Accident Analysis and Prevention 37 (2005) 699–720

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion, and spatial dependence

Shaw-Pin Miaou [a,*], Joon Jin Song [b,1]

[a] Texas Transportation Institute, Texas A&M University System, 3135 TAMU, College Station, TX 77843-3135, USA

[b] Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, USA

Received 13 March 2005; received in revised form 13 March 2005; accepted 14 March 2005

Abstract

In recent years, there has been a renewed interest in applying statistical ranking criteria to identify sites on a road network, which potentially present high traffic crash risks or are over-represented in certain type of crashes, for further engineering evaluation and safety improvement. This requires that good estimates of ranks of crash risks be obtained at individual intersections or road segments, or some analysis zones. The nature of this site ranking problem in roadway safety is related to two well-established statistical problems known as the small area (or domain) estimation problem and the disease mapping problem. The former arises in the context of providing estimates using sample survey data for a small geographical area or a small socio-demographic group in a large area, while the latter stems from estimating rare disease incidences for typically small geographical areas. The statistical problem is such that direct estimates of certain parameters associated with a site (or a group of sites) with adequate precision cannot be produced, due to a small available sample size, the rareness of the event of interest, and/or a small exposed population or sub-population in question. Model based approaches have offered several advantages to these estimation problems, including increased precision by "borrowing strengths" across the various sites based on available auxiliary variables, including their relative locations in space. Within the model based approach, generalized linear mixed models (GLMM) have played key roles in addressing these problems for many years.

The objective of the study, on which this paper is based, was to explore some of the issues raised in recent roadway safety studies regarding ranking methodologies in light of the recent statistical development in space–time GLMM. First, general ranking approaches are reviewed, which include naïve or raw crash-risk ranking, scan based ranking, and model based ranking. Through simulations, the limitation of using the naïve approach in ranking is illustrated. Second, following the model based approach, the choice of decision parameters and consideration of treatability are discussed. Third, several statistical ranking criteria that have been used in biomedical, health, and other scientific studies are presented from a Bayesian perspective. Their applications in roadway safety are then demonstrated using two data sets: one for individual urban intersections and one for rural two-lane roads at the county level. As part of the demonstration, it is shown how multivariate spatial GLMM can be used to model traffic crashes of several injury severity types simultaneously and how the model can be used within a Bayesian framework to rank sites by crash cost per vehicle-mile traveled (instead of by crash frequency rate). Finally, the significant impact of spatial effects on the overall model goodness-of-fit and site ranking performances are discussed for the two data sets examined. The paper is concluded with a discussion on possible directions in which the study can be extended.

Keywords: Roadway safety; Problem site identification; Bayesian ranking method; Small area/domain estimation problem; Disease mapping problem; Generalized linear mixed model; Decision parameter; Treatability; Statistical ranking criterion; Spatial dependence

1. Introduction

In recent years, there has been a renewed interest in applying statistical ranking criteria to identify sites on a road network, which potentially present relatively high traffic crash

* Corresponding author. Tel.: +1 979 862 8753.
E-mail addresses: s-miaou@tamu.edu (S.-P. Miaou), jsong@math.umass.edu (J.J. Song).

[1] Present address: Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003-9305, USA. Tel.: +1 413 577 0255.

0001-4575/$ – see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2005.03.012

risks or are over-represented in certain type of crashes, for further engineering evaluation and safety improvement (Persaud et al., 1999; Heydecker and Wu, 2001; Hauer et al., 2002; Kononov, 2002; Hallmark and Basavaraju, 2002; Midwest Research Institute, 2002; Agent et al., 2003). For different purposes, crash risks might be evaluated in absolute crash frequencies, crash frequency rates, or crash cost rates, and "sites" of interest could range from individual network elements, such as intersections, road segments, and ramps, to corridors or to some traffic analysis zones that satisfy certain physical, operational, and administrative specifications.

In this renewed quest to build better ranking and selection methodologies to meet the need and challenge of improving roadway safety, many basic issues, some quite sticky, have been raised and discussed. These issues include:

(1) Identifying appropriate roles and objectives of ranking methodologies in achieving the overall strategy of improving roadway safety;
(2) Understanding the context and process in which the methodology is to be applied;
(3) Developing implementation strategies with various levels of data requirements;
(4) Identifying target users and considering their resource and data constraints;
(5) Choosing ranking approaches that can properly handle the highly stochastic nature of the crash event due to low and uneven exposures by site;
(6) Selecting decision parameters for ranking, e.g., crash frequency rate or crash cost rate in number of crashes or US$ 1000, respectively, per vehicle-miles traveled (VMT) or per vehicle entry (VE);
(7) Making the use of treatability concept to assess known traffic, geometric, and other explanatory covariate effects and determining the need to include spatial, temporal, and unobserved site heterogeneities in such an assessment;
(8) Adopting absolute or relative standard for problem site selection—selecting sites based on whether they exceed a pre-determined threshold or "cut-off" point in risk levels or selecting among a group of sites based on their relative risk levels;
(9) Selecting statistical ranking criteria, e.g., ranking by some statistical means, expected ranks, or probability of being the worst site;
(10) Conducting benefit–cost analysis of applying various engineering treatment alternatives to reduce the frequency and injury severity of crashes.

