Spatiotemporal and random parameter panel data models of traffic crash fatalities in Vietnam

Article *in* Accident Analysis & Prevention · May 2016

DOI: 10.1016/j.aap.2016.05.028

3 authors, including:

Long Truong
La Trobe University

Minh Le Kieu
University of Auckland

Spatiotemporal and random parameter panel data models of traffic crash fatalities in Vietnam

Long T. Truong[1,2*], Le-Minh Kieu[3], Tuan A. Vu[4]

[1]Institute of Transport Studies, Department of Civil Engineering, Monash University, Melbourne, Australia

[2]Directorate for Roads of Vietnam, Hanoi, Vietnam

[3]Smart Transport Research Centre, School of Civil Engineering and Build Environment, Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia

[4]Vietnamese-German Transport Research Centre, Vietnamese-German University, Binhduong, Vietnam

[*] Corresponding author: Email: long.truong@monash.edu, Phone: +61 3 9905 1851, Fax: +61 3 9905 4944,

Abstract

This paper investigates factors associated with traffic crash fatalities in 63 provinces of Vietnam during the period from 2012 to 2014. Random effect negative binomial (RENB) and random parameter negative binomial (RPNB) panel data models are adopted to consider spatial heterogeneity across provinces. In addition, a spatiotemporal model with conditional autoregressive priors (ST-CAR) is utilised to account for spatiotemporal autocorrelation in the data. The statistical comparison indicates the ST-CAR model outperforms the RENB and RPNB models. Estimation results provide several significant findings. For example, traffic crash fatalities tend to be higher in provinces with greater numbers of level crossings. Passenger distance travelled and road lengths are also positively associated with fatalities. However, hospital densities are negatively associated with fatalities. The safety impact of the national highway 1A, the main transport corridor of the country, is also highlighted.

Keywords: crash, province-level, random parameter, negative binomial, spatiotemporal, conditional autoregressive

Please cite this article as: Truong, L.T., Kieu, L-M., Vu, T.A., 2016. Spatiotemporal and random parameter panel data models of traffic crash fatalities in Vietnam. Accident Analysis & Prevention 94, 153–161. http://dx.doi.org/10.1016/j.aap.2016.05.028

1. Introduction

Similar to other developing countries, rapid economic growth in Vietnam has been accompanied by an enormous increase in motorisation and a high level of traffic crashes. Between 2006 and 2014, the number of motorcycles increased with an annual growth rate of 16.1% whereas the number of cars increased with an annual growth rate of 22.6% (JICA, 2009; NTSC, 2015). By 2014, there were 43.3 million registered vehicles, of which 94.6% were motorcycles. Unsurprisingly, the number of traffic crashes increased significantly from 9,470 in 1992 to 27,993 in 2002 while the number of fatalities increased from 3,077 in 1992 to 13,186 in 2002 (NTSC, 2015). Fortunately, since 2007, there has been a small, but steady reduction in the number of fatalities, which would be attributable to stronger traffic safety programs and measures implemented by authorities (Passmore et al., 2010; Ngo et al., 2012; Nguyen et al., 2013a; Nguyen et al., 2013b). In 2014, Vietnam had 25,322 reported traffic crashes and 8,996 fatalities (NTSC, 2015). It is not surprised that motorcycles accounted for around 70% of traffic crashes (NTSC, 2015; Truong et al., 2016). The World Health Organisation (WHO) estimated that the traffic fatality rate in Vietnam was nearly 24.5 per 100,000 population, which is 44% higher than the average fatality rate in South East Asia (WHO, 2015).

Traffic crashes are one of the leading causes of deaths and disabilities in Vietnam (Nguyen et al., 2012; Tran et al., 2012). In addition, their economic impact is profound. It was estimated that the cost of traffic crashes is between 2.5% and 2.9% of the country's gross domestic product (GDP) (ADB, 2005; JICA, 2009). Traffic crashes can also cause a significant economic burden at individual and family levels. For example, a study in Thaibinh province found that the average cost of a traffic injury during hospitalisation was greater than 6 months' average salary (Nguyen et al., 2013a).

Forecasting safety impacts of alternative transport planning schemes is essential for proactive safety planning. During the last decade, there has been a growing body of research on macro-level safety analyses. In macro-level safety studies, safety performance measures, e.g. crash counts, spatially aggregated at a certain spatial unit are modelled against area-wide variables. A wide range of spatial units have been investigated, e.g. block groups (Dumbaugh and Rae, 2009), grid structure (Kim et al., 2006), census tracts (Wier et al., 2009), wards (Noland and Quddus, 2004), cantons (Aguero-Valverde, 2013), counties (Traynor, 2008; Huang et al., 2010), provinces (Erdogan, 2009; Tolón-Becerra et al., 2012), cities (Moeinaddini et al., 2014; Coruh et al., 2015), multiple provinces (Torre et al., 2007), states (Noland, 2003), countries (Kumara and Chin, 2004), and traffic analysis zones (TAZs) (Ng et al., 2002; Hadayeghi et al., 2003; Lovegrove and Sayed, 2006; Pirdavani et al., 2012; Wang et al., 2013). Effects of spatial units on modelling results have been discussed in few studies (Abdel-Aty et al., 2013; Lee et al., 2014b; Xu et al., 2014).

A variety of area-wide variables have been considered in previous macro-level safety analyses: socioeconomic variables, e.g. population density (Hadayeghi et al., 2003; Noland and Quddus, 2004; Huang et al., 2010; Tolón-Becerra et al., 2012; Lee et al., 2014b), age groups (Noland, 2003; Noland and Oh, 2004; Noland and Quddus, 2004; Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Huang et al., 2010; Aguero-Valverde, 2013; Lee et al., 2014a), income (Noland, 2003; Traynor, 2008; Pirdavani et al., 2012), GDPs (Kumara and Chin, 2004; Tolón-Becerra et al., 2012), and employment (Siddiqui et al., 2012); land use variables (Ng et al., 2002; Lovegrove and Sayed, 2006; Pulugurtha et al., 2013; Wang et al., 2013; Lee et al., 2014b); healthcare variables (Ng et al., 2002; Coruh et al., 2015); road infrastructure variables, e.g. road density, intersection density, road length (Amoros et al., 2003; Hadayeghi et al., 2003; Noland, 2003; Lovegrove and Sayed, 2006; Pirdavani et al., 2012; Tolón-Becerra et al., 2012; Jiang et al., 2016), roads with different functions (Lovegrove and Sayed, 2006; Huang et al., 2010), and road network structures (Wang et al., 2013; Moeinaddini et al., 2014); traffic pattern variables, e.g. vehicle kilometres travelled (VKT) (Dumbaugh and Rae, 2009; Abdel-Aty et al., 2013; Aguero-Valverde, 2013), highway usage (Traynor, 2008), traffic volume (Quddus, 2008; Wier et al., 2009), speed (Quddus, 2008), volume to capacity ratios (Hadayeghi et al., 2003), and trip generation and distribution (Abdel-Aty et al., 2011); and environmental variables, e.g. rainfall (Coruh et al., 2015) and snowfall (Aguero-Valverde and Jovanis, 2006).

