Direct Spatial Correlation in Crash Frequency Models

Jonathan Aguero-Valverde
Professor
Programa de Investigación en Desarrollo Urbano Sostenible
Universidad de Costa Rica
Barrio Los Profesores, Calle B, No 11, Mercedes
San Pedro, San José, 11503
Costa Rica
Email: jaguero@produs.ucr.ac.cr
Ph. 506-22834927
Fax. 506-22834815

Submitted to the 3rd International Conference on Road Safety and Simulation
Indianapolis, Indiana, USA.

Submitted July 2011

ABSTRACT

Spatial correlation has shown to be present in highway safety data, yet the distance at which sites should be considered correlated is largely unknown. The purpose of this research is to explore the effect of direct spatial correlation structures on crash frequency models at the road segment level and compare them to spatial conditional models. By using direct spatial correlation structures the "effective range" (i.e. the distance at which there is no lingering spatial correlation) is estimated.

A Full Bayes hierarchical approach is used with direct spatial correlation effects for the spatial correlation terms as well as conditional autoregressive (CAR) spatial effects. The model of crash, traffic and roadway inventory data from Pennsylvania shows an average effective range of around 168 m (0.1 mi) which confirmed previous findings from conditional models that suggested that spatial correlation is presented in segments less than a mile apart.

The direct spatial correlation model has a better goodness-of-fit than the random effects and the CAR model. In terms of posterior deviance the direct spatial correlation model performed similar to the other two models but in terms of the penalized goodness-of-fit measure Deviance Information Criteria it performed significantly better (4203, 4181, and 4109 respectively). Furthermore, the proportion of variation in the data explained by the spatial correlation term is almost the same for both spatial models (0.59 for the CAR model and 0.60 for the direct spatial correlation model). The standard deviations in coefficient estimates are slightly lower for the direct spatial correlation model compared to the random effects model but significantly lower than the CAR model.

INTRODUCTION

Spatial correlation has shown to be present in highway safety data (Miaou, Song, and Mallick, 2003; MacNab, 2004; Aguero-Valverde and Jovanis, 2006, 2010; Huang, Darwiche, and Abdel-Aty, 2010), yet the distance at which sites should be considered correlated is largely unknown. Models of spatial correlation can be divided in two main groups: direct spatial correlation and conditional models. In the former, the spatial correlation between two observations is a function of the distance between them, commonly defined though the covariance function and its correlation matrix. Conditional spatial correlation models, on the other hand, rely on the adjacency matrix that is arbitrarily defined by the modeler. By using direct spatial correlation structures the "effective range" (i.e. the distance at which there is no lingering spatial correlation) can be estimated.

Most spatial correlation models applied to highway safety used a conditional specification for the spatial effects (Miaou, Song, and Mallick, 2003; MacNab, 2004; Aguero-Valverde and Jovanis, 2006, 2009 and 2010; Huang, Darwiche, and Abdel-Aty, 2010). Conditional spatial correlation models are well suited to analyze crash frequency at segment or intersection level since a road network can be easily modeled as a lattice made off segments (edges) and intersections (nodes), forming a directed graph. Furthermore, for a road network, there is a discrete set of sites where crashes can occur, i.e. within a segment or intersection, but, by definition, never outside the road network. Apart from this theoretical advantage, conditional models have practical advantages over direct spatial correlation models. The most important of those advantages is that the expectation of a site is only conditionally dependent on its neighbors, that are defined *a priori* by the modeler; hence, model estimation is relatively straightforward within a fully Bayesian approach.

Direct spatial correlation models such as those used in geostatistics, are mostly used to model processes that occur in locations from a spatial continuum (Cressie, 1993). Direct spatial correlation models asses the spatial correlation as a function of the distance between sites through the covariance function; hence, the strength of the spatial correlation is derived directly from the data. One of the disadvantages of direct spatial correlation models is that any site can be spatially correlated to all the other sites; therefore, a correlation matrix of n x n, where n is the number of sites, need to be estimated and inversed. The number of calculations needed to inverse a matrix increases exponentially with the matrix size, which implies that with current computational power, direct spatial correlation models can be applied to datasets with only a few hundred observations at best. Furthermore, no closed form exists for Poisson models with exponential random effects; therefore, maximum likelihood estimation is not feasible and more flexible but computationally demanding approaches such as Full Bayes are needed.