The quest is timely. In the last decade or so, many highway departments have been equipped with improved desktop computing and data storage capabilities, and better tools to link computerized traffic crash, road inventory, and other related databases, which have been gradually geo-coded using Geographical Information Systems (GIS) and related technologies. On the methodological front, thanking again in part to the unprecedented computing power, the focus of the development in statistical modeling techniques have marched from the fixed-effect type of generalized linear models to random-effect models, which allow probabilistic and functional structures of complex space–time and site-specific heterogeneities to be more realistically considered and tested in statistical analyses (Besag et al., 1995; Gilks et al., 1996; Carlin and Louis, 1996; Dey et al., 2000; Goldstein, 2003). The underlying engine that powers and accelerates this statistical advancement is the Markov chain Monte Carlo (MCMC) technique, which could perform needed multidimensional numerical integration for estimating parameters of complex hierarchical models that is almost impossible to do with the traditional numerical algorithm (Gilks et al., 1996; Carlin and Louis, 1996; Robert and Casella, 1999). In turn, the development of MCMC algorithmic strategies has been mutually reinforced by the experience gained from building the Bayesian generalized linear mixed models (GLMM) with complex space–time effects (Møller, 2003). Recent applications of these models and MCMC techniques in roadway safety studies include Tunaru (1999), Miaou and Lord (2003), and Miaou et al. (2003).

The site-ranking problem requires that good estimates of ranks of crash risks be obtained at individual intersections or road segments, or some analysis zones. The nature of this problem in roadway safety is related to two well-established statistical problems known as the small area (or domain) estimation problem and the disease mapping problem (Rao, 2003; Lawson, 2001). The former arises in the context of providing estimates using sample survey data for a small geographical area or a small socio-demographic group in a large area, while the latter stems from estimating rare disease incidences for typically small geographical areas. The statistical problem is such that direct estimates of certain parameters associated with a site (or a group of sites) with adequate precision cannot be produced, due to a small available sample size, the rareness of the event of interest, and/or a small exposed population or sub-population in question. Model based approaches have offered several advantages to these estimation problems, including increased precision by "borrowing strengths" across the various sites based on available auxiliary variables, including their relative locations in space. Within the model based approach, GLMM have played key roles in addressing these problems for many years.

The objective of the study, on which this paper is based, was to explore some of the issues raised in recent roadway safety studies regarding ranking methodologies in light of the recent statistical development in space–time GLMM. Specifically, this study attempted to address some of the problems related to issues (5)–(9) as listed above. First, general ranking approaches are reviewed, which include naïve or raw crash-risk ranking, scan based (including filtering- or sliding window-based) ranking, and model based ranking. Through simulations, the limitation of using the naïve approach in ranking is illustrated. Second, following the model based approach, the choice of decision parameters and consideration of treatability are discussed. Third, several statistical ranking criteria that have been used in public health and other scientific studies are presented from a Bayesian perspective. Their applications in roadway safety are then demonstrated using two data sets: one for individual urban intersections and one for rural two-lane roads at the county level. As part of the demonstration, it is shown how multivariate spatial GLMM can be used to model traffic crashes of several injury severity types simultaneously and how the model can be used within a Bayesian framework to rank sites by crash cost rate (instead of by crash frequency rate). Finally, the significant impact of spatial effects on the overall model goodness-of-fit and site ranking performances are discussed for the two data sets examined.

Many model based ranking studies in traffic safety cited earlier have adopted the Empirical Bayes (EB) methods for site-specific parameter estimation. In this paper, we follow the full Bayesian paradigm in model construction, comparison, parameter estimation, and ranking. The advantage of full Bayes treatments is that it takes full account of the uncertainty associated with the estimates of the model parameters and can provide exact measures of uncertainty. The maximum likelihood and EB methods, on the other hand, tend to overestimate precision because they typically ignore this uncertainty (Goldstein, 2003; Rao, 2003). In practice, the estimation results between full Bayes and EB methods can be very different for individual sites under the same model structure if some model parameters are not well estimated due, e.g., to a small sample size (Miaou and Lord, 2003). Another advantage of taking the full Bayesian approach in such a study is its capability to incorporate some of the decision-making functions, such as a pre-determined loss function, in the model in a relatively natural manner when compared to other approaches.

This paper is organized as follows:

- Section 1: Introduction.
- Section 2: Naïve and scan based ranking approaches.
- Section 3: Model based ranking approach.
- Section 4: Decision parameter and treatability concept.
- Section 5: Statistical ranking criteria.
- Section 6: Considering crash severities using multivariate spatial models.
- Section 7: Data sets and illustrations.
- Section 8: Spatial dependence.
- Section 9: Future extensions.

2. Naïve and scan based ranking approaches

The model based approach is the focus of this study. For reference purposes, we provide a brief review of the naïve and scan based approaches in this section. Before the review, we note that spatial data are typically classified into three basic types: point-referenced or geostatistical data, areal or lattice data, and point pattern data (Banerjee et al., 2004). The focus of this study is on the areal data, where there are a finite number of sites under consideration and each site covers an area with a well-defined boundary. In addition, except for some situations under the scan based approach, the sites do not overlap in their areas.

