Analytic Methods in Accident Research 2 (2014) 1–11

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Incorporating spatial dependence in simultaneously modeling crash frequency and severity

Yu-Chiun Chiou*, Chiang Fu, Hsieh Chih-Wei

Department of Transportation and Logistics Management, National Chiao Tung University, 4 F, 118, Section 1, Chung-Hsiao W. Road, Taipei 100, Taiwan

ARTICLE INFO

Article history:
Received 1 September 2013
Received in revised form
14 December 2013
Accepted 14 December 2013
Available online 4 March 2014

Keywords:
Spatial dependence
Multinomial-generalized Poisson with error-components
Spatial errors
Spatial exogenous function

ABSTRACT

Estimation results obtained by models of crash frequency and severity without considering spatial dependence effects may lead to biased estimates and mis-specification of the risk factors in accident analysis. The solution developed in this study is a modification of the previously proposed multinomial-generalized Poisson with error-components (EMGP) model. Two spatial EMGP models, spatial error-EMGP and spatial exogenous-EMGP, are proposed to accommodate alternative spatial dependence structures. The spatial error-EMGP model incorporates spatial error in the structure of spatial auto-regression and spatial moving average to capture spatial correlation effects; while the spatial exogenous-EMGP model introduces the spatial exogenous functions composed of two state parameterized functions associated with traffic and geometric composite variables to explain the sources of spatial dependence. A case study of crash data for Taiwan Freeway no. 1 is performed. According to the estimation results, the spatial exogenous-EMGP model not only performs best in terms of BIC, RMSE and LR tests, it also shows the sources of spatial dependence and how spatial dependence decays as the distance to adjacent segments increases.

1. Introduction

Since crash frequency and severity are two key indices that measure road risk, numerous research contributions have attempted to explore crash frequency (the number of accidents on roadway segments or at intersections over a specified period; see Lord and Mannering, 2010 for detailed review) and severity (the percentages of accidents by severity levels; see Savolainen et al., 2011) by analyzing relevant risky factors, such as traffic conditions, roadway configuration, weather conditions, and traffic control types so as to identify key risk covariates and propose appropriate safety measures. Furthermore, many contributions in crash modeling have been devoted to jointly analyze crash frequency and severity through various multivariate models to accommodate the correlation of unobserved factors among crash counts by severity levels. Interested readers please see Mannering and Bhat (2014) for a comprehensive introduction. Additionally, there still need some continuous efforts to improve the applicability of multivariate models, for example, to accommodate the spatial dependence (Miaou and Song, 2005; Aguero-Valverde, 2013; Castro et al., 2012, 2013; Narayanamoorthy et al., 2013; Wang and Kockelman, 2013). The proposed modeling approach should be able to not only recognize the correlations among crash counts of various severity types but also to accommodate spatial dependence between adjacent analytic units.

* Corresponding author. Tel.: +8862 2349 4940.
E-mail address: ycchiou@mail.nctu.edu.tw (Y.-C. Chiou).

2213-6657/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.amar.2013.12.001

The spatial dependence of crash counts or crash severity percentages between adjacent spatial unit (intersection, road segment or area) are usually high with substitution pattern (Maher, 1990), or sustain similar accident tendency and injury risk propensity (Wang and Abdel-Aty, 2006; Quddus, 2008; Wang et al., 2009; Guo et al., 2010; Wang et al., 2011; Siddiqui et al., 2012; Castro et al., 2013), because adjacent spatial unit often bear similar crash confounding factors. The effect of spatial dependence includes the spatial spillover effect from the nature of continuity existing in some observed variables, *e.g.,* roadway configuration and traffic flow patterns between neighboring road segments (especially for the closest road segments), and the spatial correlation effect of common unobserved factors on any two adjacent segments. This problem is common in the collective modeling approach, in which the study corridor is usually subjectively divided into several segments (Miaou et al., 2003; Wang et al., 2009). Therefore, ignoring the impact of spatial dependence may results into inconsistent and inefficient parameter estimation under the multivariate models framework (Miaou and Song, 2005; Aguero-Valverde and Jovanis, 2006; Song et al., 2006).

Conventional spatial econometric models use spatial lags and spatial error specifications to accommodate spatial dependence (see Anselin, 1988, 2002). The main reason for using spatial lag is to identify spatial spillover effects based on observed variables at proximal segments. While the spatial error specification method can consider the spatial correlation effect, it does not reflect spatial spillover effects because it only accommodates the unobserved factors of proximally located neighborhoods (Bhat et al., 2012). Earlier efforts have accounted for spatial dependence effects (LaScala et al., 2000; MacNab, 2004; Ha and Thill, 2011; Delmelle et al., 2012) by converting the accident count into logarithmic form and by using regression method for parameter estimation. The main advantage of these models is the avoidance of methodological complexities when capturing spatial effects. However, these models only consider total crash counts, not the severity of crash counts. To consider the effects of spatial dependence on crash counts of varying severity, crash frequency and severity must be jointly estimated while incorporating spatial dependencies. Spatial lag dependence efficiently and intuitively captures both spatial spillover and spatial error correlation effects and is much better for justifying and accommodating spatial dependence (Sidharthan and Bhat, 2012). Recent multivariate safety studies have concerned on this issue and already made some contributions (Miaou and Song, 2005; Castro et al., 2012, 2013; Aguero-Valverde, 2006, 2013; Narayanamoorthy et al., 2013). However, the nature of spatial dependence heterogeneity has barely been discussed in more details.

As an extension of the multinomial-Poisson (MP) model (Terza and Wilson, 1990), Chiou and Fu (2013) proposed a multinomial generalized Poisson with error-components (EMGP) model for jointly estimating a generalized Poisson distribution for the marginal probability of total counts and a conditional multinomial distribution for the conditional frequency of each severity type. The applicability and performance of the proposed model has been validated. The core logic of the proposed model has been applied for joint estimation of crash frequency and severity probability by allowing common unobserved components (deemed as the mixing structure approach) for various severity counts (Narayanamoorthy et al., 2013). The EMGP model described risk factors according to their accident descriptive components (*i.e.*, severity and frequency) under an integrated framework. Two methods of modeling spatial dependence have also contributed to the development of EMGP: the spatial error-EMGP, which incorporates spatial errors (LaScala et al., 2000; Van der Kruk, 2002; Kosfeld et al., 2006; Ha and Thill, 2011), and the exogenous-EMGP model, which introduces exogenous spatial dependence functions similar to an impedance function (Mohammadian et al., 2005) to explore both spatial spillover and correlation effects.