In summary, conditional spatial models present theoretical and practical advantages, but due to their conditional nature cannot be used to estimate the strength of the spatial correlation and the distance at which two sites should be considered correlated; therefore, direct spatial correlation models where the covariance function is estimated are proposed here. The purpose of this research is to explore the effect of direct spatial correlation structures on crash frequency models at the road segment level and compare them to spatial conditional models. By using direct spatial correlation structures the "effective range" of the spatial correlation is estimated. This paper is organized as follows: a literature review is presented and the methodology is shown; then the dataset is described, followed by the discussion of results, conclusions and recommendations for future research.

Literature Review

Most of the earliest studies of spatial models in highway safety were descriptive in nature. The aim was to find spatial association between observations thought descriptive statistics and hypothesis testing. For others the objective was to account for the spatial correlation presented in the crash data by imposing a spatial model to the observations or their residuals (i.e. errors). Most recent studies generally belong to this group.

Within the spatial descriptive statistics group, one of the first studies was published by Levine, Kim, and Nitz (1995a). Crashes were geo-coded to the nearest intersection or ramp, then different spatial statistics were calculated including mean center, standard distance deviation based on "great circle" distance, the standard deviational ellipse (1^{st} and 2^{nd} principal component), and the nearest neighbor index(based on the x and y coordinate of the accidents).

Jones, Langford, and Bentham (1996) conduct a classical K-function analysis on the residuals of a logit model where the log-odds were fatalities compared to seriously injured. The variables of the model were: age, type of user (pedestrian, bicyclist, motor vehicle driver) and number of casualties. The authors found that, once the trend was removed from the data, the residuals presented spatial clustering.

Another spatial descriptive statistics study was developed by Black and Thomas (1998) who explored spatial dependency at the road segment level by using the Moran's Index (a standard statistic used to measure the strength of spatial association among area units; it is analogous to the lagged autocorrelation coefficient in time series). The study concluded that there was a significant level of positive spatial correlation in the data.

Nicholson (1999) conducted a study to identify the presence of non-random distributions of crashes by comparing actual spatial patterns to the complete spatial randomness (CSR) case. Comparisons included: stationary and isotropic (accidents not clustered but arranged regularly), non-stationary and isotropic (accidents clustered at randomly distributed points), and non-stationary and anisotropic (accident clustered along lines). Different statistical tests for spatial randomness such as quadrant methods, nearest neighbor methods, and K-function were analyzed. The author concluded that nearest neighbor methods appear more powerful and robust for detecting the kind of accident patterns that can be observed in practice and that the K-function method enabled patterns at different spatial scales to be detected.

Levine, Kim, and Nitz (1995b) also developed the first reported spatial correlation model in road crash analysis. They modeled spatial correlation at census block level, which was equivalent to a time series autoregressive lag-1 model (AR-1). The previous time was replaced by the weighted average of the neighbors. The explanatory variables included in the model were: freeway crossing the block (dummy), miles of arterials or highways, miles of minor roads, miles of freeways, population, and employment. While the model takes into account the spatial correlation of the data, its weakness is the reliance on an assumed normal distribution for crashes rather than a discrete count probability distribution such as Poisson or negative binomial.

A pioneer work in terms of spatial modeling of traffic crashes was developed by Miaou, Song, and Mallick (2003). The authors estimated a series of crash frequency models aggregated at the county level for the state of Texas. Poisson-based Bayes models of Fatal (K), incapacitating (A), and non-incapacitating (B) injuries were estimated. Conditional Auto-Regressive model (CAR) was used to model spatial correlation and Markov Chain Monte Carlo (MCMC) was used to sample the posterior probability distribution.

MacNab (2004) also performed an analysis of crashes using spatial Bayesian models. Covariates were examined using hospitalization data for 83 local health areas in British Columbia (BC), Canada, between 1990 and 1999. Socioeconomic variables like marriage and immigration were used along with medical variables like life expectancy, health care providers, and hospital beds. In addition, the age effects were modeled using a spline regression. Other variables such as miles of roads and seatbelt violations were also used for the model. Random spatial effects were included in the model and assumed to have a CAR distribution. Considerable spatial correlation was found in the data.

