Comparative Evaluation of Temporal Correlation Treatment in Crash Frequency Modeling

Journal:	*Transportmetrica A: Transport Science*
Manuscript ID	TTRA-2017-0041.R2
Manuscript Type:	Full Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Cheng, Wen; California State Polytechnic University Pomona, Department of Civil Engineering Gill, Gurdiljot ; California State Polytechnic University Pomona Choi, Simon; SOUTHERN CALIFORNIA ASSOCIATION OF GOVERNMENTS Zhou, Jiao ; California State Polytechnic University, Pomona Jia, Xudong; California State Polytechnic University, Pomona Xie, Meiquan; California State Polytechnic University, Pomona
Keywords:	Serial Correlations, Time-Varying Coefficients, Autoregressive Process, Bayesian Hierarchical Approach, Structural Heterogeneities

For Review Only

SCHOLARONE™
Manuscripts

Comparative Evaluation of Temporal Correlation Treatment in Crash Frequency Modelling

Wen Cheng[a,*], Gurdiljot Singh Gill[b], Simon Choi[c], Jiao Zhou[d], Xudong Jia[e], Meiquan Xie[f]

** Corresponding author, [a] Associate Professor, Department of Civil Engineering*

California State Polytechnic University, Pomona

3801 W. Temple Ave., Pomona, CA, USA 91768

Fax: (909) 869-4342; Tel: (909) 869-2957; Email:wcheng@cpp.edu

[b] Graduate Student, Department of Civil Engineering

California State Polytechnic University, Pomona

3801 W. Temple Ave., Pomona, CA, USA 91768; Email: gurdiljotg@cpp.edu

[c] Chief of Research and Forecasting

SOUTHERN CALIFORNIA ASSOCIATION OF GOVERNMENTS

818 West 7th Street, 12th Floor, Los Angeles, CA, USA 90017

Fax: (213) 236-1963; Tel: (213) 236-1849; Email: choi@scag.ca.gov

[d] Graduate Student, Department of Civil Engineering

California State Polytechnic University, Pomona

3801 W. Temple Ave., Pomona, CA, USA 91768

For Peer Review Only

Fax: (909) 869-4342; Tel: (909) 869-4312; Email:jiaozhou@cpp.edu

[e] *Professor, Department of Civil Engineering*

California State Polytechnic University, Pomona

3801 W. Temple Ave., Pomona, CA, USA 91768

Fax: (909) 869-4342; Tel: (909) 869-4312; Email:xjia@cpp.edu

[f] *Visiting Scholar, Department of Civil Engineering,*

California State Polytechnic University, Pomona

3801 W. Temple Ave., Pomona, CA, USA 91768

& Assistant Professor, School of Transportation and Logistics,

Central South University of Forestry and Technology,

Tianxin, Changsha, Hunan, China, 410000

Fax: (909) 869-4342; Tel: (909)282-9142; Email: mxie@cpp.edu

Words : 8140

Acknowledgements

The authors thank Southern California Association of Governments for providing covariates and geometric centroid distance data of TAZ used in this study.

Funding

This work was supported by UCCONNECT under Grant 65A0529-052.3.

For Peer Review Only

Abstract

Compared with a large amount of research using different ways of addressing serial correlations among crash data, there is relatively little research dedicated to the evaluation of the different temporal treatments on modeling performance. This study proposed two new methods which combined the strengths of linear trend and time-varying coefficients with the autoregressive process and compared their performance with seven other temporal models used in the past. Nine years of crash data and other covariates associated with traffic analysis zones were used. The Bayesian hierarchical approach was employed to account for the structural heterogeneities.

The nine models generated similar number of statistically significant variables and close variable coefficients. However, the modelling performance was different based on a set of evaluation criteria. For prediction accuracy, the model which accounts only for the autoregressive effect illustrated the superior performance in terms of both cross validation and typical assessment based on the same data used to develop models. Nonetheless, if the penalized criterion (Deviance Information Criterion) was used, both of the newly proposed models outperformed other competing models, which indicates their capability to yield similar prediction accuracy with the relatively smaller effective number of parameters. Such fact suggests that more models which combine various temporal treatments are worth further exploration. Finally, the correlations were also observed among the various modeling assessment criteria.

Keywords: Serial Correlations; Time-Varying Coefficients; Autoregressive Process; Bayesian Hierarchical Approach; Structural Heterogeneities.

For Peer Review Only

1. Introduction

Crash prediction models have been used in research and practice for determination of influential factors, planning purposes or site ranking. Models of varying complexity have been employed, ranging from very basic to sophisticated. Some studies developed univariate models by focusing on a particular crash outcome or total crashes (Ulfarsson and Shankar 2003; Lord et al. 2005; Kim et al. 2006). These models operated on the assumption that crash outcomes are independent, while they were revealed to be multivariate due to sharing of unaccounted factors (Bijleveld 2005; Park and Lord 2007). Congdon (2001) observed that ignorance of such correlation structures may reduce the efficiency of the model due to lesser precise parameters.

Lately, more advanced models have been proposed to incorporate the correlation structures to account for the unobserved heterogeneity in crash data (MacNab 2004; Miaou and Lord 2003; Aguero-Valverde and Jovanis 2006; Quddus 2008; Lord and Mannering 2010). Significant correlations were observed by some research studies which jointly considered different crash severity levels (Tunaru 2002; Ladron de Guevara et al. 2004; Miaou and Song 2005; Song et al. 2006; Ma and Kockelman 2006). Some research studies (Ma et al. 2008; Aguero-Valverde and Jovanis 2009) noted that more precise estimates were obtained for model parameters with the inclusion of correlations in multivariate crash counts. Apart from the correlations of severity levels, some studies utilized spatial random effects to explore the spatial correlations between the crash sites at different area levels like intersections (Wang and Abdel-Aty 2006; Mitra 2009), segments (Aguero-Valverde and Jovanis 2008), corridors (Abdel-Aty et al. 2006; Guo et al. 2009), Census tracts (Narayanamoorthy et al. 2013), Traffic Analysis Zones (Washington et al. 2010; Xu and Huang 2015), counties (Miaou et al. 2003; Huang et al. 2010), and so on. The studies focused on simultaneously modelling crash types (Song et al. 2006;

For Peer Review Only

Aguero-Valverde et al. 2016; Cheng et al. 2017) observed that inclusion of correlated spatial random effects in the model significantly increased the fitness of model with the crash data and superior site ranking performance.

Another dimension of studies investigated the inclusion of serial correlations to benefit from the time-dependent factors which were not incorporated in the previous models (Andrey and Yagar 1993; Hay and Pettitt 2001; Wang et al. 2013). Wang et al. (2006) did a temporal analysis of rear-end collisions at intersections by utilizing generalized estimating equations with a negative binomial link function. The three-year longitudinal data for 208 signalized intersections served for the development of four models with different correlation structures: independent, exchangeable, autoregressive (AR), and unstructured. The autoregressive structure was observed to have best goodness-of-fit and an estimated correlation of 0.4454 for each successive two years. Huang et al. (2009) employed the same AR model with a time step of one year (lag-1), along with five other models, for empirical evaluation of identification of hotspots by different approaches. The study conducted on intersection crash data revealed that the models based on Full Bayesian (FB) hierarchical approach significantly outperformed others in hotspot identification. Based on the same approach, the AR-1 model was observed to have the best fit with the crash rate-related parameters, as assessed by three goodness-of-fit criteria: DIC (Deviance Information Criterion), MAD (mean absolute deviance), and MSPE (mean-squared predictive error). Aguero-Valverde (2013) did a segment-level comparative study based on the precision of crash frequency estimates of different random effects models. Two of the models considered for comparison were Fixed-over-time and independent-over-time random effects. The results established that by fixing the random effects over time, the model parameters were able to 'pool strength' from the neighboring years as the model fit and precision of estimates

