Spatial Analysis of County Level Crash Risk in New Jersey using Severity based Hierarchical Bayesian models

Sami Demiroluk, M.Sc. (Corresponding Author)
Graduate Research Assistant,
Rutgers Intelligent Transportation Systems Laboratory
Department of Civil and Environmental Engineering,
Rutgers, The State University of New Jersey,
623 Bowser Road, Piscataway, NJ 08854 USA,
Tel: (732) 445-4012
Fax: (732) 445-0577
e-mail: dsami@eden.rutgers.edu

Kaan Ozbay, Ph.D.
Professor,
Rutgers Intelligent Transportation Systems Laboratory
Department of Civil and Environmental Engineering,
Rutgers, The State University of New Jersey,
623 Bowser Road, Piscataway, NJ 08854 USA,
Tel: (732) 445-2792
Fax: (732) 445-0577
e-mail: kaan@rci.rutgers.edu

Word count:
Text: 5,248
Tables: 4 x 250
Figures: 5 x 250
Total: 7,498
Submission Date: August 1st, 2013
This study presents an innovative hierarchical Bayesian model for spatial modeling of county level crashes in New Jersey. First, the model is estimated using raw crash counts. Then, weights are applied to crashes with different severities to obtain a weighted crash count. The goal in incorporating severities in the spatial model is to demonstrate the importance of representing spatial variation of crashes as well as their severity. As a contribution to existing literature, crash rates are also analyzed by road type. Finally, crash rate maps are developed based on modeling results to visualize the effects of spatial covariates. The results of the study indicate that the most influential covariate for the crashes is the road curvature, followed by roadway mileage and roadway defects. It is also found that it is possible to represent the crash risk better by applying severity weights to the individual crashes. The developed crash rate maps can help transportation professionals on identifying and ranking the locations at an aggregate level, which requires closer attention.
INTRODUCTION

Crashes are random, yet very common events on our roadways. Every year many lives are lost due to crashes. According to Center for Disease Control and Prevention, motor vehicle crashes are the leading cause of death among young people age 5 to 34 in the United States (1). Moreover, in 2010, 2,764,122 people were injured as a result of motor vehicles crashes that year (1). Therefore, there is a need to improve roadway safety. By better handling the crashes, the identification of high risk locations and contributing factors, roadway safety can be improved.

Mapping of spatial events is a very popular technique that is widely used in disease mapping by researchers. “Mapping transforms spatial data into a visual form, enhancing the ability of users to observe, conceptualize, validate, and communicate information” (2). After the introduction of Markov Chain Monte Carlo (MCMC) methods, hierarchical Bayesian models are employed for developing risk maps of diseases in the public health studies (3). Additionally, these models are used for ranking disease risks across regions based on spatial and temporal variations. It is possible to make an analogy between diseases and crashes since crash risks vary at different regions and there is need for ranking these regions as well. Currently, there are few documented studies that use this technique to estimate spatial crash models, even though there are several benefits of utilizing these methods in transportation. Since crashes are random events, there is almost always high variance in crash data between different regions. This is called small area problem (4). One way to overcome this problem is to use hierarchical models that account for spatial correlation. The hierarchical models are also very effective in developing models in presence of multiple sources of data and uncertainty in parameterizations (5).

Current spatial crash models generally deal with the aggregated data at the county or region level (2, 6-9). Similar to this previous research, hierarchical Bayesian models will be used in this study to analyze the factors contributing to crash risk (e.g. horizontal curves, roadway defects and wet weather) and spatial variation among different regions. County level crash data that contains all crashes from 2001 to 2010 in New Jersey will be used in the model calibration. In addition to a general model for raw crash counts, we propose the use of different models for roadways under various jurisdictions. One of the major contributions of this paper is the development of a severity based spatial model. Since different crashes have varying impacts, we propose the use of severity weights to capture the effects of the crashes more realistically.

LITERATURE REVIEW

Previous research on the spatial analysis of crashes utilizes data aggregated at different level. For example, Noland (10) analyzed the effects of infrastructure improvements on fatalities and injuries. Data from 50 states was used to develop a negative binomial regression model which includes infrastructure and socio-economic variables. Additionally, Negative binomial models are used by researchers to analyze the crashes at county level (11-14). While others utilized such models for modeling crashes for road sections (15-16).

