsymbol{\beta}\mathbf{X}_i - \phi_m) - \Phi(\mu_{i,j-1} - \boldsymbol{\beta}\mathbf{X}_i - \phi_m),\ \forall j \in \{2, \dots, J-1\}, \tag{20}$$

$$p_{i,J} = 1 - \Phi(\mu_{i,J-1} - \boldsymbol{\beta}\mathbf{X}_i - \phi_m). \tag{21}$$

The intrinsic CAR prior specified in Eq. (18) is one of the possible ways to model the spatial correlation. However, strength of correlation as specified by the intrinsic CAR prior is restrictive. For example, the spatial correlation structure remains unchanged even if the values of spatial term ϕ_m and σ_ϕ increase. It implies that the correlation structure as specified by the intrinsic CAR prior is not sensitive enough to the degree of spatial correlation. This may result in estimation bias if the degree of correlation is not considered (Lee, 2011).

3.1.3. SGOP model with Leroux CAR prior

To avoid possible drawbacks of the intrinsic CAR prior, Leroux et al. (1999) proposed a modified specification for the spatial term ϕ_m. It can be easily incorporated into the latent propensity function and formulated as,

$$\phi_m \sim N\left(\frac{\rho \sum_{n\neq m} \phi_n \omega_{m,n}}{1-\rho+\rho \sum_{n\neq m} \omega_{m,n}}, \frac{\sigma^2}{1-\rho+\rho \sum_{n\neq m} \omega_{m,n}}\right), \tag{22}$$

where ρ $(0 \leq \rho \leq 1)$ is the weight parameter that reflects the strength of correlation. $\rho = 0$ means that two crashes are spatially independent. Increase in ρ implies that the spatial correlation is stronger. $\rho = 1$ is a special case of Leroux CAR prior (i.e. intrinsic CAR prior). Therefore, the strength of spatial correlation (i.e. weak, medium and strong) can be accommodated.

3.2. Bayesian estimation and performance assessment

Before obtaining the Bayesian estimates of the parameters of interest, it is necessary to specify their prior distributions. If the prior information is available, it should be adopted to formulate the informative priors (Yu and Abdel-Aty, 2013); otherwise, uninformative prior distributions can be used. Specifically, we have used a diffused normal distribution, $\mathrm{normal}(0, 10^4)$, as the priors of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_1$. A uniform distribution, $\mathrm{uniform}(0.01, 10)$, is specified as the prior of the spatial standard deviation parameter σ. For the weight parameter, ρ, in the model with Leroux CAR prior, a uniform distribution, $\mathrm{uniform}(0, 1)$, is used as the prior.

The Bayesian posterior distributions of the parameters can be achieved using the Gibbs sampling algorithm in WinBUGS (Lunn et al., 2000). For each alternative model, we run a chain of 100,000 Markov chain Monte Carlo (MCMC) simulation iterations, with which the first 50,000 iterations are treated as burn-in to achieve the model convergence. The convergence is assessed by visual inspection using the trace plots for the model parameters, and the ratios between Monte Carlo simulation errors and standard deviations of the estimates (i.e. less than 0.05). For details, readers may refer to the *WinBUGS User Manual* (Spiegelhalter et al., 2005).

Deviance information criterion (DIC) and classification accuracy of every crash outcome and overall are adopted to assess the performances of the proposed models. DIC is deemed as a Bayesian equivalent of Akaike information criterion and Bayesian information criterion. It provides a hybrid measure of model complexity and goodness-of-fit and is defined as (Spiegelhalter et al., 2002),

$$\mathrm{DIC} = \overline{D} + pD, \tag{23}$$

where $\overline{D}$ is the posterior mean deviance that can be used as a fitness measure of the model, and pD is the effective number of parameters that are used to measure the model complexity. Generally, a model with a lower DIC value is preferred. Particularly, difference in DICs greater than 10 implies that model with higher DIC value can be ruled out; and difference between 5 and 10 indicates that improvement in model fit is considerable (Spiegelhalter et al., 2005).

The classification accuracy for crash severity level j is calculated as (Zeng et al., 2019a),

$$\mathrm{CA}_j = \frac{\sum_{\bar{y}_i = y_i = j} y_i}{\sum_{y_i = j} y_i} \times 100\%, \forall j \in \{1,2,\dots,J\}, \tag{24}$$

where $\bar{y}_i$ represents the predicted outcome of crash i.