Aguero-Valverde and Jovanis (2006) estimated Full Bayes Hierarchical models with spatial and temporal effects and space-time interactions, using injury and fatality data for Pennsylvania at county level. Independent variables include socio-demographics, weather conditions, transportation infrastructure and amount of travel. A CAR model was used for modeling spatial correlation and a time trend coefficient was included to model temporal effects. Significant spatial correlation was found in the data. Huang, Darwiche, and Abdel-Aty (2010) also analyzed crash frequencies at county level for the state of Florida using a Full Bayes spatial CAR model.

Another temporal and spatial analysis of crashes was performed by Wang and Abdel-Aty (2006). They analyzed rear-end crashes at signalized intersections using the generalized estimating equations (GEE) approach. In the study, intersections were group on clusters based on their spatial location, distances, and corridor location. Intersections within a cluster were considered correlated while intersections from different clusters were considered independent. For the spatial models, the authors explored three different correlation structures: independent correlation, exchangeable correlation (constant correlations between any two intersections within a cluster), and autoregressive (AR-1) correlation; where the correlation decreases as the gap between intersections increase. The models showed high spatial correlations between intersections for rear-end crashes.

Full Bayes Hierarchical models of spatial correlation were proposed by Aguero-Valverde and Jovanis (2008) for crash frequency at segment level. Different conditional correlation structures were proposed by the authors through the weights in the neighboring structure of the CAR model. They proposed the use of the inverse of the order (as defined by proximity) as weight in the CAR model. The simplest neighboring structure was the best in terms of the goodness-of-fit measures.

Mitra (2009) investigated the intersection-level factors that influence the concentration of fatal and injury crashes by developing Full Bayes Hierarchical models with spatial correlation. Using injury and property damage only (PDO) crash data from Tucson, Arizona, the author showed that spatial dependence plays a strong role during the analyses of road-traffic crashes. The model included both spatially structured and unstructured random effects. In the study it is suggested the use of a join spatial exponential prior for the spatial random effects, rather than use a conditional model. However, this work failed to show key posterior statistics of the spatial dependency structure of the model; hence, it is difficult to interpret the results. Guoa, Wang and Abdel-Aty (2010) also developed spatial models for intersections but with CAR priors using the inverse of the distance as the weights for the spatial correlation term.

Aguero-Valverde and Jovanis (2010) studied different neighboring structures for CAR models of crash frequency. The distance at which segments are considered correlated was studied indirectly by comparison of several conditional models group in four categories: adjacency-based models, distance-order models, distance-exponential decay models and adjacency-route information models. Pure distance-based neighboring models (i.e. exponential decay) for the weights performed poorly in comparison to adjacency-based or distance-order models. The results also suggested that spatial correlation is more important in distances of one mile or less.

METHODOLOGY

Spatial crash frequency models are implemented using a Full Bayes Hierarchical approach. At the first level of the hierarchy, the units of analysis are the road segments. The crash counts are assumed Poisson distributed:

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

where y_{it} is the observed number of crashes in segment i at time t (in years) and θ_{ilt} is the expected Poisson rate for segment i of the road type l at time t. The Poisson rate is modeled as a function of the covariates following a log-normal distribution as shown in Equation 2:

$$\log(\theta_{it}) = \beta_0 + \sum_k \beta_k x_{itk} + v_i + u_i \tag{2}$$

where β_0 is the intercept, β_k is the coefficient for the k^{th} covariate, x_{itk} is the value of the k^{th} covariate for segment i at time t, v_i captures the heterogeneity among segments, and u_i is a spatially correlated random effect for segment i.

Now, one can assume that the heterogeneity random effects follow a Normal distribution:

$$v_i \sim \text{N}(0,\ \tau_v) \tag{3}$$

where τ_v is the precision (inverse of the variance) and it controls the Poisson extra-variation due to heterogeneity.

The spatially correlated effect is modeled using a Gaussian Conditional Autoregressive (CAR) prior (*4*):

$$u_i | u_{-i} \sim N\left(\frac{\sum_{j\sim i} w_{ij} u_j}{w_{i+}}, \frac{1}{w_{i+}\tau_u} \right) \tag{4}$$

where u_{-i} means all the neighbors of i, τ_u controls the Poisson extra-variation due to clustering or spatial correlation, $j{\sim}i$ denotes that segment j is a neighbor of segment i, w_{ij} is the weight of the j^{th} neighbor of the i^{th} segment, and w_{i+} is the sum of the weights of the neighbors of segment i. Random effects are pooled over time to improve model estimation. The assumption of constant random effects over time is not restrictive, provided that the covariates explain most of this variation over time.