However, negative binomial regression models suffer from the fact that they cannot handle spatial and temporal correlation (6-7, 17). Therefore, more sophisticated methods are proposed by researchers such as hierarchical Bayesian models (2, 6-9).

Miaou et al. (2) developed a set of Poisson hierarchical Bayes models for estimating crash risk using crash frequencies for fatal, incapacitating and non-incapacitating injuries at county level on low volume roads in Texas. Spatial covariates include percentage of time that the
road surface is wet, the number of sharp curves, and roadside hazards. However, surrogate variables are used in modeling due to limitations of the data. Conditional Auto-regressive model (CAR) (18) is used to model spatial correlation. Then, the samples are drawn from posterior probability distributions using Markov Chain Monte Carlo (MCMC) method. In the study, it is recognized that although most of the disease mapping was done for area-based data, traffic crashes occur at roadway network.

MacNab (6) developed a hierarchical Bayesian model to analyze variations of accident risk factors at the regional level. The study considers ecological (regional) analysis of accident and injury variations, covariate effects, random spatial effects and age effects simultaneously. Hospital separation data for 83 local health areas in British Columbia (BC), Canada is used to investigate ecological/contextual factors of accident injury among males between 0 to 24 ages. Different spatial covariates such as socio-economic indicators, residential environment indicators (roads and parks), availability of medical services and utilizations are included in the study. The model for injury rates assumes injury hospitalization count to follow a Poisson distribution and the spatial random effects are modeled using a Markov random field (MRF) Gaussian distribution.

Aguero-Valverde and Jovanis (7) estimated hierarchical Bayesian models considering spatial and temporal effects and space–time interactions by using injury and fatality crash data from Pennsylvania at county level. Weather conditions, transportation infrastructure, socio-demographics and amount of travel are included as covariates in these models. Crashes are assumed to follow a Poisson distribution and spatial correlation; the CAR model is used to quantify spatial correlation. Bayesian fatality and injury models are also compared to negative binomial models in the study. It is reported that while negative binomial and Bayesian models are consistent for significant variables. However, marginally significant variables in negative binomial models are not found significant in Bayesian models.

Quddus (8) developed a set of negative binomial regression and Bayesian hierarchical models for London crash data. Census wards in Greater London metropolitan area are used as the spatial unit. Variables in different categories such as traffic characteristics, road characteristics and socio-economic factors are considered as the explanatory variables. The relationships between these variables and crash casualties are analyzed in the study using both techniques. It is reported that “Bayesian hierarchical models are more appropriate in developing a relationship between area wide traffic crashes and the contributing factors associated with the road infrastructure, socioeconomic and traffic conditions of the area”.

Huang et al. (9) analyzed the crashes at county level in Florida using hierarchical Bayesian models. It is argued that the previous research on aggregate crash data does not explicitly differentiate exposure variables and risk factors which might result in inconsistent results in similar studies. Hence, the exposure variables such as daily vehicle miles traveled and population is explicitly controlled in the study. Existence of spatial correlation between counties is checked by using Moran’s I. Moran’s I is a metric that range from -1 to 1 and a positive value indicates spatial correlation in the region under investigation whereas negative value indicates dispersion (19). It was found that the data was spatially correlated. They used set of road and traffic related variables, and demographic and socio-economic variables in model development.
DATA

In this study, New Jersey Department of Transportation (NJDOT)’s crash records database is used for developing the crash risk model (20). NJDOT keeps this yearly database containing five different tables: accident, driver, vehicle, pedestrian, and occupant. While accident table includes the general information about the accident, other tables such as occupant table contains detailed information on each occupant of the vehicle(s) included in the crash. The covariates considered in this study (ratio of crashes related to roadway defects, ratio of crashes occurred on the curves, and ratio of wet weather crashes) are contained in accident and vehicle tables. Severity is factored by the most severe injury in the crash which is gathered from the physical condition field in the occupant table. In the database, Physical condition of the victim is coded as complaint of pain, moderate injury, incapacitated, and killed.