The readers may wish to consult Cressie (1993) for modeling of point-referenced data and to Lawson and Denison (2002) and Diggle (2003) for recent treatments of point pattern data, which concern mainly with spatial cluster detection and modeling, where, given individual georeferenced events, the number and boundary of sites or clusters are themselves the parameters to be estimated. Also, Levine (2004) provides some exploratory tools for identification of clusters or “hotspots” under a GIS environment for the point pattern data.

2.1. Naïve or raw crash-risk based ranking approach

Under this approach, the raw crash frequency rate calculated as the number of crashes per VMT or VE is typically calculated for each site individually without using any data from other sites. Based on these raw crash frequency rates, ranks are rendered. There are several fundamental problems with this approach (some of which have been discussed in previous studies): (1) for the ranking to be valid, the linear assumption between number of crashes and VMT or VE has to hold, which has been argued in many studies to be false if the range of traffic volumes among the compared sites is wide; (2) the important interactive effects of traffic flows and other covariates and between traffic flows of major and minor approaches at intersections are ignored; and (3) the raw crash rate as an estimate of the true crash rate for a site is very crude and inefficient, which suffers from a large statistical uncertainty, when the level of vehicle exposures is low. Note that, at the segment-intersection or “link-node” level, the low exposure problem is the norm, rather than the exception, in most of the traffic crash studies that focus on individual roadway design elements.

The last problem is related to the “small area estimation problem” discussed earlier. To give an example of what this problem can do to the site ranking and selection problem of interest here, we conduct a simple computer simulation with the following scenario: consider a hypothetical stretch of roads that is 10 mile in length and is totally homogeneous in design and traffic flow, with crashes *equally likely to occur* at any point on the road. Next, randomly place 350 crashes on the road using a uniform random number generator. Independently, the 10 mile road is cut into 30 segments in a random order along the stretch with three length groups: 10 short, 10 medium-size, and 10 long segments. The lengths of the medium-size and long segments are 10 and 100 times longer than that of the short segments, respectively. In other words, the short, medium-size, and long segments are 1/1110, 1/111, and 10/111 miles in length, respectively. Then, compute raw crash frequency rates for individual segments (in number of crashes per mile), and use these rates for ranking in decreasing order, i.e., the segment with the highest rate is given the “highest rank” of 1.