At the third stage, hyperpriors are given for τ_v and τ_u. For both precision parameters an hyperprior gamma(0.1, 0.1) was selected.. The posterior proportion of variation explained by the spatial correlation term (η) is also of interest and is defined in Equation 5:

$$\eta = \frac{sd(u)}{sd(u) + sd(v)} \tag{5}$$

where $sd(\cdot)$ is the standard deviation.

For the direct spatial correlation model, a joint instead of a CAR prior is used. The prior of the geostatistical model is multivariate normal as shown:

$$P(u_1, u_2, u_3, \ldots, u_n) \sim N(\mathbf{0}, \Sigma) \tag{6}$$

The covariance matrix $\mathbf{\Sigma}$ is defined by an exponential covariogram as follows:

$$\Sigma = \sigma_u^2 R \tag{7}$$

And the adjacency matrix R is defined as:

$$R = [r_{ij}] = e^{-\phi\, d_{ij}} \tag{8}$$

where ϕ is the decay parameter and d_{ij} is the distance between the centroids of segments i and j.

It is important to introduce here the notion of effective range t_0, i.e. the distance at which there is no lingering spatial correlation (Banerjee, Carlin and Gelfand, 2004). For the exponential covariogram, the effective range is commonly defined as the distance at which this correlation has dropped to only 0.05; hence, $t_0 \approx 3/\phi$ since log(0.05) ≈ -3. An hyperprior uniform(0.0001875,0.1) was selected for the decay parameter ϕ, which is equivalent to a uniform(30, 16000) prior (in meters) for the effective range t_0. The limits of the uniform distribution were selected from 10 feet (30m) to 10 mile (16000m) approximately.

Model Comparison

Two different goodness-of-fit measures are used for model comparison and selection: posterior mean deviance and Deviance Information Criterion (DIC). The posterior mean deviance ($\bar{D}$) can be taken as a Bayesian measure of fit or 'adequacy'. To account for model complexity the Deviance Information Criterion was proposed (Spiegelhalter et al., 2002). The DIC is considered the Bayesian equivalent of the Akaike Information Criterion (AIC). DIC is defined as an estimate of fit plus twice the effective number of parameters as in Equation 9:

$$DIC = D(\bar{\theta}) + 2p_D = \bar{D} + p_D \tag{9}$$

where $D(\bar{\theta})$ is the deviance evaluated at $\bar{\theta}$, the posterior means of the parameters of interest, p_D is the effective number of parameters in the model, and $\bar{D}$ is the posterior mean of the deviance statistic $D(\theta)$. As with AIC, models with lower DIC values are preferred. For more details on the goodness of fit measures refer to Carlin and Louis (2000), Congdon (2001, 2003) Gelman et al (2003).

DATA DESCRIPTION

This dataset was introduced previously (Aguero-Valverde and Jovanis, 2008). The data for the models correspond to the state-maintained rural two-lane network of Centre County, located in Central Pennsylvania and part of the District 2-0 of the Pennsylvania Department of Transportation. The dataset is defined by segment and year, from 2003 to 2006. A total of 865 rural two-lane segments were included in the analysis. A relational database was assembled with information from the crash databases and road inventory.

Crash Data

Crash data were obtained from the PennDOT Crash Reporting System. The data includes reportable crashes for road segment locations only (i.e. those that do not occur at an intersection or ramp junction).

Road Inventory

Road data were obtained from the Pennsylvania Road Management System (RMS) for the study period. RMS includes data for each road segment such as County Number, State Route Number, Segment Number, Segment Length, Average Daily Traffic, Lane Width, Travel Lane Count, Posted Speed Limit, Divisor Type, Functional Class, and Urban/Rural Code. These data were complemented with the State Roads Digital Map from Pennsylvania Spatial Data Access (Pennsylvania State University, 2007) to be able to "map" crash locations. Summary statistics of the inventory data for the area of study are shown in Table 1.