In a similar vein, the classification accuracy for the whole dataset is computed as,

$$\mathrm{CA} = \frac{\sum_{\bar{y}_i = y_i} y_i / y_i}{\sum_i y_i / y_i} \times 100\%. \tag{25}$$

3.3. Marginal effect

Although the Bayesian estimates of the coefficients in the proposed models can directly indicate whether an explanatory variable may significantly affect the crash injury severity, it does not provide a clear sense of the direction and magnitude of the effect on the propensity of each injury severity level. Therefore, the marginal effects of significant explanatory variables should be computed. Specifically, the marginal effect of a continuous variable x is calculated using the first-order derivative with respect to x (Zeng et al., 2019a):

$$\frac{\partial p_{i,1}}{\partial x} = -\beta^x \varphi(-\boldsymbol{\beta}\mathbf{X}_i - \phi_m), \tag{26}$$

$$\frac{\partial p_{i,2}}{\partial x} = (\mu_{i,1}\alpha^x - \beta^x)\varphi(\mu_{i,1} - \boldsymbol{\beta}\mathbf{X}_i - \phi_m) + \beta^x \varphi(-\boldsymbol{\beta}\mathbf{X}_i - \phi_m), \tag{27}$$

$$\frac{\partial p_{i,3}}{\partial x} = (\beta^x - \mu_{i,1}\alpha^x)\varphi(\mu_{i,1} - \boldsymbol{\beta}\mathbf{X}_i - \phi_m), \tag{28}$$

where $\varphi(\cdot)$ is the probability density function of standard normal distribution, β^x and α^x are the parameters of covariate x in the functions of injury propensity z_i and the threshold between slight injury and FSI crashes $\mu_{i,1}$.

For a dummy variable x, the marginal effect refers to the change in the estimated probability when x changes from zero to one ($\Delta x = 1$):

$$\frac{\Delta p_{i,1}}{\Delta x} = \Phi\left(-\widetilde{\boldsymbol{\beta}}\widetilde{\mathbf{X}}_i - \beta^x - \phi_m\right) - \Phi\left(-\widetilde{\boldsymbol{\beta}}\widetilde{\mathbf{X}}_i - \phi_m\right), \tag{29}$$

$$\frac{\Delta p_{i,2}}{\Delta x} = \Phi\left(\exp\left(\widetilde{\boldsymbol{\alpha}}_1\widetilde{\mathbf{Z}}_{i,1} + \alpha^x\right) - \widetilde{\boldsymbol{\beta}}\widetilde{\mathbf{X}}_i - \beta^x - \phi_m\right) - \Phi\left(\exp\left(\widetilde{\boldsymbol{\alpha}}_1\widetilde{\mathbf{Z}}_{i,1}\right) - \widetilde{\boldsymbol{\beta}}\widetilde{\mathbf{X}}_i - \phi_m\right)$$
$$-\Phi\left(-\widetilde{\boldsymbol{\beta}}\widetilde{\mathbf{X}}_i - \beta^x - \phi_m\right) + \Phi\left(-\widetilde{\boldsymbol{\beta}}\widetilde{\mathbf{X}}_i - \phi_m\right), \tag{30}$$

$$\frac{\Delta p_{i,3}}{\Delta x} = \Phi\left(\exp\left(\widetilde{\boldsymbol{\alpha}}_1\widetilde{\mathbf{Z}}_{i,1}\right) - \widetilde{\boldsymbol{\beta}}\widetilde{\mathbf{X}}_i - \phi_m\right) - \Phi\left(\exp\left(\widetilde{\boldsymbol{\alpha}}_1\widetilde{\mathbf{Z}}_{i,1} + \alpha^x\right) - \widetilde{\boldsymbol{\beta}}\widetilde{\mathbf{X}}_i - \beta^x - \phi_m\right), \tag{31}$$

where $\widetilde{\mathbf{X}}_i$ and $\widetilde{\boldsymbol{\beta}}$ are the vectors $\mathbf{X}_i$ less x and $\boldsymbol{\beta}$ less β^x respectively; and $\widetilde{\mathbf{Z}}_{i,1}$ and $\widetilde{\boldsymbol{\alpha}}_1$ are the vectors $\mathbf{Z}_{i,1}$ less x and $\boldsymbol{\alpha}_1$ less α^x, respectively.

Noticeably, Eqs. (19)-(24) are applicable to the SGOP models only. For the conventional GOP model, the spatial term ϕ_m should be removed. In addition, the marginal effects estimated are specific to particular crash i. For the entire dataset, the average marginal effect of every significant variable would be estimated.

4. RESULTS AND DISCUSSIONS

4.1. Model comparison

Table 2 illustrates the results of model assessment and Bayesian estimates of the hyper-parameters specified to CAR priors, and Table 3 shows the results of parameter estimation of the models with which the factors that have significant effects, at the 10% level, on the latent injury propensity and the threshold between slight injury and FSI are included. Marginal effects of significant factors of the GOP model, SGOP model with intrinsic CAR prior, and SGOP model with Leroux CAR prior are presented in Table 4 to 6, respectively.

