

Research Article

Bayesian Hierarchical Modeling Monthly Crash Counts on Freeway Segments with Temporal Correlation

Qiang Zeng, Jiaren Sun, and Huiying Wen

School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, Guangdong 510641, China

Correspondence should be addressed to Huiying Wen; hywen@scut.edu.cn

Received 26 June 2017; Accepted 11 September 2017; Published 24 October 2017

Academic Editor: Francesco Bella

Copyright © 2017 Qiang Zeng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As the basis of traffic safety management, crash prediction models have long been a prominent focus in the field of freeway safety research. Studies usually take years or seasons as the observed time units, which may result in heterogeneity in crash frequency. To eliminate that heterogeneity, this study analyzes monthly crash counts and develops Bayesian hierarchical models with random effects, lag-1 autoregression (AR-1), and both (REAR-1) to accommodate the multilevel structure and temporal correlation in crash data. The candidate models are estimated and evaluated in the freeware WinBUGS using a crash dataset obtained from the Kaiyang Freeway in Guangdong Province, China. Significant temporal effects are found in the three models, and Deviance Information Criteria (DIC) results show that taking temporal correlation into account considerably improves the model fit compared with the Poisson model. The hierarchical models also avoid any misidentification of the factors with significant safety effects, because their variances are greater than in the Poisson model. The DIC value of the AR-1 model is substantially lower than that of the random effect model and equivalent to that of the REAR-1 model, which indicates the superiority of the lag-1 autoregressive structure in accounting for the temporal effects in crash frequency.

1. Introduction

The freeway has become a primary method of long-distance passenger and cargo transportation due to its high capacity and potential for high speed. The traffic on freeways consists of more motor vehicle types than urban traffic and includes passenger cars, coaches, vans, and light/medium/heavy trucks. This diversity in freeway traffic composition may result in numerous vehicle interactions, and the high speeds result in shorter driver response times when encountering emergencies. Freeways in mountainous areas may also suffer from poor geometrical design and adverse weather conditions [1, 2], which increase the crash risk. Developing a crash prediction model (or safety performance function) that provides a good understanding of the crash occurrence mechanism on a freeway is thus vital [1] when ranking sites for safety improvement and for accurately evaluating the effectiveness of countermeasures.

In the absence of detailed driving information (e.g., acceleration, braking, and steering), most studies have analyzed the relationship between the risk factors and crash frequency

[3]. The most common analytical methods are statistical count models, which explicitly illustrate the random, discrete, and nonnegative nature of crash frequency data and the safety effects of the main contributing factors [4]. Poisson regression is the basic model for crash prediction, in which crash count is assumed to follow a Poisson distribution that requires the mean to be equal to the variance [5]. To account for important issues related to crash data such as overdispersion, underdispersion, excess zero observations, spatiotemporal correlation, multilevel structures, and unobserved heterogeneity, a great many Poisson model variations have been proposed, which significantly improve model fit and predictive performance [3, 6]. With more recent advances in crash prediction modeling, Bayesian inference has been extensively applied to traffic safety analysis because of its ability to deal with complex models (often without closed-form likelihood functions) such as the hierarchical model [1, 2], spatiotemporal model [7, 8], random parameters model [9], and multivariate model [10] and through these [8]. The freeware WinBUGS provides a user-friendly platform for making Bayesian inferences using Markov chain Monte

Carlo (MCMC) techniques. The integrated nested Laplace approximation (INLA) approach has additionally been developed as a computationally efficient alternative to the MCMC methods, and an R package (R-INLA) is available to easily apply the approach [11].

Consideration of multiple levels is extremely important in freeway safety analysis because panel data are used for crash modeling [12]. Bayesian hierarchical models are thus the most widely used methods in freeway safety. However, freeway crash frequencies have previously been aggregated by year or season, which may result in information loss in time-varying explanatory variables. To avoid this phenomenon, crash data should be aggregated into small time intervals (e.g., months). This manipulation results in the same freeway section generating multiple observations, which may be correlated over time because of their shared effects of unobserved or unobservable time-dependent factors. Washington et al. [13] pointed out that ignoring temporal correlation will lead to an underestimation of the parameters' variances and thus potentially lead to the incorrect identification of the contributing factors, which has significant consequences for safety.

The widely used hierarchical Poisson model (also called the random effect model) is able to accommodate temporal correlation to some extent, but the added residual term is unstructured, which may not fully account for the temporal correlation. In addition to the random effect model, a variety of methodological approaches have been proposed to assess temporal effects in crash frequency data. These include generalized estimating equations with independent, exchangeable, autoregressive, or unstructured temporal terms [14, 15], a Bayesian hierarchical Poisson with a lag-1 autoregression (AR-1) model [7], an autoregressive integrated moving average model [16], an integer-valued autoregressive Poisson model [17], a latent variable representation of count data models with autoregressive temporal terms [18], and a multinomial generalized Poisson model with temporal dependence [19]. Of these approaches, the Bayesian hierarchical Poisson AR-1 model with the simplest formula is able to account for a multilevel structure and temporal correlation simultaneously. In both the generalized estimation equation and the hierarchical Poisson regression modeling frameworks, the autoregressive terms have been found to significantly outperform unstructured terms in model fit [7, 15].