For Peer Review Only

significantly increased. The evaluation of models also revealed that the concerned model performed consistently better at site ranking due to the notable reduction in standard errors of estimates. Jiang et al. (2014) used the site-specific fixed-over-time random effect to incorporate temporal correlations into a Poisson lognormal model for highway network screening. The comparison results with other competing models revealed that the inclusion of these correlations significantly improved the capability of the model to fit the crash data, which was in line with the observations made by a previous study (Aguero-Valverde, 2013). Moreover, the model exhibited a consistently superior performance for identification of hotspots. El-Basyouny and Kwon (2012) analyzed the yearly trend and random variation of parameters using yearly time period while investigating the effect of weather and time on crash types. Four multivariate Poisson lognormal models were developed using the Full Bayesian framework: with and without linear time trend, yearly varying intercept, and yearly varying coefficients. The results confirmed the superiority of the model with varying coefficients to possess the best fit based on DIC. This study used day-of-week as a proxy variable due to the unavailability of traffic exposure, and it was recommended to investigate different datasets with traffic exposure with the temporal models to confirm the findings. Along with a linear time trend, an ecological study by Earnest et al. (2007) used a Conditional Autoregressive (CAR) model to evaluate different neighbourhood weight matrices which incorporated quadratic time trend as well.

Compared with a large number of research studies using various temporal correlations among crash data, there is little research dedicated to the evaluation of the different temporal treatments on modelling performance. To add to the literature the much-needed research, the present study first compared seven alternate temporal models implemented in the past studies with varying complexity of random effects: (I) independent-over-time; (II) fixed-over-time; (III)

For Peer Review Only

linear time trend; (IV) quadratic time trend; (V) yearly varying intercept; (VI) yearly varying coefficients; and (VII) Autoregressive-1 (AR-1). In addition, this study also proposed two new methods which combine the strengths of previous models: the linear trend and AR-1, and the time-varying coefficients and AR-1. To the best knowledge of authors, such combination is first proposed in the traffic safety field, even though the similar combination of AR and time trend can be found in the current practice, such as the STEPAR (stepwise autoregressive method) specification provided in SAS/STAT 9.2 software package (Jones and Huddleston, 2009). It is important to note that, in addition to AR-1, there are a large number of models involving autoregressive error process (Miaou and Song 2005), such as higher order AR model AR(p), autoregressive–moving-average (ARMA), Integer valued autoregressive (INAR), Autoregressive Conditional Heteroscedastic (ARCH), and Generalized Autoregressive Conditional Heteroscedastic (GARCH), and so on. This research chose AR-1, which has often been used in traffic safety research, as a representative of the large body of AR-involved models.

As this study is primarily focused on assessment of temporal treatments, hence a same distance-based weight matrix was incorporated in all models to account for the possible spatial correlations among crash sites to ascertain that the differences observed in results are mostly influenced by the temporal random effects. However, the time and space interactions were not included which might blur the comparison of alternative temporal treatments. All models under investigation were evaluated based on the goodness-of-fit and accuracy of crash estimation. Firstly, DIC was used as a measure for assessment of the fitness of model estimates with the actual crash data. It is also utilized for determination of model complexity, or, number of effective parameters that are used for model development. Secondly, the models are evaluated based on the predictive density calculated by the leave-one-out cross-validation of LPML (log

For Peer Review Only

pseudo marginal likelihood). Third, other prediction criteria including MAD (mean absolute deviation), RMSE (root mean square error), and chi-square RSS (Residual Sum of Squares) were calculated to quantify the discrepancy in the crash estimation and observed crashes. Finally, a simple correlation analysis was done among all abovementioned assessment criteria.

The remainder of the paper first describes the development of nine models under the Full Bayesian framework, along with the convergence checks used. Then, the criteria for model fitness, validation, and empirical evaluation are described, followed by a discussion of results. Conclusions and recommendations follow.

2. Methodology

2.1 Model Development

This study used the Full Bayesian (FB) hierarchical approach to account for the structural heterogeneities, such as temporal and spatial, for development of crash frequency models. Several recent studies have revealed the capability of hierarchical modelling technique to better fit the crash count data by incorporating such heterogeneities (Song et al. 2006; MacNab 2003; Huang et al. 2008a; Huang et al. 2008b).

The model formulation is presented in order of complexity, from the independent-over-time random effects to more sophisticated proposed model, which is a combination of AR-1 and time-varying coefficient. The models under investigations are specified as below:

Model 1: Independent-over-time random effects

At the first step of the hierarchical approach, the crash count is modeled as the Poisson process:

For Peer Review Only

$$y_{it} \sim poisson\ (\theta_{it}) \tag{1}$$

Where, y_{it} is the observed crash count at zone *i* in time period *t* and θ_{it} is the mean expected crash count for site *i* in time period *t*.

The expected crash count is modeled as a function of covariates and random effects, as shown in the equation:

$$\log(\theta_{it}) = \beta_0 + \beta_k X_k + \emptyset_i + \varepsilon_{it} \tag{2}$$

Where, β_0 is the vector of intercept, β_k is the vector of independent coefficients, X_k is the vector of independent covariates, $\emptyset_i$is the spatially structured random effect, and ε_{it} is the error term for TAZ *i* at time *t*. Since the main focus of this study was to compare the influence of time on the crash count, hence the spatial random effects term ($\emptyset_i$) was separately introduced in the model (split from random effect) to avoid blurring the time comparison. The spatial correlation was accommodated by introducing the CAR (conditional autoregressive) prior for the spatial random effects (Lawson et al., 2003). A plethora of spatial weight matrices have been employed in safety research (Gill et al., 2017; Jonathan et al., 2016). The present study calculated the distance-based weight matrix where the reciprocal of the distance between two TAZs is considered.

At the second step, the coefficients of covariates (β_k) are modeled using non-informative Normal priors (i.e. $\beta_k \sim$ Normal (0, 10^6)) while the error term and associated variance are modelled with normal and gamma distributions, respectively (Bernardinelli et al. 1995):

For Peer Review Only

$$\varepsilon_{it} \sim normal\ (0, \sigma_{it}^2\ ) \tag{3}$$

$$\sigma_{it} \sim gamma\ (0.001, 0.001) \tag{4}$$

Model 2: Fixed-over-time random effects

According to Equation 3, the error term ε_{it} varies across road sites and over time. However, it can be argued that the same zone shares identical unobserved features over the years. Hence, an identical site-specific random effect across years is added taking the following form:

$$\log(\theta_{it}) = \beta_0 + \beta_k X_k + \emptyset_i + \varepsilon_i \tag{5}$$

$$\varepsilon_i \sim normal\ (0, \sigma_t^2\ ) \tag{6}$$

Model 3: Linear time trend

In this model, a linear time trend is assumed by employing time as a covariate

$$\log(\theta_{it}) = \beta_0 + \beta_k X_k + \beta_{k+1} t + \emptyset_i + \varepsilon_i \tag{7}$$

Where β_{k+1} is the scalar parameter for linear yearly trend and assigned with non-informative Normal prior (i.e., $\beta_k \sim$ Normal (0, 10^6))

Model 4: Linear as well as quadratic time trend

To explore a more complex trend of time, a quadratic temporal random effect term is also added (Cheng et al., 2017b):

For Peer Review Only

$$\log(\theta_{it}) = \beta_0 + \beta_k X_k + \beta_{k+1} t + \beta_{k+2} t^2 + \emptyset_i + \varepsilon_i \quad (8)$$

Where β_{k+2} is the parameter for quadratic yearly trend and assigned with the same non-informative Normal prior as β_{k+1}.