New Jersey has 21 counties and the location of each crash is available at county level for all crashes in the state. All crash records from 2001 to 2010 are used in the development of the model. Figure 1.a shows the raw annual crash counts by county in 2010. As seen from the figure, the counties in the central and northern part of New Jersey, close to the New York metropolitan area, have higher crash counts than the rest of the state. This could be a result of the highly developed roadway network in that region (see Figure 2). NJDOT publish roadway miles and daily vehicle miles traveled (DVMT) in each county for each year on their website. DVMT is considered as the exposure variable in this study. The crash risk (the ratio of crash frequency over the exposure) map is developed using crash counts and daily vehicle miles traveled (DVMT) in each county, which is shown in Figure 1.b. From this crash risk map, it can be observed that the crash count in a county is highly related to exposure to traffic (DVMT) as Figure 1.b generates a smoother map.

![Figure 1](image-url)

**FIGURE 1** (a) Annual crash counts by county in 2010 (b) Crash risks by county in 2010 (per thousand DVMT).
In this study, the goal is to develop a crash model that depends on spatially varying roadway characteristics. This task was challenging due to the limitations of the crash data since it does not include any spatial covariates. Only covariate directly available from NJDOT is the roadway mileage in each county. Instead, surrogate variables approach, similar to the approach proposed in Miaou (2), is followed to represent the spatial covariates. The proportion of curve crashes is considered as surrogate variables to represent the number of horizontal curves. The detailed roadway data which includes the number of horizontal curves on the roadway is not available. On the other hand, when a crash occurs, the road characteristics are coded as straight or curve. Therefore, the ratio of the crashes that occurred on the curves is included as a surrogate variable. The detailed weather data is also not available; however, when a crash occurs, the road surface condition is recorded. The proportion of crashes that occurred on a wet roadway surface condition is used to indicate the percentage of time that the roadway is wet due to rain or snow, ice, etc. Another covariate considered in the study is the number of roadway hazards in each county. Equivalently to the other covariates, this data was not available. However, when a crash occurs the apparent contributing circumstances are recorded in the database in one of four categories: driver actions, vehicle factors, pedestrian factors, and roadway/environmental factors. Roadway factors include obstruction/debris in road, physical obstructions, etc. Surrogate variable proportion of crashes related to roadway defects is devised to indicate the number of roadway defects in a county. Miaou (2) explains in detail that surrogate variables approach above “mixes several possible relationships and has limited explanatory power”. Although these surrogate variables cannot quantify the impacts spatial covariates fully, they are related to the desired covariates and it is essential to include them in the model. Surrogate variables are widely and
successfully used in medical sciences since outcome cannot be measured during the treatment and surrogate outcomes are necessary to evaluate the effectiveness of treatment in advance. In transportation many surrogate measures for safety are also defined such as aggressive lane merging, speed, accepted gaps, shock-waves (21). The detailed review of the surrogate safety measures can be found in (21). Thus, the use of similar surrogate measures in this in order to compensate for the unavailability of certain data is consistent with the past practice. The descriptive statistics of the variables included in the study is given in Table 1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>All crashes</td>
<td>Frequency of all crashes</td>
<td>14716.64</td>
<td>1718</td>
<td>57047</td>
<td>11429.75</td>
</tr>
<tr>
<td>Fatal</td>
<td>Frequency of fatal crashes</td>
<td>30.33</td>
<td>5</td>
<td>70</td>
<td>14.956</td>
</tr>
<tr>
<td>Incapacitating</td>
<td>Frequency of incapacitating injury crashes</td>
<td>22.6</td>
<td>4</td>
<td>57</td>
<td>11.207</td>
</tr>
<tr>
<td>Injury</td>
<td>Frequency of other injury crashes</td>
<td>3504.54</td>
<td>424</td>
<td>10041</td>
<td>2276.659</td>
</tr>
<tr>
<td>Property damage</td>
<td>Frequency of property damage crashes</td>
<td>11159.16</td>
<td>1268</td>
<td>27723</td>
<td>7226.155</td>
</tr>
<tr>
<td>DVMT</td>
<td>Daily vehicle miles traveled (in thousands)</td>
<td>9486.79</td>
<td>2160</td>
<td>21587</td>
<td>5367.45</td>
</tr>
<tr>
<td>Roadway mileage</td>
<td>Total length of roadways (in miles)</td>
<td>1823.34</td>
<td>620</td>
<td>3509</td>
<td>758.089</td>
</tr>
<tr>
<td>Wet roadway</td>
<td>Proportion of wet roadway crashes</td>
<td>0.24</td>
<td>0.12</td>
<td>0.38</td>
<td>0.049</td>
</tr>
<tr>
<td>Curve crashes</td>
<td>Proportion of curve crashes</td>
<td>0.13</td>
<td>0.06</td>
<td>0.3</td>
<td>0.05</td>
</tr>
<tr>
<td>Roadway Defects</td>
<td>Proportion of roadway related crashes</td>
<td>0.08</td>
<td>0.02</td>
<td>0.25</td>
<td>0.054</td>
</tr>
</tbody>
</table>