After reviewing the development of these spatial-EMGP models and their estimation results, this study further considers how spatial dependence affects crash frequency and severity. The remainder of this paper is structured as follows. Section 2 briefly introduces the EMGP and spatial-EMGP models. Section 3 presents the descriptive statistics of the empirical data. Section 4 examines and discusses the estimation results for the proposed models. Finally, Section 5 concludes the study and recommends further research.

2. The model

As mentioned above, the proposed spatial-EMGP model is mainly based the EMGP model proposed by Chiou and Fu (2013). The EMGP model is an extension of the multinomial-Poisson regression model (Terza and Wilson, 1990) by relaxing the assumption on marginal (unconditional) probability using the generalized Poisson (GP) model (Famoye et al., 2004; Dissanayake et al., 2009) and by incorporating the error components structure. In accordance with the derivations reported in previous research, the MGP model can be formulated as follows:

$$p(\tilde{y}) = \frac{\prod_{j=1}^{J} \pi_j(Z)^{y_j}}{\prod_{j=1}^{J} y_j!}[1+\eta N]^{N-1}\left(\frac{\lambda(X)}{1+\eta\lambda(X)}\right)^N \exp\left[\frac{-\lambda(X)(1+\eta N)}{1+\eta\lambda(X)}\right] \tag{1}$$

$$\lambda = \exp(\beta' X + \varepsilon) \tag{2}$$

$$S_j = \gamma' Z + v_j \tag{3}$$

$$g(\cdot) = \left(\frac{\lambda}{1+\eta\lambda}\right)^N \frac{(1+\eta N)^{N-1}}{N!} \exp\left(\frac{-\lambda(1+\eta N)}{1+\eta\lambda}\right) \tag{4}$$

$$\pi_j = \frac{\exp(S_j)}{\sum_i^J \exp(S_i)} \tag{5}$$

where Eq. (1) integrates two descriptive model components (*i.e.*, frequency and severity) based on the product of conditional probability and marginal probability inspiration. $p(\tilde{y})$ is the resulting joint probability density function of crash frequencies of various severity levels. N is the total number of accidents across different severity levels at certain segment within a given period (*e.g.*, 1 year). y_j is the crash frequencies at j severity level given in certain total number of accidents, N. Thus, $N = \sum y_j$, which is the sum of crash counts of all severities for a road segment over a unit of time.

The expected number of crash frequencies is measured by the function λ, which is specified as Eq. (2). π_j is the probability of severity at level j and measured by the severity function S_j as shown in Eq. (3), which assuming as a linearly additive function for measuring risk of severity level j. ε is the random error term of λ which is the generalized Poisson (GP) model derived by Famoye et al. (2004) as expressing in Eq. (4). v_j is the random error term of S_j follows the Gumbel distribution, thus the severity probability (Milton et al., 2008) is determined by the multinomial logit (MNL) as depicted in Eq. (5).

X and Z are two vectors of non-random explanatory variables (such as roadway geometrics, traffic factors, land use, weather condition and so on) for the above crash measuring functions λ and S_j, respectively. β is a vector of the estimated parameters of λ, and γ is the parameter vector of S_j. Finally, η is the dispersion parameter derived from generalized Poisson distribution assumption in the function λ. If $\eta > 0$ indicates the over-dispersion in the empirical case. In contrast, $\eta < 0$ denotes the under-dispersion. While $\eta = 0$, the probability function degenerated to the Poisson distribution.

The structure of an error component specified by the MGP model may be an innovative choice compared to the formulation of a pseudo-covariance-variance matrix (Ye et al., 2009, 2013). The above expected frequency λ and severity function S_j are stitched together through accommodating the common error components. In this case, four random coefficients are specified for the expected frequency λ and severity function S_j ($j = 1, 2, 3$) to derive the following covariance structure:

$$\lambda = \exp(\beta' X + \varepsilon + \sigma_\lambda u) \tag{6}$$

$$S_j = \gamma' Z + v_j + \sigma_{sj} u \tag{7}$$

where u is an independent random variable assumed to be normally distributed. σ_λ, σ_{sj} are corresponding coefficients to be estimated. For estimation, $j-1$ parameters are estimated by arbitrarily setting one parameter to 1. The cumulative probability function conditional on this random variable $p(\tilde{y}|u)$ is expressed as Eq. (8) and named as the multinomial generalized Poisson model with error components (EMGP) to distinguish it from Eq. (1).

$$p(\tilde{y}|u) = \frac{\prod_{j=1}^J \pi_j(Z,u)^{y_j}}{\prod_{j=1}^J y_j!} [1+\eta N]^{N-1} \left(\frac{\lambda(X,u)}{1+\eta\lambda(X,u)} \right)^N \exp\left[\frac{-\lambda(X,u)(1+\eta N)}{1+\eta\lambda(X,u)} \right] \tag{8}$$

2.1. Spatial error-EMGP

A conventional econometric method of modeling spatial dependence is the spatial error model, which expresses each residual as a function of adjacent residuals (Van der Kruk, 2002; Kosfeld et al., 2006; Eckey et al., 2007). The EMGP model considering the spatial error is named as the spatial error-EMGP. Meanwhile, a spatial error can be classified as a spatial autocorrelation (SAR) error or as a spatial moving average (SMA) error, and the general formulas for the unmeasured error term in the expected crash function λ and crash severity function S_j can be expressed as Eqs. (9) and (10), respectively:

$$\varepsilon = \sum_{k=1}^{p} \rho_{\lambda k} W \varepsilon_k + \sum_{m=1}^{q} \theta_{\lambda m} W \tilde{e}_{\lambda m} + e_\lambda \tag{9}$$

$$v_j = \sum_{k=1}^{p} \rho_{sk} W v_{jk} + \sum_{m=1}^{q} \theta_{sm} W \tilde{e}_{sm} + e_s \tag{10}$$

where ε_k and v_{jk} are the spatial autocorrelation error vectors depicted by any two segments being the kth order spatial neighbors for the functions λ and S_j, respectively. $\tilde{e}_{\lambda m}$ and $\tilde{e}_{sm}$ denote the moving average error vectors depicted by any two segments being the mth order spatial neighbors for the functions λ and S_j, respectively. $\rho_{\lambda k}$, ρ_{sk}, $\theta_{\lambda k}$, and θ_{sk} are associated estimating coefficient vectors reflecting spatial autocorrelation and spatial moving average effects. p and q represent the chosen order of the spatial autocorrelation and moving average parts, respectively. Different from time series models, W is the spatial weights matrix. e_λ and e_s are the white error terms associated with functions λ and S_j, respectively.