Model 5: Varying intercept

This model has an intercept which varies with the yearly time period for every site

$$\log(\theta_{it}) = \beta_{0t} + \beta_k X_k + \emptyset_i + \varepsilon_i \quad (9)$$

Where β_{0t} is the vector of yearly-varying intercept. Each intercept was assigned with a non-informative Normal prior.

Model 6: Varying coefficients

This model includes the yearly varying coefficients for the intercept as well as covariates.

$$\log(\theta_{it}) = \beta_{0t} + \beta_{kt} X_{kt} + \emptyset_i + \varepsilon_i \quad (10)$$

Where, β_{kt} is the matrix of time-varying coefficients of independent covariates. Each element of the matrix was assigned with a non-informative Normal prior (i.e. $\beta_{kt} \sim$ Normal (0, 10^6)).

Model 7: Autoregressive-1 (AR-1)

This model accounts for the autoregressive safety effect by specifying the distribution of ε_{it} as a lag-1 dependence on errors, where lag-1 means that the time is varying yearly. It incorporates the weighted sum of the past one year of values together with a random term. We chose AR-1 based

For Peer Review Only

on the assumptions of stationarity restriction.

$$\log(\theta_{it}) = \beta_0 + \beta_k X_k + \emptyset_i + \varepsilon_{it} \tag{11}$$

The weighted sum is fixed, and the random terms change at every time step following the same distribution, which means this model is homoscedastic. The distributions are given by:

$$\varepsilon_{it} \sim normal\left(0, \frac{\sigma_{it}^2}{(1-\gamma^2)}\right) \tag{12}$$

$$\varepsilon_{it} \sim normal\left(\gamma\varepsilon_{i,t-1}, \sigma_{it}^2\right) \quad \text{for } t > 1 \tag{13}$$

Where γ is the autocorrelation coefficient with the range of -1< γ < 1 (Jung, Kukuk, and Liesenfeld, 2006). If γ is close to 0, then the random process looks like white noise, but as $|\gamma|$ approaches 1, the value of present year gets a larger contribution from the value of previous year relative to the noise. The coefficient γ was assigned with a uniform prior (Johnson and Hoeting, 2003).

Model 8: Combination of AR-1 and linear time trend

Based on the modelling results pertaining to the posterior deviance and effective parameters employed for all models, the authors developed two more models by combining the strengths of different specifications. The autoregressive specification (Model 7) employs lag-1 dependence of errors where the prior year is taken into consideration. The linear trend (Model 3) portrays a different treatment of time by considering the yearly time interval as an explanatory variable which may be a viable significant factor for the crash occurrence. Model 8 combines the

For Peer Review Only

capabilities of both the aforementioned models and the model takes the following form:

$$\log(\theta_{it}) = \beta_0 + \beta_k X_k + \beta_{k+1} t + \emptyset_i + \varepsilon_{it} \quad (14)$$

Where the error term (ε_{it}) follows the distribution from Model 7 and rest of the variables follow the specifications as previously stated.

Model 9: Combination of AR-1 and time-varying coefficients

Ae mentioned previously, Model 8 is similar to the STEPAR specification provided in SAS/STAT 9.2 software package, which combines a time trend regression with an autoregressive model for departures from trend (Lunn et al., 2000). To further explore the performance of the combination of different temporal treatments, the authors also proposed a new model, or, Model 9, which is the mixture of AR-1 and time-varying coefficients. The common trends of the proposed models are expected to yield more reliable performance of the combination models relative to the conventional models. Likewise, the form of Model 9 can be expressed as follows:

$$\log(\theta_{it}) = \beta_{0t} + \beta_{kt} X_{kt} + \emptyset_i + \varepsilon_{it} \quad (15)$$

Where the error term (ε_{it}) follows the distribution from Model 7 and rest of the variables follow the specifications as previously stated.

2.2 Convergence Checking

These nine models were estimated with the Full Bayesian techniques using the open source software WinBUGS (Brooks and Gelman 1998). While fitting these models to the crash data, summary statistics of the posterior inference of parameters were obtained via two chains with 30,000 iterations, initial 10,000 of which were discarded as a burn-in sample. Convergence was

For Peer Review Only

monitored to ensure that posterior distribution has been found and sampling should be initiated (El-Basyouny and Kwon 2012). Convergence was assessed by visual inspection of the Markov chains trace plots for the model parameters. Moreover, the number of iterations was selected such that the ratios of Monte Carlo error for each parameter in the model relative to standard deviations would be less than would be less than 0.05. Also, the convergence of multiple chains was assessed by ensuring the value of Brooks–Gelman–Rubin (BGR) statistic (Spiegelhalter et al. 2002) to be less than 1.2.

2.3 Model Comparison

The present study employed a set of evaluation criteria to compare the modelling performance which contains the penalized goodness-of-fit (GOF), cross validation as well as prediction assessment based on in-sample date. The purpose is two-fold: the various criteria not only demonstrate the performance of models from different angles, but also allow us to check the possible correlation of different assessment measures. The details of each criterion are described in order as follows.

2.3.1 Deviance Information Criterion (DIC)

This study used the penalized goodness-of-fit criterion of DIC for model comparison as it accounts for the complexity of the model inferred from the use of effective number of parameters. DIC is a hierarchical Bayesian equivalent of Akaike's Information Criteria (AIC), which was proposed by Spiegelhalter et al. (2003) to account for model fit and complexity. DIC is especially useful for hierarchical model structures since the number of parameters employed for model development is not always evident. Specifically, DIC is defined as an estimate of fit plus twice the effective number of parameters:

For Peer Review Only

$$DIC = D(\bar{\theta}) + 2p_D = \bar{D} + p_D \tag{16}$$

Where $D(\bar{\theta})$ is the deviance evaluated at the posterior means of the parameters of interest ($\bar{\theta}$), and posterior mean deviance $\bar{D}$ can be taken as a Bayesian measure of fit or "adequacy". p_D denotes the effective number of parameters in a model, which reflects the complexity of the 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 by Spiegelhalter et al. (2003), a difference of 7+ points in the DIC is treated as significant for modelling performance, and the lower value of DIC indicates superior goodness-of-fit.

2.3.2 Log Pseudo Marginal Likelihoods (LPML)

This criterion computes the conditional predictive ordinates (CPOs) by using the cross-validation predictive densities (Gelfand, 1996; Mukhopadhyay and Gelfand, 1997). The CPOs allow the estimation of LPML and pseudo Bayes factor (PBF), which gives the measure of superiority among two competing models (Ntzoufras, 2009; Heydari et al., 2016). The predictive capability was checked based on a special case of cross-validation, that is, leave-one-out cross-validation (LOOCV). This technique is similar to the conventional cross-validation as the dataset comprising of crash counts at individual locations are split up into two subsets, where one acts as validation set and the other as training set. However, in the case of LOOCV, the validation set only contains one observation while rest of them make up the training set. This is contrary to the conventional approach where the original dataset is usually divided into equal subsets. LOOCV approach allows a significant reduction of test errors and leads to relatively unbiased estimates. This approach also allows the capability to handle the randomness of results usually associated

For Peer Review Only

with the selection bias introduced in the conventional method since LOOCV repeatedly fits the entire dataset by choosing a different observation as the validation set till each location has been incorporated (James et al., 2013). In general, Under the MCMC framework, calculation of LPML can be done through the following expression:

$$\begin{cases} CPO = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{p(Y_i|\beta^{(t)})}\right)^{-1} \\ LPML = \sum_{i=1}^{n}\log(CPO_i) \end{cases} \tag{17}$$

Where CPO is conditional predictive ordinate; Y_i is the ith observation (i = 1, 2, 3, . . ., n) for all TAZs; β is the vector of estimated model parameters; $p(Y_i|\beta^{(t)})$ is the likelihood; t represents the number of iterations in MCMC simulation. The larger the LPML value, the better the predictive performance tends to be.