**METHODOLOGY**

In this study, the hierarchical Bayesian generalized linear model is considered for estimating the model. The full hierarchical Bayesian modeling requires a three step approach. At the first step, conditional on mean, $\mu_i$, weighted crash counts, $Y_{it}$ are assumed to follow a Poisson distribution:

$$Y_{it} \sim Po(\mu_{it})$$

where $Y_{it}$ is the observed number of crashes in county $i$ at time $t$ (years), $i=1,...,N$ and $t=1,...,T$; and $\mu_{it}$ is the mean of the Poisson process for county $i$ at time $t$. The mean is further formulated as:

$$\mu_{it} = e_{it} \lambda_{it}$$

In the above equation, crash risk rate, $\lambda_{it}$, is assumed to be proportional to traffic exposure, $e_{it}$, hence DVMT is considered as an offset. Then, the crash risk rate, $\lambda_{it}$, is modeled as:

$$\log(\lambda_{it}) = \alpha + \sum_k \beta_k X_{i,k} + \theta_i + \phi_i$$

where $\alpha$ represents the intercept term, $\beta_k$ is the coefficient for spatial covariate $k$, $X_{i,k}$ is the observed value of $k$th covariate for county $i$ at time $t$, $\theta_i$ stands for region-wide or statewide global heterogeneity, and $\phi_i$ represents spatially correlated random effects for county $i$.

Next, coefficients are modeled by using non-informative priors. For the random effect
terms, second level of hierarchy is defined. The heterogeneity term is modeled using a normal prior:

\[ \theta_i \sim N\left(0, \frac{1}{\tau_h}\right) \]

where \( \tau_h \) is a precision parameter that controls the amount of \( \theta_i \) among municipalities. The heterogeneity term enables us to include extra-Poisson variability due to unobserved variables over the entire state. For modeling the spatially correlated random effects, a CAR prior is adopted, proposed by Besag (18):

\[ \phi_i | \phi_{-i} \sim N\left(\sum_{j=1}^{n} \left(\frac{w_{ij}}{w_i}\right)\phi_j, \frac{1}{\tau_c w_i}\right) \]

where \( \tau_c \) is a precision parameter that controls clustering, \( ij \) is a neighbor municipality adjacent to municipality \( i \), \( w_{ij} \) is the weight of the neighbor \( j \), and \( w_i \) represents the sum of the weights of the neighbors of municipality \( i \). Note that, in this study, it is assumed that all neighbors have equal weight. By including a spatially correlated random effects term, extra-Poisson variability in the log-relative risk which varies from municipality to municipality can be modeled in such a way that nearby municipalities will have more similar rates.