As shown in Table 2, values of $\overline{D}$ of the two SGOP models are much lower (differences greater than 10), compared to conventional GOP model. This indicates that the former is superior in term of the improvement of goodness-of-fit when modeling the crash injury severity. This is consistent to that of extant research (Meng et al., 2017; Xu et al., 2016; Zeng et al., 2019a) which suggest that it is necessary to account for the spatial correlation of crashes that are in close proximity. On the other hand, the classification accuracies of SGOP models (with intrinsic CAR prior and Leroux CAR prior) are higher than that of conventional GOP model. Particularly, the classification accuracies of FSI crashes (which have much higher economic and social implications than slight injury and no injury crashes) of SGOP models are 4 times greater than that of GOP model. As also shown in Table 2, the spatial standard deviation parameters σ of SGOP models are statistically significant, both at the 5% level. Significant spatial effects are anticipated. It may be attributable to the unobserved factors (i.e. terrain and road layout) which are spatially clustered and have common effects on the outcome of crashes that are in close proximity (Zeng et al., 2019a). In addition, proposed model with which the spatial correlations are accommodated can improve the accuracy of crash hotspot identification (Huang et al., 2016).

While the spatial correlations are found significant in the two spatial models, as indicated by the Bayesian estimates of σ. Posterior mean of the weight parameter ρ in the SGOP model with Leroux CAR prior is estimated at 0.67. This suggests that the spatial correlation is moderate. In contrast, SGOP model with intrinsic CAR prior is not capable of accommodating the strength of spatial correlation. As mentioned, SGOP model with intrinsic CAR prior can only indicate the presence of (strong) spatial correlation, but SGOP model with Leroux CAR prior can generalize the formulation by stratifying the spatial correlation into weak, medium, and strong (Lee, 2011). Therefore, the possible identifiability and convergence problems of Bayesian inference can be resolved (Eberly and Carlin, 2000). In addition, improvements in model fit (difference in DICs of seven) and overall classification accuracy are considerable. This justifies the suitability of SGOP model with Leroux CAR prior, particularly when the strength of spatial correlation of crashes that are in close proximity can vary.

As also shown in Table 3, results of parameter estimation of SGOP model with Leroux CAR prior are comparable to that of SGOP model with intrinstic CAR prior. However, there are slight derivations of the results of SGOP model with Leroux CAR prior (with greater posterior standard deviations) from that of GOP model, even not all. This is consistent to the findings of previous studies that suggest the precision of parameter estimation could have been overestimated if the spatial correlation was ignored (Quddus, 2008; Zeng et al., 2019a). This again justifies the needs of considering the spatial correlation when modeling the crash severity. Nonetheless, directions of the effects of all parameters are consistent among the proposed models. In addition, as shown in Table 4-6, differences in the marginal effects of significant factors that affect the crash injury severity, across the three models, are considerable. For example, probability of no injury is reduced by 6.7% for the crashes that involve truck as indicated in the SGOP model with Leroux CAR prior, however, the probability reduction of no injury is magnified to 8.0% (magnified by 19.4%) as indicated in the GOP model. On the other hand, probability of no injury is reduced by 11.7% for the angled crashes as indicated in the SGOP model with Leroux CAR prior, however, the probability reduction of no injury is magnified to 13.3% (magnified by 13.7%) as indicated in the SGOP model with intrinsic CAR prior. In general, effects of significant factors on crash injury severity are seemingly overestimated as indicated in the conventional spatial models, as compared to the proposed SGOP model with Leroux CAR prior.

Table 2. Performance assessment of proposal models

Assessment criterion	Model 1: GOP	Model 2: SGOP with intrinsic CAR prior	Model 3: SGOP with Leroux CAR prior
$\overline{D}$	790	735	**733**
pD	**49**	90	85
DIC	839	825	**818**
CA_1	**98.4%**	96.9%	97.5%
CA_2	6.3%	**11.6%**	**11.6%**
CA_3	2.5%	**12.5%**	**12.5%**

CA	80.2%	80.3%	**80.7%**
σ	—	0.44 (0.06, 0.99)[a]	0.49 (0.21, 0.79) [a]
ρ	—	—	0.67 (0.13, 0.99) [a]

[a] Estimated mean and 95% Bayesian credible interval for the hyper-parameters.