2.3.3 Mean Absolute Deviation (MAD)

MAD is a popular criterion frequently employed by the researchers to check the fitness of data irrespective of its distribution. It is simply based on the model deviation or residue. The corresponding expression is shown in Equation 18.

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|Y_i - O_i| \tag{18}$$

Where Y_i is the Bayesian-estimated crash frequency and O_i is the observed crash count for TAZ i by a model during the same time period. The smaller the MAD value, the better fitness to the data.

2.3.4 Root Mean Square Error (RMSE)

RMSE essentially represents the sample standard deviation for the discrepancy between

For Peer Review Only

Bayesian estimated crash counts and the observed crashes for the specific TAZ. Specifically, RMSE assumes the form shown below:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - O_i)^2} \tag{19}$$

Where terms are as defined previously. Again, the smaller RMSE is desirable as it indicates superior performance due to lesser discrepancy between observed and estimated crash counts.

2.3.5 Chi-squared Residual Sum of Squares (RSS)

The Chi-squared RSS is defined as:

$$RSS = \sum_{i=1}^{n} \frac{(O_i - Y_i)^2}{Y_i} \tag{20}$$

The TAZs with considerably larger crash counts may be penalized more under the MAD and RMSE due to greater residues. RSS (Cheng and Washington, 2008) tends to remove such bias by calculating the squared residual relative to the estimated number of crashes. The smaller RSS value indicates reliability in terms of accuracy.

3. Data Description

This study was conducted on the data of 203 TAZs from the city of Irvine, California. Various other studies have focused on the TAZs for planning-level analysis of crashes as they have benefits of better homogeneity and easy integration into the transportation planning process (Siddiqui et al. 2012; Abdel-Aty et al. 2013; Pulugurtha et al. 2013). As the objective of this study required the development of crash prediction models, while accounting for spatial and

For Peer Review Only

temporal correlations, an array of variables was employed. In specific, 9-year crash data (2003-2011) used as independent variables were collected from Statewide Integrated Traffic Records System (SWITRS), and other covariates containing socioeconomic factors, transportation-related information and road environment factors covering the same period were provided by SCAG (Southern California Association of Governments). The relatively long period was selected to ensure a clean serial trend in crashes since AR-1 model is reliant on the previous one year's crash count. The crash dataset was comprised of fatal and injury collisions only, without the inclusion of property damage only (PDO) due to the underreporting issue related to those crashes (Ye and Lord 2011). The map showing the TAZs and the distribution of hotspot and cold spots based on observed crash count is presented in Figure 1. Daily Vehicle Miles Travel (DVMT) was incorporated as the traffic exposure factor (Miaou et al. 2003). As the models also incorporated the spatial correlations, the distance between the geometric centroids of TAZs was also obtained from SCAG. As the city of Irvine has 203 TAZs, so the distance matrix had the size of 203x202 with a minimum value of 0.16 miles and maximum of 13.21 miles. Summary information for the various dependent and independent variables for the TAZs is shown in Table 1. As this research was done at the planning level, so we used similar a set of variables in accordance with the SCAG. The statistics reflect the respective measure (maximum, minimum, and so on) for a particular year while considering the whole dataset of 9 years (2003-2011). It is important to note that many of the variables have 'zero values' for the minimum statistics which may be attributed to the fact that some of the areas (TAZs) differed from the rest with respect to the 'area type'. For example, some areas may be completely vacant, farms, or only serve as office destinations without any permanent residents.

For each model, preliminary multi-collinearity tests and stepwise selection methods were

For Peer Review Only

employed in selecting the best subset of predictors for final model development. To obtain a parsimonious model and ensure the most appropriate subset of variables was entered for model development, various techniques were employed for detection of correlated variables. First, the correlation tests were conducted using the Harrell Miscellaneous package in R software which allowed the calculation of Pearson correlation coefficient. The variables observed to be correlated at a significance level of 0.05 were eliminated in multiple steps. Second, the autocorrelation plots generated for variables under the sample monitor menu were also checked by visual inspection to ensure each variable followed an expected shape which inferred absence of any significant autocorrelation among the two chains and ensured structural robustness of the model. Only the variables which satisfied the criteria were incorporated for model development.

[Figure 1 about here]

[Table 1 about here]

4. Results

4.1 Model Estimates

Only the statistically significant variables were shown in Table 2. It should be noted that since the main objective of this study was to explore the similarities and differences among the alternate modelling approaches rather than investigation of influential factors, hence the explanation of results is focused in that direction hereafter. As evident from Table 2, the models had a similar set of variables that were identified to have a significant impact on crash risk. The traffic exposure precursor of VMT and household density were consistently observed to be the two significant factors across all models. Such consistency demonstrates the robustness of all the developed models. Models 3, 4, and 8 captured an additional influential factor (linear trend)

For Peer Review Only

which was consistently observed to have a negative correlation with crash frequency. This capability of the concerned models allowed the quantification of this relationship, i.e., the crashes are expected to decrease with the passage of time, hence resulting in less biased estimates for the other explanatory factors. This clarity will help generate more meaningful and accurate inferences for regulating the influential factors. In the case of models with AR specification (Models 7~9), the autocorrelation coefficient (γ) was statistically significant in Models 7 and 9. The exception of Model 8 may be attributed to the inclusion of linear trend which may have captured the serial correlation to some degree and hence rendering the γ non-significant. The largest value of γ was observed for Model 9 (γ = 0.86), which is closer to 1 while the other two models' γ values incline more towards zero. This indicates that in the case of the model with time-varying coefficients (Model 9), the AR specification generated a larger contribution from the previous year for estimation of crash counts for the present year whereas for the other two models it resembled the white noise.

[Table 2 about here]

4.2 Model Fitness and Predictive Accuracy

Table 3 shows the comparison results for model fitness ($\bar{D}$) and the effective number of parameters (Pd). A substantial model-fitting improvement was observed in the models which accounted for serial autocorrelations (Models 7~9). The models with independent and fixed-time random effects (Models 1 and 2) had a 55-point difference of DIC value compared to the best performing model (Model 8). Contrary to the expectation that fixed over time random effects (Model 2) may improve model fit by reducing the degree of freedom, the similar DIC as well as $\bar{D}$ and p_D indicate that this specification employed similar number of effective parameters and

For Peer Review Only

lend no significant advantage at model fit. The addition of linear or quadratic trend improved model fit without any substantial increase in the number of model parameters, as reflected by similar values of p_D for Models 1~4. The time-varying intercept or coefficients seemed to employ larger number of effective parameters (significantly larger in the case of Model 6), but the corresponding reduction in posterior deviance balanced the overall goodness-of-fit performance (DIC). The AR models had the lowest DIC even though they employed the largest number of effective parameters, as reflected by their highest values of p_D. On closer scrutiny, it was revealed that the incorporation of γ significantly reduced the deviance ($\bar{D}$), which compensated for the increase in DIC due to the inclusion of more effective parameters. The worst model fitness was exhibited by Models 1 and 2, while Models 3, 4, and 5 had almost similar DIC values, considering that the difference of 7 points in DIC reflects equivalent goodness-of-fit. It is noteworthy that the inclusion of more time-varying coefficients (Model 6) significantly reduced the deviance (average difference of around 40) but the effective parameters increased due to the addition of yearly-varying coefficients for each covariate. However, the remarkable reduction of posterior deviance due to the addition of more parameters eventually compensated for the potentially raised complexity and led to a superior overall fit, with a difference of 9 points in the DIC compared to Model 5. This demonstrates the advantage associated with such models to fit the crash data better as they allow the flexibility to accommodate the yearly-varying influential factors. Given the consideration for fit and complexity, the proposed model of AR-1 and linear trend (Model 8) exhibited the best performance as assessed by the penalized DIC. The second proposed model (Model 9) exhibited similar advantage at overall fit with a difference of 1 point in DIC. It is noteworthy, though, that the two proposed models seem to adopt different approaches for generating superior fit. Model 8

For Peer Review Only

employed lesser number of effective parameters as it incorporated only the linear trend as an addition to the regular AR specification. On the other hand, Model 9 employed the advantage of time-varying coefficients which allowed significant reduction of posterior deviance (13-point difference in $\overline{D}$), while simultaneously utilizing larger number of parameters (15-point difference in Pd) for estimation. These models offer the balance between fit and complexity as they borrow the strengths from different specifications. The linear trend and time-varying coefficients allowed the maintenance of an optimum number of effective parameters (considerably lower than traditional AR specification), while adopting the advantage of the lowest deviance from AR (Model 7). Clearly, such combinations keep a check on effective number of parameters while maintaining sufficient deviance to reach the lowest DIC indicating the superior flexibility to fit the crash data.