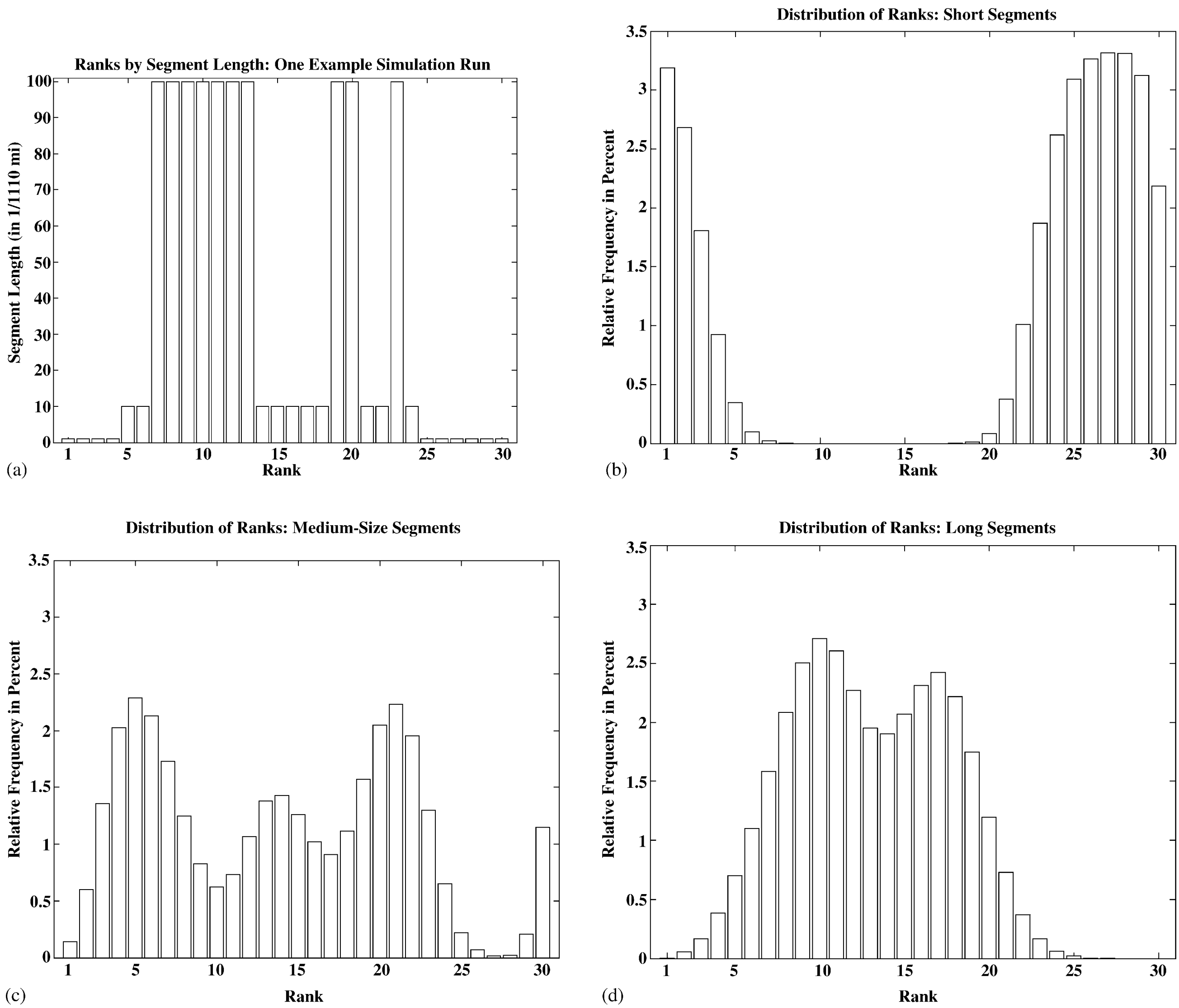


Fig. 1. Simulation results from a naïve ranking approach.

As an example, Fig. 1(a) shows the relationship between segment length and rank from one simulation run. Given that all simulated segments, regardless of their lengths, have the same crash risks, we expect to see the length–rank relationship to be random if a sensible approach is taken to rank them. The simulation scenario as described above was conducted 10,000 times and the relative frequencies of each rank being assigned to short, medium-size, and long segments are calculated and presented in Fig. 1(b–d), respectively. The relative frequencies in these figures sum to 100%. In addition, the maximum relative frequency possible for a size group to occupy a particular rank is about 3.33% (=1/30). For example, in Fig. 1(b), the relative frequency for the short segment to be ranked as number one is about 3.2%, which almost reaches the maximum.

Because of the high statistical uncertainty associated with the estimates for small segments, they tend to produce extreme high and extreme low raw crash rates. As a result, using such estimates in site ranking produces extreme high and extreme low ranks as seen in Fig. 1(b). On the other hand, the ranks so obtained for the long segments tend to be shrunk to the middle, as the bell-like shape shown in Fig. 1(d). The intended lesson from this simulated example is that the naïve ranking method tends to improperly select sites with low exposures, due either to their small physical sizes or low traffic volumes, as the potential problem sites (or as the safest sites for the same reason).

2.2. Scan based ranking approach

This approach is also called filtering- or sliding window-based approach. It could be a scan of road segments over a network with a pre-determined window size, say, 1 mi, and a moving step size of, say 0.1 mi. It could also be a cell- or grid-based analysis, using a pre-determined size of cells or grids to aggregate and produce averages for crashes, vehicle-miles, and other explanatory covariate values. The most primitive scan based analysis would just compute crashes (of certain type) per VMT for each window or grid independently, without making any reference to the statistics associated with the background population, and observe the changes as the window or grid moves systematically in a pre-determined step size. More sophisticated scan could adjust the estimate based on statistics observed for some background population or a pre-screened reference group with similar key characteristics. The most sophisticated scan could be assisted by some form of the Poisson-based regression models with some parameters fixed based on a background population and some parameters adjusted based on "local" data. This last method would allow adjustments of a number of explanatory covariates provided by the user. An example software package for this type of analysis is SaTScan$^{\text{TM}}$—Spatial and Space–Time Scan Statistics—of National Cancer Institute (for reference, see http://srab.cancer.gov/satscan/).

The main difference among the three levels of the scan based approach described above is the extent and manner in which the background population data is utilized, which affects the statistical uncertainties of the estimated parameters at each scan. The "small area estimation" problem, as illustrated in the naïve ranking approach, can potentially bias the ranking results when no or limited background population data is utilized (especially when the size of the sliding window is small or when scanning over a low traffic volume area). The exact statistical uncertainty of the parameter estimated at each window or grid is usually difficult to produce for this approach. Selection of best window or grid size and moving step size for picking up "local patterns" is the strength of this approach, but it can also be difficult to do in practice and the analysis results can be sensitive to the selection.

3. Model based ranking approach

The model based approach is the focus of this paper. As mentioned earlier, this approach offers several advantages to the small area estimation problem, including increased precision. Two other commonly cited advantages in the literature are the derivation of "optimal" estimates and associated measures of variability under an assumed model, and the ability to validate models from the sample data (Rao, 2003).

3.1. General modeling framework

This approach has been used for analysis of traffic crash data at individual intersections and segments, as well as at an areal or zonal level such as county. To set the stage for our discussion, we first present a framework that is quite general in modeling vehicle crashes and define necessary terms and statistical notations. Consider the following functional and probabilistic structures: the number of crashes at the ith site and tth time period, Y_{it}, when conditional on its mean μ_{it}, is assumed to be a random variable which is Poisson and independently distributed overall sites and time periods:

$$Y_{it}|\mu_{it} \overset{\text{ind}}{\sim} \text{Po}(\mu_{it}) \quad \text{where} \quad i = 1, 2, \ldots, m \quad \text{and} \quad t = 1, 2, \ldots, T \tag{1}$$

The mean of the Poisson is structured as:

$$\mu_{it} = f(F_{1,it}, F_{2,it}, x_{it}; \beta)\, \exp(\delta_t + \phi_i + \varepsilon_{it}) \tag{2}$$

where $f(\cdot)$ is a positive-valued function of traffic flows from the major and minor approaches of the intersection in, say, 1000 vehicles/day, represented by $F_{1,it}$ and $F_{2,it}$, respectively, and other explanatory covariates indicated by vector variables x_{it}, for non-intersections $F_{2,it}$ is set to zero; β a vector of unknown "fixed-effect" parameters; δ_t represents global time effects due, e.g., to statutory changes in posted speed limits and changes in weather and socio-economic conditions; ϕ_i a spatial random effect which will be described in more detail shortly; and ε_{it} is an independent random effect that is unstructured in its mean and variance and is identically distributed over space. Furthermore, it is commonly assumed that δ_t, ϕ_i, and ε_{it} are mutually independent.