In the next step, the hyperpriors are defined for structured random effect terms which are heterogeneity and clustering precision parameters. Setting same priors for \( \tau_h \) and \( \tau_c \) to give same prior emphasis on them is incorrect due to two reasons. First, \( \tau_h \) uses a marginal specification, whereas, \( \tau_c \) is specified conditionally. Second, \( \tau_c \) is multiplied by number of neighbors before it is involved in prior specification. Therefore, hyperpriors for the precision parameters, \( \tau_h \) and \( \tau_c \), are defined based on a fair priors criteria approximated by Bernardinelli et al. (22) as follows:

\[ sd(\theta_i) = \frac{1}{\tau_h} \approx \frac{1}{0.7 \sqrt{\bar{w} \tau_c}} \approx sd(\phi_i) \]

where \( sd(\theta_i) \) and \( sd(\phi_i) \) are standard deviations of \( \theta_i \) and \( \phi_i \) respectively, and \( \bar{w} \) average number of fair neighbors. The graphical representation of the hierarchical model is shown in Figure 3.
FIGURE 3 Graphical representation of the hierarchical Bayesian crash model

The proportion of variability in the random effects is another metric which is included in modeling to analyze clustering later and it is defined as:

\[
\psi = \frac{sd(\phi_i)}{sd(\phi_i) + sd(\theta_i)}
\]

Several parameterization approaches of the proposed hierarchical model are considered in the study and to evaluate the model fit deviance information criterion (DIC) is used. DIC is proposed by Spiegelhalter et al. (23) to compare the fit and complexity of hierarchical models in which the number of parameters is not clearly defined. DIC is calculated using the posterior distribution of deviance and can be considered as more general form of Akaike’s information criterion (AIC). DIC is defined as in the following equation:

\[
DIC = D(\bar{\theta}) + 2p_D
\]

where \(D(\bar{\theta})\) is the classical estimate of fit (posterior mean of deviance) and \(p_D\) is the effective number of parameters. It should be noted that similar to AIC, the models with lower DIC are preferred.

DISCUSSION OF RESULTS

The hierarchical Bayesian model, which takes into account spatial correlation and unobserved heterogeneity, was estimated using the software package called WinBUGS (24). WinBUGS uses MCMC methods to sample from the joint posterior probability distribution of the multiple variables. Starting with only an intercept, different subsets of variables were considered in the
model development. Multiple chains were simulated for these models. 50,000 iterations were performed for each model and the first 10,000 iterations were considered as burn-ins and were not included in the sampling. The model with the lowest DIC was selected. The convergence of the model was evaluated using the mixing of MCMC chains and Gelman-Rubin statistics in WINBUGS.

Although MCMC methods are very effective in estimation of hierarchical models, sometimes they can be very slow to converge due to high correlation between parameters or due to non-informative priors (25). Several methods are proposed to overcome this issue and increase the efficiency of MCMC estimation of hierarchical models such as hierarchical centering (25), orthogonal polynomials (26) and parameter expansion (27). In order to improve convergence of the model in this study, covariates were standardized (i.e. mean 0 and standard deviation 1) and hierarchical centering was used for random effects. After these two procedures, the MCMC chains are mixed well and the convergence is reached rather quickly based on the visual inspection of monitoring plots in WINBUGS.

Table 2 shows the estimated hierarchical crash risk model for raw county level crash counts. All variables considered in the study are found to be statistically significant. The results indicate that based on the estimates of covariate effects, the “curve crashes” is the most significant variable in explaining crash risk over space. This result confirms the findings of Miaou et al. (2) in which the curve crashes were also the most significant covariate effects. “Roadway mileage” is found to be the second most influential variable. This result shows the general trend that as the roadway mileage increases in a county, if all other variables are controlled in the model, the number of crashes increases. Previous research found that the roadway mileage increases the crash risk (7-8). “Roadway defects” are found to be the third most influential variable. This result implies that although a surrogate variable was used in the study to represent the number of roadway defects, roadway defects elevate the crash risk. The least influential variable in the model is found to be wet roadway. This was the case in (7) as well. However, this might stem from the fact that during the wet weather, traffic volume might be lower. Furthermore, drivers might be more careful and potentially drive more slowly during the wet conditions.