The cross-validation criterion of LPML and three other criteria were employed to evaluate the model accuracy from different perspectives. Relatively higher LPML and lower MAD, RMSE, and RSS values indicate better capability to predict the crash count in an area. This reflects the accuracy of these methods to predict crashes with least deviation from the observed count. Model 7 consistently outperformed the competing models by demonstrating higher accuracy for crash prediction. This superiority may be attributed to the lowest posterior deviance of Model 7 which enabled it to fit the observed crash data in a more refined manner. Overall, the AR models performed better than the rest across all evaluation criteria. The worst values were observed for Model 1 and Model 2, while the addition of linear or quadratic trend improved the prediction capabilities, which could be attributed to the inclusion of quadratic time trend for providing a subtler fitting to the data. A similar trend of enhancement in performance was observed with the incorporation of time varying intercept or coefficients, in that order,

For Peer Review Only

indicating that the flexibility of model specifications seems to improve the prediction capabilities.

The superior performance demonstrated by the AR models indicated that the advantages of model fit may be transferable to prediction accuracy. Considering this possibility, the authors were interested in exploring if some correlation may exist among the various prediction accuracy criteria and the GOF measures. As shown in Figure 2, the posterior deviance ($\bar{D}$) exhibited a strong positive dependency with LPML, MAD, RMSE, and RSS. Moreover, as expected, the increased effective number of parameters appeared to decrease the posterior deviance. It is also interesting to note that DIC seemed to have the same trend with posterior deviance, which indicates that the prediction accuracy had a larger influence on the DIC than did the effective number of parameters, in the case of this study. This relationship between the components of GOF and evaluation measures pertaining to prediction performance may be beneficial to safety professionals for model selection as they can make the educated judgement without the comprehensive evaluation, which may not be feasible under some conditions such as financial constraints or time inadequacy.

[Table 3 about here]

[Figure 2 about here]

5. Conclusions and Recommendations

This study was mainly focused on the evaluation of alternative ways of addressing serial correlation in crash prediction models, which include two newly proposed models which combine two types of temporal treatments. Nine-year traffic data of 203 TAZs from the city of Irvine were used to develop nine statistical models based on the Full Bayesian framework. The models varied in their complexity, with the simplest model being independent over time and the

For Peer Review Only

proposed models (most sophisticated) being a combination of AR-1 and time-varying coefficients or linear trend. It is noteworthy that a distance-based spatial random effect term was intentionally introduced so as not to blur the probable advantages of temporal random effects. Comparisons were made among these models based on various evaluation criteria. Some important findings are shown as follows:

1. The number of statistically significant variables being identified was generally similar across all models which indicates their robustness. However, the treatment of time as a covariate in the form of linear trend proved to be an important addition as it was observed to be statistically significant with a negative correlation with crash frequency.

2. The results from DIC revealed significant differences in modelling performance amongst the competing models. The models with independent and fixed-time random effects had 55 points of DIC higher than the AR models. The inclusion of time-varying coefficients was observed to result in significant reduction in the posterior deviance. Although such models employed higher number of effective parameters, but remarkable advantage at deviance generated superior overall fit. Overall, the two proposed models outperformed the competing models with superior fit with the observed crash data. These models were developed by combining advantage of lowest deviance from AR specification with the linear trend (Model 8) and time-varying coefficients (Model 9), while employing the optimum number of effective parameters to generate lowest DIC.

3. For prediction accuracy, the model which accounted only for the autoregressive effect (Model 7) illustrated the superior performance in terms of both cross validation and typical assessment based on the same data used to develop models.

For Peer Review Only

4. Strong correlations were observed among the various prediction assessment criteria including $\overline{D}$, LPML, MAD, RMSE, and RSS. In addition, the increased effective number of parameters was accompanied with the increase of prediction accuracy. Finally, given that the DIC and $\overline{D}$ had the same trend generally, it can be concluded that the $\overline{D}$ had a larger impact on DIC then did p_D in the case of this study.
5. In sum, Model 7 (AR-1) was ranked the first with respect to the predictive capability. However, the performance can be improved by combining the autoregressive effect with other temporal treatment (as illustrated in Model 8 and Model 9) if the penalized criterion, DIC, was utilized, as the latter ones could yield similar prediction accuracy using relatively smaller effective number of parameters.

Although the research here reflects an improved understanding of how various temporal treatments perform, further work is still needed. Some of the recommendations are mentioned below:

1. The study used the zonal level data for evaluation of veracious temporal treatments of crash data. The studies based on different geospatial units are needed to verify the results obtained in this study.
2. Only AR-1 was chosen to represent the large body of models involving the autoregressive process. Those models are also worthwhile to explore and check whether there are better models than the proposed ones of this study.
3. Only two combination models were proposed for illustrative purpose. More ways of temporal combination are preferred to check the performance of these models relative to the typical ones.

For Peer Review Only

4. The correlation analysis of the evaluation criteria was based on a small number of observations (9 models), more models are expected to yield more reliable result of the correlation analysis.
5. The Full Bayesian hierarchical approach was employed for incorporating heterogeneities in crash data. Other frameworks may also be explored.

Acknowledgements

Support for this study was provided by 2017 UCCONNECT Research Program (Grant number: 65A0529-052.3).The authors thank Southern California Association of Governments for providing covariates and geometric centroid distance data of TAZ used in this study. The authors also wish to thank the anonymous reviewers for providing constructive criticism for improvement of the paper.

References

1. Abdel-Aty, M., Wang, X., 2006. Crash estimation at signalized intersections along corridors: analysing spatial effect and identifying significant factors. *Transportation Research Record: Journal of the Transportation Research Board, No.* 1953. *Transportation Research Board of the National Academies, Washington, DC, pp*. 98–111.
2. 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.
3. Aguero-Valverde, J. 2013. Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates. *Accident Analysis & Prevention*, 50, 289-297.

For Peer Review Only

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

5. Aguero-Valverde, J., Jovanis, P.P., 2008. Analysis of road crash frequency with spatial models. *Transp. Res. Rec.* 2061, 55–63.

6. Aguero-Valverde, J., & Jovanis, P. 2009. Bayesian multivariate poisson lognormal models for crash severity modelling and site ranking. *Transportation Research Record*, *2136*(2136), 82-91.

7. Aguero-Valverde, J., Wu, K. F. K., & Donnell, E. T. 2016. A multivariate spatial crash frequency model for identifying sites with promise based on crash types. *Accident Analysis & Prevention*, *87*, 8-16.

8. Andrey, J., & Yagar, S. 1993. A temporal analysis of rain-related crash risk. *Accident Analysis & Prevention*, *25*(4), 465-472.

9. 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.

10. Bijleveld, F. D. 2005. The Covariance between the Number of Accidents and the Number of Victims in Multivariate Analysis of Accident Related Outcomes. *Accident Analysis and Prevention,* Vol. 37, No. 4, pp. 591–600.