The negative (positive) sign of ρ values represent negative (positive) spatial autocorrelation with a range from "1" (perfect correlation) to "-1" (perfect dispersion). While values of θ represent the mean effects of very nearer neighboring spatial units, as compared with the long-term effects from ρ. Additionally, the spatial weights matrix (W) is specified as follows:

$$\mathbf{W}=\begin{bmatrix} 0 & w_{12} & \cdots & w_{1n} \\ w_{21} & 0 & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ w_{n1} & w_{n2} & \cdots & 0 \end{bmatrix}; \quad w_{ii}=0, w_{ij}=(1/\overset{2}{\underset{ij}{d}})/\sum_{k=1}^{n}(1/\overset{2}{\underset{ik}{d}}); \quad \sum_{k=1}^{n} w_{ik}=1, \forall i \tag{11}$$

Of course, the spatial matrix can be changed according to different research needs. In the proposed spatial error-EMGP, the square of the geographical distance (d^2) of the road segments is used to determine the spatial weighted matrix. In doing this, each of road segments is formed by two adjacent interchanges and d_{ij} is measured by the distance between the corresponding end interchanges of segment i and j.

Consequently, assuming that the error type is Spatial-ARMA(p=1,q=1), namely SARMA(1,1), the settings for the exponentiated crash frequency function λ and crash severity function S_j can be expressed as follows:

$$\lambda=\exp(\beta' X+\sigma_\lambda u+\rho_{\lambda 1}\, W\varepsilon_1+\theta_{\lambda 1} W\tilde{e}_{\lambda 1}+e_\lambda) \tag{12}$$

$$S_j=\gamma' Z+\sigma_{sj} u+\rho_{s1}\, W v_{j1}+\theta_{s1}\, W\tilde{e}_{s1}+e_s \tag{13}$$

Eq. (14) is the spatial error-EMGP model after modification from the EMGP model in Eq. (8) to incorporate spatial error.

$$p(\tilde{y}|u)=\frac{\prod_{j=1}^{J}\pi_j(Z,u|\rho_s,\theta_s)^{y_j}}{\prod_{j=1}^{J} y_j!}[1+\eta N]^{N-1}\left(\frac{\lambda(X,u|\rho_\lambda,\theta_\lambda)}{1+\eta\lambda(X,u|\rho_\lambda,\theta_\lambda)}\right)^N \exp\left[\frac{-\lambda(X,u|\rho_\lambda,\theta_\lambda)(1+\eta N)}{1+\eta\lambda(X,u|\rho_\lambda,\theta_\lambda)}\right] \tag{14}$$

2.2. Spatial exogenous-EMGP

Another approach to modeling spatial dependence, which was proposed by Mohammadian et al. (2005), is to incorporate a spatial dependence term φ to measure spatial dependence in the discrete choice model as

$$\varphi=\sum_{j=1}^{n}\tau\exp\,(d_{nj}/\kappa) \tag{15}$$

where τ and k are the state parameters of φ.

This study similarly uses two state parameters, τ and κ, to reflect the impact of spatial dependence. This model can be viewed as an alternative specification of traditional spatial exogenous model. Furthermore, in order to explain how the crash frequency and severity of a spatial analytic unit are affected by geographically adjacent analytic units in terms of their geometric and traffic characteristics. The spatial dependence term φ can be further explained by extending two state functions $\tilde{\tau}$ and $\tilde{\kappa}$, *i.e.* the spatial exogenous (state) functions $\varphi(.)$, instead of specifying the above two estimated state parameters τ and κ. Hence, crash frequency function λ and crash severity function S_j incorporate lagged explanatory variables (*i.e.*, explanatory variables of adjacent analytic units) associating with the sources of spatial heterogeneity into the spatial exogenous function as shown in Eqs. (16) and (17), respectively:

$$\lambda=\exp(\beta' X+\varepsilon+\sigma_\lambda u+\varphi(\dot{X})) \text{ and } \varphi(\dot{X})=\sum_{j=1}^{n}\tilde{\tau}(\dot{X})\exp(d_{nj}/\tilde{k}(\dot{X})) \tag{16}$$

$$S_j=\gamma' Z+v_j+\sigma_{sj} u+\varphi(\dot{Z}) \text{ and } \varphi(\dot{Z})=\sum_{j=1}^{n}\tilde{\tau}(\dot{Z})\exp(d_{nj}/\tilde{k}(\dot{Z})) \tag{17}$$

where $\dot{X}$ is a vector of normalized lagged variables for explaining the spatial dependence of crash frequency. $\dot{Z}$ is a vector of normalized lagged variables for explaining the spatial dependence of crash severity.

The values of function $\varphi(.)$ in Eqs. (16) and (17) represent the level of spatial dependence of various road segments. Extending two state parameters into two state parameterized functions allows the spatial dependence to vary across segments in order to explain the impact of spatial dependence and to relax the assumption of similar spatial dependence in all analytic units. The estimation is simplified by assuming that the state functions have a linear form. The concept of Euclidean distance is applied to compute the lagged variables. Additionally, prevent collinearity problems from affecting the distance measurement, all variables in the state functions are normalized as

$$\dot{x}_n=\sqrt{\sum_l^L(x'_{ln}-x'_{ln\,-1})^2}, x'_{ln}=\frac{(x_{ln}-x_l^{\min})}{(x_l^{\max}-x_l^{\min})} \tag{18}$$

where $\dot{x}_n$ is a vector of the lagged variables of segment n to represent the Euclidean distance measured by the variables of x_1, x_2, …, x_L. $\dot{x}_n$ is one of elements of $\dot{X}$. L is the number of variables used to measure the Euclidean distance. $x_l^{\min}$ is the minimal value of x_l among all analytic units. $x_l^{\max}$ is the maximal value of x_l among all analytic units.