Spatial effects, i.e. spatial dependence or correlation and spatial heterogeneity (Anselin, 1988), have been considered in macro-level safety analyses. For example, Bayesian spatial approaches have been used to account for possible spatial correlation between areas (Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Siddiqui et al., 2012; Wang et al., 2013; Xu et al., 2014; Dong et al., 2015; Lee et al., 2015; Song et al., 2015; Siddiqui and Watkins, 2016). To consider spatial heterogeneity, previous macro-level safety studies have adopted the geographically weighted regression (GWR) models (Hadayeghi et al., 2003; Erdogan, 2009; Hadayeghi et al., 2010; Li et al., 2013) and random parameter models (Coruh et al., 2015; Xu and Huang, 2015). Space-time interaction has also been considered by Aguero-Valverde and Jovanis (2006).

A number of studies have focused on macro-level safety analyses, which however is mainly Western-based. There is a need to understand the safety effects of area-wide characteristics in the context of developing countries, including South East Asia countries and Vietnam specifically. In addition, the issue of level crossings are not considered in existing macro-level safety studies. In Vietnam, most of railway crashes occurred at level crossings (JICA, 2009). Level crossing fatalities represent about 1.5% of all traffic fatalities in Vietnam (NTSC, 2015), which is higher when compared to Australia and the European Union with 0.5% and 1.3% respectively (ITSR, 2011; ERA, 2014). Note that Vietnam has one of the densest level crossing systems and the lowest percentages of protected level crossings in Asia (ESCAP, 2000). This paper investigates the relationships between various area-wide variables and traffic crash fatalities in Vietnam at the province level. Random effect and random parameter negative binomial panel data models are utilised to account for spatial heterogeneity. In addition, a conditional autoregressive (CAR) model is utilised to account for spatiotemporal autocorrelation. The number of level crossings in each province is included in the analysis. A better understanding of safety effects of area-wide variables is critical to safety planning and policy in Vietnam.

2. Data

The number of traffic crash fatalities in 63 provinces from 2012 to 2014, was obtained from the National Transportation Safety Committee. Level crossing data were collected from the Vietnam Railway Administration (VNRA, 2015). Road network data, consisting of lengths of national highways, provincial roads, district roads, commune roads, and urban roads, were obtained from the Directorate for Roads of Vietnam. Expressways are not considered in this study due to the limited length of the current network, i.e. only 700km of the planned 6,000km expressway network have been in operation recently (VEA, 2015). A dummy variable is added to investigate the safety effect of the national highway 1A, which is the main transport corridor of the country with over 2,200km long travelling through 30 provinces (DRVN, 2015). Note that level crossing and road length data were only reported in 2015. This study reasonably assumes that changes in the number of level crossings and road lengths from 2012 to 2015 were minor and could be ignored.

Socio-demographic panel data, e.g. population, population density, residential area ratio, and passenger distance travelled (PDT), in each province from 2012 to 2014 were downloaded from the website of the General Office of Statistics of Vietnam. In addition, medical-related panel data (e.g. the number of hospitals) were also downloaded, which is then used to calculate the hospital density in each province.

Fig. 1a shows the number of traffic crash fatalities by province in 2014. Hanoi and Ho Chi Minh City, two major cities, had the highest numbers of traffic crash fatalities. It is clear that high values were clustered around Ho Chi Minh City. Moreover, results of the Moran's I test (Moran, 1950; Truong and Somenahalli, 2011) indicated evidence of spatial autocorrelation with the Moran's I statistic of 0.28 and p-value<0.001. Fatality rates by million passenger km travelled by province in 2014 are presented in Fig. 1b. It can be seen that high values were clustered in north-west provinces, which were in mountainous areas and had low PDTs. Spatial autocorrelation was also evident with the Moran's I statistic of 0.16 and p-value<0.01.

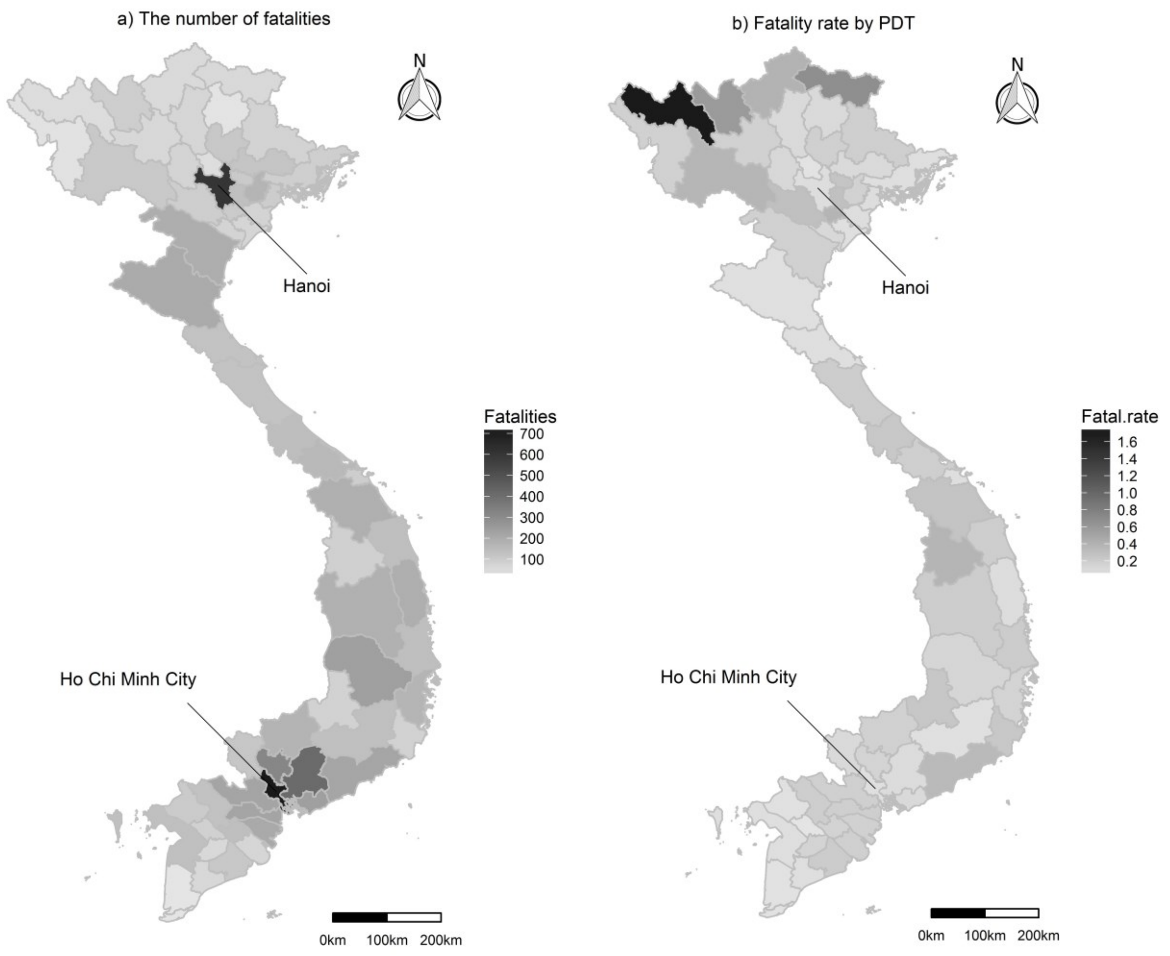

Fig. 1 The number of traffic crash fatalities and fatality rate by million passenger km travelled by province in 2014