In this study, the key objective is to develop a hierarchical temporal model to analyze freeway crash frequency aggregated by month, which accommodates a multilevel structure in panel crash data and temporal correlation across observations at the same site. Bayesian hierarchical Poisson and hierarchical Poisson AR-1 models are the two candidate methods. An approach integrating the two methods is also proposed, to simultaneously account for the structured and unstructured temporal effects. A Poisson model is used as a benchmark to demonstrate these temporal models, and they are calibrated and compared in the Bayesian context using a year's worth of crash data from the Kaiyang Freeway in China.

The remainder of this paper is as follows. In the next section, the alternative models and a model comparison

criterion are specified. Section 3 describes the collected data for model demonstration. Detailed estimations of the models are introduced in Section 4, and the results of model comparison and parameter estimation are discussed. Section 5 concludes and presents directions for future research.

2. Methodology

In this section, the structure of the Poisson model is formulated. The formulations of the three Bayesian hierarchical models for predicting crash frequency with temporal correlation are then specified in order of complexity. Finally, the Deviance Information Criteria (DIC) is introduced for the purpose of model comparison.

2.1. Model Specification

2.1.1. Poisson Model. In the Poisson model, the crash occurrence is assumed to be a Poisson process. That is, the crash count $Y_{i,t}$ on freeway segment i during month t is assumed to follow a Poisson distribution [5]:

$$P(Y_{i,t} = k) = \frac{(\lambda_{i,t})^k}{k!} \exp(-\lambda_{i,t}), \quad (1)$$

$$k = 0, 1, 2, \dots, \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T,$$

where N and T are the number of observed sites and periods, respectively, and $\lambda_{i,t}$ is the underlying Poisson mean of $Y_{i,t}$. Conceptually, the expected crash count $\lambda_{i,t}$ is modeled as the product of crash exposure $e_{i,t}$ and crash risk $\mu_{i,t}$ [8]:

$$\lambda_{i,t} = e_{i,t} \mu_{i,t}. \quad (2)$$

Crash exposure is defined as the number of opportunities for crashes in a given time in a given area. The crash exposure of a roadway segment is generally associated with its length and the traffic volume. Forms proposed in previous studies include annual average daily traffic [20] and vehicle miles traveled [1]. In the current research, the observational time unit is month. As the numbers of days in certain months differ, the monthly total traffic (MTT) is used as a crash exposure variable to specify the traffic volume precisely. The freeway crash exposure is formulated by the product of a power of MTT and of segment length, which reveal the potential nonlinear relationship between crash frequency and traffic volume [8]:

$$e_{i,t} = (\text{MTT}_{i,t})^{\alpha_1} (\text{Length}_i)^{\alpha_2}, \quad (3)$$

in which Length_i is the length of freeway segment i and $\text{MTT}_{i,t}$ is its MTT during month t . The two parameters to be estimated are α_1 and α_2 .

A generalized linear function is assumed between the crash risk $\mu_{i,t}$ and the observed risk factors $\mathbf{X}_{i,t}$:

$$\log \mu_{i,t} = \mathbf{X}'_{i,t} \boldsymbol{\beta}, \quad (4)$$

where $\boldsymbol{\beta}$ are the coefficients corresponding to the risk factors.

2.1.2. Random Effect Model. The monthly crash counts may be affected by unobserved or unobservable factors related to the freeway section, resulting in site-specific effects [1]. The shared site-specific effects of the crash counts on the same freeway section during different months are referred to as unstructured temporal effects [3]. To account for the site-specific/unstructured temporal effects in the random effect model, a residual term θ_i is added to the generalized linear function for modeling crash risk:

$$\log \mu_{i,t} = \mathbf{X}'_{i,t} \boldsymbol{\beta} + \theta_i, \quad (5)$$

where θ_i is assumed to follow a normal distribution with mean 0 and standard deviation σ ($\sigma > 0$):

$$\theta_i \sim \text{normal}(0, \sigma^2). \quad (6)$$

2.1.3. Autoregression-1 (AR-1) Model. The AR-1 model accounts for the temporal correlation among crash frequencies during successive months by specifying a residual term $\delta_{i,t}$ with lag-1 dependence [7], where lag-1 means that the temporal effect on a specific freeway section during a month is affected by its counterpart during the previous month:

$$\log \mu_{i,t} = \mathbf{X}'_{i,t} \boldsymbol{\beta} + \delta_{i,t}, \quad (7)$$

where the temporal terms $\delta_{i,t}$ ($t = 1, 2, \dots, T$) are assumed to follow the normal distributions, which are based on the stationarity assumption [21]

$$\delta_{i,1} \sim \text{normal}\left(0, \frac{\sigma_\delta^2}{1 - \gamma^2}\right), \quad (8)$$

$$\delta_{i,t} \sim \text{normal}(\gamma \delta_{i,t-1}, \sigma_\delta^2), \quad \text{for } t > 1. \quad (9)$$