Two lagged variables are assumed in these state functions: a geometric composite variable (LGC) and traffic composite variable (LTC). The geometric composite variable represents the distance (dissimilarity) between two adjacent units measured by geometric variables, *e.g.*, curvature, slope, number of lanes, *etc.* The traffic composite variable is measured by traffic variables, *e.g.*, traffic volumes of passenger cars, large vehicles, and tractor-trailers. Including spatial exogenous functions in the EMGP model changes Eqs. (8)–(19).

$$p(\tilde{y}|u) = \frac{\prod_{j=1}^{J} \pi_j(Z, u, \varphi(\dot{Z}))^{y_j}}{\prod_{j=1}^{J} y_j!} [1+\eta N]^{N-1} \left(\frac{\lambda(X, u, \varphi(\dot{X}))}{1+\eta\lambda(X, u, \varphi(\dot{X}))} \right)^N \exp\left[\frac{-\lambda(X, u, \varphi(\dot{X}))(1+\eta N)}{1+\eta\lambda(X, u, \varphi(\dot{X}))} \right] \tag{19}$$

2.3. Estimation and comparison

In this study, the models are estimated by maximizing the log-likelihood function by using GAUSS software (Aptech Systems, 2012) which is a matrix programming language for solving numerical problems such in econometrics, mathematics and statistics. The maximum likelihood module of GAUSS software (Aptech Systems, 2008) was used to estimate the EMGP and two spatial-EMGP models. Additionally, Estimation is performed using simulation-based maximum likelihood method with Halton draws, which obtain a more efficient distribution for numerical integration compared to purely random draws (Train, 2003). For better accuracy, the estimations by the proposed models are obtained for 200 Halton draws. In this empirical study, however, the final convergence did not vary markedly once the number of replications exceeded 150 times.

To determine the performance of EMGP with and without spatial dependence, commonly used performance indicators, such as the adjusted likelihood ratio index (ρ^2), Bayesian Information Criterion (BIC), and root mean square error (RMSE) are reported and expressed as follows:

$$\text{Adjusted } \rho^2 = \frac{\text{LL}(M) - \text{LL}(\beta)}{\text{LL}(M)} \tag{20}$$

where LL(M) is the value of the log-likelihood function with three constants for crash severity (market share) and a single constant for generalized Poisson model. LL(β) is the value of the log-likelihood function at final convergence.

$$\text{BIC} = -2 \times \text{LL}(\beta) + K \times \ln(M) \tag{21}$$

where k is the number of parameters and N is the sample size.

$$\text{RMSE} = \sqrt{\frac{\sum_j (\hat{y}_j - y_j)^2}{N}} \tag{22}$$

where $\hat{y}_j$ is the number of crashes at severity level j predicted by the model. y_j is real number of crashes.

3. Empirical data

The dataset included accidents that occurred on the Taiwan No. 1 freeway during year of 2005. This north-south freeway is 373.3 km long, and has 63 interchanges. To facilitate model estimation, the analytic spatial units were limited to two adjacent interchanges. Separately counting the number of units in the north-bound and south-bound directions resulted in 124 analytical units. Table 1 shows the descriptive statistics for these segments. Since the segments lengths markedly differed, the dependent variable represented by the crash count was divided by the segment length (GL) to provide a better reflection of crash risk (for additional details, see Chiou and Fu, 2013). Fig. 1 presents the distribution of crash counts/GL (*i.e.*, the crash counts/km) from the northernmost segment (No. 1) to the southernmost segment (No. 62). It also shows that the crash counts markedly differed by direction, even in the same segment. Therefore, units at the same road segment for different directions were analyzed as different samples. Moreover, the number of crashes at some analytic segments is similar to their adjacent segments, suggesting potential existence of spatial dependence.

Finally, when developing the spatial exogenous-EMGP model, the spatial exogenous functions were divided into two major composite variables: geometric variables and vehicle variables. The five geometric variables included number of lanes, degree of curvature, degree of upward/downward slope, and clothoid values. The three vehicle variables included trailers, large vehicles, and small vehicles.

4. Results and discussion

To determine whether spatial-EMGP models were needed to explain the spatial dependence effects, this study compared the EMGP model without spatial dependence proposed by Chiou and Fu (2013) and the two spatial-EMGP models, spatial error-EMGP and spatial exogenous-EMGP. Moreover, all explanatory variables in the spatial-EMGP models, except for those related to spatial dependence, were identical to those used in the EMGP model introduced in Chiou and Fu (2013) for simplifying the comparisons. Table 2 compares the performance test results for the three models. Notably, the estimation results indicated that, in terms of the performance indicators BIC and RMSE, the spatial exogenous-EMGP model was the best model, and the spatial error-EMGP model was the second best model, which suggested the need to consider spatial dependence. The performance of the three models was also ranked by LR-tests.

Table 1
Descriptive statistics of 124 segments.

Variable	Description	Mean	SE	Min	Max
Dependent variables					
y_3	PDO crashes/km(=Crash counts/segment length)	16.4	16.3	1.2	83.5
y_2	Injury crashes/km(=Crash counts/segment length)	0.8	0.6	0.0	2.8
y_1	Fatal crashes/km(=Crash counts/segment length)	0.1	0.2	0.0	1.5
$\acute{N}$	Total crashes/km(=Crash counts/segment length)	17.3	16.6	2.2	85.3
Independent variables (continuous)					
GN	Number of lanes	2.6	0.6	2.0	4.0
GL	Segment length (km)	5.9	4.4	0.8	22.4
GC	Curvature (‰)	0.7	1.1	0.0	7.1
GU	Maximum upward slope (%)	1.3	2.1	0.0	13.7
GD	Maximum downward slope (%)	1.2	1.4	0.0	5.2
GO	Clothoid curve value (1000°)	0.9	0.9	0.0	3.2
GS	Speed limit (GS=1 for 110 km, GS=0 else)	0.5	0.5	–	–
RF	Annual rainfall (100 mm)	21.1	7.5	11.1	38.9
TTV	Total traffic (thousand passenger car units)	69.2	28.7	10.8	157.0
PSV	Percentage of small vehicles (%)	51.4	10.1	31.9	70.5
PLV	Percentage of large vehicles (%)	23.8	4.2	15.5	34.2
PKV	Percentage of trailer-tractors (%)	24.8	8.7	9.2	41.0
Independent variables (dummy)					
PT	Presence of toll station (yes=1; no=0)	0.2	0.4	–	–
PR	Presence of rest area (yes=1; no=0)	0.1	0.3	–	–
PS	Presence of posted speed camera (yes=1; no=0)	0.3	0.5	–	–
AM	Adjacent to metropolitan (yes=1; no=0)	0.5	0.5	–	–
AP	Adjacent to airport, seaport or industry area (yes=1; no=0)	0.2	0.4	–	–