Descriptive statistics of variables are presented in Table 1. Two dummy variables, i.e. year 2013 and year 2014, were included to compare with the base year (2012). The number of traffic crash fatalities was selected as the dependent variable while other variables were considered as independent variables. Multicollinearity was checked by calculating the variance inflation factors (VIFs) for all independent variables. As a common rule of thumb, a VIF value of larger than 5 indicates high multicollinearity. As a result, population, population density, and residential area ratio were omitted. VIF values of ten variables included in the models are presented in Table 1.

Table 1 Descriptive statistics of variables

Variables	Notations	Mean	Std.	Min	Max	VIF
Traffic crash fatalities		148.63	123.82	18.00	787.00	
Population (1000 persons)		1,424.86	1,236.05	303.00	7,981.90	
Population density (persons per km2)		476.06	578.67	43.80	3,809.00	
Residential area ratio (%)		3.37	3.07	0.40	12.20	
Passenger distance travelled (million passenger km)	PDT	1,392.40	2,033.27	22.90	13,137.90	4.37
Hospital density (number of hospitals per 1000km^2)	HDEN	5.07	4.92	0.99	24.81	1.90
Length of national highways (km)	LNH	337.11	225.40	0.00	1,146.00	2.18
Length of provincial roads (km)	LPR	423.17	200.55	0.00	999.00	1.51
Length of district and commune roads (km)	LDCR	3,688.34	2,431.18	110.74	14,375.20	1.77
Length of urban roads (km)	LUR	339.90	378.63	42.34	2,320.65	3.52
On national highway 1A (dummy)	NH1A	0.48		0.00	1.00	1.45
Number of level crossings	LX	23.75	32.19	0.00	181.00	1.81
Year 2013 (dummy)	Y13	0.33		0.00	1.00	1.34
Year 2014 (dummy)	Y14	0.33		0.00	1.00	1.34

Table 2 presents correlation coefficients for variables selected for modelling. There was a positive correlation between passenger distance travelled and length of urban roads (correlation coefficient = 0.797). However, the VIF results suggested no evidence of multicollinearity, in which, passenger distance travelled had the highest VIF value of 4.37, followed by length of urban roads with a VIF value of 3.52. Nevertheless, models with and without passenger distance travelled were compared in the following analysis.

Table 2 Pearson correlation matrix for variables selected for modelling

	PDT	HDEN	LNH	LPR	LDCR	LUR	NH1A	LX	Y13	Y14
PDT	1.00									
HDEN	0.49	1.00								
LNH	-0.17	-0.50	1.00							
LPR	-0.10	-0.36	0.53	1.00						
LDCR	0.28	-0.23	0.45	0.24	1.00					
LUR	0.797	0.29	0.11	0.12	0.35	1.00				
NH1A	0.24	0.09	0.03	0.01	0.03	0.12	1.00			
LX	0.39	0.11	0.19	-0.02	0.38	0.38	0.48	1.00		
Y13	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	1.00	
Y14	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.50	1.00

3. Models

3.1 Random effect and random parameter negative binomial models

Since data contain province-specific characteristics and annual traffic crash fatalities in each province are likely to be correlated, panel count models should be employed. To account for spatial heterogeneity, random effect negative binomial (RENB) and random parameter negative binomial (RPNB) models have been applied in recent macro-level safety studies (Coruh et al., 2015; Xu and Huang, 2015). Therefore, RENB and RPNB models for panel data were used in this paper.

The negative binomial (NB) model has been widely used for crash frequency analysis, particularly for over-dispersed data (Lord and Mannering, 2010; Washington et al., 2011). Let y_{it} denote the observed number of traffic crash fatalities in province i and year t, X_{itk} is the k^{th} variable for province i and year (time period) t, β_k is the coefficient to be estimated, p is the number of variables, n is the number of provinces (zones), and T is the number of years (time periods). The NB model is derived by assuming:

$$y_{it} \sim Poisson(\lambda_{it}) \quad (1)$$

$$\lambda_{it} = exp(\beta_0 + \sum_{k=1}^{p} \beta_k X_{itk} + \varepsilon_{it}) \quad (2)$$

where λ_{it} is the Poisson parameter, which is the expected number of fatalities in province i and year t, $exp(\varepsilon_{it})$ is a gamma-distributed error term with mean one and variance α. With the addition of this term, the variance can be different to the mean as $VAR(y_{it}) = \lambda_{it} + \alpha\lambda_{it}{}^2$.

To account for heterogeneity across individuals, e.g. provinces, the RPNB model can be written as:

$$\lambda_{it} = exp\left((\beta_0 + \omega_{i0}) + \sum_{k=1}^{p}(\beta_k + \omega_{ik})X_{itk} + \varepsilon_{it}\right) \quad (3)$$

where ω_{ik} is a randomly distributed term, e.g. a normally distributed term with mean zero and variance $\sigma_k{}^2$ (Washington et al., 2011; Greene, 2012). In practice, a random parameter will be used if its standard deviation is significantly larger than zero; otherwise, the parameter is fixed across individuals.

It is noteworthy that a RENB model is equivalent to a RPNB model with the intercept term being the only random parameter (Anastasopoulos and Mannering, 2009; Washington et al., 2011; Chen and Tarko, 2014). In this paper, RENB and RPNB models are estimated using NLOGIT 5 (Econometric Software, 2012).

To compare models, the Likelihood Ratio (LR) test is used. The LR test statistic is calculated as:

$$LR = -2[LL_0 - LL_1] \quad (4)$$

where LL_0 and LL_1 are the log likelihood at convergence for null and alternative models respectively. This statistic is χ^2-distributed with degrees of freedom equal to the difference in the numbers of parameters between these models.