In the above two equations, γ is the autocorrelation coefficient and σ_δ is the standard deviation of the temporal terms.

2.1.4. Random Effect with Autoregression-1 (REAR-1) Model. As mentioned above, the random effect and AR-1 models account for unstructured and structured temporal effects, respectively. To combine the strengths of the two models, both the unstructured and structured residual terms, θ_i and $\delta_{i,t}$, are added to the crash risk modeling function, resulting in the REAR-1 model:

$$\log \mu_{i,t} = \mathbf{X}'_{i,t} \boldsymbol{\beta} + \theta_i + \delta_{i,t}. \quad (10)$$

2.2. Model Comparison. The DIC is commonly used for measuring the goodness-of-fit of the models inferred by the Bayesian method [7, 8, 22]. As in previous research, it is used here to compare the above formulated models. The DIC is intended to be a Bayesian generalization of Akaike's Information Criteria that penalizes models with more parameters. Specifically, it provides a Bayesian measure of model complexity and fitting and is given by [23]

$$\text{DIC} = \overline{D(\boldsymbol{\theta})} + pD, \quad (11)$$

where $\overline{D(\boldsymbol{\theta})}$ is the posterior mean deviance that can be taken as a Bayesian measure of fitting and pD is a complexity measure for the effective number of parameters. Generally, models with lower DIC values are preferable. However, any critical difference in DIC is very difficult to determine. According to Spiegelhalter et al. [24], roughly over 10 differences may rule out the model with the higher DIC; differences between 5 and 10 are considered substantial; and if the DIC difference is less than 5 and the parameter estimation results are significantly different then it could be misleading to simply report the model with the lowest DIC.

3. Data Preparation and Preliminary Analysis

To calibrate the candidate models and compare their performances on model fit, the crash, traffic, and roadway data on Kaiyang Freeway in Guangdong Province, China, in 2014 were collected. Kaiyang Freeway has four lanes and a median barrier. Its total length is about 125 km and the posted speed limit is 120 km/h. The disaggregated crash data are obtained from the Highway Maintenance and Administration Management Platform of the Guangdong Transportation Group. The traffic data are acquired from the Guangdong Freeway Networked Toll System, and the roadway data are extracted from the Horizontal and Longitudinal Profile, designed by Guangdong Province Communication Planning and Design Institute Co., Ltd.

The first and essential step in data preparation is roadway segmentation. With reference to the previous studies on freeway traffic analysis [1], the major criterion used for segmenting the freeway is homogeneity in roadway horizontal and vertical alignments. In addition, the minimum length of each segment is set to 150 m, to eliminate the low exposure issue and the high statistical uncertainty of the crash risk on short segments. Segments shorter than 150 m are combined with proximal segments that have similar roadway features where possible. According to the two segmentation criteria, Kaiyang Freeway is divided into 154 segments ($I = 154$).

Crashes are mapped to these segments, based on their locations along the freeway recorded in the collected crash data. The crash counts are aggregated by segment and month ($T = 12$). In the Networked Toll System, vehicles are classified into five categories, according to their head height, axis number, wheel number, and wheelbase. The classification criteria are listed in Table 1. The initial traffic volumes in the system are recorded for each vehicle category. For freeway segment i and month t , the weighted average traffic $\text{MTT}_{i,t}$ is calculated by using the weights 1, 1.5, 2, 3, and 3.5 for vehicle categories (1)–(5), respectively:

$$\text{MTT}_{i,t} = V_{i,t}^{(1)} + 1.5V_{i,t}^{(2)} + 2V_{i,t}^{(3)} + 3V_{i,t}^{(4)} + 3.5V_{i,t}^{(5)}, \quad (12)$$

where $V_{i,t}^{(1)}$, $V_{i,t}^{(2)}$, $V_{i,t}^{(3)}$, $V_{i,t}^{(4)}$, and $V_{i,t}^{(5)}$ are the traffic volumes for vehicle categories (1)–(5), respectively, on freeway segment i during month t . The traffic composition, that is, the

TABLE I: Vehicle categorization.