Table 1: Summary Statistics of the Data by Segment and Year

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Crashes	0.310	0.691	0	7
Volume (AADT)	2636.4	3197.7	45	18749
Length (miles)	0.464	0.107	0.039	0.751
Indicators				
Functional Class Expressway and Arterial	0.295	0.456		
Functional Class Collector and Local	0.705	0.456		
Speed Limit <= 35 MPH	0.331	0.471	20	55
Speed Limit > 35 MPH	0.669	0.471		
Lane width < 10'	0.616	0.486	6	23.5
Lane width > 10' and < 12'	0.098	0.298		
Lane width 12'	0.017	0.128		
Lane width > 12' and <14'	0.011	0.106		
Lane width >= 14'	0.095	0.294		
Shoulder width < 4'	0.608	0.488	0	14
Shoulder width > 4' and < 6'	0.161	0.367		
Shoulder width 6'	0.096	0.295		
Shoulder width > 6' and <10'	0.099	0.299		
Shoulder width >= 10'	0.036	0.187		

RESULTS

Models were estimated using the open source software OpenBUGS (Thomas *et al*, 2006). For the models, 1000 iterations were discarded as burn-in. The following 10000 iterations were used to obtain summary statistics of the posterior distribution of parameters. Convergence was assessed by visual inspection of the Markov chains for the parameters. Furthermore, the number of iterations was selected such that the Monte Carlo error for each parameter in the model would be less than 10% of the value of the standard deviation of that parameter.

Table 2 presents the estimates for the Poisson random effects, CAR and direct spatial correlation models. Both spatial models performed better than the non-spatially correlated random effects model in terms of DIC. The uncorrelated random effects model presented a DIC of 4203 while the CAR model has a DIC of 4181 and the direct spatial correlation model showed a DIC of 4109, significantly lower than the other two models. Interestingly, the CAR model outperformed the other two models in terms of the Posterior Deviance.

Table 2 Models for Rural Two-Lane Centre County Roads

Variable	Uncorrelalted Random effects model					CAR model					Direct Spatial Correlation model				
	mean	sd	MC error	Confidence Interval		mean	sd	MC error	Confidence Interval		mean	sd	MC error	Confidence Interval	
				2.5	97.5				2.5	97.5				2.5	97.5
Intercept	-6.007	0.508	0.012	-6.981	-5.028	-6.122	0.746	0.036	-7.603	-4.642	-6.029	0.498	0.047	-7.017	-4.911
Volume (AADT)	0.715	0.062	0.002	0.593	0.834	0.659	0.092	0.005	0.479	0.838	0.718	0.060	0.006	0.591	0.836
Functional Class Expressway and Arterial	-0.035	0.144	0.006	-0.317	0.247	0.244	0.253	0.016	-0.241	0.750	-0.041	0.142	0.009	-0.309	0.238
Speed Limit > 35 MPH	-0.226	0.093	0.001	-0.408	-0.046	-0.302	0.117	0.004	-0.529	-0.069	-0.225	0.091	0.003	-0.401	-0.045
Lane width < 10'	-0.517	0.191	0.004	-0.891	-0.143	-0.374	0.265	0.013	-0.900	0.154	-0.511	0.186	0.008	-0.881	-0.150
Lane width > 10' and < 12'	-0.041	0.114	0.003	-0.261	0.186	0.114	0.192	0.011	-0.274	0.492	-0.036	0.110	0.005	-0.252	0.182
Lane width > 12' and <14'	-0.156	0.267	0.006	-0.690	0.358	-0.073	0.369	0.015	-0.814	0.637	-0.154	0.264	0.007	-0.679	0.361
Lane width >= 14'	0.421	0.272	0.004	-0.113	0.955	0.593	0.313	0.010	-0.033	1.204	0.422	0.271	0.006	-0.119	0.936
Shoulder width < 4'	0.239	0.146	0.004	-0.049	0.522	0.554	0.215	0.010	0.132	0.980	0.245	0.141	0.007	-0.027	0.517
Shoulder width > 4' and < 6'	0.110	0.144	0.003	-0.173	0.390	0.309	0.207	0.009	-0.097	0.710	0.112	0.142	0.006	-0.159	0.392
Shoulder width > 6' and <10'	0.083	0.142	0.003	-0.195	0.365	0.410	0.206	0.009	0.010	0.815	0.090	0.139	0.005	-0.184	0.361
Shoulder width >= 10'	0.087	0.203	0.004	-0.313	0.479	0.268	0.273	0.011	-0.264	0.797	0.091	0.195	0.006	-0.294	0.464
sd(u)						0.617	0.058	0.003	0.519	0.749	0.430	0.103	0.009	0.215	0.601
sd(v)	0.563	0.048	0.002	0.472	0.660	0.422	0.049	0.002	0.330	0.519	0.293	0.127	0.012	0.080	0.532
sigma2.u						0.077	0.017	0.001	0.050	0.116	0.196	0.086	0.008	0.046	0.364
sigma2.v	0.322	0.056	0.002	0.222	0.444	0.183	0.043	0.002	0.110	0.275	0.102	0.079	0.007	0.007	0.284
eta						0.594	0.035	0.002	0.528	0.665	0.601	0.157	0.015	0.303	0.870
t_0											168.7	370.6	30.5	30.7	1300.0
phi											0.050	0.029	0.001	0.002	0.098
deviance	4003	33.42	0.83	3938	4069	3975	29.62	0.83	3916	4032	4013	34.77	1.57	3944	4081
DIC	4203					4181					4109				