The component $\exp(\varepsilon_{it})$ has a typical independent one-parameter gamma distribution assumption with mean equal to one and variance $1/\psi_t$ for all i (with $\psi_t > 0$ for all t). It can be viewed as un-modeled variations in its mean due to isolated site-time specific heterogeneities. The key assumptions here are that (1) $\exp(\varepsilon_{it})$ are independent (or, more strictly and statistically speaking, exchangeable) across all i and have a fixed variance at a particular time period t and (2) ε_{it} are independent of all covariates, including flows and x_{it}. In this paper, we will call ψ_t the "inverse dispersion parameter" of the model. As ψ_t increases, the amount of over dispersion (over the Poisson model), when conditional on other fixed and random effect components, decreases. Without much loss of generality, we assume one time period to be 1 year in the following discussion. Note that for a Bayesian interpretation of "fixed" and "random" effects, the readers are referred to the book "*Bayesian Data Analysis*" (Gelman et al., 1995).

Variants of the space–time model presented in Equation (2) have been used in Miaou and Lord (2003) and Miaou et al. (2003). For urban intersections, one of the functional forms, which was found to perform well statistically when compared to several other popular forms, is slightly modified and adopted in this study as follows:

$$\mu_{it} = (F_{1,it}\lambda'_{1,it} + F_{2,it}\lambda'_{2,it})\exp(\beta_3 \text{FR}_{it} + \phi_i + \varepsilon_{it}) \tag{3}$$

where $\lambda'_{1,it} = \exp(\beta_{0,t} + \beta_1 F_{2,it})$, $\lambda'_{2,it} = \exp(\beta^*_{0,t} + \beta_2 F_{1,it})$, FR_{it} stands for flow ratios ($=F_{2,it}/F_{1,it}$ and typically $F_{2,it} \leq F_{1,it}$), and, as before, β's are unknown regression parameters. The part of the functional form pertaining to traffic flows is based on a logic that vehicles entering from the major and minor approaches may have different risks, characterized by their "partial" crash rates $\lambda'_{1,it}$ and $\lambda'_{2,it}$, respectively. Note that the flows in the equation can be multiplied by number of days per year, e.g., 365, to obtain annual rates for the two approaches. This concept of differential risks was first introduced in Miaou and Lord (2003) and was elaborated as follows: take a vehicle entering the intersection from the major approach as an example. The functional form of the "partial" crash rate $\lambda'_{1,it}$ postulates first that this vehicle is exposed to a certain level of crash risks involving the vehicle itself and vehicles in the same approach, which is captured by parameters $\beta_{0,t}$. Second, this vehicle is exposed to risks due to vehicles entering from the minor approach, which is captured by the term $\beta_1 F_{2,it}$. The same logic is then applied to a vehicle entering from the minor approach.

In a county level crash data analysis, which included both intersection and segment crashes on rural two-lane low-volume state roadways, Miaou et al. (2003) used a functional form similar to the following equation:

$$\mu_{it} = v_{it} \exp\left(\beta_0 + \sum_{k=1}^{K} \beta_k x_{k,it} + \delta_t + \phi_i + \varepsilon_{it}\right) \quad (4)$$

where the total amount of VMT, represented by v_{it}, is treated as an offset in the model, and $x_{k,it}$, $k = 1, 2, \ldots, K$, are a set of explanatory covariates. This particular functional form was adopted in this study for a county level data analysis and ranking. Also, in Miaou et al. (2003), the time effect δ_t was treated in two ways: as fixed effects varying by t (or a "year-wise" fixed-effect model) and as an order one autoregressive model (AR(1)) with the same coefficient for all t. For the model to be identifiable, in the fixed-effect model δ_1 was set to zero and in the AR(1) model δ_1 was set to be an unknown fixed constant. In this study, two temporally short data sets, three time periods each, were used and the "year-wise" fixed-effect model was adopted. The importance of including a spatial random component, ϕ_i, in modeling vehicle crashes was suggested in Miaou and Lord (2003) and further demonstrated in Miaou et al. (2003). In the following discussion, we provide some background information about this random spatialcomponent.

3.2. Modeling spatial effects

In 1988, Anselin and Griffith (1988) reminded the statistical research community with an old and tacit issue regarding regression analysis: "Do spatial effects really matter in regression analysis?" They demonstrated that spatial effects really do matter in regression analysis and ignoring spatial dependence, structure, and heterogeneity may lead to biased estimates and serious errors in the interpretation of familiar regression diagnostics and mis-specification tests. Since then, there has been a steady pace of research and applications in spatial regressions in social, environmental, agricultural, public health, and physical sciences. Also, many books on spatial modeling have been published, e.g., Haining (1990), Cressie (1993), Gatrell and Bailey (1996), Lawson (2001), Lawson and Denison (2002), Fotheringham et al. (2002), Diggle (2003), and Banerjee et al. (2004).