Table 3. Results of parameter estimation

	GOP		SGOP with intrinsic CAR prior		SGOP with Leroux CAR prior	
Variable	**Latent injury propensity**	**Threshold between slight injury and FSI**	**Latent injury propensity**	**Threshold between slight injury and FSI**	**Latent injury propensity**	**Threshold between slight injury and FSI**
Constant	-3.35 (0.86)*		-12.2 (0.90)*	—	-3.14 (0.72)*	—
Truck	0.32 (0.14)*	—	0.31 (0.15)*	—	0.30 (0.16)*	—
Others	—	-0.71 (0.43)^	—	-0.76 (0.46)^	—	-0.80 (0.45)^
Non-local vehicle	0.27 (0.14)*	—	0.33 (0.16)*	—	0.33 (0.14)*	—
Horizontal curvature	—	-0.19 (0.10)^	—	-0.20 (0.11)^	—	-0.19 (0.11)^
Vertical grade	—	-0.86 (0.24)*	—	-0.86 (0.25)*	—	-0.89 (0.24)*
% of Vehicle class III	0.05 (0.02)*	—	0.04 (0.02)*	—	0.04 (0.02)*	—
Precipitation	—	0.25 (0.18)*	—	0.26 (0.20)*	—	0.28 (0.21)*
Visibility	0.007 (0.004)^	—	0.007 (0.005)^	—	0.008 (0.004)^	—
EMS response time	0.008 (0.003)*	—	0.01 (0.004)*	—	0.009 (0.004)*	—
Weekend	—	0.60 (0.33)^	—	0.57 (0.30)^	—	0.54 (0.30)^
Rear-end crash	0.75 (0.15)*	—	0.77 (0.17)*	—	0.77 (0.17)*	—
Angled crash	0.46 (0.16)*	—	0.65 (0.19)*	—	0.57 (0.18)*	—

Note: Standard deviation of coefficients are shown in parentheses.

[*] Statistically significant at the 5% level.

^ Marginally significant at the 10% level.

Table 4. Marginal effects of significant factors in GOP model

Scope of work	Variable	No injury	Slight injury	FSI

Vehicle involved	Truck	-8.0%	+4.6%	+3.4%
	Other vehicle	0.0%	-3.7%	+3.7%
	Non-local vehicle	-6.8%	+3.9%	+2.9%
Road geometry	Horizontal curvature	0.0%	-5.9%	+5.9%
	Vertical grade	0.0%	-26.9%	+26.9%
Traffic composition	% of Vehicle Class III	-6.9%	+4.0%	+2.9%
Weather condition	Precipitation	0.0%	+7.8%	-7.8%
	Visibility	-1.0%	+0.6%	+0.4%
EMS response	EMS response time	-1.2%	+0.7%	+0.5%
Day of the week	Weekend	0.0%	+2.9%	-2.9%
Crash type	Rear-end crash	-18.5%	+10.4%	+8.1%
	Angled crash	-10.3%	+6.1%	+4.2%

Table 5. Marginal effects of significant factors in SGOP model with intrinsic CAR prior

Scope of work	Variable	No injury	Slight injury	FSI
Vehicle involved	Truck	-6.9%	+3.9%	+3.0%
	Other vehicle	0.0%	-4.0%	+4.0%
	Non-local vehicle	-7.3%	+4.1%	+3.2%
Road geometry	Horizontal curvature	0.0%	-6.5%	+6.5%
	Vertical grade	0.0%	-27.4%	+27.4%
Traffic composition	% of Vehicle Class III	-5.1%	+2.9%	+2.2%
Weather condition	Precipitation	0.0%	+8.5%	-8.5%
	Visibility	-0.9%	+0.5%	+0.4%
EMS response	EMS response time	-1.3%	+0.7%	+0.6%
Day of the week	Weekend	0.0%	+2.8%	-2.8%
Crash type	Rear-end crash	-16.3%	+9.3%	+7.0%
	Angled crash	-13.3%	+7.7%	+5.6%

Table 6. Marginal effects of significant factors in SGOP model with Leroux CAR prior

Scope of work	Variable	No injury	Slight injury	FSI
Vehicle involved	Truck	-6.7%	+3.9%	+2.8%
	Other vehicle	0.0%	-4.2%	+4.2%
	Non-local vehicle	-7.6%	+4.3%	+3.3%
Road geometry	Horizontal curvature	0.0%	-6.0%	+6.0%
	Vertical grade	0.0%	-28.3%	+28.3%
Traffic composition	% of Vehicle Class III	-5.9%	+3.4%	+2.5%

Weather condition	Precipitation	0.0%	+9.0%	-9.0%
	Visibility	-1.0%	+0.6%	+0.4%
EMS response	EMS response time	-1.3%	+0.7%	+0.6%
Day of the week	Weekend	0.0%	+2.6%	-2.6%
Crash type	Rear-end crash	-16.9%	+9.6%	+7.3%
	Angled crash	-11.7%	+6.8%	+4.9%

4.2. Parameter and mariginal effect interpretation

As the SGOP model with Leroux CAR prior is superior, implications of Bayesian parameter estimates (Table 3) and marginal effects of significant factors (Table 4) are elaborated.