3.2 Spatiotemporal model with conditional autoregressive priors (ST-CAR)

In a previous study, Aguero-Valverde and Jovanis (2006) extended the model proposed by Bernardinelli et al. (1995), which specified a space-time interaction term, to include covariates. However, these models strictly assume linear temporal trends. In this paper, the model proposed by Rushworth et al. (2014) is utilised to account for spatiotemporal autocorrelation:

$$y_{it} \sim Poisson(\lambda_{it}) \tag{5}$$

$$\lambda_{it} = exp(\beta_0 + \sum_{k=1}^{p} \beta_k X_{itk} + \phi_{it}) \tag{6}$$

$$\phi_{i1}|\phi_{-i1} \sim N\left(\frac{\rho_S \sum_{j=1}^{n} w_{ij}\phi_{j1}}{\rho_S \sum_{j=1}^{n} w_{ij}+1-\rho_S}, \frac{\tau^2}{\rho_S \sum_{j=1}^{n} w_{ij}+1-\rho_S}\right) \tag{7}$$

$$\phi_t|\phi_{t-1} \sim N(\rho_T \phi_{t-1}, \tau^2 Q(W, \rho_S)^{-1}) \quad t = 2, \ldots, T \tag{8}$$

where ϕ_{it} are random effects that account for residual spatiotemporal autocorrelation in the data after the effects of covariates have been removed, ϕ_{-i1} is the vector of random effects for time period 1 except for ϕ_{i1}, ϕ_t is the vector of random effects for time period t, $W = \{w_{ij}\}$ is the $n \times n$ adjacent matrix (w_{ij}=1 if provinces i and j are adjacent or 0 otherwise), ρ_S is the spatial parameter, ρ_T is the temporal parameter, and τ^2 is the parameter controlling the variance of random effects. The precision matrix $Q(W, \rho_S)$ corresponds to the conditional autoregressive (CAR) prior proposed by Leroux et al. (2000) and is given by $Q(W, \rho_S) = \rho_S(diag(W\mathbf{1}) - W) + (1 - \rho_S)I$, where $\mathbf{1}$ is the $n \times 1$ vector of ones, I is the $n \times n$ identity matrix.

Eq. (7) corresponds to the intrinsic CAR prior (Besag et al., 1991) for strong spatial correlation if ρ_S=1. On the contrary, ρ_S=0 suggests independent random effects with constant mean and variance. In Eq. (8), spatial and temporal autocorrelation are induced by the variance and mean respectively. Strong temporal autocorrelation is suggested by ρ_T=1 whereas temporal independence is indicated by ρ_T=0.

Model parameters were estimated in a Bayesian setting using Markov Chain Monte Carlo (MCMC) simulation. A burn-in period was set as 10,000 iterations. Model estimates are then based on 40,000 samples. Convergence of the model was examined by visual diagnostic and Geweke convergence diagnostic. Deviance Information Criteria (DIC) (Spiegelhalter et al., 2002) was used to provide a measure of model fit. Data analyses were performed using the CARBayesST package in the R Statistical Environment (R Development Core Team, 2015).

3.3 Measures of model prediction performance

To compare model prediction performance, the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) are adopted.

$$MAE = \frac{1}{n_o}\sum_{j=1}^{n_o}\left|O_j - P_j\right| \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n_o}\sum_{j=1}^{n_o}\left(O_j - P_j\right)^2} \tag{10}$$

$$MAPE = \frac{1}{n_o}\sum_{j=1}^{n_o}\left|\frac{O_j - P_j}{O_j}\right| \tag{11}$$

Where O_j is the observed value, P_j is the predicted value from the model, and n_o is the number of observations.

4. Results

4.1 RENB and RPNB models

Given the correlation between PDT and length of urban roads, four models, including random effect with PDT (RENB-1), random effect without PDT (RENB-2), random parameter with PDT (RPNB-1), and random parameter without PDT (RPNB-2), were compared. Estimation results using 200 Halton draws are presented in Table 3.

In the RENB-1 and RENB-2 models, all variables were significant at $p<0.01$, except for the year 2013 variable. The standard deviation of the intercept distribution was significantly different to zero, indicating that modelling the intercept as random parameter was appropriate. Similarly, in the RPNB-1 and RPNB-2 models, all variables were significant at $p<0.05$, apart from the variable for year 2013. In addition, the intercept, length of urban roads, and the number of level crossings resulted in random parameters. Length of provincial roads resulted in a random parameter in the RPNB-1 model, but its standard deviation of parameter distribution was not significantly different to zero in the RPNB-2 model. The signs of parameters were consistent among the models. Dispersion parameters for these models were significantly different to zero, suggesting the use of the negative binomial model over the Poisson model was appropriate.

Similar parameters and associated significance levels between the RENB-1 and RENB-2 models confirmed that there is no multicollinearity when both PDT and length of urban roads were considered. Compared to the RENB-2 model, the RENB-1 model had better log likelihood; however, the LR test was not significant (LR = 2.32, df = 1, p-value = 0.127). Likewise, estimates were consistent between the RPNB-1 and RPNB-2 models, suggesting no multicollinearity when both PDT and length of urban roads were considered. Moreover, the RPNB-1 model was significantly better than the RPNB-2 model as shown by the LR test (LR = 4.67, df = 1, p-value = 0.031) and AIC. In addition, the RENB-1 model had slightly better AIC compared to the RPNB-1 model, which could be attributed to fewer parameters in the RENB-1 model. Although the RPNB-1 model had the best log likelihood, results of the LR test showed it was not significantly better than the RENB-1 model (LR = 5.15, df = 3, p-value = 0.161) or the RENB-2 model (LR = 7.47, df = 4, p-value = 0.112).

Fig. 2 presents estimates of local parameters for variables associated with length of provincial roads and length of urban roads, obtained from the RPNB-1 model. It can be seen that the parameters for these variables varied across provinces. These variations may be attributed to different levels of traffic composition, traffic congestion, enforcement, and safety measures among provinces. Spatial variation patterns were however different between these two variables. Table 3 suggests coefficients of variations of local parameters for length of provincial roads and length of urban roads were 12% (0.00004/0.00034) and 50% (0.00037/0.00074) respectively. Fig. 2b shows that the effect of length of urban roads seemed to be smaller in major cities, i.e. Hanoi and Ho Chi Minh City. This may be due to the fact that urban roads in major cities have better traffic safety systems, e.g. modern signal controls and car/motorcycle lane separation, and are strongly enforced.

