Evaluating Influence of Neighboring Structures on Spatial Crash Frequency Modeling and Site-Ranking Performance

Gurdiljot Singh Gill, Wen Cheng, Meiquan Xie, Tom Vo, Xudong Jia, and Jiao Zhou

Many neighborhood weight matrices have been adopted for modeling crash spatial heterogeneity. However, there has been little evaluation of their influence on crash prediction modeling performance. This study investigated 17 spatial-proximity matrices for development of spatial crash prediction models and site ranking with county-level data in California. Of the group of matrices being evaluated, traffic exposure–weighted and population-weighted distance-based matrices were first proposed in the traffic safety field. Bayesian spatial analysis was conducted with a combination of a first-order autoregressive error process and time trend for crashes to address the serial correlation of crashes in successive years. Two diagnostic measures were used for assessment of goodness of fit and complexity of models, and seven evaluation criteria were employed to assess the benefits associated with better-fitting models in site ranking. The results showed that modeling performance improved with an increase in number of neighbors considered in the weight matrix. However, a larger number of neighbors also led to greater variability of modeling performance. Specifically, Queen-2 and Decay-50 models proved to be superior among the adjacency- and distance-based models, respectively. The significance of incorporating spatial correlations was highlighted by the consistently poor performance of the base model, which included only the heterogeneity random effect. Finally, model-fitting performance seems to be strongly correlated with site-ranking performance. The models with closer goodness of fit tend to yield more consistent ranking results.

Previous research studies explored the spatial component of crashes as an advancement to the existing crash prediction models (*1–3*). Numerous spatial units were considered to understand the implications of crash-causing factors that operate at a spatial scale (e.g., urban planning policy, census characteristics, and highway classification). Depending on the purpose of the study, the sites of interest could range from microscopic locations, such as block groups (*4*), intersections (*5*), road segments (*6*), and corridors (*7, 8*), to macroscopic areas, such as census tracts (*9*), local health areas (*10*), traffic analysis zones (TAZs) (*11*), or counties (*12–15*). Comparatively speaking, microscopic analysis is primarily centered on investigating geometric or traffic characteristics that influence the safety on a network. In contrast, macroscopic safety analysis concentrates on quantifying the impact of socioeconomic and demographic characteristics, transportation demand, and network attributes so as to provide countermeasures from a planning perspective such as enactments of traffic rules, police enforcement, safety campaigns, and areawide engineering treatments.

The literature review demonstrates that a wide range of neighborhood weight matrix structures have been proposed to model crash spatial heterogeneity. Aguero-Valverde and Jovanis (*16*) explored the effect of spatial correlation in models of crash frequency at the segment level by using a full Bayesian approach with conditional autoregressive (CAR) effects (*17*). Three adjacency-based weight matrices were developed for first-, second-, and third-order neighbors; these models showed a significantly better fit than the Poisson lognormal model, which considered only heterogeneity. Guo et al. developed models to incorporate the spatial proximity at the corridor level between intersections due to similarity in road design and environmental characteristics (*8*). The modeling results demonstrated that the Poisson spatial model provided the best fit. Recently, Aguero-Valverde et al. used a multivariate spatial model to account for spatial correlation among adjacent sites (road segments) to enhance model prediction for different crash types (*18*). The multivariate CAR model was used with the first-order adjacency-based weight matrix, which was observed to have the best fit due to spatial and multivariate correlation.

Of the macrolevel studies, Best et al. investigated the risk of leukemia in children at three levels of data aggregation: local authority districts, census wards, and 1-km^2 grid squares (*19*). They examined adjacency- versus distance-based neighborhood spatial weights for each analysis. To explore the significant variables influencing crashes, Rhee et al. used spatial variables developed by geographic information systems to prepare a database of traffic crashes at the TAZ level (*20*). The Rook adjacency-based weight matrix was used for analysis of the spatial component of crash heterogeneity. Results showed that the spatial error model was better than the spatial lag model. Aguero-Valverde and Jovanis applied a univariate space–time model to analyze county-level crash counts (*14*). The first-order adjacency matrix was utilized for the CAR error term. The results demonstrated the existence of spatial correlation in the crash data. Huang et al. proposed a Bayesian spatial model to account for county-level variations of crash risk in Florida (*15*). A CAR prior was specified to accommodate for the spatial autocorrelations of adjacent counties. The results exhibited little difference between the safety effects of risk factors on all crashes and those on severe crashes.

G. S. Gill, W. Cheng, M. Xie, X. Jia, and J. Zhou, Department of Civil Engineering, College of Engineering, California State Polytechnic University, Pomona, 3801 West Temple Avenue, Pomona, CA 91768. T. Vo, Southern California Association of Governments, 818 West 7th Street, 12th Floor, Los Angeles, CA 90017. Corresponding author: W. Cheng, wcheng@cpp.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2659, 2017, pp. 117–126.
http://dx.doi.org/10.3141/2659-13

Compared with the large amount of research dedicated to modeling spatial heterogeneity in crash counts with various weight matrix structures, there is little evaluation of the influence of these different weight matrices on crash prediction modeling performance. Aguero-Valverde and Jovanis evaluated the effects of different neighboring structures on the spatial correlation in crash frequency models using the CAR model (*16*). The weight structures being investigated include exponential decay; adjacency-based, adjacency-route information; and distance order structures. Modeling results showed relatively inferior performance by using exponential decay models. Another study presented an evaluation of crash prediction models at the TAZ level with alternative types of spatial-proximity structures containing 0–1 first-order adjacency, common-boundary length, and centroid distance-based models (*21*). The results confirmed the extensive existence of cross-zonal spatial correlation in crash occurrence.

This study compares alternative spatial-proximity structures and represents a natural continuation of the foregoing two studies, with a number of important differences and unique contributions. First, more comprehensive weight matrices are evaluated that include two orders of Queen adjacency-based, two orders of Rook adjacency-based, common boundary length adjacency-based, five exponential decay, five pure distance order, population-weighted, and traffic exposure–weighted distance order matrices. Of the 17 neighboring structures, the last two were first proposed in the traffic safety field. Second, the serial correlation of the county-level crash count was taken into account by using the first-order autoregressive error process and global time trend combined. Third, the relationship between the crash frequency modeling performance and the number of neighbors in the weight matrices was explored in greater detail. Moreover, the site-ranking performance of different weight matrices was assessed by seven evaluation criteria: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), Cohen's kappa, total ranking difference (TRD), and mean absolute deviation (MAD).