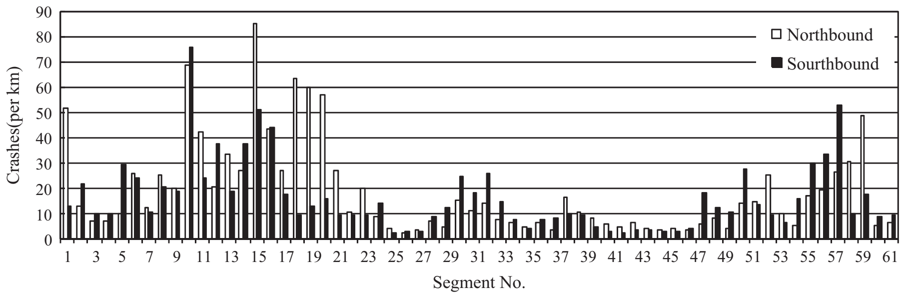

Fig. 1. Distribution of crash counts/km on the study freeway.

Tables 3–4 show the results for the two spatial-EMGP models. The estimation results for the EMGP model are given in Chiou and Fu (2013). Compared with the EMGP model, the coefficients of many explanatory variables are markedly reduced and influence their significance levels with their previous signs. For example, for the geometric variable GO, the coefficient and significance are considerable reduced in the spatial error-EMGP. While it still retain in the spatial exogenous-EMGP, its magnitude is also reduced, which suggests the overestimation of marginal impact of the EMGP model. Moreover, the exogenous errors σ_s and σ_λ integrating frequency and severity functions are clearly affected by spatial dependencies. The coefficients of σ_λ in the spatial models are also lower than those of the EMGP model. Finally, changes in coefficient σ_s in the spatial models indicate the effects of spatial dependence on the error structure. These findings reveal that spatial dependence should be considered when modeling crash frequency and severity.

Since it is no priori reason that how to choose a proper SARMA error specification initially due to the nature of data driven. One way out is to estimate different combinations of Spatial error-EMGP models, such as SARMA (2,1), SARMA(1,2), SARMA(1,0) and SARMA(0,1). After comparing among above combinations through the indicators as described in Section 2.3, the results show that for both the crash frequency and severity submodels, SARMA(1,1) performs best. Therefore, only the estimated results of SARMA(1,1) are reported in Table 3.

According to the estimation results in Table 3, the SAR(1) error structure expressed that crashes among whole freeway road segments are correlated in certain steady manner and at relative high degrees, since both SAR(1) values are 0.957 and 0.754 in frequency and severity submodels, respectively. Also, the SMA(1) indicated that the crash frequencies within the very limited distance are correlated, but the spatial dependence of severity is ambiguous. As noted in the “Introduction” section, the spatial error-EMGP model adequately reflects spatial correlation effects but does not explain them. Therefore, one inference is that only parts of the unobserved factors overlap.

Table 2
Prediction accuracy and statistical test.

Model	Severity type	Mean crash counts[a]	RMSE	Final Log-likelihood Value (BIC) [b]	LR-test	
					Against *EMGP*	Against *spatial error*
Actual	Fatality	0.10(0.82%)				
	Injury	0.78(7.48%)				
	PDO	16.44(91.70%)				
	Overall (Std.)	17.32(16.55)				
EMGP	Fatality	0.12(0.69%)	0.25	−1037.935 (2215.659)		
	Injury	1.07(6.18%)	0.83			
	PDO	16.18(93.14%)	15.22			
	Overall (Std.)	17.37(14.52)	15.20			
Spatial error-EMGP	Fatality	0.12(0.67%)	0.20	−1019.540 (2198.150)	$36.79>\chi^2_{(4,0.01)}=13.28$	
	Injury	1.31(7.49%)	1.60			
	PDO	16.03(91.85%)	12.90			
	Overall (Std.)	17.45(13.32)	12.82			
Spatial exogenous-EMGP	Fatality	0.14(0.78%)	0.43	−988.838 (2175.307)	$98.19>\chi^2_{(12,0.01)}=26.22$	$61.40>\chi^2_{(8,0.01)}=20.09$
	Injury	1.07(6.18%)	1.09			
	PDO	16.17(93.04%)	12.07			
	Overall (Std.)	17.38(13.63)	12.32			

Note: The percentages (%) of mean predicted crash counts/km (crash counts/segment length) at three severity levels are given in parentheses.

Table 3
Estimation results of the spatial error-EMGP.

Variable		Fatality		Injury		PDO	
		Para.	*t*-Stat	Para.	*t*-Stat	Para.	*t*-Stat
Logit crash severity model component							
Constant (for severity submodel)		–		1.851	3.78	4.825	9.26
GN	Number of lanes	–		–		−0.049	−1.92
Exp(*GU*)	Exponential of maximum upward slope	−0.221	−7.10	–		–	
GD	Maximum downward slope	0.256	4.89	0.209	5.12	–	
GO	Clothoid parameter	0.012	0.15	0.111	2.31	–	
GS	Speed limit	–		0.441	3.40	–	
TTV	Total traffic	−0.568	−2.94	−0.777	−6.34	–	
PLV	Percentage of large vehicles	6.656	7.64	6.656	7.64	–	
AM	Adjacent to metropolitan	–		–		0.149	4.58
PS	Presence of posted speed camera	–		−0.324	−1.90	−0.222	−6.51
PR	Presence of rest area	−0.564	−6.79	−1.125	−5.80	–	
Error component in crash severity(σ_s)		1.000	–	0.475	5.89	0.991	14.14
Spatial error of severity submodel							
Spatial-AR(1)		0.754	11.78				
Spatial-MA(1)		−0.512	−10.62				
Generalized Poisson crash frequency model component							
Constant (for frequency submodel)		0.466	3.11				
η	Dispersion parameter	0.066	7.06				
GC	Curvature	0.066	1.99				
GD	Maximum downward slope	−0.394	−5.37				
GD^2	Square of maximum downward slope	0.037	2.14				
PSV	Percentage of small vehicles	1.536	3.64				
AM	Adjacent to metropolitan	0.310	3.69				
AP	Adjacent to airport, seaport or industry area	0.177	5.53				
PT	Presence of toll station	−0.309	−3.16				
Error component in crash frequency(σ_λ)		0.113	2.99				
Spatial error of crash frequency submodel							
Spatial-AR(1)		0.957	64.30				
Spatial-MA(1)		0.171	2.64				
Log-likelihood (Null model)		−1298.488					
Log-likelihood (Full model)		−1019.540					
Adjusted ρ^2		0.215					
Samples sizes (Included variable)		124 (32)					