### Table 2 Modeling results of hierarchical model for raw crash counts

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D.</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.17730</td>
<td>0.01619</td>
<td>0.00006</td>
<td>0.17730</td>
</tr>
<tr>
<td>Roadway mileage</td>
<td>0.07259</td>
<td>0.00834</td>
<td>0.00037</td>
<td>0.07207</td>
</tr>
<tr>
<td>Curve crashes</td>
<td>0.11170</td>
<td>0.00667</td>
<td>0.00021</td>
<td>0.11170</td>
</tr>
<tr>
<td>Wet roadway</td>
<td>0.02788</td>
<td>0.00195</td>
<td>0.00004</td>
<td>0.02787</td>
</tr>
<tr>
<td>Roadway defects</td>
<td>0.05564</td>
<td>0.00065</td>
<td>0.00001</td>
<td>0.05564</td>
</tr>
<tr>
<td>$\sigma_\phi$</td>
<td>0.47370</td>
<td>0.01637</td>
<td>0.00077</td>
<td>0.47580</td>
</tr>
<tr>
<td>$\sigma_\psi$</td>
<td>0.06362</td>
<td>0.02647</td>
<td>0.00129</td>
<td>0.06010</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.88320</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$D: 2554.63, D(\bar{\theta}) = 2343.62, p_D = 24.816, DIC = 2765.65$

The model also includes terms that capture the global heterogeneity and spatial correlation. The spatial correlation is significantly higher than the global heterogeneity. The
standard deviation of $\sigma_\phi$ is 0.4737 while the standard deviation of $\sigma_\theta$ is 0.06362. Based on the proportion of variability in the spatial random effects, $\psi$, it can be implied that there is a strong spatial correlation in crash risk between neighboring counties which will be discussed more in detail in the conclusion section of the paper.

In the literature, there is a lack of analysis of crash rates according to road types using this kind of hierarchical Bayesian models. To address this gap in the literature, crashes are classified by the type of roadways under different jurisdictions to re-estimate the model. In New Jersey, there are four levels of jurisdiction for roadways: state (also include interstate highways maintained by NJDOT), authority (toll roads), county, and municipal. Although the roadway jurisdiction is provided in the crash records, DVMT for the roadways was not available; therefore, the county DVMT values are used in the each model. On the other hand, the roadway mileage by jurisdiction is available and is used for model estimations. Six counties are excluded from the analysis of authority roadways since the length of authority roadways are less than five miles. Table 3 shows the estimated hierarchical crash risk model for roadways by jurisdiction. As expected, the effects of the contributing factors differ by the roadway jurisdiction. While roadway mileage is the most influential factor on state and authority roadways, it is the second most influential factor on county and municipal roadways. Moreover, it is only negatively related to crash risk on state roadways. This result might be related to low average speed of vehicles due to high congestion on the state roadways. The curve crash is also another significant variable for all roadway jurisdictions. However, it is positively related to the crash risk for state roadways. This might suggest that more safety precautions are needed to be taken in the vicinity of sharp curves on state roadways. The effect of wet roadways was found to be significant and positive on authority roadways, and municipal roadways. This can be attributed to higher speed limits on authority roadways. On the other hand, municipal roadways are generally two-lane roadways and there is only a small margin for error for drivers when the roadway is wet. The roadway defect is found to be a factor for all road types except for county roadways. Based on the proportion of variability in the spatial random effects, $\psi$, the spatial correlation is strong for all jurisdictions. The highest value of $\psi = 0.82$ is found for municipal roadway, which indicate that they are highly clustered in the neighboring regions.