As shown in Table 3, for vehicle type, involvement of truck significantly increases the crash injury propensity, as compared to passenger car, at the 5% level. Involvement of truck is correlated with 3.9% and 2.8% increases in slight injury crash and FSI crash respectively. This could be attributed to the increase in the amount of energy dissipation for crashes involving trucks. Therefore, harms on the casualties, especially in other vehicles involving in the same crash, can increase (Huang et al., 2011; Zeng et al., 2016). Involvement of other vehicle type (i.e. vehicle with a trailer) marginally affect the propensity of FSI crash at the 10% level. In particular, crashes involving other vehicle type are correlated with 4.2% reduction in the probability of slight injury crash and 4.2% increase in the probability of FSI crash. In addition, there is no change in the probability of no injury crash. This demonstrates the relaxation of the monotonic effects of explanatory factors on crash severity. Moreover, involvement of non-local vehicle (not registered in the province where crashes occurred) significantly increase the crash injury propensity at the 5% level. Specifically, when a non-local vehicle involves, probabilities of slight injury crash and FSI crash can increase by 4.3% and 3.3% respectively. This may be attributed to the route unfamiliarity of non-local drivers. Hence, defensive action would be absent when one is in emergency and/or driving in the adverse road environments. Hence, propensities of severe crashes increase (Wen and Xue, 2020; Zeng et al., 2020a).

For the road geometry, horizontal curvature and vertical grade significantly affect the threshold between slight injury and FSI levels at the 10% and 5% levels respectively. In particular, per unit increase (i.e. 10^{-1} km) in horizontal curvature is correlated with 6.0% increase in the probability of FSI crash, and 1% increase in vertical grade is correlated with 28.3% increase in the probability of FSI crash respectively. Such findings are reasonable and consistent to that of previous studies (Wang and Prato, 2019; Zeng et al., 2020a). In particular, Labi's study (2011) indicates that greater horizontal curvature would result in stronger centrifugal force exerted on the vehicle and harsher transition between tangent sections. Therefore, likelihood of run-off crash and overturn crash would increase. Also, increase in vertical grade would result in shorter sight distance. This would reduce the time available for emergency response before collision (Christoforou et al., 2010).

For the traffic composition, increase in the percentage of Class III vehicle significantly increase the crash injury propensity at the 5% level. 1% increase in Class III vehicles is correlated with 3.4% and 2.5% increases in the probabilities of slight injury crash and FSI crash respectively. A plausible reason is that the dimensions of Class III vehicles (constitute to 21.3% of overall traffic) are substantially greater than that of Class I vehicles (42.2% of overall traffic). Presence of Class III vehicles may obstruct the field of vision of the drivers of following vehicles. Therefore, likelihoods of severe crashes increase (Zeng et al., 2019a).

For the weather condition, visibility and precipitation significantly affect the crash injury severity at the 5% and 10% level respectively. Per unit increase in precipitation (i.e. 1 mm) is correlated with 0.9% reduction in FSI crash and per unit reduction in visibility (i.e. 1 km) is correlated with 0.4% reduction in FSI crash respectively. It could be attributed to the risk compensation behaviors of drivers when driving in the inclement weather conditions (e.g., rainy and low visibility) (Christoforou et al., 2010; Quddus et al., 2002). Drivers could adapt to the adverse driving environments by slowing down and being more cautious. Hence, the crash severity may reduce (Zeng et al., 2020a). This is indicative to the deployment of effective traffic engineering and traffic control measures, i.e. variable message sign, that can increase the safety awareness of drivers when driving in the rainy and foggy days. Additionally, the transport agencies can launch targeted road safety campaign in rainy season.

For the crash time, probabilities of FSI crashes on weekdays are 2.6% lower than that on weekends.

It could be because majority of drivers driving on the weekdays are commuters (Christoforou et al., 2010; Quddus et al., 2010). They are more familiar with the driving routes. Therefore, risk of FSI crash reduces. However, risk of slight injury crash can increase, as the commuter drivers tend to be less attentive and over-confident. They could overestimate their own driving capability and undermine the potential hazards on the roads (Christoforou et al., 2010; Quddus et al., 2010).