Category	Categorization criteria				Representative vehicle types
	Head height	Number of axes	Number of wheels	Wheelbase	
(1)	<1.3	2	2–4	<3.2	Passenger car, jeep, taxi-head truck
(2)	≥1.3	2	4	≥3.2	Minibus, minivan, light truck
(3)	≥1.3	2	6	≥3.2	Medium bus, large normal bus, medium truck
(4)	≥1.3	3	6–10	≥3.2	Large luxury bus, large truck, large trailer, 20-foot container vehicle
(5)	≥1.3	>3	>10	≥3.2	Heavy truck, heavy trailer, 40-foot container vehicle

percentages of each vehicle category $Veh_{i,t}^{(1)}$, $Veh_{i,t}^{(2)}$, $Veh_{i,t}^{(3)}$, $Veh_{i,t}^{(4)}$, and $Veh_{i,t}^{(5)}$, is calculated as

$$\begin{aligned}
 Veh_{i,t}^{(1)} &= \frac{V_{i,t}^{(1)}}{V_{i,t}^{(1)} + 1.5V_{i,t}^{(2)} + 2V_{i,t}^{(3)} + 3V_{i,t}^{(4)} + 3.5V_{i,t}^{(5)}}, \\
 Veh_{i,t}^{(2)} &= \frac{1.5V_{i,t}^{(2)}}{V_{i,t}^{(1)} + 1.5V_{i,t}^{(2)} + 2V_{i,t}^{(3)} + 3V_{i,t}^{(4)} + 3.5V_{i,t}^{(5)}}, \\
 Veh_{i,t}^{(3)} &= \frac{2V_{i,t}^{(3)}}{V_{i,t}^{(1)} + 1.5V_{i,t}^{(2)} + 2V_{i,t}^{(3)} + 3V_{i,t}^{(4)} + 3.5V_{i,t}^{(5)}}, \\
 Veh_{i,t}^{(4)} &= \frac{3V_{i,t}^{(4)}}{V_{i,t}^{(1)} + 1.5V_{i,t}^{(2)} + 2V_{i,t}^{(3)} + 3V_{i,t}^{(4)} + 3.5V_{i,t}^{(5)}}, \\
 Veh_{i,t}^{(5)} &= \frac{3.5V_{i,t}^{(5)}}{V_{i,t}^{(1)} + 1.5V_{i,t}^{(2)} + 2V_{i,t}^{(3)} + 3V_{i,t}^{(4)} + 3.5V_{i,t}^{(5)}}.
 \end{aligned} \tag{13}$$

Table 2 illustrates the definitions and descriptive statistics of the variables used in the model development. Correlation tests and multicollinearity diagnoses for the risk factors are conducted. Table 3 shows the results of Pearson correlation tests. From the results, we find that the two variable pairs, $Veh_{i,t}^{(2)}$ and $Veh_{i,t}^{(5)}$ and $Veh_{i,t}^{(4)}$ and $Veh_{i,t}^{(5)}$, are significantly correlated with coefficients over 0.6 or below -0.6 . To avoid the adverse effect of significant correlation $Veh_{i,t}^{(5)}$ is excluded from the models. The results of the diagnoses indicate that there is no significant collinearity in the other factors.

4. Model Estimation and Result Analysis

4.1. Model Estimation. All candidate models are programmed, estimated, and evaluated in WinBUGS, where Bayesian inference can be easily implemented. In the absence of sufficient prior knowledge, noninformative priors are specified for the parameters and the hyperparameters. Specifically, a diffused normal distribution $N(0, 10^4)$ is used for the priors of α_1 , α_2 , γ , and each element of β , and a diffused gamma distribution $\text{gamma}(0.001, 0.001)$ is used for the priors of precisions of the normal distributions, $1/\sigma^2$ and $1/\sigma_\delta^2$. For each model, a chain of 500,000 iterations of the MCMC simulation are made, with the first 4,000 acting as burn-ins. The Gelman-Rubin statistics available in WinBUGS are used to evaluate the MCMC convergence.

4.2. Result Analysis. The results of the model estimation and comparison are summarized in Table 4. The hyperparameters (σ , γ , and σ_δ) in the three Bayesian hierarchical models (i.e., random effect, AR-1, and REAR-1), which measure the magnitudes of temporal effects, are all statistically significant with at least 90% of their Bayesian credible intervals away from zero. These results indicate that the monthly crash counts on the freeway sections are significantly correlated. The values of DIC also demonstrate that these hierarchical regressions substantially outperform the Poisson regression on model fit, demonstrating that accommodating temporal effects could considerably improve model fit [25]. Temporal correlation can pool strength from neighboring periods for parameter estimation and be a surrogate for unknown and time-dependent covariates, thereby reducing model misspecification [26]. After accounting for temporal effects, the standard deviations of the explanatory variables' coefficients increase, as expected, which eliminates the misidentification issue on factors contributing to crash frequency [13]. A representative example in this empirical analysis is curvature. The estimates of its effect are significant at a 90% credible level in the Poisson model, while being insignificant (less than 80% credible) in the hierarchical models.