Traditionally, analysts tend to pay more attention to autocorrelation in time than that in space due, perhaps, to the lack of exposure to spatial statistics and limited availability of georeferenced data and spatial statistical tools. With the advances in spatial statistics research, we know the role and impact of spatial dependence in regression analysis are actually similar to those of temporal dependence. In addition, there has been a steady trend of convergence in the methods used to model them (Congdon, 2001; Diggle, 2003). As a result, in recent years, a more balanced treatment of space–time data seems to have emerged in some of the scientific fields mentioned above.

Two particular formulations of spatial models have dominated the spatial statistics literature: Gaussian conditional autoregressive (CAR) model and simultaneous autoregressive (SAR) model. Banerjee et al. (2004) and Griffith (1988) provide some review and discussion of the relationship between these two types of spatial models, their variants, and their applications in various scientific fields. Here we only give a brief introduction of the Gaussian CAR model, which was used by this study. One general formulation of the joint probability distribution of ϕ under the Gaussian CAR model is a pairwise difference specification as follows (Ghosh et al., 1999; Sun et al., 2000):

$$p(\phi_1, \phi_2, \ldots \phi_m) \propto \exp\left(-\frac{1}{2\sigma_\phi^2} \sum_{\{(i,j):1\le i<j\le m\}} c_{ij}[\phi_i - \rho\phi_j]^2\right) \quad (5)$$

where "$\propto$" stands for "proportional to"; ρ is a parameter between -1 and 1 and determines the direction and magnitude of the overall spatial dependence; $c_{ij} \ge 0$ ($i, j = 1, 2, \ldots, m$) are proximity measures that determine the relative influence of site j on site i, with c_{ii}, $i = 1, 2, \ldots, m$, usually set to 0 to preclude "self predicting" the summation is taken over all possible pairs of i and j (with $j \ne i$); and σ_ϕ^2 is a parameter controlling the variance of the distribution function.

Two types of proximity measures are commonly used in the literature: adjacency-based and distance-based measures. A typical adjacency-based measure has $c_{ij} = 1$ or 0 depending on whether sites i and j are "neighboring" sites or not. Many distance-based weighting functions have been proposed for use in modeling c_{ij} (LeSage, 2001). Two popular weighting functions are: inverse distance weighting function and exponential distance-decay weighting function. In this study, we adopted the latter in modeling spatial effects: $c_{ij} = \exp(-\alpha d_{ij})$, where d_{ij} is the straight-line distance from the centroid of site i to the centroid of site j. The parameter α (>0) is a fixed parameter controlling the rate of decline of correlation with distance, with larger α values indicating more rapid decay and thus smaller neighboring area of influence.

This joint distribution of the spatial random effect as expressed in Equation (5) can be shown to be uniquely determined by a set of full conditional distributions as follows:

$$p(\phi_i|\phi_{-i}) \propto \exp\left(-\frac{c_{i+}}{2\sigma_\phi^2}\left[\phi_i - \rho\sum_{j\neq i}\frac{c_{ij}}{c_{i+}}\phi_j\right]^2\right),\quad i = 1, 2, \ldots, m \tag{6}$$

where ϕ_{-i} indicates all ϕ's except ϕ_i; $c_{i+} = \sum_{j\neq i} c_{ij}$, which sums over all j except i. This formulation can be shown to give a conditional normal density for ϕ_i with mean $\rho\sum_{j\neq i}(c_{ij}/c_{i+})\phi_j$ and variance σ_ϕ^2/c_{i+}. Although the proof that Equation (5) is uniquely determined by Equation (6) was introduced over 30 years ago, the great computational savings of using Equation (6) over Equation (5) in modeling spatial effects was unleashed for only about a decade or so, after the rediscovery of MCMC methods by statisticians, particularly the Gibbs sampling method. More detailed introduction of the history of CAR models and the associated statistical theories can be found in Chapter 3 of Banerjee et al. (2004).

In practice, the value of spatial correlation parameter ρ is expected to be between 0 and 1. In other words, we expect the un-modeled spatial effects for areal units that are near to each other to take more similar values than those for units that are far from each other. An immediate interpretation for $\rho = 0$ in Equation (6) is that, ϕ_i, $i = 1, 2, \ldots, m$, are independent and normally distributed with mean 0 and variance σ_ϕ^2/c_{i+}. However, we need to proceed with caution in placing a direct interpretation of the parameter ρ as the "strength of spatial dependence" in the sense of common spatial descriptive measures, such as Moran's *I* or Geary's *C*. For example, through simulation examples, Banerjee et al. (2004) warned that, for the cell units in a 10×10 regular grid under a simple adjacency-based proximity measure, a CAR model for such an areal data can have a very high ρ value (close to 0.9) even when the Moran's *I* value is low (less than 0.25). Note that in this study we did not use the adjacency-based proximity measure.