For the medical service and emergency response, increase in EMS response time significantly increases the crash injury propensity at the 5% level. In particular, one-minute increase in EMS response time is correlated with 0.7% and 0.6% increases in slight injury crash and FSI crash respectively. As expected, improvement in emergency medical service upon crash (i.e. on-site first-aid treatments and transport of victims to nearby hospitals) can reduce the probabilities of severe injury and mortality of casualties (Lee et al., 2018; Ma et al., 2019; Zeng et al., 2020a).

Last but not least, for the crash type, injury propensities of rear-end crashes and angled crashes are significantly higher than that of single vehicle crashes, both at the 5% level. Specifically, probabilities of slight injury and FSI of rear-end crashes are 9.6% and 7.3% higher than that of single vehicle crashes respectively. On the other hand, probabilities of slight injury and FSI of angled crashes are 6.8% and 4.9% higher than that of single vehicle crashes respectively. This could be because of the number of casualties involved in multiple vehicle crash, regardless of collision direction, tends to be higher than that involved in single vehicle crash (Castro et al., 2013; Meng et al., 2017).

5. CONCLUSIONS AND FUTURE RESEARCH

This study proposes a Bayesian SGOP model with Leroux CAR prior for crash injury severity analysis. Proposed model accounts for the ordinal nature of crash injury severity, relaxes the assumption of monotonic effects of explanatory factors on crash injury severity, and more importantly, provides a generalized structure of spatial correlation of the crashes that are in close proximity.

A comprehensive traffic, weather and crash dataset of Kaiyang Freeway in Guangdong Province of China in 2014 is used. Factors including traffic flow, traffic composition, road geometry, weather condition, crash time, and emergency medical services are considered. In addition, prediction performances of conventional GOP model, SGOP model with intrinsic CAR prior and SGOP model with Leroux CAR prior are compared. Results indicate that goodness-of-fit of SGOP model with Leroux CAR prior is better than that of the counterparts. SGOP model with Leroux CAR prior is capable of capturing the statistically significant while moderate spatial correlation of the crashes that are in close proximity. In addition, Bayesian parameter estimates of SGOP model with Leroux CAR prior are generally more precise. More importantly, current results generally conform with the road safety literatures. They all justify the suitability of the Leroux CAR prior formulation for the accommodation of spatial correlation in crash severity models.

To sum up, this study demonstrates the advancement of Bayesian SGOP model with Leroux CAR prior and underscores the importance of considering the strength of spatial correlation when modeling crash severity. In addition to spatial correlation, accounting for spatial heterogeneity (i.e., variations in the safety effects of contributing factors) using the methods like geographically weighted regression would be beneficial (Wang et al., 2016). Furthermore, it is worth exploring the use of advanced model formulations that can address other issues including underreporting (Yamamoto et al., 2008), unobserved heterogeneity (Dong et al., 2016; Mannering et al., 2016), and temporal correlation (Cheng et al., 2018; Meng et al., 2017; Zeng et al., 2018) simultaneously, when more comprehensive dataset is available in the extended study. Moreover, application of machine learning technique for crash severity models would be an interesting extension. Nevertheless, results of parameter estimation are based on the crash data of one freeway in China in 2014. Transferability of the results by time and location should deserve further investigation.

Acknowledgments

This work was jointly supported by the International Science & Technology Cooperation Program of China [No. 2017YFE0134500], the Natural Science Foundation of China [No. 71801095], the Research Grants Council of Hong Kong (Project No. 25203717), and the Fundamental Research Funds for the Central Universities (No. 2020ZYGXZR007).

References

Abdel-Aty, M., & Wang, X. (2006). Crash estimation at signalized intersections along corridors: Analyzing spatial effect and identifying significant factors. Transportation Research Record, 1953, 98-111.

Ahmed, M., Huang, H., Abdel-Aty, M., & Guevara, B. (2011). Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. Accident Analysis & Prevention, 43(4), 1581-1589.

Aguero-Valverde, J., & Jovanis, P. P. (2008). Analysis of road crash frequency with spatial models. Transportation Research Record, 2061, 55-63.

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.

Cai, Q., Abdel-Aty, M., Lee, J., & Huang, H. (2019). Integrating macro-and micro-level safety analyses: A Bayesian approach incorporating spatial interaction. Transportmetrica A: Transport Science, 15(2), 285-306.

Castro, M., Paleti, R., & Bhat, C. R. (2013). A spatial generalized ordered response model to examine highway crash injury severity. Accident Analysis & Prevention, 52, 188-203.

Cheng, W., Gill, G. S., Choi, S., Zhou, J., Jia, X., & Xie, M. (2018). Comparative evaluation of temporal correlation treatment in crash frequency modelling. Transportmetrica A: Transport Science, 14(7), 615-633.