Although the temporal correlation is found to be significant in the random effect model and the AR-1 model, the $D(\theta)$ value of the AR-1 model (=2682) is much smaller than that of the random effect model (=2790), suggesting that the AR-1 model fits the freeway crash data much better than the random effect model. While there are more effective parameters (as reflected by pD) in the AR-1 model, which increase the complexity, its lower DIC indicates that it outperforms the random effect model substantially. As previously stated, the temporal correlation is mainly derived from the effects of unobserved or unobservable time-dependent factors. For the collected crash data, environmental factors (e.g., precipitation and temperature) may contribute to the temporal effects. For example, precipitation, which has been found to increase crash frequency [27], varies by month on this freeway according to the Guangdong meteorological statistics. With the lag-1 autoregressive terms, the AR-1 model can approximate the varied and serially correlated effects of the unobserved factors more effectively than the random effect model, which assumes that different months share the same temporal effects. The result is in line with previous findings showing that the lag-1 autoregressive structure is

TABLE 2: Descriptive statistics of variables in the model.

Variables	Description	Mean	SD	Min.	Max.
<i>Response variable</i>					
Crash	Crash count per segment per month	0.372	0.672	0	5
<i>Exposure variables</i>					
Length	Segment length (m)	813	303	154.4	2000
MTT	Monthly total traffic (10^3 pcu ^a)	1851	513	1327	3759
<i>Risk factors</i>					
Veh _{i,t} ⁽¹⁾	The percentages of vehicles of Category (1) vehicles accounted for (reference case)	38.1	5.76	31.4	53.5
Veh _{i,t} ⁽²⁾	The percentages of vehicles of Category (2) vehicles accounted for	2.56	0.513	1.97	5.33
Veh _{i,t} ⁽³⁾	The percentages of vehicles of Category (3) vehicles accounted for	21.9	2.15	15.7	25.5
Veh _{i,t} ⁽⁴⁾	The percentages of vehicles of Category (4) vehicles accounted for	6.59	0.766	4.57	7.80
Veh _{i,t} ⁽⁵⁾	The percentages of vehicles of Category (5) vehicles accounted for	30.9	5.57	20.2	39.4
Curvature	Segment curvature (0.1 km^{-1})	1.77	1.27	0	4.35
Gradient	Segment gradient (%)	0.741	0.568	0	2.91
Bridge	A part of a bridge: yes = 1, no = 0	0.500	0.502	0	1
Access	Presence of roadway access: yes = 1, no = 0	0.210	0.407	0	1

^apcu: passenger car units.

TABLE 3: Pearson correlation coefficients between explanatory variables.

	Length	MTT	Veh _{i,t} ⁽²⁾	Veh _{i,t} ⁽³⁾	Veh _{i,t} ⁽⁴⁾	Veh _{i,t} ⁽⁵⁾	Curvature	Gradient	Bridge	Access
Length	1	0.13	-0.09	-0.04	-0.08	0.02	0.10	-0.25	0.23	0.18
MTT	0.13	1	0.18	-0.10	-0.00	-0.15	-0.02	-0.13	0.06	0.17
Veh _{i,t} ⁽²⁾	-0.09	0.18	1	-0.44	-0.12	-0.63	0.00	0.03	-0.04	-0.00
Veh _{i,t} ⁽³⁾	-0.04	-0.10	-0.44	1	-0.26	-0.16	0.02	0.02	-0.01	0.00
Veh _{i,t} ⁽⁴⁾	-0.08	-0.00	-0.12	-0.26	1	0.69	0.03	0.04	0.00	0.01
Veh _{i,t} ⁽⁵⁾	0.02	-0.15	-0.63	-0.16	0.69	1	0.01	0.00	0.02	-0.01
Curvature	0.10	-0.02	0.00	0.02	0.03	0.01	1	-0.06	0.08	-0.06
Gradient	-0.25	-0.13	0.03	0.02	0.04	0.00	-0.06	1	-0.00	-0.04
Bridge	0.23	0.06	-0.04	-0.01	0.00	0.02	0.08	-0.00	1	0.00
Access	0.18	0.17	-0.00	0.00	0.01	-0.01	-0.06	-0.04	0.00	1

a very good form for accounting for serial correlation in crash data [7, 15].