As expected, the variance and standard deviation of the uncorrelated random effects were significantly reduced when the spatial effects were introduced, since the spatial effects could explain some of the extra-variation in the data previously explained by the uncorrelated random effects. The variance dropped from 0.322 to 0.183 and 0.102 for the CAR and the direct spatial correlation models respectively. The standard deviation presented important reductions as well, going from 0.563 to 0.422 for the CAR model and 0.293 for the direct spatial correlation model. For both variables the reduction is more prominent for the direct spatial model.

The spatially correlated random effects present significant differences between the spatial models as well. While the expectation of the variance increased from 0.077 to 0.196 the expectation of the standard deviation decreased from 0.617 to 0.430 for the CAR and direct spatial correlation models respectively. This is somewhat unexpected since the standard deviation is the square root of the variance but it can be possible due to the skewness of the distributions and the fact than the mean is not a good indicator of central tendency for skewed distributions.

As measured by the standard deviation, the marginal variation in the data explained by the random effects is lower in the direct spatial correlation model in comparison with the CAR model. The sum of standard deviations for the CAR model is around 1.04 while for the direct spatial correlation model is 0.723; nevertheless, both are higher than the 0.563 of the standard deviation for the uncorrelated random effects model.

The proportion of the variation explained by the spatial correlation term η is almost the same for the CAR (0.594) and the direct spatial correlation (0.601) models. This happens even though the standard deviations in the CAR model are higher than those of the direct spatial correlation model; however, both change at a similar rate which resulted in a very similar η for both models.

The expected values of the coefficients for the covariates are similar for the three models; however, the estimates for the direct spatial correlation model are much closer to the uncorrelated estimates than those of the CAR model. While the coefficients for the Intercept, AADT and Speed limit > 35MPH are -6.007, 0.715 and -0.226 for the uncorrelated model, the same coefficients for the CAR and direct spatial correlation models are -6.122, 0.659, -0.302 and -6.029, 0.718, -0.225 respectively. The standard deviations of the coefficients present a similar picture between models with the uncorrelated and direct spatial correlation models, i.e. very close estimates, and higher values of variation for the CAR model.

The effective range t_0 was found to be relatively small for the data analyzed, i.e. only 168.7 meters; however, this value is in line with previous findings (Aguero-Valverde and Jovanis, 2008 and 2010) that suggested that spatial correlation was important between segments less than 1.6 km (approximately 1 mile) apart. From the practical applications point-of-view, this result suggests that, for crash frequency models at road segment level, controlling for spatial correlation between direct neighbors is sufficient.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

The direct spatial correlation model presents better goodness-of-fit than the random effects and the CAR model. In terms of posterior deviance the direct spatial correlation model performed similar to the other two models (4003, 4009, and 4013 for the random effects, the CAR and the direct spatial correlation model respectively) but in terms of the penalized goodness-of-fit measure DIC it performed significantly better (4203, 4181, and 4109 respectively). Furthermore, while the DIC of the CAR model was only around 20 points lower than the uncorrelated model, the DIC of the direct spatial correlation model was almost 100 points lower. This finding is also in line with previous works that found spatial correlation to be significant in highway safety models.