METHODOLOGY

Spatial autocorrelations were explored among 58 counties in California with a wide array of weight matrices ranging from simple to more sophisticated ones. This study analyzed 17 neighborhood matrices and applied them to the county-level data sets. In addition, the model without accounting for spatial heterogeneity was also developed for comparison with spatial models. Hence, the process involved development of 18 models using WinBUGS for estimation of crash rate, evaluation of their goodness of fit, and finally evaluation of the relative site-ranking performance. This study used the full Bayesian hierarchical approach to account for structural heterogeneities such as temporal and spatial ones. The model is of the form developed by Besag et al. (*22*):

$$y_{it} \sim \text{Poisson}(e_{it}\theta_{it}) \tag{1}$$

where

y_{it} = observed crash count at county i in time period t,
θ_{it} = mean expected crash rate for site i in time period t, and
e_{it} = exposure in county i of time period t.

In this case, the exposure is the total daily vehicle miles (DVMT) by county. The crash rate is modeled as follows:

$$\log(\theta_{it}) = \beta_0 + \beta_1 t_k + \varnothing_i + u_i \tag{2}$$

where

β_0 = intercept,
β_1 = fixed coefficient for linear trend term t_k,
$\varnothing_i$ = spatially structured random effect, and
u_i = spatially unstructured random effect (heterogeneity).

This study did not incorporate any covariates because the major focus was to investigate the spatial correlations. A similar framework can be found in other studies (*23, 24*). Likewise, the interaction between time and space was not included since it might have potentially blurred the comparison of different weight structures. In addition to the global time trend, this model also accounts for the autoregressive safety effect by specifying the distribution of u_i as a lag-1 dependence in errors. Lag-1 is the correlation 1 year apart in this study (hence $k = 1$). The first-order autoregressive error process was chosen to capture the departure from the trend and it was based on the assumption of stationarity restriction:

$$u_i \sim \text{normal}\left(0, \frac{\sigma_i^2}{(1-\gamma^2)}\right) \tag{3}$$

$$u_{it} \sim \text{normal}\left(\gamma\varepsilon_{i,t-1}, \sigma_{it}^2\right) \qquad \text{for } t > 1 \tag{4}$$

where ε is the uncorrelated heterogeneity and γ is the autocorrelation coefficient with the following range: $0 < \gamma < 1$. The combination of first-order autoregressive and deterministic time trend for serial correlation can be found in current practice, such as stepwise autoregressive method specification (*25*).

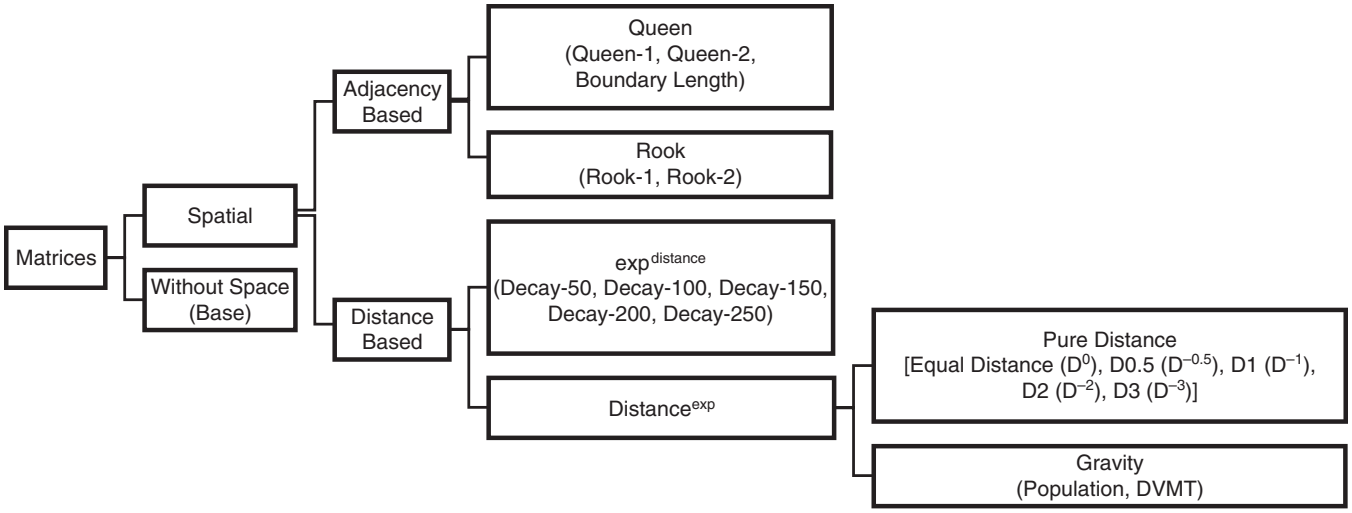
To accommodate the spatial correlation in the model, the CAR prior was introduced for the spatial random effects. The formulation of the CAR model used in these analyses is as follows (*26*):

$$\left[\varnothing_i \middle| \varnothing_j, i \neq j, \tau_{\varnothing}^2\right] \sim N\left(\varnothing_j, \tau_i^2\right) \tag{5}$$

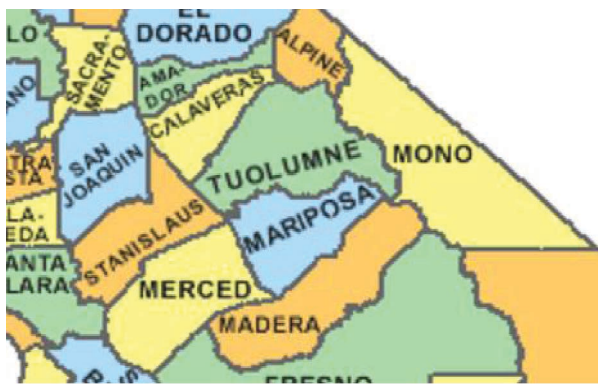
The foregoing equations show that the neighboring sites influence the crash risk associated with an area. Subscripts i and j represent a county and its neighbor, respectively, and $j \in N_i$, where N_i represents the neighbor set for region i. The weights are included because they also influence the risks, besides the neighbors. The weights for the adjacency and distance models are given by weights ij (w_{ij}) = 1 if i, j are adjacent, and 0 otherwise. Apart from the adjacency-based models, different weights were used for other models; these weights are explained in detail in the following subsections. For ease of illustration, all weight matrices being evaluated in the study are classified in Figure 1*a*.

Adjacency-Based Models

Adjacency-based models ignore the distance between sites of interest and focus only on neighboring structures based on proximity in space; these models give dichotomous outcomes. Five neighborhood adjacency-based weight matrices were developed, namely, Queen-1, Queen-2, Rook-1, Rook-2, and boundary length (BL). The difference between Queen and Rook is the criterion of assignment of neighbors. Queen uses the common boundaries as well as vertices to determine the adjacent neighbors, whereas Rook considers only the common boundaries. For example, in Figure 1*b*, in the case of Mariposa County, Rook neighbors are Tuolumne, Merced, and Madera



(a)



(b)

FIGURE 1 (*a*) Types of weight matrices and (*b*) Queen and Rook neighbors.