Table 4 shows the estimation of the spatial exogenous-EMGP which combines two state functions $\tilde{\tau}$ and $\tilde{\kappa}$, associated with lagged explanatory variables $\dot{X}$ and $\dot{Z}$ for crash frequency and severity, respectively. With the presuming specifications of $\tilde{\tau}$ and $\tilde{\kappa}$ in the spatial dependence φ, the value $\tilde{\kappa}$ on φ is in the impact convergence region and its high value indicates more road segments involving and longer distances extending. Regarding the impact of value $\tilde{\tau}$ on φ, a high value indicates

Table 4
Estimation results of the spatial exogenous-EMGP.

Variable			Fatality		Injury		PDO	
			Para.	*t*-Stat	Para.	*t*-Stat	Para.	*t*-Stat
Logit crash severity model component								
Constant (for severity submodel)			–		2.801	43.92	2.226	71.04
GN	Number of lanes		–		–		−0.176	−6.25
Exp(*GU*)	Exponential of maximum upward slope		−0.562	−17.65	–		–	
GD	Maximum downward slope		0.410	11.65	0.107	3.52	–	
GO	Clothoid parameter		0.154	3.80	0.150	3.40	–	
GS	Speed limit		–		−0.140	−4.12	–	
TTV	Total traffic		−0.800	−19.01	−1.130	−18.72	–	
PLV	Percentage of large vehicles		4.728	96.07	4.728	96.07	–	
AM	Adjacent to metropolitan		–		–		0.119	3.04
PS	Presence of posted speed camera		–		−1.011	−15.90	−0.569	−15.60
PR	Presence of rest area		−0.928	−29.06	−0.484	−12.75	–	
Error component in crash severity(σ_s)			–		0.463	14.65	−0.062	−1.89
Spatial exogenous function of severity submodel $\varphi(\dot{z})$								
$\tilde{k}(\dot{z})$	Constant ($\tilde{\kappa}$)		0.332	10.53	1	–	0.012	7.53
	LGC	Lagged geometrics composite	–		–		0.022	6.39
$\tilde{\tau}(\dot{z})$	Constant ($\tilde{\tau}$)		−0.110	−4.93	–		2.150	39.40
	LTC	Lagged traffic composite	0.877	17.27	0.239	17.10	6.681	182.11
Generalized Poisson crash frequency model component								
Constant (for frequency submodel)			−2.362	−73.24				
η	Dispersion parameter		0.062	7.27				
GC	Curvature		0.133	3.65				
GD	Maximum downward slope		−0.376	−11.50				
GD^2	Square of maximum downward slope		0.024	2.28				
PSV	Percentage of small vehicles		3.103	95.62				
AM	Adjacent to metropolitan		0.412	8.60				
AP	Adjacent to airport, seaport or industry area		0.437	12.60				
PT	Presence of toll station		−0.269	−4.28				
Error component in crash frequency (σ_λ)			0.108	3.26				
Spatial exogenous function of frequency submodel $\varphi(\dot{x})$								
$\tilde{k}(\dot{x})$	Constant ($\tilde{\kappa}$)		0.005	2.07				
	LGC	Lagged geometrics composite	0.016	3.12				
$\tilde{\tau}(\dot{x})$	Constant ($\tilde{\tau}$)		3.151	91.92				
	LTC	Lagged traffic composite	0.910	23.80				
Log-likelihood (Null model)			−1298.488					
Log-likelihood (Full model)			−988.838					
Adjusted ρ^2			0.238					
Samples sizes (Included variable)			124 (41)					

a strong correlation and dependence on the following road segment. According to the final estimation results, a lagged geometric composite (LGC) variable is a better indicator of the state function of $\tilde{\kappa}$ and is used in the expected frequency function (λ) and PDO crash function of the severity sub-model. Meanwhile, the lagged traffic composite variable (LTC) is determined by $\tilde{\tau}$, and all crash severity functions of the severity sub-model can significant confirm as well as the function λ in crash frequency sub-model.

According to abovementioned findings, the effects of geometric characteristics are a better indicator of spatial impact convergence of spatial-dependence. The empirical data confirm that the effects on crash frequency are larger than the effects on crash severity. The effects of traffic factors on spatial-dependence are relatively larger in terms of impact magnitude, and the designation is better regardless of frequency and severity; therefore, the above results are rational since the continuity of geometric design in freeway is usually subject to geographic and environmental considerations, which may be involving longer distance and many spatial units. Traffic factors include the vehicle fleet and crash exposure. Therefore, the perceived impact of spatial dependence is larger compared to other environment factors. Moreover, the marginal effect of function $\tilde{\tau}$ in the case study is larger than that of function $\tilde{\kappa}$; therefore, the spatial dependencies for a given distance should be substantial even if the convergence is restrictive.

The spatial exogenous-EMGP model not only describes the effects of spatial error correlation (exhibited in the constants of the function), it also describes spatial spillover effects (reflected by the lagged explanatory variables) through estimated spatial exogenous functions. Thus, it provides better estimation results than the spatial error-EMGP does. Moreover, the spatial exogenous-EMGP reveals the causes of spatial dependence in terms of the significance of lagged variables. Given the superior performance of the spatial exogenous-EMGP model, following discussion based on it.