11. Brooks, S.P., Gelman, A., 1998. Alternative methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. 7, 434–455.

12. Castro, M., Paleti, R., & Bhat, C. R. 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. *Transportation research part B: methodological*, *46*(1), 253-272.

For Peer Review Only

13. Congdon, P., 2001. Bayesian Statistical Modelling. *John Wiley & Sons, West Sussex, United Kingdom*

14. Cheng, W., & Washington, S. New criteria for evaluating methods of identifying hot spots. Transportation Research Record: *Journal of the Transportation Research Board*, (2083), 2008, pp.76-85.

15. Cheng, W., Gill, G.S., Dasu, R., Xie, M., Jia, X. and Zhou, J., 2017a. Comparison of Multivariate Poisson lognormal spatial and temporal crash models to identify hot spots of intersections based on crash types. *Accident Analysis & Prevention*, *99*, pp.330-341.

16. Cheng, W., Gill, G.S., Sakrani, T., Dasu, M. and Zhou, J., 2017b. Predicting motorcycle crash injury severity using weather data and alternative Bayesian multivariate crash frequency models. *Accident Analysis & Prevention*, *108*, pp.172-180.

17. Delen, D., Sharda, R. and Bessonov, M., 2006. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. *Accident Analysis & Prevention*, *38*(3), pp.434-444.

18. Earnest, A., Morgan, G., Mengersen, K., Ryan, L., Summerhayes, R., & Beard, J. 2007. Evaluating the effect of neighbourhood weight matrices on smoothing properties of Conditional Autoregressive (CAR) models. *International journal of health geographics*, *6*(1), 1.

19. El-Basyouny, K., & Kwon, D. W. 2012. Assessing time and weather effects on collision frequency by severity in Edmonton using multivariate safety performance functions. In *Transportation Research Board 91st Annual Meeting* (No. 12-0494).

For Peer Review Only

20. Gelfand, A., 1996. Model determination using sampling-based methods, in W. Gilks, S. Richardson, and D. Spiegelhalter, eds., Markov Chain Monte Carlo in Practice, Chapman & Hall, Suffolk

21. Gill, G.S., Cheng, W., Vo, T., Xie, M., Jia, X. and Zhou, J., 2017. Evaluating the Influence of Neighboring Structures on Spatial Crash Frequency Modeling and Site Ranking Performance, *Transportation Research Record* (No. 17-04119).

22. Guo, F., X. Wang, and M. A. Abdel-Aty. 2009. Corridor-Level Signalized Intersection Safety Analysis Using Bayesian Spatial Models. *Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C.*

23. Hay, J. L., & Pettitt, A. N. 2001. Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease. *Biostatistics*, *2*(4), 433-444.

24. Heydari, S., Fu, L., Lord, D., Mallick, B.K., 2016. Multilevel Dirichlet process mixture analysis of railway grade crossing crash data. *Analytic Methods in Accident Research* 9, 27-43.

25. Huang, H., Darwiche, A.L., Abdel-Aty, M.A., 2010. County-level crash risk analysis in Florida: Bayesian spatial modelling. *Transportation Research Record* 2148, 27–37.

26. Huang, H., Chin, H., Haque, M., 2009. 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* 2103, 32–41

27. Huang, H. L., H. C. Chin, and M. M. Haque. 2008. Severity of Driver Injury and Vehicle Damage in Traffic Crashes at Intersections: A Bayesian Hierarchical Analysis. *Accident Analysis and Prevention,* Vol. 40, pp. 45–54.

For Peer Review Only

28. Huang, H., H. C. Chin, and M. M. Haque. 2008. Bayesian Hierarchical Analysis of Crash Prediction Models. *Presented at 87th Annual Meeting of the Transportation Research Board*, Washington, D.C.

29. James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. *An introduction to statistical learning* (Vol. 112). New York: springer.

30. Jiang, X., Abdel-Aty, M., & Alamili, S. 2014. Application of Poisson random effect models for highway network screening. Accident Analysis & Prevention, 63, 74-82.

31. Johnson, D. S. and Hoeting, J. A. (2003). Autoregressive Models for Capture-Recapture Data: A Bayesian Approach. Biometrics, 59(2), 341-350.

32. Jonathan, Aguero-Valverde, Kun-Feng Ken Wu, and Eric T. Donnell. "A multivariate spatial crash frequency model for identifying sites with promise based on crash types." Accident Analysis & Prevention 87 (2016): 8-16.

33. Jones, A., & Huddleston, E. 2009. SAS/STAT 9.2 user's guide. *Cary, NC: SAS Institute Inc*.

34. Jung, R. C., Kukuk, M., & Liesenfeld, R. Time series of count data: modeling, estimation and diagnostics. Computational Statistics & Data Analysis, 51(4), 2006, pp. 2350-2364

35. Madsen, H. 2007. *Time series analysis*. CRC Press.

36. Kim, D. G., Washington, S., & Oh, J. 2006. Modelling crash types: New insights into the effects of covariates on crashes at rural intersections. *Journal of Transportation Engineering*, *132*(4), 282-292.

37. Ladron de Guevara, F., Washington, S., & Oh, J. 2004. Forecasting crashes at the planning level: simultaneous negative binomial crash model applied in Tucson, Arizona. *Transportation Research Record: Journal of the Transportation Research Board*, (1897), 191-199.

For Peer Review Only

38. Lawson, A. B., Browne, W. J., & Rodeiro, C. L. V. Disease mapping with WinBUGS and MLwiN (Vol. 11). John Wiley & Sons, 2003.

39. Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A* 44, 291–305.

40. Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention* 37 (1), 35–46.

41. Lunn, D., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Stat. Comput*. 10, 325–337.

42. Ma, J., Kockelman, K.M., 2006. Bayesian multivariate Poisson regression for models of injury count by severity. *Transportation Research Record* 1950, 24–34.

43. Ma, J., Kockelman, K., & Damien, P. 2008. A multivariate poisson-lognormal regression model for prediction of crash counts by severity, using bayesian methods. *Accident Analysis and Prevention*, *40*(3), 964-975

44. MacNab, Y.C., 2004. Bayesian spatial and ecological models for small-area crash and injury analysis. *Accident Analysis and Prevention* 36 (6), 1019–1028.

45. MacNab, Y. C. 2003. A Bayesian Hierarchical Model for Accident and Injury Surveillance. *Accident Analysis and Prevention,* Vol. 35, pp. 91–102.

46. Miaou, S.-P., Lord, D., 2003. Modelling traffic crash–flow relationships for intersections: dispersion parameter, functional form and Bayes versus empirical Bayes. *Transportation Research Record* 1840, 31–40.

47. Miaou, S. P., Song, J. J., & Mallick, B. K. 2003. Roadway traffic crash mapping: a space-time modelling approach. *Journal of Transportation and Statistics*, *6*, 33-58.

For Peer Review Only

48. Miaou, S. P., & Song, J. J. 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence. *Accident Analysis & Prevention*, *37*(4), 699-720.

49. Mitra, S. 2009. Spatial autocorrelation and Bayesian spatial statistical method for analysing intersections prone to injury crashes. *Transportation Research Record: Journal of the Transportation Research Board*, (2136), 92-100.

50. Mukhopadhyay, S., Gelfand, A.E., 1997. Dirichlet process mixed generalized linear models. *Journal of the American Statistical Association* 92 (438), 633-639.

51. Narayanamoorthy, S., Paleti, R., & Bhat, C. R. 2013. On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. *Transportation research part B: methodological*, *55*, 245-264.

52. Ntzoufras, I., 2009. Bayesian Modeling Using WinBUGS. John Wiley & Sons.

53. Park, E., & Lord, D. 2007. Multivariate poisson-lognormal models for jointly modelling crash frequency by severity. *Transportation Research Record*, *2019*(2019), 1-6

54. 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.