**TABLE 3 Modeling results of hierarchical model for crashes by jurisdiction**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D.</th>
<th>2.50%</th>
<th>97.50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>State¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.9493</td>
<td>0.03721</td>
<td>-1.026</td>
<td>-0.8723</td>
</tr>
<tr>
<td>Roadway mileage</td>
<td>-0.2091</td>
<td>0.02726</td>
<td>-0.2651</td>
<td>-0.1549</td>
</tr>
<tr>
<td>Curve crashes</td>
<td>0.03301</td>
<td>0.006056</td>
<td>0.0212</td>
<td>0.04499</td>
</tr>
<tr>
<td>Wet roadway</td>
<td>0.003339</td>
<td>0.003007</td>
<td>-0.00255</td>
<td>0.009248</td>
</tr>
<tr>
<td>Roadway defects</td>
<td>0.03294</td>
<td>0.006668</td>
<td>0.01983</td>
<td>0.04603</td>
</tr>
<tr>
<td>$\sigma_\phi$</td>
<td>0.24</td>
<td>0.04201</td>
<td>0.1567</td>
<td>0.3136</td>
</tr>
<tr>
<td>$\sigma_\theta$</td>
<td>0.1508</td>
<td>0.05595</td>
<td>0.03403</td>
<td>0.2423</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.6206</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TRB 2014 Annual Meeting  
Paper revised from original submittal.
Not all crashes have the same impact. The more severe crashes result in higher costs and potential fatalities. Consequently, instead of only using a frequency based approach, it is possible to assign different weights to the crashes with varying severity. FHWA Five Percent Report proposes a severity weighing scheme for crashes in which the weights for fatality is 15, incapacitating injury is 7, non-incapacitating injury is 4, possible injury is 2 and property damage is only 1 (28). Although the theoretical background behind the weighing scheme is not explained in the report, it suggests a guideline for weighing the crashes according to their severity. The severity weights in this report will be adopted to re-estimate the previous model. Since non-incapacitating injury and possible injury crashes are not differentiated in the data, for all injury
crashes, except incapacitating injury, 3 is assigned as the weight. Table 4 shows the results of
modeling when the severity weights are applied to crash counts. The results indicate a similar
trend to the model based on raw crash counts. Note that the spatial correlation decreased in this
case while global unobserved heterogeneity increases. This might suggest that there are other
unobserved underlying factors affecting the severity of the crashes.

### Table 4: Modeling results based on severity weighted crash counts

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D.</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.56660</td>
<td>0.02632</td>
<td>0.51050</td>
<td>0.62290</td>
</tr>
<tr>
<td>Roadway mileage</td>
<td>0.09741</td>
<td>0.00657</td>
<td>0.08404</td>
<td>0.11010</td>
</tr>
<tr>
<td>Curve crashes</td>
<td>0.12100</td>
<td>0.00532</td>
<td>0.11070</td>
<td>0.13160</td>
</tr>
<tr>
<td>Wet roadway</td>
<td>0.02312</td>
<td>0.00157</td>
<td>0.02005</td>
<td>0.02622</td>
</tr>
<tr>
<td>Roadway defects</td>
<td>0.06717</td>
<td>0.00053</td>
<td>0.06611</td>
<td>0.06821</td>
</tr>
<tr>
<td>$\sigma_\phi$</td>
<td>0.47050</td>
<td>0.02996</td>
<td>0.40940</td>
<td>0.52900</td>
</tr>
<tr>
<td>$\sigma_\theta$</td>
<td>0.10300</td>
<td>0.04928</td>
<td>0.02962</td>
<td>0.18780</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.82420</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\hat{D} : 2643.32, D(\tilde{\theta}) = 2432.57, p_D = 24.646, DIC = 2854.06$

Figure 4 shows the crash rate maps for 2010 which are developed based on the model
findings of mean crash rates from raw crash counts (Figure 4.a) and from crash counts when
severity weights are applied (Figure 4.b). In Figure 4.a, the natural breaks in Figure 1.b is used to
make the comparison easy between the crash rates from the model and the real data. It can be
observed that the model estimates are proximate to the real crash rates, which confirms the fit of
the model to the data. The comparison of Figure 4.a and 4.b indicates that the crash severity is
not uniformly distributed in New Jersey. Figure 4.c presents the differences in crash rates
between both models. Three of the counties (Passaic, Essex and Hudson) have the highest
difference when the severity weights are considered. Since these are neighboring counties, the
site specific factors due to shared roadways might play a role in the higher rankings.
Additionally, the five counties that follow the top three counties have a similar trend. While
Union county neighbors the three counties mentioned earlier, four southern counties, Ocean,
Atlantic, Camden, and Cumberland are also neighbors of each other. This trend is not followed
for the remaining counties, except for Warren, Hunterdon, and Morris counties which have the
smallest change between two models. Based on the results, it can be recommended that the
counties that have higher rankings when severity weights are applied need to be further
researched. It is necessary to investigate the factors of the likelihood of occurrence of more
severe crashes at these locations and the cause behind the neighboring counties to have similar
trends.
FIGURE 4 Mean crash rates (a) from raw crash counts (b) from severity weighted crash counts (c) difference between raw and severity weighted crash counts by county in 2010 (per thousand DVMT).
Crash rate maps are also developed for the different road types as shown in Figure 5. Since DVMT data is not available at the jurisdiction level, county DVMT miles are used for estimating the road based on crash rates. However, the roadway length by jurisdiction is available and the model was estimated using this data. The limitations of the data might have a negative effect on the accuracy of the results as the exposure to traffic is different for various road types. The model results reveal lower crash rates than the actual rates for each road type, but the sum of the crash rates for different road types is equal to the crash rate of a county for all crashes. This is due to the fact that the same county level DVMT is used for all models. These results indicate that crash rates vary for different counties in terms of road types. Thus, crash rates are not spatially consistent within each county independent of the road type. This can have major policy implications because one county that appears to have lower overall crash rates can have higher crash rates in terms of its county or municipal roadways. Moreover, it is found that the state and authority roadways have lower crash rates than the county and municipal roadways. This result can be used to address possible funding and other geometric and operational improvement issues for roads where there is room for improvement compared to other roads located within the same county.