The superiority of the AR-1 model can be further confirmed by its equivalent $\overline{D(\theta)}$ and DIC values to the REAR-1 model, which simultaneously accounts for unstructured and structured temporal correlation. Specifically, the equivalent model fit performance reveals that the unstructured residual terms do not account for the possible temporal effects omitted by the lag-1 autoregression or that the AR-1 model may have fully accommodated the temporal dependence. Nonetheless, the estimates of the structured temporal terms' standard deviation σ_δ are almost equal (mean = 0.37; standard deviation = 0.13/0.14) in these two models, while the autocorrelation coefficient γ decreases dramatically from 0.72 (standard deviation = 0.14) in the AR-1 model to 0.46 (standard deviation = 0.23) in the REAR-1 model. Although the numbers of effective parameters are the same in the two models, in theory the formulation of the AR-1 model is less complex than that of the REAR-1 model. The AR-1 model also consumes less time

than the REAR-1 model in Bayesian parameter estimation processes. Therefore, it is reasonable to conclude that the AR-1 model is to an extent superior to the REAR-1 model.

To comprehensively justify the model validity, the results for explanatory factors that have statistically significant effects on crash frequency are also discussed in this section. Overall, the plus or minus signs of the parameters in all four models are generally consistent. They conform to engineering intuition and to the findings in the literature, but the magnitude of the parameters in the Bayesian hierarchical models is more consistent and has substantial differences from those in the Poisson model, indicating the effects of taking temporal correlation into consideration.

Specifically, in all of the crash prediction models the length variable is consistently significant at a 95% credible level, with positive coefficients. The value of 1.0 is within 95% Bayesian credible intervals. That is, the coefficients are not significantly different from 1.0, which confirms that the monthly crash counts may be linearly associated with the

TABLE 4: Model estimation results.

	Poisson	Random effect	Autoregression-1	Random effect with autoregression-1
Constant	-10.25 (1.39) ^{a,*}	-10.82 (2.19) [*]	-10.4 (1.56) [*]	-10.34 (2.39) [*]
Length	1.04 (0.11) [*]	1.00 (0.15) [*]	1.02 (0.15) [*]	0.99 (0.16) [*]
MTT	0.35 (0.15) [*]	0.42 (0.26) ^{***}	0.37 (0.18) ^{**}	0.41 (0.21) ^{**}
Veh _{<i>i,t</i>} ⁽²⁾	0.08 (0.08)	0.11 (0.10)	0.10 (0.10)	0.08 (0.10)
Veh _{<i>i,t</i>} ⁽³⁾	-0.01 (0.02)	-0.003 (0.02)	-0.008 (0.02)	-0.01 (0.03)
Veh _{<i>i,t</i>} ⁽⁴⁾	-0.08 (0.05) ^{**}	-0.09 (0.06) ^{***}	-0.10 (0.06) ^{**}	-0.11 (0.06) ^{**}
Gradient	0.25 (0.07) ^{**}	0.17 (0.09) ^{**}	0.18 (0.09) [*]	0.17 (0.09) ^{**}
Curvature	0.06 (0.03) ^{**}	0.04 (0.04)	0.04 (0.04)	0.04 (0.04)
Bridge	0.02 (0.08)	0.01 (0.11)	0.01 (0.10)	0.02 (0.10)
Access	0.07 (0.09)	0.03 (0.13)	0.04 (0.12)	0.04 (0.12)
σ	—	0.39 (0.06) [*]	—	0.27 (0.12) [*]
γ	—	—	0.72 (0.14) [*]	0.46 (0.23) ^{**}
σ_δ	—	—	0.37 (0.13) [*]	0.37 (0.14) [*]
$D(\theta)$	2883	2790	2682	2683
pD	10	65	155	155
DIC	2893	2854	2837	2838

^aEstimated mean (standard deviation) for the parameter; ^{*}significant at the 95% credible level; ^{**}significant at the 90% credible level; ^{***}significant at the 80% credible level.

freeway-section length. A linear relationship between crash frequency and segment length has been found in numerous functional class roadways, ranging from urban major and minor arterials [8] and rural minor arterials and major collectors [28] to rural frontage roads [29].

Unsurprisingly, the coefficients of the MTT variable are positive in all four models, because a higher traffic volume brings about more crash exposure [8]. However, its credible level varies from 95% in the Poisson model to 90% in the AR-1 and REAR-1 models and 80% in the random effect model. The coefficients with their 95% Bayesian credible intervals away from 1.0 are substantially different from 1.0, suggesting that the relationship between crash frequency and traffic volume is nonlinear. This is congruent with the model assumption and the findings of numerous studies [8, 30].

With respect to traffic composition, only the percentage of Category (4) vehicles is shown to have a significant effect on crash frequency (at a 90% credible level in the Poisson, AR-1, and REAR-1 models and at an 80% credible level in the random effect model). The negative coefficients indicate that increasing the percentages of Category (4) vehicles would decrease the crash risk. Several factors may account for this: (1) drivers of large buses/trucks/trailers (Category (4) vehicles) generally hold a higher level of driving license and have better driving skills than passenger car (Category (1) vehicles) drivers; (2) large buses and trucks are usually owned by professional transport companies and their daily traveling routes are relatively constant, so the drivers tend to be more familiar with the freeway conditions and are more experienced in handling emergencies; and (3) the Ministry of Transport in China has recently issued a series of policies and laws aimed at regulating bus and truck driving behavior, which may reduce the occurrence of crashes related to these vehicle types.