Table 3 Estimation results for RENB and RPNB models

	RENB-1		RENB-2		RPNB-1		RPNB-2	
Variables	Estimate	z value	Estimate	z value	Estimate	z value	Estimate	z value
Intercept	4.29200 ***	132.73	4.24300 ***	130.70	4.31572 ***	128.03	4.36099 ***	132.45
Standard deviation of parameter distribution	*0.46823 ****	*47.11*	*0.46752 ****	*47.41*	*0.45708 ****	*44.03*	*0.48287 ****	*45.69*
Passenger distance travelled (million passenger km)	0.00007 ***	8.62			0.00009 ***	10.32		
Hospital density (number of hospitals per 1000km^2)	-0.01320 ***	-5.56	-0.00887 ***	-3.91	-0.01535 ***	-6.42	-0.01425 ***	-6.18
Length of national highways (km)	-0.00055 ***	-10.86	-0.00068 ***	-14.00	-0.00073 ***	-13.64	-0.00077 ***	-15.12
Length of provincial roads (km)	0.00055 ***	10.86	0.00042 ***	8.39	0.00034 ***	6.48	0.00046 ***	8.95
Standard deviation of parameter distribution					*0.00004 ***	*2.07*	*0.00002*	*1.26*
Length of district and commune roads (km)	0.00001 ***	2.81	0.00003 ***	7.95	0.00003 ***	6.30	0.00002 ***	3.76
Length of urban roads (km)	0.00061 ***	15.19	0.00098 ***	40.16	0.00074 ***	17.61	0.00085 ***	33.40
Standard deviation of parameter distribution					*0.00037 ****	*18.43*	*0.00041 ****	*20.30*
On national highway 1A (dummy)	0.14345 ***	6.87	0.17940 ***	8.88	0.09170 ***	4.36	0.11718 ***	5.64
Number of level crossings	0.00561 ***	17.25	0.00439 ***	13.51	0.00706 ***	20.19	0.00634 ***	18.21
Standard deviation of parameter distribution					*0.00042 ***	*2.00*	*0.00103 ****	*5.13*
Year 2013 (dummy)	-0.01278	-0.54	-0.00392	-0.16	-0.01513	-0.67	-0.00275	-0.12
Year 2014 (dummy)	-0.10533 ***	-6.04	-0.09603 ***	-5.60	-0.10785 ***	-6.42	-0.09525 ***	-5.78
Dispersion parameter for negative binomial distribution	217.868 ***	4.82	214.795 ***	4.84	241.085 ***	4.64	235.527 ***	4.71
Number of observations	189		189		189		189	
Log likelihood	-882.29811		-883.22039		-879.72426		-882.07340	
Akaike information criterion (AIC)	1790.6		1790.4		1791.4		1794.1	

Note: * p<0.1; ** p<0.05; *** p<0.01

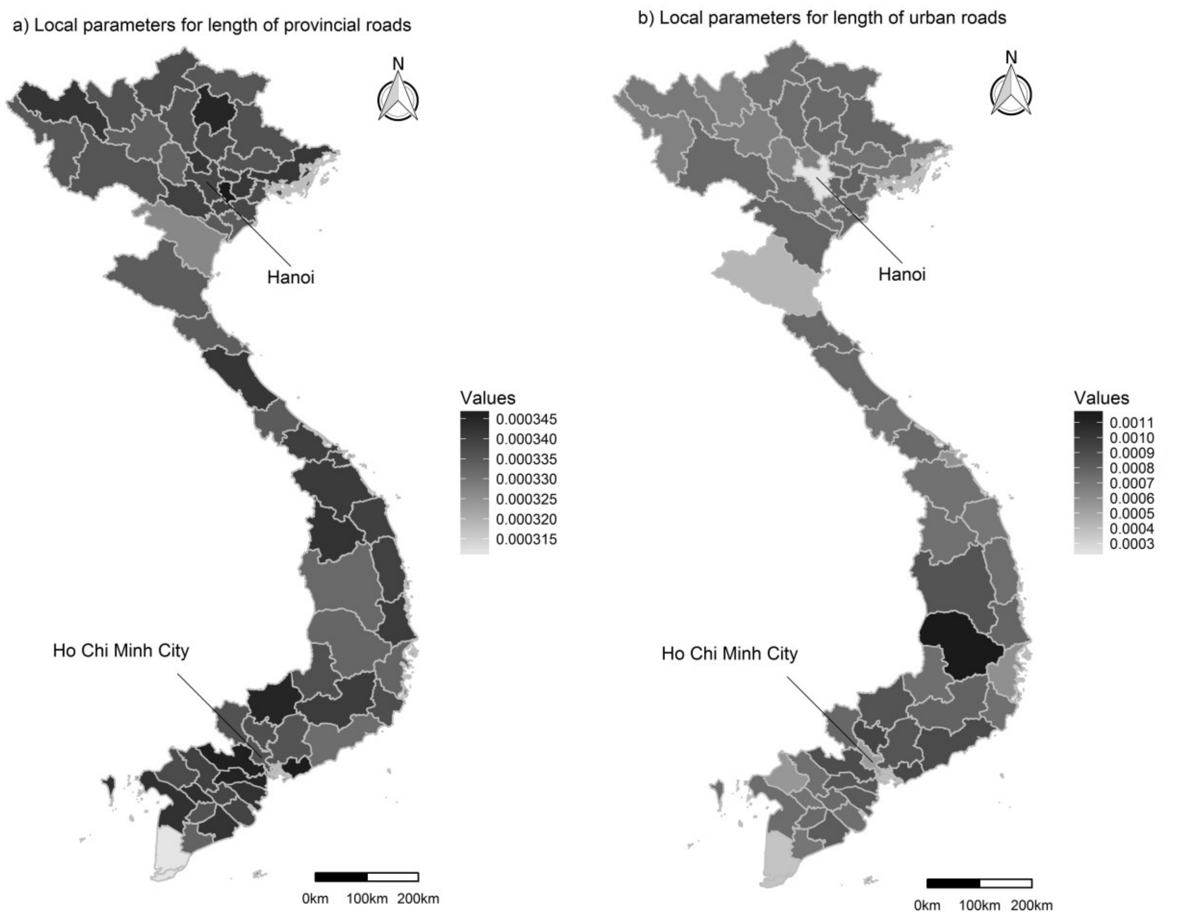

Fig. 2 Parameters obtained from the RPNB-1 model for length of provincial roads and length of urban roads by province

4.2 ST-CAR model

Estimation results for the ST-CAR model are presented in Table 4. Note that the year 2013 and year 2014 variables were excluded since temporal autocorrelation was already considered in the ST-CAR model. It is clear that 95% Bayesian credible intervals (BCIs) of all parameters had the same sign (or did not contain zero). As the variance parameter τ^2 was 0.16 (95%BCI: 0.12 – 0.21), spatial dependence parameter ρ_S was 0.55 (95%BCI: 0.31 – 0.79), and temporal parameter ρ_T was 0.95 (95%BCI: 0.86 – 0.998), the spatiotemporal autocorrelation in the data was evident. In general, the high temporal parameter is consistent with a decreasing trend of fatalities during the three-year study period suggested by the RENB and RPNB models. For example, in the RENB and RPNB models, both year 2013 and year 2014 variables resulted in negative coefficients where the year 2014 variable was significant at p<0.01. The effect of including the temporal parameter was further examined by estimating a model with spatial autocorrelation only (ρ_T was set as 0). The DIC of the model with spatial autocorrelation only was considerably larger than that of the ST-CAR model with spatiotemporal autocorrelation (1650.892 versus 1586.463). This suggests the ST-CAR model was the better model and considering spatiotemporal autocorrelation improved model fit. In addition, the signs of parameters in the ST-CAR model were in accordance with those in the RENB and RPNB models.