Counties, and Queen neighbors are Tuolumne, Stanislaus, Merced, Madera, and Mono Counties. The numbers for each term reflect the order of contiguity; this means that the difference between Queen-1 and Queen-2 (and corresponding Rook-1 and Rook-2) is that Queen-1 includes the direct neighbors that share common points, whereas Queen-2 also includes the further neighbors of neighbors. Finally, another weight matrix (BL) was developed for immediate neighbors based on the length of the boundary shared between counties.

Distance-Based Models

To account for a variety of scenarios, 12 models based on distance matrices were also developed. The simplest model, equal distance (ED), assigned equal weight for the weight matrix because it included all the counties as neighbors. The other models placed different weights on neighbors. For distance 0 (ED), 0.5, 1, 2, and 3, the following formulations were used: $w_{ij} = 1/\text{dist}_{ij}^{0}$, $w_{ij} = 1/\text{dist}_{ij}^{0.5}$, $w_{ij} = 1/\text{dist}_{ij}$, $w_{ij} = 1/\text{dist}_{ij}^{2}$, and $w_{ij} = 1/\text{dist}_{ij}^{3}$, respectively. These models only accounted for the relative distance between neighboring counties and placed more weight on counties that were closer together. Another similar set of models was developed (Decay-50, -100, -150, -200, and -250) that was also based on the distance between neighbors. However, these decay models were essentially different from regular distance models since there was a drastic reduction of weights as the distance between neighbors increased (*27*). The corresponding weight matrix for the decay model was defined as follows:

$$w_{ij} = e^{-\text{dist}_{ij}/\delta_0} \tag{6}$$

where

w_{ij} = weight of jth neighbor of ith county,
dist_{ij} = geographic centroid distance between counties i and j, and
δ_0 = a constant.

Five decays were chosen to incorporate different distances between the counties because the range of geometric centroid distances between counties was from 25 to 962 mi. In addition to the two types of distance-based matrices, two gravity models based on distance were developed; these models borrow the idea of a gravity model from the standard "four-step" travel demand modeling process. The corresponding weight matrices were defined by $w_{ij} = p_i p_j/\text{dist}_{ij}$, where p_i and p_j are normalized populations and DVMTs

of two counties, respectively. To the best of the authors' knowledge, such gravity models are the first applied in the traffic safety field. Similar neighboring structures can be found in the public health field (*23*).

Goodness of Fit of Crash Frequency Models

This study used the deviance information criterion (DIC) to assess the complexity and goodness of fit of the models. DIC is a hierarchical modeling generalization of the Akaike information criterion, which was proposed by Spiegelhalter et al. to account for model fit and complexity (*28*). Specifically, DIC is defined as

$$\mathrm{DIC} = D\left(\bar{\theta}\right) + 2p_D = \bar{D} + p_D \tag{7}$$

where $D(\bar{\theta})$ is the deviance evaluated at the posterior means of estimated unknowns ($\bar{\theta}$), and the posterior mean deviance $\bar{D}$ can be taken as a Bayesian measure of fit or "adequacy." The term p_D denotes the effective number of parameters in a model as the difference between $D(\bar{\theta})$ and $\bar{D}$ (i.e., mean deviance minus the deviance of the means). Generally, smaller values of DIC are preferred. As a general guideline, a difference of 7+ points in the DIC is treated as significant for modeling performance (*17*).

To account for the random spatial effect explained by the model, the fraction of total random variation was computed as a ratio of the empirical variance of the spatial component versus the total variance (*15*). The formulas are

$$\mathrm{var}(u) = \sum_i \frac{(u_i - \bar{u})^2}{n-1} \tag{8}$$

$$\mathrm{var}(v) = \sum_i \frac{(v_i - \bar{v})^2}{n-1} \tag{9}$$

$$\mathrm{fraction} = \frac{\mathrm{var}(u)}{\left(\mathrm{var}(u) + \mathrm{var}(v)\right)} \tag{10}$$

where u_i is a spatially structured random effect and v_i is a spatially unstructured random effect, with i ranging from 1 to $n = 58$.

Site-Ranking Evaluation

Criteria Based on Binary Partitions of Data

Under the criteria based on binary partitions of the data, the counties were divided into two groups based on estimated crash rates from various models with a certain threshold. In the current study, a cut line of 10% was used. This cutoff means that the top 10% counties were considered as high risk, and the remaining counties were treated as low risk. The common binary diagnostic criteria including sensitivity, specificity, PPV, and NPV were calculated with the following formulas (*29*):

$$\mathrm{sensitivity} = \frac{TP}{TP + FN} \tag{11}$$

$$\mathrm{specificity} = \frac{TN}{TN + FP} \tag{12}$$

$$\mathrm{PPV} = \frac{TP}{TP + FP} \tag{13}$$

$$\mathrm{NPV} = \frac{TN}{TN + FN} \tag{14}$$

where

TP = number of truly high-risk counties correctly identified as unsafe,
FP = number of truly low-risk counties identified as unsafe,
TN = number of truly low-risk counties identified as safe, and
FN = number of truly high-risk counties identified as safe.

Because this study is only focused on the relative site-ranking performance of various models, the model with the best predictive performance (the lowest DIC) is chosen to establish the "truly" high-risk or low-risk counties. Briefly, sensitivity is the ability of a model to correctly classify an individual as truly high risk. Specificity measures the ability of a model to correctly classify an individual as low risk. PPV represents the percentage of counties identified by one model as high risk that are truly high risk, whereas NPV indicates the percentage of counties identified by one model as low risk that are truly low risk. For all criteria, if the value is higher (as close to 100 as possible), this finding suggests that the model of interest is doing as well as the model with the lowest DIC.

Criteria Based on Multiple Partitions of Data

Aside from the foregoing criteria based on the dichotomy of counties, the counties can also be divided into multiple groups to get more precise results. Cohen's kappa was chosen to determine the magnitude of county-ranking agreement based on three groups: high (top 10th percentile), medium (11th to 89th percentile), and low (90th percentile and lower) (*30*). For each group, the number of common identified counties by two models was tallied and the kappa statistic was calculated with the following equation:

$$K = \frac{(p_o - p_e)}{(1 - p_e)} \tag{15}$$

where p_o is the proportion of observations in agreement, and p_e represents the proportion in agreement due to chance. The kappa statistic can ensure that only truly high-risk counties (excluding the random entries) were included for their respective groups. A relatively higher value of kappa indicates a larger number of sites (agreement) commonly identified as unsafe between the crash prediction models.