55. Quddus, M.A., 2008. Time series count data models: empirical application to traffic accidents. *Accident Analysis and Prevention* 40 (5), 1732–1741.

56. Shankar, V., F. Mannering, and W. Barfield 1995. Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies. *Accident Analysis and Prevention,* Vol. 27, pp. 371–389.

For Peer Review Only

57. Siddiqui, C., Abdel-Aty, M., & Choi, K. 2012b. Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis & Prevention*, *45*, 382-391.

58. Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van der Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 64 (4), 583–616.

59. Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003. WinBUGS Version 1.4 User Manual. *MRC Biostatistics Unit, Cambridge, http://www.mrc-cam.ac.uk/bugs*

60. Song, J. J., M. Ghosh, S. Miaou, and B. Mallick 2006. Bayesian Multivariate Spatial Models for Roadway Traffic Crash Mapping. *Journal of Multivariate Analysis,* Vol. 97, pp. 246–273.

61. Tunaru, R. 2002. Hierarchical Bayesian Models for Multiple Count Data. *Austrian Journal of Statistics,* Vol. 31, Nos. 2 and 3, pp. 221–229

62. Ulfarsson, G.F., Shankar, V.N., 2003. An accident count model based on multi-year cross-sectional roadway data with serial correlation. *Transportation Research Record* 1840, 193–197.

63. Wang, X., Abdel-Aty, M., & Brady, P. 2006. Crash estimation at signalized intersections: significant factors and temporal effect. *Transportation Research Record: Journal of the Transportation Research Board*, (1953), 10-20.

64. Washington, S., Schalkwyk, I.V., You, D., Shin, K., Samuelson, J.P., 2010. PLANSAFE User Manual: Forecasting the Safety Impacts of Socio-Demographic Changes and Safety Countermeasures. *NCHRP* 8-44(2)

65. Wang, C., Quddus, M., & Ison, S. 2013. A spatio-temporal analysis of the impact of congestion on traffic safety on major roads in the UK.*Transportmetrica A: Transport Science*, *9*(2), 124-148.

For Peer Review Only

66. Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. *Accident Analysis & Prevention* 38 (6), 1137–1150.

67. Xu, P., Huang, H., 2015. Modelling crash spatial heterogeneity: Random parameter versus geographically weighting. *Accident Analysis and Prevention*. 75, 16-25.

68. Ye, F., & Lord, D. 2011. 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*, (2241), 51-58.

For Peer Review Only

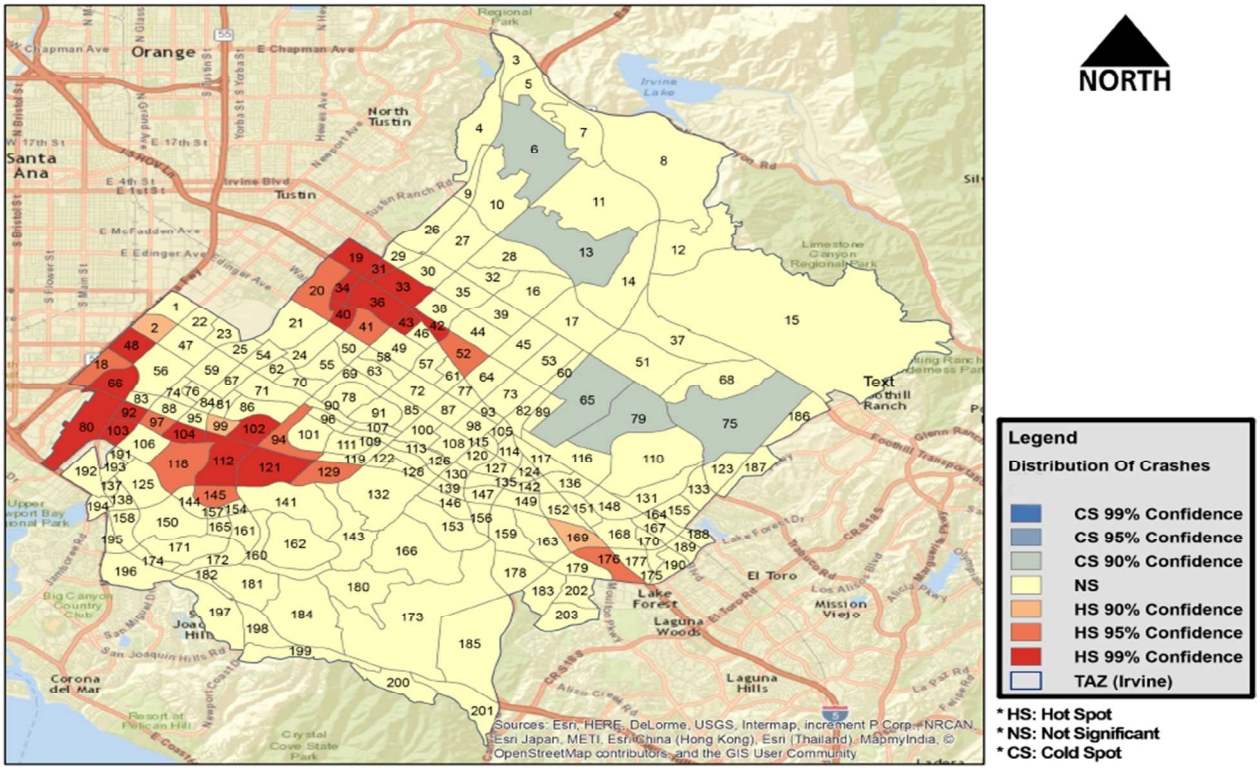

FIGURE 1 Crash distributions at TAZ level in the City of Irvine, California.

Review Only

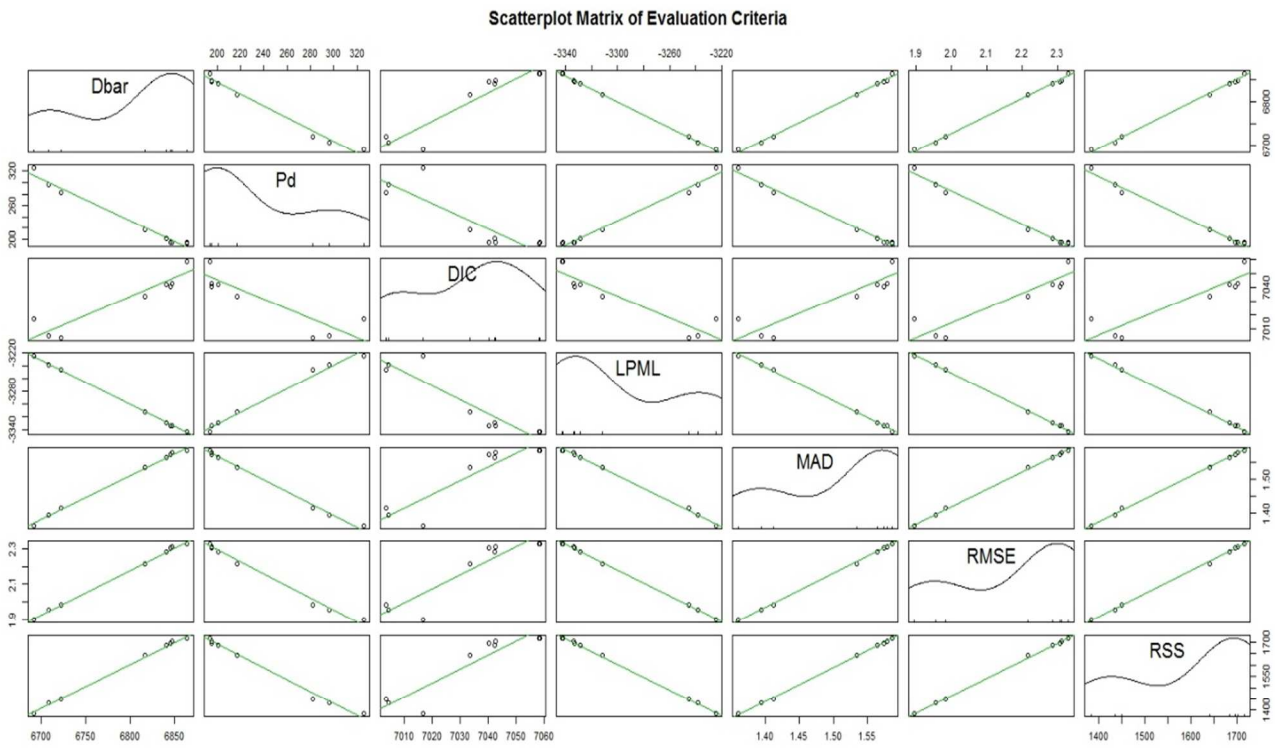

FIGURE 2 Correlation among the evaluation criteria

Notes: 1. The curve in the principal diagonal represents the density plot of each variable;