Table 4 Estimation results for the ST-CAR model

Variables	Mean	s.d.	Bayesian credible interval	
			2.5%	97.5%
Intercept	4.22784	0.06447	4.10487	4.36023
Passenger distance travelled (million passenger km)	0.00007	0.00001	0.00005	0.00010
Hospital density (number of hospitals per 1000km^2)	-0.00308	0.00065	-0.00435	-0.00178
Length of national highways (km)	-0.00034	0.00012	-0.00059	-0.00013
Length of provincial roads (km)	0.00024	0.00011	0.00002	0.00045
Length of district and commune roads (km)	0.00004	0.00001	0.00002	0.00005
Length of urban roads (km)	0.00042	0.00007	0.00030	0.00056
On national highway 1A (dummy)	0.24537	0.02983	0.18438	0.29819
Number of level crossings	0.00259	0.00071	0.00115	0.00397
τ^2	0.15937	0.02350	0.11801	0.21046
ρ_S	0.55164	0.12338	0.31036	0.78930
ρ_T	0.95150	0.03766	0.85880	0.99827
DIC	1586.463			

4.3 Model comparison

A summary of model comparison statistics is shown in Table 5. NB is the fixed parameter negative binomial model. All RENB and RPNB models provided similar MAE, RMSE, and MAPE, which were significantly better than those in the NB model. This suggests the consideration of random effects or random parameters improved model fit considerably. It is clear that the ST-CAR model, which accounted for spatiotemporal autocorrelation, outperformed other models as its MAE, RMSE, and MAPE were significantly lower. The performance of the ST-CAR model is further illustrated in Fig. 3 that compares predicted versus observed values by NB, RENB-1, RPNB-1, and ST-CAR models. In short, the ST-CAR model was the most favourable model.

Table 5 Model comparison statistics

Statistics	NB	RENB-1	RENB-2	RPNB-1	RPNB-2	ST-CAR
MAE	51.96	9.57	9.55	9.56	9.54	3.23
RMSE	78.78	13.56	13.68	13.56	13.69	4.53
MAPE	46.63%	7.44%	7.45%	7.43%	7.42%	2.74%

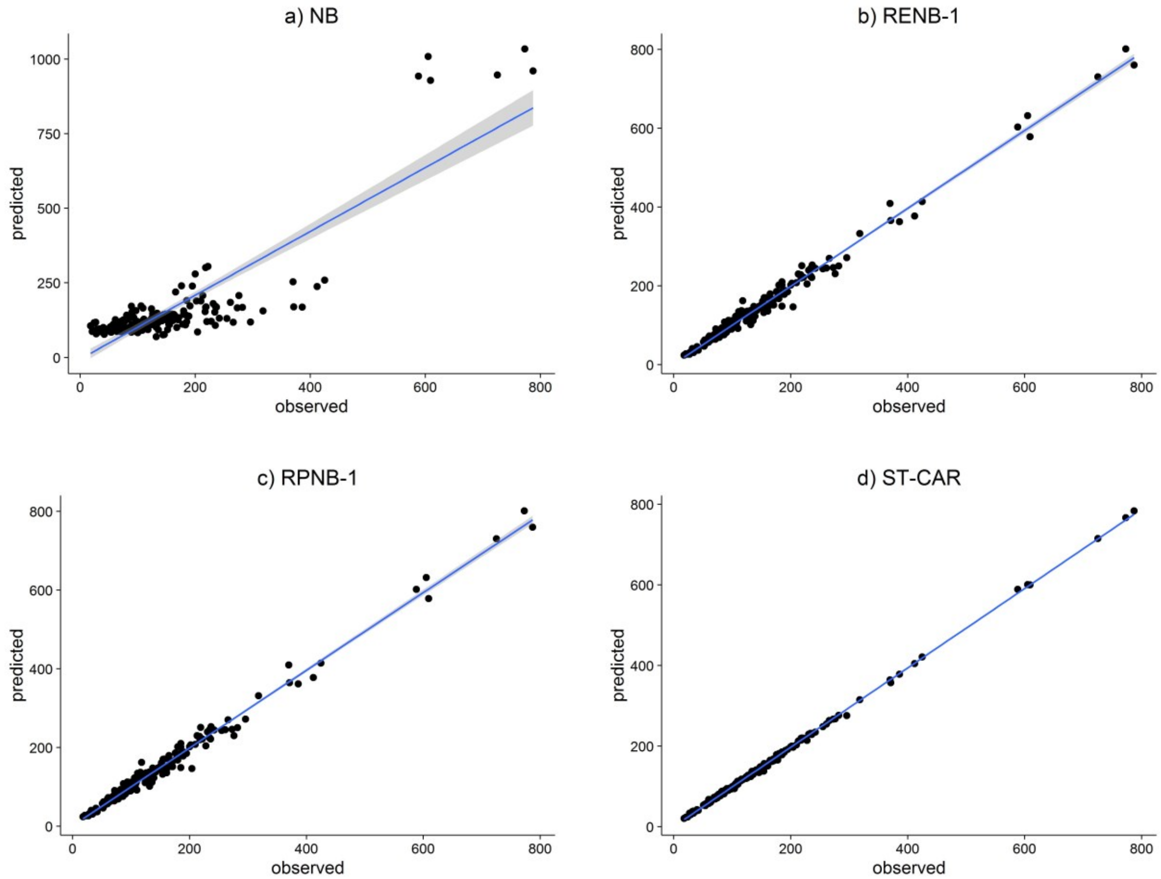

Fig. 3 Predicted versus observed traffic crash fatalities

5. Discussion

The discussion in this section is focused on estimates of the ST-CAR model. Passenger distance travelled had a positive parameter of 0.00007 (95%BCI: 0.00005 – 0.0001), suggesting greater passenger distance travelled values are associated with increasing traffic fatalities. The positive effect of passenger distance travelled is expected, given that it is considered as a measure of exposure to safety risk. Previous studies also indicated the positive effect of VKT, another exposure measure, on traffic crashes (Dumbaugh and Rae, 2009; Abdel-Aty et al., 2013; Aguero-Valverde, 2013).

Hospital density resulted in a negative parameter of -0.00308 (95%BCI: -0.00435 to -0.00178). This indicates provinces with higher hospital densities are more likely to have lower fatalities. The availability of hospitals related to hospital density is essential to post crash care since a faster access to medical care could prevent deaths. This result is in line with findings of previous research, which found that reduced crash fatalities are associated with higher hospital bed densities (Castillo-Manzano et al., 2013) and the availability of Magnetic Resonance Imaging (MRI) scans (Torre et al., 2007). An implication of this result is that enhancing local emergency medical services, e.g. ambulances and first-aid training, could also contribute to limiting the severity of traffic crashes.

Length of national highways was negatively associated with fatalities with a parameter of -0.00034 (95%BCI: -0.00059 to -0.00013). This should be carefully interpreted since important factors such as intersection density were not included in the model. Previous studies showed that higher intersections densities are associated with increasing crashes (Hadayeghi et al., 2003; Pirdavani et al., 2012). However, provinces in mountainous areas tend to have longer national highways, but lower intersection densities. The impact of length of national highways should be investigated in future work by enhancing the variety of area-wide factors. Nevertheless, fatalities in a province would increase if it contains the national highway 1A, indicated by a positive parameter of 0.24537 (95%BCI: 0.18438 – 0.29819). This is expected given that the national highway 1A is the backbone (north-south) transport corridor with heavy traffic travelling through 30 provinces. This finding supports authorities' ongoing programs to upgrade and expand the national highway 1A.