The most recognized possibility for having a correlated short-memory time effect in time series data analysis is the so-called omitted variable problem. It occurs when temporal variations of omitted covariates cause the dependent variable to vary over time, which cannot be fully accounted for in the model with the available covariates. By the same token, the most popular interpretation for having a correlated "local" spatial effect in spatial data analysis is also the omitted covariates—both observables and unobservables, which vary in space, usually smoothly over small distance bands. Interesting mechanisms postulated in the literature regarding this effect include: (1) interaction, clustering, spillover, externality, diffusion, attraction, and copycat effects via some mechanisms through which actions or phenomena at a given location affect actions or properties of the phenomena at other locations; (2) varying degrees of measurement errors over space; and (3) mis-specification of functional forms for the mean of the response in the model. The concept of a phenomenon spreading through geographical space has been considered in many diverse subject areas, such as the spread of infectious disease, growth of an urban center, the spread of wildfires, diffusion of innovation, and ripple, or chain effects in an ecological system. However, a thorough understanding of the mechanisms of spread remains elusive due to limitations both with the data and the state of development in spatial models (Anselin, 2003; Holloway et al., 2002).

An interesting example of the spillover effect in road safety is the traffic crash "migration effect" observed in Boyle and Wright (1984). In general, this is a phenomenon whereby the crash rate rises at sites that are untreated but that are "neighbors" to treated sites. Specifically, in Boyle and Wright's study it was observed that, after engineering treatments of a number of selected "blackspots," the traffic crash rate in the neighbors of treated sites increased.

Maher (1990) studied the phenomenon observed by Boyle and Wright analytically where the spatial correlation was modeled using a bivariate negative binomial model. He found that the size of the migration effect could not be explained by the site selection effect alone due to the conditioning, which is known to generate upward bias from the regression-to-the-mean (RTM) effect on the treated sites and reverse RTM effect to the untreated neighboring sites. In other words, he tried to find if reverse RTM effect gave an apparent migration of crashes from treated to untreated sites, even when the treatment was ineffective. He concluded that the magnitude of this reverse RTM effect was not sufficient in practice to provide a satisfactory explanation for the migration observed by Boyle and Wright. Further, he indicated that if this migration effect were a genuine effect, it would have serious implications for the assessment of remedial treatments. A more detailed study of the migration effect exemplified here seems to call for some type of spatial random effect models as presented above.

3.3. Bayesian inferential and computational methods

As indicated earlier, a full Bayesian approach was adopted in this study. To complete the full Bayesian specification, non-informative priors were used for all the hyperparameters involved. In addition, the specification of hyperpriors followed closely to those used in Miaou et al. (2003). For example, in specifying priors, β, σ_ϕ^2, and ψ were assumed to be mutually independent, with β having rather "flat" independent normal priors and both σ_ϕ^2 and ψ having rather diffused gamma distributions. The spatial correlation parameter ρ in the CAR component model was assumed to have a uniform prior between 0 and 1. The prior for the exponential distance-decay parameter, α, was assumed to be uniformly distributed with lower and upper bounds calculated as Thomas et al. (2002) suggested.

For all the models presented in this paper, parameter estimates, inferences, and site rankings were obtained using programs coded in the WinBUGS language (Version 1.4; Spiegelhalter et al., 2003). These programs were executed on a Pentium IV 2.8 GHz notebook computer. Typically, 12,000–15,000 MCMC simulation iterations were used with the first 4000 iterations as burn-ins. In addition, Gelman–Rubin statistics available in WinBUGS were checked to ensure MCMC convergence. In developing the models, we usually started with very simple models and then used the estimates from the simple model as initials for a more complex model. For example, we usually took the estimated posterior means from a simpler model as initial estimates in a slightly more complex model. For the data sets examined in this study, we found the MCMC simulation runs converged rather quickly under such model development strategy. As in other iterative parameter estimation approaches, providing good initial estimates is usually a key to quick convergence.

The deviance information criterion (DIC) is a Bayesian generalization of the Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC) for hierarchical models where structured random effects are present (Spiegelhalter et al., 2002). When a model contains only fixed effects, DIC value of the model is close to its AIC value. Conceptually, this is because one "apparent" number of model parameters in a fixed effect model is equal to one "effective" number of model parameters. When a model contains random effects, the "apparent" number of model parameters is no longer equal to its "effective" number of model parameters because of the hierarchy. Simply put, one "apparent" number of parameters in the model could equal to, say, just 0.25 "effective" number of parameters, depending on the level to which this parameter belongs in the hierarchy and some other factors. DIC takes into account of this "effective" number of model parameters in its complexity measure for random effect models. Note that a Bayesian hierarchical model can have the "apparent" number of model parameters exceeds its total number of observations. We define standardized DIC as DIC/n, where n is the sample size of the data set. Both DIC/n and the inverse dispersion parameter values ψ_t were employed by this study in assessing the overall goodness-of-fit of models at various levels of complexities. The smaller the DIC/n value, the better the overall model performance, and the larger the ψ_t value, the smaller the unexplained and isolated site-specific heterogeneities (and thus the better the model performance).

4. Decision parameter and treatability concept

The decision parameter provides a link between the model, and thus the data, and the decision to be made. Therefore, the choice of decision parameters needs to take into consideration the context under which the rank is to be used, especially the range of treatments to be performed. In this regard, making an effort to obtain pertinent inputs from decision makers can be very useful.