The significantly positive (at least a 90% credibility level) coefficients for segment gradient in all four models suggest that steeper slopes experience more crashes. The result is in line with previous findings and empirical engineering judgments: higher vertical grades reduce stopping sight distance, thus increasing the crash risk [27]. As safety considerations are vital in freeway vertical design, all design manuals recommend avoiding or minimizing the use of steep slopes [1].

5. Conclusions and Future Research

This study develops three Bayesian hierarchical models for predicting monthly crash counts on freeway segments, which simultaneously account for hierarchical structure and temporal correlation in crash data using unstructured terms, lag-1 autoregressive terms, and a combination of both. The models are estimated and evaluated by Bayesian methods via programming in the freeware WinBUGS. The crash data used for model calibration and comparison are collected from Kaiyang Freeway in Guangdong Province, China, in 2014.

The inference results show that both the unstructured and AR-1 temporal terms are significant with at least 90% Bayesian credible intervals away from 0, which indicates that the monthly crash counts at a certain freeway section are correlated. The results of DIC reveal that the Bayesian hierarchical models have substantially better fit than the Poisson model, suggesting that temporal correlation is able to improve the model specification and parameter estimation, thereby improving the goodness-of-fit. The underestimation phenomenon of parameter variance caused by ignoring temporal correlation is also observed in the parameter estimation results. Of the Bayesian hierarchical models, the AR-1 model is found to outperform the random effect model,

which demonstrates that the structured AR-1 terms fit the temporal effects better than the unstructured terms. Its merit is further confirmed by the equivalent DIC values of the AR-1 and REAR-1 models. The covariate estimates show that crash frequency may be linearly associated with segment length while being nonlinearly associated with traffic volume. Increasing the percentages of large buses/trucks/trailers and decreasing the segment gradient would significantly reduce crash occurrence on the freeway.

In summary, the empirical analysis demonstrates the significance of temporal correlation in the monthly freeway crash counts and the considerable advantage of the Bayesian hierarchical AR-1 model for analyzing the crash frequency with temporal correlation. However, if over one year of monthly crash data are collected, seasonal effects may exist across the crash counts in the same months of different years. More complicated models are expected to accommodate the autoregressive and seasonal effects simultaneously. Further research efforts could also be devoted to extending the univariate model to its multivariate form, to simultaneously predict crash frequency in specific categories (e.g., injury severity, crash type, or the number of vehicles involved), and applying this to identifying sites that could potentially benefit from engineering improvements in freeway safety management.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was jointly supported by the Natural Science Foundation of China (nos. 51378222 and 51578247), China Postdoctoral Science Foundation (no. 2017M610529), the Natural Science Foundation of Guangdong Province (no. 2017BQ026), and the Fundamental Research Funds for the Central Universities (no. 2017BQ026). The authors would like to sincerely thank Professor S. C. Wong at the University of Hong Kong for proofreading this paper.