The proportion of variation in the data explained by the spatial correlation term is almost the same for both spatial models (0.59 for the CAR model and 0.60 for the direct spatial correlation model), even though the standard deviations in the CAR model are higher than those of the direct spatial correlation model; however, both changed at a similar rate which resulted in a very similar η for both models. Clearly, most of the variation in the data explained by the random effects is spatial in nature.

The standard deviations in coefficient estimates are slightly lower for the direct spatial correlation model compared to the random effects model but significantly lower than the CAR model. This finding suggests that by controlling for spatial correlation throughout a direct spatial correlation model the precision of coefficient estimates is also improved.

The effective range t_0 was found to be relatively small (168m) but consistent with previous findings that suggested that spatial correlation was important between segments less than 1.6 km apart. From the practical applications point-of-view, this result suggests that, for crash frequency models at road segment level, controlling for spatial correlation between direct neighbors is sufficient. More datasets should be tested to corroborate these preliminary findings.

An exponential covariagram function was implemented in this work for its simplicity but many other covariagrams exist and can be tested. In particular, spherical, power, Guassian, and Matérn covarigrams can be used to model spatial correlation in highway safety data.

The way the distance between segments is measured has potential for future research as well. Here, it was proposed to use the “aerial” distance (i.e. as the bird flies) between segment centroids but the network distance (i.e. as the car travels) between segment centroids can be used as well. Another option is to use the shortest network distance, between segment extremes rather than segment centroids.

REFERENCES

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

Aguero-Valverde, J., Jovanis, P.P. (2008). Analysis of Road Crash Frequency with Spatial Models. *Transportation Research Record* 2061, 55-63.

Aguero-Valverde, J., Jovanis, P.P. (2010). Spatial correlation in multilevel crash frequency models: Effects of different neighboring structures. *Transportation Research Record* 2165, 21-33.

Banerjee, B.P., Bradley P. C., Gelfand A.E. (2004). *Hierarchical Modeling and Analysis for Spatial Dat*a. Chapman and Hall/CRC.

Black, W.R., Thomas, I. (1998). Accidents on Belgium's motorways: a network autocorrelation analysis. *Journal of Transport Geography* 6 (1), 23-31.

Carlin, B.P., Louis, T.A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall/CRC.

Congdon, P. (2001). *Bayesian Statistical Modelling*. John Wiley & Sons, England.

Congdon, P. (2003). *Applied Bayesian Modelling*. John Wiley & Sons, England.

Cressie, N.A.C. (1993). *Statistics for Spatial Data*. John Wiley & Sons, New York.

Gelman, A., Carlin, J., Stern, H.S., Rubin, D.B. (2003). *Bayesian Data Analysis*, Chapman & Hall/CRC, Second Edition.

Guoa, F., Wang, X., Abdel-Aty, M.A. (2010). Modeling signalized intersection safety with corridor-level spatial correlations. *Accident Analysis and Prevention* 42 (1), 84–92.

Huang H., Darwiche, A.L., Abdel-Aty M.A. (2010). County-Level Crash Risk Analysis in Florida: Bayesian Spatial Modeling. *Transportation Research Record* 2148, 27-37.

Jones, A.P., Langford, I.H., Bentham, G. 1996. The application of K-function analysis to the geographical distribution of road traffic accident outcomes in Norfolk, England. S*ocial Science and Medicine* 42 (6), 879-885.

Levine, N., Kim, K.E., Nitz, L.H. (1995). Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns. *Accident Analysis and Prevention* 27 (5), 663-674.

Levine, N., Kim, K.E., Nitz, L.H. (1995b). Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. *Accident Analysis and Prevention*, 27 (5), 675-685.

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

Miaou, S-P., Song, J.J., Mallick, B.K. (2003). Roadway traffic crash mapping: A space-time modeling approach. *Journal of Transportation and Statistics* 6 (1), 33- 57.

Mitra, S. (2009). Spatial Autocorrelation and Bayesian Spatial Statistical Method for Analyzing Intersections Prone to Injury Crashes. *Transportation Research Record* 2136, 92-100.

Nicholson, A. (1999). Analysis of spatial distributions of accidents. *Safety Science* 31, 71-91.

Spiegelhalter, D., Best N., Carlin, B.P., Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society B* 64 (4) 583–639.

Thomas, A., O'Hara, B., Ligges, U., Sturtz, S. (2006). Making BUGS Open. *R News* 6 (1), 12-17.

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