CONCLUSION

This paper presents a county level hierarchical Bayesian model framework for New Jersey. Unlike regular NB models, use of hierarchical Bayesian models enables us to include spatial correlation and global heterogeneity simultaneously. To develop a truly spatial model, only spatial covariates related to roadway conditions are considered in the model development stage. However, due to limitations of the data, spatial covariates are generated from a set of surrogate variables. It is recognized that the surrogate variables may not fully represent the effects of the considered spatial covariates which is a limitation of this study. The spatial variation of crashes is also analyzed by roadway type. However, due to data limitations, DVMT in a county is used instead of the DVMT by roadway type. In addition to common practice of modeling based on raw crash counts, we used of crash counts with severity weights in the hierarchical crash models. Severity weights proposed in FHWA’s “Five Percent Report” is used to represent the impact of the different severity crashes. Finally, the crash rate maps are developed from modeling results for raw crash counts, weighted crash counts, and crashes by roadway type. These maps can be valuable tools to visualize the spatial meaning of models that are not always easy to understand by just observing estimated model equations. Furthermore, the maps based on the spatial models developed in this study enables a fast and intuitive understanding of model predictions especially for non-statisticians.

The results of the study indicate that the most influential covariate for the crashes is the roadway curvature, followed by roadway mileage and roadway defects. “Wet roadway” is found to be the least influential factor. By applying severity weights to the crashes in the hierarchical models, it is found that it is possible to represent the crash risk better using this new model. The developed maps based on the raw as well as the weighted crash rates from the model identified counties which are prone to more severe crashes. It is believed that the results of this study will help transportation professionals on identifying and ranking the locations at an aggregate level, which requires closer attention. The crash rate maps can be used to pin-point local needs and allocate funds by the state or federal governments. The maps of crash rate by roadway type can also be used to understand the effectiveness of local governments in terms of highway safety.
FIGURE 5  Mean crash rates for (a) state (b) authority (six counties are excluded) (c) county (d) municipal roadways by county in 2010 (per thousand DVMT).
For future studies, it is recommended that the methodology described in this paper is extended to analyze more disaggregate level data such as crashes at the municipality level or crashes on individual roadway segments. It will be helpful to investigate the differences in the crash rates and spatial correlation in smaller regions. Although a model for various roadway characteristics is developed in the study, it is also possible to examine the effects of socio-economic, demographic and land use variables. Additionally, more accurate results can be obtained from the model if data from surrogate variables can be replaced with real data. Bayesian hierarchical models offer a flexible framework to model temporal random effects in crash data as well as spatial random effects. Future efforts should focus on the investigation of temporal effects and space-time interactions in crash data. For the severity weights, different weighing approaches such as economic impact approach in which property damage only equivalents of accidents are considered (29) can be investigated.

ACKNOWLEDGEMENTS
The contents of this paper reflect views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents of the paper do not necessarily reflect the official views or policies of the agencies.

REFERENCES