Length of provincial roads, length of district and commune roads, and length of urban roads resulted in positive parameters of 0.00024 (95%BCI: 0.00002 – 0.00045), 0.00004 (95%BCI: 0.00002 – 0.00005), and 0.00042 (95%BCI: 0.0003 – 0.00056) respectively. These results suggest that increases in lengths of district and commune roads, provincial roads, and urban roads are associated with increases in fatalities, which is in accordance with findings of previous studies (Quddus, 2008; Tolón-Becerra et al., 2012).

Number of level crossings resulted in a positive parameter of 0.00259 (95%BCI: 0.00115 – 0.00397), suggesting that provinces with a greater number of level crossings tend to have more traffic fatalities. This is an important result that should be considered by transport planners and authorities. Countermeasures might include removing level crossings at critical intersections, reducing the number of unprotected level crossings, and enhancing enforcement and safety education.

In the RENB and RPNB models, the year 2014 variable was significant at $p<0.01$ with a negative parameter. Although the variable for year 2013 was not significant, but it also resulted in a negative parameter. In general, a decreasing trend in fatalities during the period from 2012 to 2014 was evident.

6. Conclusion

This paper has explored factors associated with traffic crash fatalities in 63 provinces of Vietnam during the period from 2012 to 2014. The RENB and RPNB panel data models were adopted to consider spatial heterogeneity across provinces, which can arise from observed and unobserved factors. In addition, the ST-CAR model was utilised to account for spatiotemporal autocorrelation in the data. The statistical comparison indicated that the signs of estimated parameters were consistent among these models and the ST-CAR model outperformed the RENB and RPNB models.

Estimation results provide several significant findings. For example, traffic crash fatalities were positively associated with the number of level crossings, passenger distance travelled, length of provincial roads, length of district and commune roads, and length of urban roads. In addition, traffic crash fatalities were positively associated with the presence of the national highway 1A. However, hospital densities were negatively associated with traffic crash fatalities.

Although the models performed well in exploring effects of area-wide factors on traffic crash fatalities, there are several areas that can be improved. For example, other important factors, e.g. urbanisation rate, motorcycle to car ratio, intersection density, and education level, should be investigated. Expressways may also need to be considered in future as the length of the expressway network is growing. Future work should also investigate factors associated with fatal and injury crashes at different spatial levels, e.g. districts and suburbs. Nevertheless, this study gives an important contribution towards understanding safety effects of area-wide factors in the context of South East Asia countries and Vietnam specifically.

Acknowledgements

The authors would like to thank Hoang Anh Tuan, Vu Duc Phuc, Tran Minh Thu, and Cao Hoang Can for their support in data collection. The authors are also grateful to the three anonymous reviewers for their constructive comments to improve the manuscript.

References

Abdel-Aty, M., Lee, J., Siddiqui, C., Choi, K., 2013. Geographical unit based analysis in the context of transportation safety planning. Transportation Research Part A: Policy and Practice 49, 62-75.

Abdel-Aty, M., Siddiqui, C., Huang, H., Wang, X., 2011. Integrating Trip and Roadway Characteristics to Manage Safety in Traffic Analysis Zones. Transportation Research Record: Journal of the Transportation Research Board 2213, 20-28.

ADB, 2005. Arrive Alive: ASEAN Commits to Cutting Road Deaths-Association of Southeast Asian Nations Regional Road Safety Strategy and Action Plan (2005—2010). Asian Development Bank.

Aguero-Valverde, J., 2013. Multivariate spatial models of excess crash frequency at area level: Case of Costa Rica. Accident Analysis & Prevention 59, 365-373.

Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis & Prevention 38(3), 618-625.

Amoros, E., Martin, J.L., Laumon, B., 2003. Comparison of road crashes incidence and severity between some French counties. Accident Analysis & Prevention 35(4), 537-547.

Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis & Prevention 41(1), 153-159.

Anselin, L., 1988. Spatial econometrics: methods and models. Kluwer Academic, Dordrecht.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., Songini, M., 1995. Bayesian analysis of space—time variation in disease risk. Statistics in Medicine 14(21-22), 2433-2443.

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.

Castillo-Manzano, J.I., Castro-Nuño, M., Fageda, X., 2013. Can health public expenditure reduce the tragic consequences of road traffic accidents? The EU-27 experience. The European Journal of Health Economics 15(6), 645-652.

Chen, E., Tarko, A.P., 2014. Modeling safety of highway work zones with random parameters and random effects models. Analytic Methods in Accident Research 1, 86-95.

Coruh, E., Bilgic, A., Tortum, A., 2015. Accident analysis with aggregated data: The random parameters negative binomial panel count data model. Analytic Methods in Accident Research 7, 37-49.

Dong, N., Huang, H., Zheng, L., 2015. Support vector machine in crash prediction at the level of traffic analysis zones: Assessing the spatial proximity effects. Accident Analysis & Prevention 82, 192-198.

DRVN, 2015. National highway system statistics. Directorate for Roads of Vietnam, Hanoi.

Dumbaugh, E., Rae, R., 2009. Safe Urban Form: Revisiting the Relationship Between Community Design and Traffic Safety. Journal of the American Planning Association 75(3), 309-329.

Econometric Software, 2012. NLOGIT. Version 5 Reference Guide.

ERA, 2014. Railway Safety Performance in the European Union 2014. European Railway Agency.

Erdogan, S., 2009. Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. Journal of Safety Research 40(5), 341-351.

ESCAP, 2000. Evaluation of cost-effective systems for railway level-crossing protection. United Nations Economic and Social Commission for Asia and the Pacific, New York.

Greene, W.H., 2012. Econometric Analysis, 7th ed. Pearson Prentice Hall, Upper Saddle River, NJ.

Hadayeghi, A., Shalaby, A., Persaud, B., 2003. Macrolevel Accident Prediction Models for Evaluating Safety of Urban Transportation Systems. Transportation Research Record: Journal of the Transportation Research Board 1840, 87-95.

Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2010. Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. Accident Analysis & Prevention 42(2), 676-688.

Huang, H., Abdel-Aty, M., Darwiche, A., 2010. County-Level Crash Risk Analysis in Florida. Transportation Research Record: Journal of the Transportation Research Board 2148, 27-37.

ITSR, 2011. Transport safety bulletins: Level crossing accidents in Australia. Independent Transport Safety Regulator, Sydney, Australia.

Jiang, X., Abdel-Aty, M., Hu, J., Lee, J., 2016. Investigating macro-level hotzone identification and variable importance using big data: A random forest models approach. Neurocomputing 181, 53-63.

JICA, 2009. Research Master Plan for Road Safety in Viet Nam. Japan International Cooperation Agency, Hanoi.

Kim, K., Brunner, I., Yamashita, E., 2006. Influence of Land Use, Population, Employment, and Economic Activity on Accidents. Transportation Research Record: Journal of the Transportation Research Board 1953, 56-64.

Kumara, S., Chin, H., 2004. Study of Fatal Traffic Accidents in Asia Pacific Countries. Transportation Research Record: Journal of the Transportation Research Board 1897, 43-47.