Criteria Based on Continuous Comparison of Data

Finally, in addition to categorizing counties into multiple groups, the relative site-ranking performance of various models can be evaluated by comparing all counties in a continuous fashion. Two criteria of this

category were implemented: TRD (*31*), and MAD (*32*). TRD is of the following form:

$$\text{TRD} = \sum_{i=1}^{n} \left| R_{i,j} - R_{i,k} \right| \tag{16}$$

where $R_{i,j}$ is the rank of county i by model j, and $R_{i,k}$ is the rank of county i by model k. The smaller the TRD, the closer the two models are in terms of site ranking.

Alternatively, the estimated crash rates by various models can be compared to determine MAD:

$$\text{MAD} = \frac{1}{n} \sum_{i=1}^{n} \left| \text{CR}_{i,j} - \text{CR}_{i,k} \right| \tag{17}$$

where $\text{CR}_{i,j}$ is the estimated crash rate of county i by model j, and $\text{CR}_{i,k}$ is the estimated crash rate of county i by model k. This measure may be regarded as the safety component that represents the crash rate prediction performance of the models relative to the measure of "true safety." The degree of deviation from the expected crash rate would be reflected by the MAD value for a particular model: a relatively greater value (compared with other competing models) would mean a greater deviation from the expected "true" crash rate; this deviation reflects inferior prediction performance.

DATA PREPARATION

The data used in this study were collected from multiple sources in California: Statewide Integrated Traffic Records System, Highway Performance Monitoring System, California Department of Finance, and Southern California Association of Governments (SCAG). Collisions from various counties in California that occurred from 2008 to 2013 were obtained from the Statewide Integrated Traffic Records System, which contains different severity levels. Given the underreporting issue related with property-damage-only collisions, the study focused only on the total fatal and injury collisions of the counties (*33*). In addition, a main exposure-related factor of county safety performance, DVMT (*12*), was collected from the Highway Performance Monitoring System for the same time periods. Furthermore, a main demographic factor, population, was gained from the California Department of Finance for the use in population-based gravity model. In addition, the data for BL and the geometric centroid distances between all counties were obtained from SCAG. All the distance-based models were based on the other data set, which included the distances between centroids of various counties. Table 1 shows the characteristics of first- and second-order adjacency-based matrices, BL, and distance-based matrices. Table 2 depicts the summary information for the various counties in terms of population, DVMT, BL, distance, and total number of fatal and injury collisions.

TABLE 1 Characteristics of Neighborhood Weight Matrices

Weight Matrix	Min.	Max.	Median	Mean	SD	Sum
Queen-1	2	8	5	4.91	1.32	285
Queen-2	4	21	14	13.1	3.54	762
BL	2	8	5	4.91	1.32	285
Rook-1	2	8	4	4.29	1.23	249
Rook-2	4	17	13	12.1	3.36	700
Distance based	57	57	57	57	na	3,306

NOTE: na = not applicable.

RESULTS

The models were implemented in the freeware WinBUGS package using a Markov chain Monte Carlo algorithm (*17*). In the absence of strong prior information for factor effects and dispersion

TABLE 2 Descriptive Statistics of Collected Data of Various Counties

Variable	Year	Min.	Max.	Median	Mean	SD
Collisions	2008	34	52,896	791	2,993	7,299
	2009	31	51,371	769	2,868	7,076
	2010	34	50,683	669	2,821	6,985
	2011	26	50,989	730	2,789	7,010
	2012	25	51,207	697	2,801	7,065
	2013	30	51,502	689	2,755	7,101
Population	2008	1,214	10,347,422	180,923	656,696	1,469,310
	2009	1,194	10,398,067	182,519	662,962	1,478,749
	2010	1,177	9,840,555	179,588	644,265	1,408,182
	2011	1,113	9,866,172	179,134	647,470	1,413,526
	2012	1,088	9,923,806	180,800	652,028	1,422,391
	2013	1,078	10,002,804	181,150	657,967	1,434,566
DVMT (mi)	2008	168,265	214,971,058	5,005,121	15,387,562	31,617,871
	2009	170,690	214,236,853	4,836,964	15,317,670	31,469,076
	2010	169,420	211,876,665	5,448,915	15,482,772	31,148,547
	2011	164,587	214,458,135	4,761,505	15,353,471	31,594,730
	2012	166,923	214,482,442	4,551,148	14,768,115	31,478,320
	2013	165,180	215,817,520	4,462,740	14,924,626	31,747,694
BL (mi)	na	0.04	204	37	45	34
Distance between county centroids (mi)	na	25	962	227	273	176

parameters, uninformative priors were assumed with normal distribution (0, 1,000) for all regression coefficients, and with gamma distribution (0.001, 0.001) for precision estimates. In the model calibration, the first 5,000 samples were discarded as a burn-in and a further 25,000 iterations were run; these iterations were used in the calculation of the posterior estimates. Convergence was assessed with the Gelman–Rubin convergence statistic and visual inspection of the trace plot (*34*). Plus, the sample Markov chain errors of all parameters were less than 5% of the associated standard deviation. For the gravity matrix, the neighbors were weighted according to the size of the county population and DVMT, but they were not used as independent variables for development of the models.

Modeling Results

Mean and Significance

Table 3 shows the posterior mean and standard deviation of the coefficient estimates (intercept, β_1, and γ) for all 18 models; these values indicate their quantitative proportionality on the crash rate by DVMT. The coefficient for yearly time trend, β_1, is statistically significant for all models except the base one. The base model, which included only heterogeneity random effects, was observed to have a four times larger standard deviation for β_0 and β_1 than rest of the spatial models. Apparently, the loss of precision due to greater variance was so high that β_1 was rendered insignificant considering the 95% confidence interval. Such a significant loss of precision could be attributed to the exclusion of spatial correlations because all other spatial models had a remarkably higher precision. Moreover, the autocorrelation coefficient (γ) appeared to be statistically significant for all models; this finding indicates the strong serial correlation between crashes of successive years.

The lower value of DIC indicates a better fit of the model. As shown in Table 3, the values for the 18 matrices range from the lowest (2,825.25) for Decay-50 to the highest (3,046.76) for the base. The value of fit ($\bar{D}$) is mostly similar for all the matrices, but DIC varies greatly since it is governed mainly by the complexity (p_D), which has a wide range across the models with the lowest for Decay-50 (59.9) and highest for the base (280.5). The two gravity models also exhibit complexities comparable with the base model. The increase in complexity reflects more effective parameters in the model. Usually, such an increase in complexity is compensated for by better fitness to accomplish a lower DIC value. But the $\bar{D}$-values of these three matrices are similar to the rest; this finding indicates that better fitness was achieved at the cost of a much larger number of effective parameters.