Christoforou, Z., Cohen, S., & Karlaftis, M. G. (2010). Vehicle occupant injury severity on highways: An empirical investigation. Accident Analysis & Prevention, 42(6), 1606-1620.

Cressie, N. (1993). Statistics for spatial data. John Wiley & Sons.

Dong, C., Clarke, D. B., Nambisan, S. S., & Huang, B. (2016). Analyzing injury crashes using random-parameter bivariate regression models. Transportmetrica A: Transport Science, 12(9), 794-810.

Eberly, L. E., & Carlin, B. P. (2000). Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. Statistics in Medicine, 19, 2279-2294.

Eluru, N., Bhat, C. R., & Hensher, D. A. (2008). A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis & Prevention, 40(3), 1033-1054.

Huang, H., Siddiqui, C., & Abdel-Aty, M. (2011). Indexing crash worthiness and crash aggressivity by vehicle type. Accident Analysis & Prevention, 43(4), 1364-1370.

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.

IBM Corp. 2017. IBM SPSS Statistics for Windows, Version 25.0. Released. IBM Corp., Armonk, NY.

Labi, S. (2011). Efficacies of roadway safety improvements across functional subclasses of rural two-lane highways. Journal of Safety Research, 42(4), 231-239.

Lee, D. (2011). A comparison of conditional autoregressive models used in Bayesian disease mapping. Spatial and Spatio-Temporal Epidemiology, 2(2), 79-89.

Lee, J., Abdel-Aty, M., Cai, Q., & Wang, L. (2018). Effects of emergency medical services times on traffic injury severity: A random effects ordered probit approach. Traffic Injury Prevention, 19(6), 577-581.

Leroux, B. G., Lei, X., & Breslow, N. (2000). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In Statistical models in epidemiology, the environment, and clinical trials (pp. 179-191). Springer, New York, NY.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-a Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing, 10(4), 325-337.

Malyshkina, N. V., & Mannering, F. L. (2009). Markov switching multinomial logit model: An application to accident-injury severities. Accident Analysis & Prevention, 41(4), 829-838.

Mannering, F. L., Shankar, V., & Bhat, C. R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research, 11, 1-16.

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.

Meng, F., Xu, P., Wong, S. C., Huang, H., & Li, Y. C. (2017). Occupant-level injury severity analyses for taxis in Hong Kong: A Bayesian space-time logistic model. Accident Analysis & Prevention, 108, 297-307.

Milton, J. C., Shankar, V. N., & Mannering, F. L. (2008). Highway accident severities and the mixed logit model: An exploratory empirical analysis. Accident Analysis & Prevention, 40(1), 260-266.

Naik, B., Tung, L. W., Zhao, S., & Khattak, A. J. (2016). Weather impacts on single-vehicle truck crash injury severity. Journal of Safety Research, 58, 57-65.

Prato, C. G., Kaplan, S., Patrier, A., & Rasmussen, T. K. (2018). Considering built environment and spatial correlation in modeling pedestrian injury severity. Traffic Injury Prevention, 19(1), 88-93.

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.

Quddus, M. A., Noland, R. B., & Chin, H. C. (2002). An analysis of motorcycle injury and vehicle damage severity using ordered probit models. Journal of Safety Research, 33(4), 445-462.

Quddus, M. A., Wang, C., & Ison, S. G. (2010). Road traffic congestion and crash severity: econometric analysis using ordered response models. Journal of Transportation Engineering, 136(5), 424-435.

Savolainen, P. T., Mannering, F. L., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. Accident Analysis & Prevention, 43(5), 1666-1676.

Soroori, E., Mohammadzadeh Moghaddam, A., & Salehi, M. (2019). Application of local conditional autoregressive models for development of zonal crash prediction models and identification of crash risk boundaries. Transportmetrica A: Transport Science, 15(2), 1102-1123.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(4), 583-639.

Spiegelhalter, D. J., Thomas, A., Best, N., & Lunn, D. (2005). WinBUGS user manual. MRC Biostatistics Unit, Cambridge, United Kingdom.

Public Security Ministry of China (2019). Annual Statistical Report on Roadway Traffic Accidents (2018). Wuxi: Traffic Management Research Institute of the Public Security Ministry, China.

Wang, X., Liu, J., Khattak, A. J., & Clarke, D. (2016). Non-crossing rail-trespassing crashes in the past decade: A spatial approach to analyzing injury severity. Safety Science, 82, 44-55.

Wang, Y., & Prato, C. G. (2019). Determinants of injury severity for truck crashes on mountain expressways in China: A case-study with a partial proportional odds model. Safety Science, 117, 100-107.