Fig. 2 shows the mean distance-decay curve for all segments depicted by the spatial exogenous-EMGP to illustrate the spatial dependences, $\varphi(\dot{X})$ and $\varphi(\dot{Z})$. The result tells that the $\varphi(\dot{Z})$ values of PDO crashes (bold dotted line) decrease from

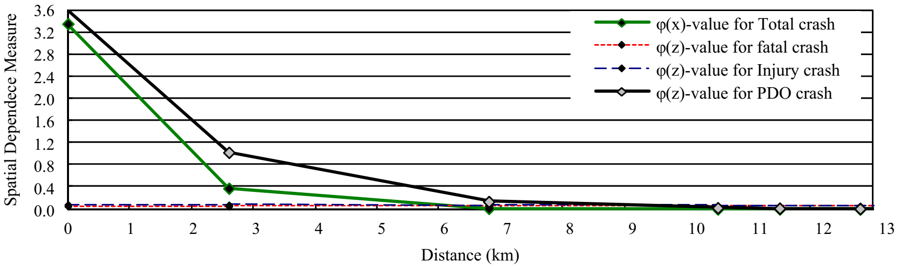

Fig. 2. Mean distance-decay curve for each segments.

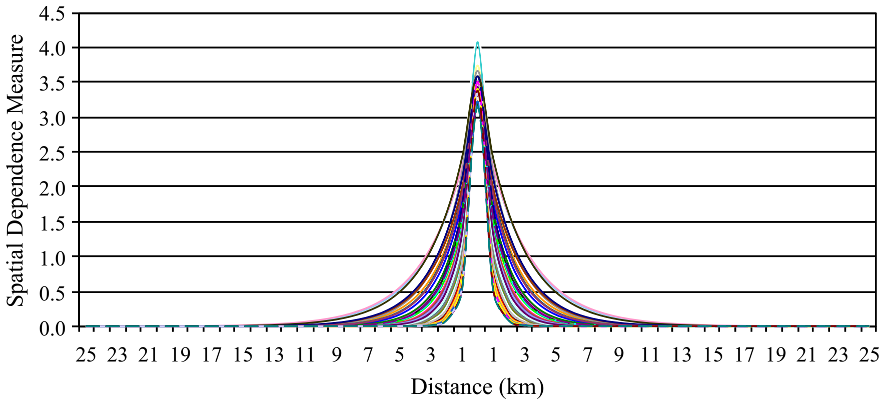

Fig. 3. Spatial dependence measure of crash counts.

3.6 at 0 km to 0.2 at 7 km while the $\varphi(\dot{X})$ values decreases from 3.4 at 0 km to 0.01 at 7 km. The $\varphi(\dot{Z})$ values for the other two severity levels are relatively low (*i.e.*, less than 0.1, lower two dotted lines). Thus, random effects on the highest severity levels are likely. The PDO accidents of lower severity are less affected by chance due to their high occurrence frequencies. Thus, the spatial relationship is relatively easy to interpret. Fig. 2 also shows the curve for total crash (accident) frequency. The patterns resemble those in the curve for PDO accidents. It is anticipated that about 95% crashes are of the PDO accidents. For this reason, the pattern for the impact of spatial-dependence on crash frequency is apparently the weighted sum of various crash severity levels.

Since the realized values of lagged covariates (*i.e.*, geometric and vehicle composite variables) vary, the spatial dependence measure of 62 road segments also vary (Fig. 3). In some road segments, the spatial dependence measure is greatly impacted by the adjacent road segment. Therefore, the $\varphi(\dot{X})$ values are relatively high, and impact convergence is relatively long (>20 km) in some road segments. Restated, the spatial dependence on each segment is heterogeneous. Generally, short segments with high crash frequency tend to have high spatial dependence.

In sum, the spatial dependences sharply decrease when the neighboring segment is more than 7 km away, which suggests that spatial dependence only occurs in adjacent segments. This finding in consistence with the spatial error-EMGP model with spatial error of SARMA(1,1). The study freeway has 19 segments longer than 7 km and 43 segments shorter than 7 km. Notably, the segments with the nine highest crash frequencies are all shorter than 7 km. It is because that short segments (formed by two close interchanges) tend to present frequent traffic weaving behaviors resulting in a high crash potential. Safety measures that should be considered include avoiding designs with short segments and strictly enforcing traffic control strategies within short segments.

5. Conclusions

Based on the EMGP model proposed by Chiou and Fu (2013), this study proposes two alternative spatial dependence structures to accommodate spatial dependences: the spatial error-EMGP, which considers a preset spatial error structure and the spatial exogenous-EMGP, which incorporates spatial exogenous functions associated with the lagged explanatory variables related to traffic and geometric conditions. Performance was evaluated using the same dataset and the same determinants applied in Chiou and Fu (2013). The estimation results show that the spatial exogenous-EMGP model is superior in terms of BIC, RMSE and likelihood ratio-test, which suggests the need to consider spatial dependencies when modeling crash frequency and severity jointly. The comparisons also showed that the estimated coefficients of the spatial-EMGP models tended to be smaller than those estimated by the EMGP model, which suggests that overestimation may be problematic in the EMGP model without spatial dependence. Notably, the spatial dependence effect was larger in PDO accidents compared to the other two severity levels and was larger in crash frequency than in crash severity.

Two main drawbacks of the spatial error-EMGP are noted. Firstly, as in time series models, the spatial error structures must be preset either subjectively or empirically according to the autocorrelation function or partial autocorrelation function. Here, comparisons of different spatial error structures of the spatial error-EMGP model showed that the SARMA (1,1) error structure fits best. Another limitation is the difficulty to identify the source of spatial dependence.

In this study, the spatial exogenous-EMGP is preferable for modeling spatial dependence because of its better explanatory capability. Its estimation results reveal the different spatial dependence effects in each segment. In a word, the spatial dependencies sharply decreases at distances exceeding 7 km, and shorter segments with high crash frequency tend to have high spatial dependence, which suggests that the study network should be carefully segmented. The best solutions are avoiding designs with short segments and strictly enforcing traffic control strategies within the short segments.

To avoid model estimation complexity just as the proposed spatial EMGP models, it is suggested to optimally divide the study network to minimize the potential spatial dependence in the first place. Secondly, this study did not consider how different spatial matrix specifications affect the spatial error-EMGP, which requires further study. Finally, further studies are needed to develop techniques for rigorous analysis of pre-processing techniques for lagged variables in spatial exogenous functions of the spatial exogenous-EMGP.

Acknowledgments

The authors are indebted to anonymous reviewers for their insightful comments and constructive suggestions, which help clarify several points made in the original manuscript. This manuscript is moderately revised from an early work (Chiou et al., 2013a) presented at the 4th International Conference on Road Safety and Simulation. The study was partially sponsored by the National Science Council, Taiwan (NSC 101-2628-E-009-018-MY3).