Because this study employed both adjacency- and distance-based matrices, the authors investigated the possible influence of the number of neighbors involved in a matrix on the goodness of fit. The DIC and the average number of neighbors for all matrices were plotted (Figure 2*a*). There was a significant difference between the number of neighbors for adjacency- and distance-based matrices and also within the adjacency matrices themselves (Table 1). This difference is clearly shown in Figure 2*a*, where similar matrices are aggregated. A drop in DIC is observed among the adjacency-based matrices, with Rook-2 and Queen-2 demonstrating better fitness of the model with the real data. Both these matrices had twice the number of neighbors than their corresponding first-order matrices. As for the Queen-1 and BL, BL has a lower DIC even though they have the same number of neighbors. This finding is probably due to the increased information embedded in the BL matrix or the length

TABLE 3 Parameter Estimates and Comparison of Model Fit

Model	β_0 Mean (SD)	β_1 Mean (SD)	γ Mean (SD)	Goodness of Fit $\bar{D}$	p_D	DIC	Average Predicted Crash Rate (10^{-4})	Fraction
Base	−8.702 (0.058)	−0.022 (0.016)	0.238 (0.012)	2,766.23	280.53	3,046.76	1.66674	NA
Queen-1	−8.701 (0.015)	−0.022 (0.006)	0.062 (0.019)	2,764.72	219.74	2,984.46	1.66404	0.629
Queen-2	−8.699 (0.014)	−0.023 (0.006)	0.058 (0.021)	2,764.42	168.70	2,933.12	1.66426	0.629
Rook-1	−8.699 (0.013)	−0.023 (0.006)	0.058 (0.019)	2,765.18	236.28	3,001.47	1.66379	0.631
Rook-2	−8.697 (0.014)	−0.023 (0.006)	0.056 (0.021)	2,764.42	178.44	2,942.86	1.66413	0.629
BL	−8.699 (0.014)	−0.023 (0.006)	0.058 (0.023)	2,755.21	207.30	2,972.51	1.66381	0.629
D0 (equal distance)	−8.698 (0.014)	−0.023 (0.006)	0.051 (0.020)	2,765.58	230.39	2,995.98	1.66452	0.629
D0.5 (1/distance$^{0.5}$)	−8.698 (0.014)	−0.023 (0.006)	0.053 (0.022)	2,765.31	206.32	2,971.63	1.66444	0.629
D1 (1/distance)	−8.695 (0.014)	−0.024 (0.006)	0.053 (0.020)	2,765.10	167.78	2,932.88	1.66431	0.629
D2 (1/distance^2)	−8.701 (0.013)	−0.022 (0.006)	0.057 (0.022)	2,765.25	198.28	2,963.54	1.66416	0.631
D3 (1/distance^3)	−8.702 (0.013)	−0.022 (0.006)	0.056 (0.022)	2,765.57	190.15	2,955.73	1.66398	0.631
Decay-50	−8.699 (0.013)	−0.023 (0.006)	0.057 (0.022)	2,765.33	59.91	2,825.25	1.66412	0.629
Decay-100	−8.701 (0.014)	−0.022 (0.006)	0.054 (0.021)	2,765.06	218.85	2,983.91	1.66426	0.629
Decay-150	−8.699 (0.014)	−0.023 (0.006)	0.054 (0.022)	2,765.01	232.48	2,997.50	1.66437	0.629
Decay-200	−8.696 (0.014)	−0.024 (0.007)	0.056 (0.022)	2,764.91	147.48	2,912.39	1.66439	0.629
Decay-250	−8.697 (0.014)	−0.023 (0.006)	0.052 (0.021)	2,765.21	186.63	2,951.84	1.66445	0.629
Gravity–population	−8.705 (0.014)	−0.021 (0.005)	0.077 (0.015)	2,764.96	269.72	3,034.69	1.66401	0.627
Gravity–DVMT	−8.706 (0.017)	−0.021 (0.007)	0.077 (0.015)	2,764.62	265.82	3,030.45	1.66419	0.626

NOTE: Nonsignificant variable is shown in shaded cell. Base model is one without considering spatial correlation among counties. Refer to Figure 1 for details of weight matrices classification.