Wen, H., & Xue, G. (2020). Injury severity analysis of familiar drivers and unfamiliar drivers in single-vehicle crashes on the mountainous highways. Accident Analysis & Prevention, 144, 105667.

Wen, H., Zhang, X., Zeng, Q., Lee, J., & Yuan, Q. (2019a). Investigating spatial autocorrelation and spillover effects in freeway crash-frequency data. International Journal of Environmental Research and Public Health, 16(2), 219.

Wen, H., Zhang, X., Zeng, Q., & Sze, N. N. (2019b). Bayesian spatial-temporal model for the main and interaction effects of roadway and weather characteristics on freeway crash incidence. Accident Analysis & Prevention, 132, 105249.

Winston, C., Maheshri, V., & Mannering, F. (2006). An exploration of the offset hypothesis using disaggregate data: The case of airbags and antilock brakes. Journal of Risk and Uncertainty, 32(2), 83-99.

World Health Organization, (2018). Global Status Report on Road Safety 2018. World Health Organization, Geneva, Switzerland.

Xie, Y., Zhang, Y., & Liang, F. (2009). Crash injury severity analysis using Bayesian ordered probit models. Journal of Transportation Engineering, 135(1), 18-25.

Xu, P., Huang, H., Dong, N., & Wong, S. C. (2017). Revisiting crash spatial heterogeneity: a Bayesian spatially varying coefficients approach. Accident Analysis & Prevention, 98, 330-337.

Xu, X., Xie, S., Wong, S. C., Xu, P., Huang, H., & Pei, X. (2016). Severity of pedestrian injuries due to traffic crashes at signalized intersections in Hong Kong: a Bayesian spatial logit model. Journal of Advanced Transportation, 50(8), 2015-2028.

Yamamoto, T., Hashiji, J., & Shankar, V. N. (2008). Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. Accident Analysis & Prevention, 40(4), 1320-1329.

Yasmin, S., Eluru, N., Bhat, C. R., & Tay, R. (2014). A latent segmentation based generalized ordered logit model to examine factors influencing driver injury severity. Analytic Methods in Accident Research, 1, 23-38.

Yu, R., & Abdel-Aty, M. (2013). Investigating different approaches to develop informative priors in hierarchical Bayesian safety performance functions. Accident Analysis & Prevention, 56, 51-58.

Yu, R., & Abdel-Aty, M. (2014). Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. Safety Science, 63, 50-56.

Zeng, Q., Wen, H., & Huang, H. (2016). The interactive effect on injury severity of driver-vehicle units in two-vehicle crashes. Journal of Safety Research, 59, 105-111.

Zeng, Q., Sun, J., & Wen, H. (2017). Bayesian hierarchical modeling monthly crash counts on freeway segments with temporal correlation. Journal of Advanced Transportation, Article ID: 5391054, DOI:10.1155/2017/5391054.

Zeng, Q., Wen, H., Huang, H., Pei, X., & Wong, S. C. (2018). Incorporating temporal correlation into a multivariate random parameters Tobit model for modeling crash rate by injury severity. Transportmetrica A: Transport Science, 14(3), 177-191.

Zeng, Q., Gu, W., Zhang, X., Wen, H., Lee, J., & Hao, W. (2019a). Analyzing freeway crash severity using a Bayesian spatial generalized ordered logit model with conditional autoregressive priors. Accident Analysis & Prevention, 127, 87-95.

Zeng, Q., Guo, Q., Wong, S. C., Wen, H., Huang, H., & Pei, X. (2019b). Jointly modeling area-level crash rates by severity: a Bayesian multivariate random-parameters spatio-temporal Tobit regression. Transportmetrica A: Transport Science, 15(2), 1867-1884.

Zeng, Q., Hao, W., Lee, J., & Chen, F. (2020a). Investigating the impacts of real-time weather conditions on freeway crash severity: a Bayesian spatial analysis. International Journal of Environmental Research and Public Health, 17(8), 2768.

Zeng, Q., Wang, X., Wen, H., & Yuan, Q. (2020b). An empirical investigation of the factors contributing to local-vehicle and non-local-vehicle crashes on freeway. Journal of Transportation Safety and Security, DOI: 10.1080/19439962.2020.1779422. (In press)

Zhai, X., Huang, H., Xu, P., & Sze, N. N. (2019). The influence of zonal configurations on macro-level crash modeling. Transportmetrica A: Transport Science, 15(2), 417-434.

Ziakopoulos, A., & Yannis, G. (2020). A review of spatial approaches in road safety. Accident Analysis & Prevention, 135, 105323.

Zou, W., Wang, X., & Zhang, D. (2017). Truck crash severity in New York city: an investigation of the spatial and the time of day effects. Accident Analysis & Prevention, 99, 249-261.