Lee, J., Abdel-Aty, M., Choi, K., 2014a. Analysis of residence characteristics of at-fault drivers in traffic crashes. Safety Science 68, 6-13.

Lee, J., Abdel-Aty, M., Choi, K., Huang, H., 2015. Multi-level hot zone identification for pedestrian safety. Accident Analysis & Prevention 76, 64-73.

Lee, J., Abdel-Aty, M., Jiang, X., 2014b. Development of zone system for macro-level traffic safety analysis. Journal of Transport Geography 38, 13-21.

Leroux, B.G., Lei, X., Breslow, N., 2000. Estimation of Disease Rates in Small Areas: A new Mixed Model for Spatial Dependence, in: Halloran, M.E., Berry, D. (Eds.), Statistical Models in Epidemiology, the Environment, and Clinical Trials. Springer New York, New York, NY, pp. 179-191.

Li, Z., Wang, W., Liu, P., Bigham, J.M., Ragland, D.R., 2013. Using Geographically Weighted Poisson Regression for county-level crash modeling in California. Safety Science 58, 89-97.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice 44(5), 291-305.

Lovegrove, G.R., Sayed, T., 2006. Macro-level collision prediction models for evaluating neighbourhood traffic safety. Canadian Journal of Civil Engineering 33(5), 609-621.

Moeinaddini, M., Asadi-Shekari, Z., Zaly Shah, M., 2014. The relationship between urban street networks and the number of transport fatalities at the city level. Safety Science 62, 114-120.

Moran, P.A.P., 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37(1/2), 17-23.

Ng, K.-s., Hung, W.-t., Wong, W.-g., 2002. An algorithm for assessing the risk of traffic accident. Journal of Safety Research 33(3), 387-410.

Ngo, A.D., Rao, C., Phuong Hoa, N., Hoy, D.G., Thi Quynh Trang, K., Hill, P.S., 2012. Road traffic related mortality in Vietnam: Evidence for policy from a national sample mortality surveillance system. BMC Public Health 12(1), 1-9.

Nguyen, H., Ivers, R.Q., Jan, S., Martiniuk, A.L.C., Li, Q., Pham, C., 2013a. The economic burden of road traffic injuries: evidence from a provincial general hospital in Vietnam. Injury Prevention 19(2), 79-84.

Nguyen, H.T., Passmore, J., Cuong, P.V., Nguyen, N.P., 2012. Measuring compliance with Viet Nam's mandatory motorcycle helmet legislation. International Journal of Injury Control and Safety Promotion 20(2), 192-196.

Nguyen, N.P., Passmore, J., Tran, L.T.N., Luong, A.M., 2013b. Role of Alcohol in Hospitalized Road Trauma in Viet Nam. Traffic Injury Prevention 14(4), 329-334.

Noland, R.B., 2003. Traffic fatalities and injuries: the effect of changes in infrastructure and other trends. Accident Analysis & Prevention 35(4), 599-611.

Noland, R.B., Oh, L., 2004. The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data. Accident Analysis & Prevention 36(4), 525-532.
Noland, R.B., Quddus, M.A., 2004. A spatially disaggregate analysis of road casualties in England. Accident Analysis & Prevention 36(6), 973-984.
NTSC, 2015. Traffic safety statistics report. National Transportation Safety Committee of Vietnam, Hanoi.
Passmore, J., Tu, N.T.H., Luong, M.A., Chinh, N.D., Nam, N.P., 2010. Impact of Mandatory Motorcycle Helmet Wearing Legislation on Head Injuries in Viet Nam: Results of a Preliminary Analysis. Traffic Injury Prevention 11(2), 202-206.
Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., Wets, G., 2012. Application of Different Exposure Measures in Development of Planning-Level Zonal Crash Prediction Models. Transportation Research Record: Journal of the Transportation Research Board 2280, 145-153.
Pulugurtha, S.S., Duddu, V.R., Kotagiri, Y., 2013. Traffic analysis zone level crash estimation models based on land use characteristics. Accident Analysis & Prevention 50, 678-687.
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.
R Development Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rushworth, A., Lee, D., Mitchell, R., 2014. A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London. Spatial and Spatio-temporal Epidemiology 10, 29-38.
Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accident Analysis & Prevention 45, 382-391.
Siddiqui, C., Watkins, K., 2016. Evaluating Long-Range Regional Safety with Scenario Planning Analysis. Transportation Research Record: Journal of the Transportation Research Board 2563, 18-28.
Song, B., Huang, H., Zeng, Q., Deng, Q., Abdel-Aty, M., 2015. A Comparative Analysis of Macro and Micro Models for Zonal Crash Prediction, Transportation Research Board 94th Annual Meeting, Washington DC.
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.
Tolón-Becerra, A., Lastra-Bravo, X., Flores-Parra, I., 2012. National and Regional Analysis of Road Accidents in Spain. Traffic Injury Prevention 14(5), 486-495.
Torre, G.L., Beeck, E.V., Quaranta, G., Mannocci, A., Ricciardi, W., 2007. Determinants of within-country variation in traffic accident mortality in Italy: a geographical analysis. International Journal of Health Geographics 6(49).
Tran, N.T., Bachani, A.M., Pham, V.C., Lunnen, J.C., Jo, Y., Passmore, J., Nguyen, P.N., Hyder, A.A., 2012. Drinking and Driving in Vietnam: Public Knowledge, Attitudes, and Practices. Traffic Injury Prevention 13(sup1), 37-43.
Traynor, T.L., 2008. Regional economic conditions and crash fatality rates – a cross-county analysis. Journal of Safety Research 39(1), 33-39.
Truong, L.T., Nguyen, H.T.T., De Gruyter, C., 2016. Mobile phone use among motorcyclists and electric bike riders: A case study of Hanoi, Vietnam. Accident Analysis & Prevention 91, 208-215.
Truong, L.T., Somenahalli, S.V.C., 2011. Using GIS to Identify Pedestrian-Vehicle Crash Hot Spots and Unsafe Bus Stops. Journal of Public Transportation 14(1), 99-114.

VEA, 2015. Expressway statistics. Vietnam Expressway Administration, Hanoi.
VNRA, 2015. Level crossings on the railway network. Report No. 2314. Vietnam Railway Administration, Hanoi.
Wang, X., Wu, X., Abdel-Aty, M., Tremont, P.J., 2013. Investigation of road network features and safety performance. Accident Analysis & Prevention 56, 22-31.
Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2011. Statistical and Econometric Methods for Transportation Data Analysis. Chapman & Hall/CRC, Boca Raton, FL.
WHO, 2015. Global status report on road safety 2015. World Health Organization, Geneva.
Wier, M., Weintraub, J., Humphreys, E.H., Seto, E., Bhatia, R., 2009. An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. Accident Analysis & Prevention 41(1), 137-145.
Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. Accident Analysis & Prevention 75, 16-25.
Xu, P., Huang, H., Dong, N., Abdel-Aty, M., 2014. Sensitivity analysis in the context of regional safety modeling: Identifying and assessing the modifiable areal unit problem. Accident Analysis & Prevention 70, 110-120.