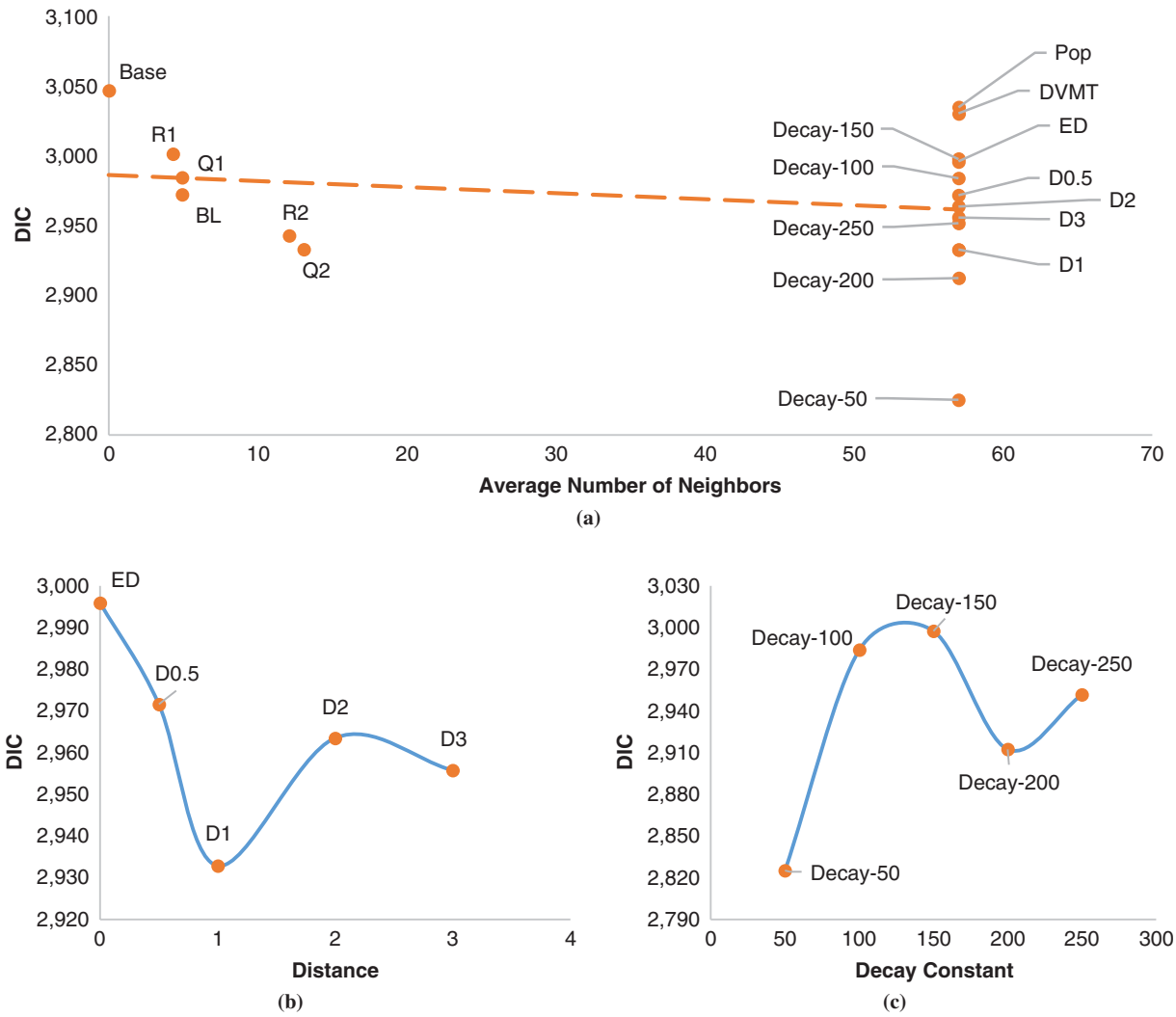


FIGURE 2 DIC trend (*a*) with corresponding number of neighbors by matrix, (*b*) by pure distance matrix, and (*c*) by decay matrix (R = rook; Q = queen).

of the common boundary. An overall downward linear trend is observed with an increase in the number of neighbors, indicating that distance-based models had a better fitness collectively. However, it is clearly shown in Figure 2*a* that the distance models were scattered in a wide range on the DIC scale. In order to better understand their behavior, they were split into two parts—pure distance and decay models—and their DIC trend is shown in Figure 2, *b* and *c*, respectively. As shown, both groups have a variable trend. In the case of pure distance matrices, ED had the worst fit and then the DIC score dropped steeply until D1, which had a difference of 63 points from ED. The model fitness again drops for D2 and finally improves for D3. A similar highly variable trend was exhibited by the decay matrices, with the decay constant of 50 revealing the best fit, which abruptly drops (DIC rises) for Decay-100 and then seems to oscillate until Decay-250. Overall, the trend implies that careful consideration should be placed when dealing with distance-based weight matrices when the distance has a wide range and high variance. It is therefore highly recommended that a sensitivity analysis be imperative for spatial models using distance-based matrices to ensure that the appropriate distance-based weight structures are selected.

Average Predicted Crash Rate

As shown in Table 3, the base model significantly deviates from the rest in terms of model predicted crash rate. In the case of distance-based models, a linear trend was observed. For decay models, the predicted crash rate reflected an increase for each increment of decay constant; it went from 1.66412 for Decay-50 to an upward limit of 1.66445 for Decay-250. A similar linear trend was noted from Models D0 to D3, where the average predicted crash rate decreased with an increase in the exponential power of the distance function. This variation among the estimation results of the models exhibits the tendency of the prediction performance to be based on the type of distance weight matrices.

Fraction

In terms of the fraction, which is a measure of the spatial random effect explained by the model, the decay models had marginal inferiority over pure distance as D2 and D3 were able to attribute a

relatively higher percentage of the variation to spatial effects. The gravity models had the lowest scores. It seems that the advantage of distance-based modeling was diluted because of the inclusion of extra information on populations and traffic exposure of the two counties under consideration.

Site-Ranking Evaluation Results

The 18 models were subjected to the aforementioned criteria for assessment of complexity and fitness. In accordance with previous studies (*35*), this study conducted the evaluation tests by comparing the site ranking of all models with Decay-50, since it displayed significantly better model fit and complexity. Table 4 exhibits the corresponding evaluation results.

Sensitivity, Specificity, PPV, and NPV

This study utilized a 10% threshold for filtering out the hazardous sites by considering the budget constraints used in practice. Table 4 demonstrates that except for the gravity models, all other models have values of 1 for both sensitivity and specificity. This finding means that these models screen out the same top 10% high-risk counties as did the Decay-50 matrix, and the proportion of counties identified by these models as high risk that are truly high risk is 100%. For the criteria of specificity and PPV, the base model has the smallest values, 0.8182 and 0.9853, respectively. These numbers suggest that about 18% of the counties identified by the base model as low risk are not truly safe, and the percentage of counties identified by the base model as high risk that are truly high risk is 98.53%. Overall, the base model and gravity models have the lowest site-ranking agreement with the Decay-50 model. It is interesting to note that the three models have the largest DIC difference with Decay-50. Such a phenomenon indicates that the model with closer predictive capability tends to flag more common high-risk counties.

Cohen's Kappa, TRD, and MAD

Kappa is another quantitative measure of agreement between Decay-50 and other models. Because all the models revealed 83% agreement, further scrutiny was required for assessment of the relative site-ranking agreement. The base model had the lowest agreement, closely followed by the gravity models. Again, such a trend was also witnessed for the model fitness assessment.

TRD and MAD are more sophisticated tools since they compare each county in a continuous fashion. Lower scores indicate greater agreement between the models and Decay-50. From Table 4, it is seen that under MAD, the base model has a significantly larger value of 2.2342, followed by the gravity models. A similar trend was found for TRD. Once more, these criteria show that the base model has the most ranking difference with Decay-50. Another noteworthy trend observed for TRD and MAD is the superior performance of most of the distance-based models with relatively lower values of TRD and MAD. For example, the lowest TRD score (52 for D2) belongs to a distance-based model, which is three times less than the worst adjacency-based model (176 for BL). The same set of models exhibit similar performance for MAD; the deviation observed for BL is twice that of D2. The potential rationale may be explained by the fact that distance-based weight matrices give a continuous output depending on the proximity between sites whereas the adjacency-

TABLE 4 Site-Ranking Evaluation Results of Alternate Models Compared with Decay-50

Model	Sensitivity	Specificity	PPV	NPV	Cohen's Kappa	TRD	MAD
Base	1	0.8182	0.9853	1	0.8365	517	2.2342
Queen-1	1	1	1	1	0.8380	130	0.5115
Queen-2	1	0.9545	0.9963	1	0.8384	116	0.4490
Rook-1	1	1	1	1	0.8380	136	0.4927
Rook-2	1	0.9545	0.9963	1	0.8384	124	0.4562
BL	1	1	1	1	0.8380	176	0.6213
D0 equal distance	1	0.9545	0.9963	1	0.8384	110	0.4708
D0.5 (1/distance$^{0.5}$)	1	0.9545	0.9963	1	0.8384	90	0.4380
D1 (1/distance)	1	0.9545	0.9963	1	0.8384	71	0.3579
D2 (1/distance^{2})	1	1	1	1	0.8380	52	0.3085
D3 (1/distance^{3})	1	1	1	1	0.8380	94	0.3934
Decay-100	1	1	1	1	0.8380	57	0.3458
Decay-150	1	1	1	1	0.8380	72	0.3259
Decay-200	1	0.9545	0.9963	1	0.8384	81	0.3724
Decay-250	–1	0.9545	0.9963	1	0.8384	85	0.4137
Gravity–population	0.9963	1	1	0.9565	0.8379	170	0.9970
Gravity–DVMT	0.9963	1	1	0.9565	0.8379	170	0.8968
Correlation coefficient between each criterion and DIC	**–0.55 (0.0224)**	–0.13 (0.6187)	–0.13 (0.624)	**–0.55 (0.0224)**	**–0.71 (0.0015)**	**0.60 (0.01)**	**0.68 (0.003)**

NOTE: Base model is one without considering spatial correlation among counties. Refer to Figure 1 for details of weight matrices classification. Statistical significances of correlation coefficients shown in parentheses. Statistically significant correlation coefficients shown in bold.

based matrices are restricted to a binary output. Also, distance-based matrices incorporated more neighbors.