2. The linear line results from the simple linear regression of each evaluation pair.

Review Only

TABLE 1 Descriptive Statistics of Collected Data of Various TAZs

Variables	Description	Minimum	Maximum	Median	Mean	S.D.
Collision	Total Annual Fatal and Injury Collisions	0	56	3	5.06	6.6
VMT_per_day	Vehicle Miles Travelled per day	112.57	276079.92	34795.66	54262.44	56156.84
Acre	TAZ area in acre	0.69	5062.95	183.35	282.90	431.75
Med-inc	Median household income ($)	0.00	183347.00	41581.00	48440.78	50635.10
Pop_den	Population density (persons/acre)	0.00	32.40	0.79	6.19	7.96
HH_den	Household density (hh/acre)	0.00	13.62	0.10	2.34	3.15
Emp_den	Employment density (jobs/acre)	0.00	121.10	2.01	10.34	17.43
Ret_den	Retail job density	0.00	17.45	0.08	0.79	2.02
RetSer_den	Retail+ Service (retail + FIRE + ArtsFood + Other Serv.) job density	0.00	50.60	0.44	2.99	6.29
Jobmix13	Job mix (13sectors); 1 = highest mix (jobs are equal for all sectors)	0.00	0.93	0.62	0.54	0.28
Int34_Den	Intersection density (3- and 4-legs)	0.00	0.62	0.08	0.12	0.12
BKlnAcc	Bike lane access (1=if a TAZ has bike lane)	0.00	1.00	1.00	0.92	0.28
Rail	1=at least one rail station in a TAZ	0.00	1.00	0.00	0.01	0.10
TTbus_D	Total Bus Stop Density	0.00	0.53	0.01	0.05	0.09
WalkAcc	Walk Accessibility	0.00	74.53	0.42	3.87	9.46
Pct_Art	Percent of main arterial (45-55 mph) of TAZ	0.00	0.80	0.00	0.11	0.17
Distance	Distance among centroids of TAZs (unit: miles)	0.16	13.21	4.24	4.42	2.23

For Peer Review Only

TABLE 2 Estimates of Variable Coefficients Obtained by various Models

Models	Variables and Parameters					
	β_0 (Intercept)	β_1 (VMT per day, in thousands)	β_2 (Household Density)	β_{t1} (linear trend)	β_{t2} (quadratic trend)	γ (autocorrelation coefficient)
1	-5.296 (0.6338)	0.5843 (0.06031)	0.090 (0.02464)			
2	-5.259 (0.4876)	0.5811 (0.046)	0.0896 (0.02409)			
3	-5.631 (0.6935)	0.6242 (0.06575)	0.0924 (0.02394)	-0.0167 (0.004)		
4	-5.306 (0.5162)	0.5985 (0.04858)	0.0913 (0.02261)	-0.0474 (0.01764)	*0.0031 (0.00173)*	
5	-4.881 (0.403)					
	-5.008 (0.4031)					
	-4.931 (0.4034)					
	-4.974 (0.4034)					
	-5.036 (0.4033)	0.5564 (0.0377)	0.0868 (0.0262)			
	-5.065 (0.4032)					
	-5.039 (0.4034)					
	-5.011 (0.4035)					
	-5.048 (0.4031)					
6	-4.843 (0.7103)	0.5561 (0.06606)	0.08022 (0.0251)			
	-5.372 (0.7149)	0.5904 (0.06625)	0.08851 (0.02546)			
	-5.097 (0.6457)	0.577 (0.06021)	0.06779 (0.02533)			
	-5.742 (0.7055)	0.6296 (0.06538)	0.07294 (0.02549)			
	-5.162 (0.6687)	0.5713 (0.06225)	0.07796 (0.02542)			
	-4.585 (0.6869)	0.5097 (0.06397)	0.1085 (0.02512)			
	-5.125 (0.709)	0.5637 (0.06598)	0.09617 (0.02509)			
	-5.521 (0.6829)	0.6011 (0.06347)	0.09905 (0.02515)			
	-6.037 (0.6774)	0.6391 (0.06294)	0.1199 (0.02522)			

For Peer Review Only

7	-5.407 (0.6191)	0.5944 (0.05932)	0.08915 (0.02295)		0.108 (0.03005)
8	-4.888 (0.6716)	0.5533 (0.06367)	0.0864 (0.02315)	-0.01694 (0.004494)	*0.06384* *(0.05583)*
9	-5.086 (0.7401)	0.5787 (0.06936)	0.08244 (0.0256)		
	-5.556 (0.7747)	0.6077 (0.07237)	0.09007 (0.02596)		
	-5.359 (0.7681)	0.601 (0.07182)	0.07007 (0.02593)		
	-5.998 (0.8074)	0.653 (0.0751)	0.07515 (0.02617)		
	-5.382 (0.7727)	0.5915 (0.07224)	0.07931 (0.02622)		0.8625 (0.1238)
	-4.756 (0.7805)	0.5257 (0.07317)	0.1089 (0.02578)		
	-5.36 (0.7745)	0.5854 (0.07234)	0.09688 (0.02619)		
	-5.73 (0.8338)	0.6205 (0.0776)	0.09954 (0.0262)		
	-6.187 (0.7965)	0.6534 (0.07423)	0.1196 (0.02604)		

Notes: 1. Numbers in parentheses represent uncertainty estimates, or, posterior standard deviations.

2. The statistically non-significant (at level of 0.05) variable coefficients are shown in *italics*.

For Review Only

TABLE 3 Goodness-of-fit and Prediction Performance of Alternate Models

Models	$\bar{D}$	P_d	DIC	LPML	MAD	RMSE	RSS
1	6864.41	193.941	7058.35	-3342.1	1.586	2.330	1716.8
2	6864.44	**193.822**	7058.26	-3342.3	1.586	2.330	1716.6
3	6847.49	195.113	7042.6	-3333.7	1.579	2.311	1702.2
4	6845.02	195.35	7040.37	-3332.7	1.574	2.308	1696.6
5	6841.19	201.052	7042.24	-3328.7	1.565	2.285	1684.9
6	6816.33	217.323	7033.65	-3311.7	1.534	2.217	1640.9
7	**6691.19**	325.544	7016.74	**-3224.6**	**1.361**	**1.896**	**1382.5**
8	6721.79	281.835	**7003.62**	-3245.2	1.413	1.983	1450.4
9	6708.18	296.338	7004.51	-3238.3	1.394	1.956	1434.6

Notes: 1. The values shown in bold represent the best performance under each criterion.

2. $\bar{D}$-posterior mean deviance; p_D-effective number of parameters; DIC - Deviance Information Criterion; LPML- Log Pseudo Marginal Likelihoods; MAD- Mean Absolute Deviation; RMSE- Root Mean Square Error; RSS- Residual Sum of Squares.

For Peer Review Only