References

- [1] M. Ahmed, H. Huang, M. Abdel-Aty, and B. Guevara, "Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway," *Accident Analysis & Prevention*, vol. 43, no. 4, pp. 1581–1589, 2011.
- [2] R. Yu, M. Abdel-Aty, and M. Ahmed, "Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors," *Accident Analysis & Prevention*, vol. 50, pp. 371–376, 2013.
- [3] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 291–305, 2010.
- [4] Q. Zeng, H. Huang, X. Pei, and S. C. Wong, "Modeling nonlinear relationship between crash frequency by severity and contributing factors by neural networks," *Analytic Methods in Accident Research*, vol. 10, pp. 12–25, 2016.
- [5] P. P. Jovanis and H.-L. Chang, "Modeling the relationship of accidents to miles traveled," *Transportation Research Record*, pp. 42–51, 1986.
- [6] F. L. Mannering and C. R. Bhat, "Analytic methods in accident research: methodological frontier and future directions," *Analytic Methods in Accident Research*, vol. 1, pp. 1–22, 2014.
- [7] H. Huang, H. C. Chin, and M. Haque, "Empirical evaluation of alternative approaches in identifying crash hot spots," *Transportation Research Record*, no. 2103, pp. 32–41, 2009.
- [8] Q. Zeng and H. Huang, "Bayesian spatial joint modeling of traffic crashes on an urban road network," *Accident Analysis & Prevention*, vol. 67, pp. 105–112, 2014.
- [9] Q. Zeng, H. Wen, H. Huang, and M. Abdel-Aty, "A Bayesian spatial random parameters Tobit model for analyzing crash rates on roadway segments," *Accident Analysis & Prevention*, vol. 100, pp. 37–43, 2017.
- [10] S. Barua, K. El-Basyouny, and M. T. Islam, "Multivariate random parameters collision count data models with spatial heterogeneity," *Analytic Methods in Accident Research*, vol. 9, pp. 1–15, 2016.
- [11] M. Blangiardo and M. Cameletti, *Spatial and spatio-temporal Bayesian models with R-INLA*, John Wiley & Sons, Ltd., 2015.
- [12] H. Huang and M. Abdel-Aty, "Multilevel data and Bayesian analysis in traffic safety," *Accident Analysis & Prevention*, vol. 42, no. 6, pp. 1556–1565, 2010.
- [13] S. P. Washington, M. G. Karlaftis, and F. L. Mannering, *Statistical and Econometric Methods for Transportation Data Analysis*, CRC Press, Boca Raton, Fla, USA, 2nd edition, 2011.
- [14] X. Wang and M. Abdel-Aty, "Temporal and spatial analyses of rear-end crashes at signalized intersections," *Accident Analysis & Prevention*, vol. 38, no. 6, pp. 1137–1150, 2006.
- [15] W. Xuesong, M. Abdel-Aty, and P. A. Brady, "Crash estimation at signalized intersections: significant factors and temporal effect," *Transportation Research Record*, no. 1953, pp. 10–20, 2006.
- [16] R. B. Noland, M. A. Quddus, and W. Y. Ochieng, "The effect of the London congestion charge on road casualties: An intervention analysis," *Transportation*, vol. 35, no. 1, pp. 73–91, 2008.
- [17] M. A. Quddus, "Time series count data models: an empirical application to traffic accidents," *Accident Analysis & Prevention*, vol. 40, no. 5, pp. 1732–1741, 2008.
- [18] M. Castro, R. Paleti, and C. R. Bhat, "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: Methodological*, vol. 46, no. 1, pp. 253–272, 2012.
- [19] Y.-C. Chiou and C. Fu, "Modeling crash frequency and severity with spatiotemporal dependence," *Analytic Methods in Accident Research*, vol. 5–6, pp. 43–58, 2015.
- [20] S.-P. Miaou, "The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions," *Accident Analysis & Prevention*, vol. 26, no. 4, pp. 471–482, 1994.
- [21] P. Congdon, *Applied Bayesian Modelling*, John Wiley and Sons, 2nd edition, 2014.
- [22] X. Xu, S. Xie, S. C. Wong, P. Xu, H. Huang, and X. Pei, "Severity of pedestrian injuries due to traffic crashes at signalized intersections in Hong Kong: a Bayesian spatial logit model," *Journal of Advanced Transportation*, vol. 50, no. 8, pp. 2015–2028, 2016.
- [23] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde, "Bayesian measures of model complexity and fit," *Journal of the*

- Royal Statistical Society: Series B (Statistical Methodology)*, vol. 64, no. 4, pp. 583–639, 2002.
- [24] D. J. Spiegelhalter, A. Thomas, N. G. Best, and D. Lunn, “WinBUGS user manual,” *MRC Biostatistics Unit*, 2005.
- [25] Q. Zeng, H. Wen, H. Huang, X. Pei, and S. Wong, “A multivariate random-parameters Tobit model for analyzing highway crash rates by injury severity,” *Accident Analysis & Prevention*, vol. 99, pp. 184–191, 2017.
- [26] J. Aguero-Valverde, “Full Bayes Poisson gamma, Poisson log-normal, and zero inflated random effects models: comparing the precision of crash frequency estimates,” *Accident Analysis & Prevention*, vol. 50, pp. 289–297, 2013.
- [27] Q. Zeng, H. Huang, X. Pei, S. C. Wong, and M. Gao, “Rule extraction from an optimized neural network for traffic crash frequency modeling,” *Accident Analysis & Prevention*, vol. 97, pp. 87–95, 2016.
- [28] S. Labi, “Efficacies of roadway safety improvements across functional subclasses of rural two-lane highways,” *Journal of Safety Research*, vol. 42, no. 4, pp. 231–239, 2011.
- [29] Y. Xie, D. Lord, and Y. Zhang, “Predicting motor vehicle collisions using Bayesian neural network models: an empirical analysis,” *Accident Analysis & Prevention*, vol. 39, no. 5, pp. 922–933, 2007.
- [30] X. Qin, J. N. Ivan, and N. Ravishanker, “Selecting exposure measures in crash rate prediction for two-lane highway segments,” *Accident Analysis & Prevention*, vol. 36, no. 2, pp. 183–191, 2004.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