Correlation Between Each Criterion and DIC

As discussed earlier, the foregoing criteria seemed to have some correlation with the modeling goodness-of-fit measure: DIC. The authors were encouraged to run the Pearson's correlation between each site-ranking criterion and DIC. The correlation coefficients and associated p-values are shown in the last row of Table 4. It is clearly shown that, except for specificity and PPV, all other criteria demonstrated a statistically significant correlation with DIC. Plus, the associated coefficients were high in magnitude as well. This finding suggests that the models with closer fitness and predictive performance tend to have more similar site-ranking performance.

CONCLUSIONS AND RECOMMENDATIONS

The results from the comparison of different weight matrix structures yielded the following major findings:

1. The highest values of $\bar{D}$, p_D, and DIC indicated that the inclusion of the spatially structured random effect can fit the data better and reduce the effective number of parameters in the model, thereby leading to higher predictive capability.
2. The modeling performance appeared to increase with the increase in number of neighbors in the weight matrices. For example, Queen adjacency performed better than Rook adjacency, the second-order adjacency performed better than first-order adjacency, and Decay-50 claimed the remarkably lower DIC than others.
3. The distance-based models had more neighbors in the weight matrices and tended to have better modeling performance than others. However, they also demonstrated larger variability of DIC compared with adjacency-based models.
4. The newly proposed gravity models appeared to fit the data slightly better than other distance-based models based on the $\bar{D}$-value. However, such a benefit was accompanied by a much larger effective number of parameters and therefore led to a higher DIC.
5. Most of the site-ranking evaluation criteria used in the study had a statistically significant correlation with DIC. This finding suggests that models with similar predictive capabilities tend to yield similar site-ranking performance.

Some recommendations for future research are the following:

1. In the current study, the different spatial-proximity matrices were compared by using the intrinsic Gaussian CAR prior. The other formulations of the CAR model such as proper Gaussian CAR, multivariate, and mixture models are worthwhile to examine and see if the results reported here can be replicated in those models.
2. Because the focus of this study was the comparison of different spatial-proximity matrices, no covariates were considered as influential factors for model development. But the authors acknowledge that the incorporation of significant covariates at the county level may affect the performance of the spatial models.
3. A sensitivity analysis is recommended for the distance-based matrices because an interesting (though not straightforward) trend was obtained in this study with variation of the constant in the case of decay models and the exponent of distance in the case of pure distance models.
4. Finally, such an investigation of spatial weight matrices should be conducted for different area levels (such as TAZs, census tracts, land assembly districts) so that the results of this study may be verified.

ACKNOWLEDGMENTS

Support for this study was provided by the 2016 UCCONNECT Summer Research Program at the University of California Polytechnic, Pomona. The authors thank SCAG for providing the geometric data for the counties used in this study. Finally, the authors also thank the anonymous reviewers for their constructive comments and suggestions on an earlier version of the paper.

REFERENCES

1. Song, J. J., M. Ghosh, S. Miaou, and B. Mallick. Bayesian Multivariate Spatial Models for Roadway Traffic Crash Mapping. *Journal of Multivariate Analysis,* Vol. 97, 2006, pp. 246–273. https://doi.org/10.1016/j.jmva.2005.03.007.
2. Guo, F., X. Wang, and M. A. Abdel-Aty. Modeling Signalized Intersection Safety with Corridor-Level Spatial Correlations. *Accident Analysis and Prevention,* Vol. 42, No. 1, 2010, pp. 84–92. https://doi.org/10.1016/j.aap.2009.07.005.
3. Cheng, W., G. S. Gill, R. Dasu, M. Xie, X. Jia, and J. Zhou. Comparison of Multivariate Poisson Lognormal Spatial and Temporal Crash Models to Identify Hot Spots of Intersections Based on Crash Types. *Accident Analysis and Prevention,* Vol. 99, 2017, pp. 330–341. https://doi.org/10.1016/j.aap.2016.11.022.
4. Levine, N., K. E. Kim, and L. H. Nitz. Spatial Analysis of Honolulu Motor Vehicle Crashes: I. Spatial Patterns. *Accident Analysis and Prevention,* Vol. 27, No. 5, 1995, pp. 663–674. https://doi.org/10.1016/0001-4575(95)00017-T.
5. Wang, X., and M. Abdel-Aty. Temporal and Spatial Analyses of Rear-End Crashes at Signalized Intersections. *Accident Analysis and Prevention,* Vol. 38, No. 6, 2006, pp. 1137–1150. https://doi.org/10.1016/j.aap.2006.04.022.
6. Shankar, V., F. Mannering, and W. Barfield. Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies. *Accident Analysis and Prevention,* Vol. 27, No. 3, 1995, pp. 371–389. https://doi.org/10.1016/0001-4575(94)00078-Z.
7. Abdel-Aty, M., and X. Wang. Crash Estimation at Signalized Intersections Along Corridors: Analyzing Spatial Effect and Identifying Significant Factors. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1953, 2006, pp. 98–111. https://doi.org/10.3141/1953-12.
8. Guo, F., X. Wang, and M. A. Abdel-Aty. Corridor-Level Signalized Intersection Safety Analysis Using Bayesian Spatial Models. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
9. Noland, R. B., and M. A. Quddus. A Spatially Disaggregate Analysis of Road Casualties in England. *Accident Analysis and Prevention,* Vol. 36, No. 6, 2004, pp. 973–984. https://doi.org/10.1016/j.aap.2003.11.001.
10. MacNab, Y. C. Bayesian Spatial and Ecological Models for Small-Area Accident and Injury Analysis. *Accident Analysis and Prevention,* Vol. 36, No. 6, 2004, pp. 1019–1028. https://doi.org/10.1016/j.aap.2002.05.001.
11. Abdel-Aty, M., C. Siddiqui, H. Huang, and X. Wang. Integrating Trip and Roadway Characteristics to Manage Safety in Traffic Analysis Zones. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2213, 2011, pp. 20–28. https://doi.org/10.3141/2213-04.
12. Miaou, S. P., J. J. Song, and B. K. Mallick. Roadway Traffic Crash Mapping: A Space-Time Modeling Approach. *Journal of Transportation and Statistics,* Vol. 6, 2003, pp. 33–58.
13. Noland, R. B., and L. Oh. The Effect of Infrastructure and Demographic Change on Traffic-Related Fatalities and Crashes: A Case Study of Illinois County-Level Data. *Accident Analysis and Prevention,* Vol. 36, No. 4, 2004, pp. 525–532. https://doi.org/10.1016/S0001-4575(03)00058-7.