References

Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis and Prevention 38 (3), 618–625.
Aguero-Valverde, J., 2013. Multivariate spatial models of excess crash frequency at area level: Case of Costa Rica. Accident Analysis and Prevention 59, 365–373.
Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic, The Netherlands.
Anselin, L., 2002. Under the hood issues in specification and interpretation of spatial regression models. Agricultural Economics 27, 247–267.
Aptech Systems, 2008. GAUSS Version 9.0 User Guide, Aptech Systems, Inc.
Aptech Systems, 2012. Online Official web site: 〈http://www.aptech.com/〉 (accessed October, 2013).
Bhat, C.R., Paleti, R., Singh P., 2012. A spatial multivariate count model for firm location decisions (Technical paper), Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin. Retrieved from: 〈http://www.caee.utexas.edu/prof/bhat/PAPERS.htm〉.
Castro, M., Paleti, R., Bhat, C.R., 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: application to predicting crash frequency at intersections. Transportation Research Part B 46 (1), 253–272.
Castro, M., Paleti, R., Bhat, C.R., 2013. A spatial generalized ordered response model to examine highway crash injury severity. Accident Analysis and Prevention 52, 188–203.
Chiou, Y.C., Fu, C., 2013. Modeling crash frequency and severity using multinomial generalized Poisson model with error components. Accident Analysis and Prevention 50, 73–82.
Chiou, Y.C., Fu, C., Hsieh, C.W., 2013a. Incorporating spatial dependence in simultaneously modeling crash frequency and severity. In: Proceedings of the 4th International Conference on Road Safety and Simulation, October 23–25, Rome, Italy.
Delmelle, E.C., Thill, J.-C., Ha, H.-H., 2012. Spatial epidemiologic analysis of relative collision risk factors among urban bicyclists and pedestrians. Transportation 39 (2), 433–448.
Dissanayake, D., Aryaijab, J., Wedagamac, P., 2009. Modeling the effects of land use and temporal factors on child pedestrian casualties. Accident Analysis and Prevention 41 (4), 1016–1024.
Eckey, H.F., Kosfeld, R., Türck, M., 2007. Regional convergence in Germany: a geographically weighted regression approach. Spatial Economic Analysis 2 (1), 45–64.
Famoye, F., Wulu, J.T., Singh, K.P., 2004. On the generalized Poisson regression model with an application to accident data. Journal of Data Science 2, 287–295.
Guo, F., Wang, X., Abdel-Aty, M., 2010. Modeling signalized intersection safety with corridor spatial correlations. Accident Analysis and Prevention 42 (1), 84–92.
Ha, H.H., Thill, J.C., 2011. Analysis of traffic hazard intensity: A spatial epidemiology case study of urban pedestrians. Computers, Environment and Urban Systems 35 (3), 230–240.
Kosfeld, R., Eckey, H.F., Dreger, C., 2006. Regional productivity and income convergence in the unified Germany, 1992-2000. Regional Studies 47, 755–767.
LaScala, E.A., Gerber, D., Gruenewald, P.J., 2000. Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. Accident Analysis and Prevention 32 (5), 651–658.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A 44 (5), 91–305.
Miaou, S.-P., Song, J., Mallick, B., 2003. Roadway traffic crash mapping: a space-time modeling approach. Journal of Transportation and Statistics 6 (1), 33–58.
Miaou, S.-P., Song, J., 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion and spatial dependence. Accident Analysis and Prevention 37 (4), 699–720.
MacNab, Y., 2004. Bayesian spatial and ecological models for small-area crash and injury analysis. Accident Analysis and Prevention 36 (6), 1019–1028.
Maher, M.J., 1990. A bivariate negative binomial model to explain traffic accident migration. Accident Analysis and Prevention 22 (5), 487–498.
Milton, J., Shankar, V., Mannering, F., 2008. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accident Analysis and Prevention 40 (1), 260–266.
Mannering, F., Bhat, C., 2014. Analytic methods in accident research: methodological frontier and future directions. Analytic Methods in Accident Research 1, 1–22.

Mohammadian, A., Haider, M., Kanaroglou, P., 2005. Incorporating observed spatial dependencies in random parameter discrete choice models. In: Proceedings of 84th Annual Meeting of the Transportation Research Board, Washington, D.C., U.S.
Narayanamoorthy, S., Paleti, R., Bhat, C.R., 2013. On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. Transportation Research Part B 55, 245–264.
Quddus, M.A., 2008. Modeling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. Accident Analysis and Prevention 40, 1486–1497.
Savolainen, P., Mannering, F., Lord, D., Quddus, M., 2011. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. Accident Analysis and Prevention 43 (5), 1666–1676.
Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accident Analysis and Prevention 45, 382–391.
Sidharthan, R., Bhat, C.R., 2012. Incorporating spatial dynamics and temporal dependency in land use change models. Geographical Analysis 44 (4), 321–349.
Song, J.J., Ghosh, M., Miaou, S., Mallick, B., 2006. Bayesian multivariate spatial models for roadway traffic crash mapping. Journal of Multivariate Analysis 97 (1), 246–273.
Terza, J.V., Wilson, P.W., 1990. Analyzing frequencies of several types of events: a mixed multinomial-Poisson approach. The Review of Economics and Statistics 72 (1), 108–115.
Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge, UK.
Van der Kruk. R., 2002. A General Spatial ARMA Model: Theory and Application. Conference Paper of ERSA (Europian Regional Science Association, Dortmund, Germany), pp. 110–131.
Wang, C., Quddus, M.A., Ison, S.G., 2009. Impact of traffic congestion on road accidents: a spatial analysis of the M25 motorway in England. Accident Analysis and Prevention 41 (4), 798–808.
Wang, C., Quddus, M.A., Ison, S.G., 2011. Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. Accident Analysis and Prevention 43 (6), 1979–1990.
Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis and Prevention 38 (6), 1137–1150.
Wang, Y., Kockelman, K.M., 2013. A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. Accident Analysis and Prevention 60 (1), 71–84.
Ye, X., Pendyala, R.M., Washington, S.P., Konduri, K., Oh, J., 2009. A simultaneous equations model of crash frequency by collision type for rural intersections. Safety Science 47 (3), 443–452.
Ye, X., Pendyala, R.M., Shankar, V., Konduri, K.C., 2013. A simultaneous equations model of crash frequency by severity level for freeway sections. Accident Analysis and Prevention 57 (1), 140–149.