14. Aguero-Valverde, J., and P.P. Jovanis. Spatial Analysis of Fatal and Injury Crashes in Pennsylvania. *Accident Analysis and Prevention,* Vol. 38, No. 3, 2006, pp. 618–625. https://doi.org/10.1016/j.aap.2005.12.006.
15. Huang, H., M. Abdel-Aty, and A. Darwiche. County-Level Crash Risk Analysis in Florida: Bayesian Spatial Modeling. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2148, 2010, pp. 27–37. https://doi.org/10.3141/2148-04.
16. Aguero-Valverde, J., and P. Jovanis. Analysis of Road Crash Frequency with Spatial Models. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2061, 2008, pp. 55–63. https://doi.org/10.3141/2061-07.
17. Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn. *WinBUGS User Manual,* Version 1.4. MRC Biostatistics Unit, Institute of Public Health, Cambridge, United Kingdom, 2003.
18. Aguero-Valverde, J., K.F.K. Wu, and E.T. Donnell. A Multivariate Spatial Crash Frequency Model for Identifying Sites with Promise Based on Crash Types. *Accident Analysis and Prevention,* Vol. 87, 2016, pp. 8–16. https://doi.org/10.1016/j.aap.2015.11.006.
19. Best, N., S. Cockings, J. Bennett, J. Wakefield, and P. Elliott. Ecological Regression Analysis of Environmental Benzene Exposure and Childhood Leukaemia: Sensitivity to Data Inaccuracies, Geographical Scale and Ecological Bias. *Journal of the Royal Statistical Society. Series A (Statistics in Society),* Vol. 164, No. 1, 2001, pp. 155–174. https://doi.org/10.1111/1467-985X.00194.
20. Rhee, K.A., J.K. Kim, Y.I. Lee, and G.F. Ulfarsson. Spatial Regression Analysis of Traffic Crashes in Seoul. *Accident Analysis and Prevention,* Vol. 91, 2016, pp. 190–199. https://doi.org/10.1016/j.aap.2016.02.023.
21. Dong, N., H. Huang, P. Xu, Z. Ding, and D. Wang. Evaluating Spatial-Proximity Structures in Crash Prediction Models at the Level of Traffic Analysis Zones. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2432, 2014, pp. 46–52. https://doi.org/10.3141/2432-06.
22. Besag, J., J. York, and A. Mollié. Bayesian Image Restoration with Two Applications in Spatial Statistics (with discussion). *Annals of the Institute of Statistical Mathematics,* Vol. 43, 1991, pp. 1–59. https://doi.org/10.1007/BF00116466.
23. Earnest, A., G. Morgan, K. Mengersen, L. Ryan, R. Summerhayes, and J. Beard. Evaluating the Effect of Neighbourhood Weight Matrices on Smoothing Properties of Conditional Autoregressive (CAR) models. *International Journal of Health Geographics,* Vol. 6, No. 1, 2007. https://doi.org/10.1186/1476-072X-6-54.
24. Best, N., S. Richardson, and A. Thomson. A Comparison of Bayesian Spatial Models for Disease Mapping. *Statistical Methods in Medical Research,* Vol. 14, No. 1, 2005, pp. 35–59. https://doi.org/10.1191/0962280205sm388oa.
25. Jones, A., and E. Huddleston. *SAS/STAT 9.2 User's Guide.* SAS Institute Inc., Cary, N.C., 2009.
26. Lawson, A.B., W.J. Browne, and C.L.V. Rodeiro. *Disease Mapping with WinBUGS and MLwiN,* Vol. 11. John Wiley & Sons, Inc., New York, 2003. https://doi.org/10.1002/0470856068.
27. Best, N.G., R.A. Arnold, A. Thomas, L.A. Waller, and E. Conlon. Bayesian Models for Spatially Correlated Disease and Exposure Data. In *Bayesian Statistics 6: Proceedings of the Sixth Valencia International Meeting,* Oxford University Press, 1999, pp. 131–156.
28. Spiegelhalter, D.J., N. 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, 2002, pp. 583–639. https://doi.org/10.1111/1467-9868.00353.
29. Parikh, R., A. Mathai, S. Parikh, G. Chandra Sekhar, and R. Thomas. Understanding and Using Sensitivity, Specificity and Predictive Values. *Indian Journal of Ophthalmology,* Vol. 56, No. 1, 2008, pp. 45–50. https://doi.org/10.4103/0301-4738.37595.
30. Byrt, T., J. Bishop, and J.B. Carlin. Bias, Prevalence and Kappa. *Journal of Clinical Epidemiology,* Vol. 46, No. 5, 1993, pp. 423–429. https://doi.org/10.1016/0895-4356(93)90018-V.
31. Cheng, W., and S. Washington. New Criteria for Evaluating Methods of Identifying Hot Spots. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2083, 2008, pp. 76–85. https://doi.org/10.3141/2083-09.
32. Huang, H., H. Chin, and M. Haque. Empirical Evaluation of Alternative Approaches in Identifying Crash Hot Spots: Naive Ranking, Empirical Bayes, and Full Bayes Methods. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2103, 2009, pp. 32–41. https://doi.org/10.3141/2103-05.
33. Ye, F., and D. Lord. Investigation of Effects of Underreporting Crash Data on Three Commonly Used Traffic Crash Severity Models: Multinomial Logit, Ordered Probit, and Mixed Logit. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2241, 2011, pp. 51–58. https://doi.org/10.3141/2241-06.
34. Brooks, S.P., and A. Gelman. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics,* Vol. 7, No. 4, 1998, pp. 434–455.
35. Aguero-Valverde, J. Full Bayes Poisson Gamma, Poisson Lognormal, and Zero Inflated Random Effects Models: Comparing the Precision of Crash Frequency Estimates. *Accident Analysis and Prevention,* Vol. 50, 2013, pp. 289–297. https://doi.org/10.1016/j.aap.2012.04.019.

The Standing Committee on Safety Data, Analysis, and Evaluation peer-reviewed this paper.