In roadway safety studies, the site-specific decision parameter used in ranking, denoted by Θ_i for site i, is usually some function of μ_{it}, which can include traffic flow, covariate, space, time, and the un-modeled or exchangeable effects (as discussed under Equation (2)). For example, if the interest is to rank sites by crash frequency rate for the years for which the data are available, we may set $\Theta_i = \sum_{t=1}^{T}\mu_{it}/\sum_{t=1}^{T}365(F_{1,it} + F_{2,it})$ for the ith intersection and $\Theta_i = \sum_{t=1}^{T}\mu_{it}/\sum_{t=1}^{T}365(F_{1,it}\ell_i)$ for the ith road segment where ℓ_i is the length of the segment. As an alternative, one may want to rank sites based on predicted values of a particular out year, given predicted flows and predicted covariate values for each site. In some situations, it is also possible that one may not want to consider all the effects in μ_{it} when devising the decision parameter. For example, we may be interested in identifying sites that exhibit unusually high unobserved random effects so that we can conduct some field studies to find out what might have contributed to these effects, an appropriate decision parameter may be $\Theta_i = \sum_{t=1}^{T}\exp(\delta_t + \phi_i + \varepsilon_{it})$.

In the illustration section to be presented later in this paper, we have used $\Theta_i = \sum_{t=1}^{T}\mu_{it}/\sum_{t=1}^{T}365(F_{1,it} + F_{2,it})$ for ranking the crash frequency rate of individual urban intersections, $\Theta_i = \sum_{t=1}^{T}\mu_{it}/\sum_{t=1}^{T}v_{it}$ for ranking the crash frequency rate at the county level, and $\Theta_i = (\sum_{t=1}^{T}\sum_{s=1}^{S}\mu_{s,it}\text{cost}_s)/\sum_{t=1}^{T}v_{it}$ for ranking the crash cost rate when a multivariate version of the spatial sub-model is used to model crash frequencies of different crash severities simultaneously, where cost_s is the average economic cost of a crash with a severity level of s and $s = 1, 2, \ldots, S$.

Treatability is another concept that can be used to help select appropriate decision parameters when treatment options are known to be limited. In medical diagnostics and clinical profiling studies, the effect of explanatory covariates that could not be altered to improve the patient's condition (and have no interactive effects with other potentially treatable covariates or effects), e.g., patient's age, gender, and race, are sometimes adjusted out of the decision parameter (e.g., p. 33, Conlon and Louis, 1999). By adjusting out a particular covariate, we mean to set the value of that covariate to zero or a constant in μ_{it} across patients or sites. Ideally, we would like to devise a decision parameter that is most indicative of the type of crash risks that can potentially be treated by a planned project. In other words, the choice of the decision parameter should reflect the type of safety treatments that are within the scope of the planned project, which could, e.g., be a resurfacing project or a specific type of hazard elimination projects.

Let us consider a hypothetical scenario where the mean function μ_{it} includes three covariate effects due to geometries, say, paved shoulder width, horizontal curvature, and lane width. When devising a decision parameter for ranking sites for a pure resurfacing project where roadway geometries are not to be changed, these geometric effects should be adjusted out of the decision parameter. On the other hand, if edgeline or shoulder rumble strips are to be part of the treatment to reduce run-off-road crashes during resurfacing, then paved shoulder width should be considered in devising the decision parameter. In this scenario, whether the horizontal curvature should be adjusted out is admittedly subject to debate. If we choose to adjust out the horizontal curvature and lane width effects, we would essentially accept that, everything else being equal, a road segment with sharper horizontal curvatures and narrower lane width should not be given higher ranks over a segment with a mild curvature and wider lane width under such a resurfacing project where roadway geometries are not to be treated.

5. Statistical ranking criteria

Given a decision parameter, two types of ranking and selection standards have been used: absolute standard and relative standard (as described earlier). Most of the ranking and selection studies that have been performed in roadway safety, public health, and biomedical studies are considered relative standard (Morris and Christiansen, 1996; Schluter et al., 1997; Shen and Louis, 1998; Conlon and Louis, 1999; Heydecker and Wu, 2001). In this study, we only focused on the relative standard, i.e., ranking and selection among a predetermined group of sites based on their relative risk levels.

Our discussion of the statistical criteria for ranking will be from a Bayesian perspective. Given a decision parameter of interest, say Θ_i, the popular criteria that have been considered in the literature include:

- Ranking by posterior mean (PM) of the decision parameter: symbolically, it is expressed as $E_{\text{post}}[\Theta_i|y]$, i.e., the expected value of Θ_i taken over the posterior distribution of Θ_i given all data y, i.e., $p(\Theta_i|y)$.
- Ranking by posterior expected rank (PER) of the decision parameter: expressed as $E_{\text{post}}[R(\Theta_i)|y]$, the expected value of the true rank order of Θ_i, indicated as $R(\Theta_i)$, taken over the posterior distribution of $R(\Theta_i)$ given all data y, i.e., $p(R(\Theta_i)|y)$.
- Ranking by the probability that the site is the worst among all sites considered (PrWORST) in terms of the decision parameter: expressed as $p(\Theta_i > \Theta_j$, for all $j \neq i|y)$, the posterior probability that site i has the largest decision parameter value than any other site, given all data y.

Other less popular criteria include histogram or empirical distribution function (EDF) of the site-specific parameters, and some composite criteria that were formulated to perform well under two or three different criteria simultaneously (Shen and Louis, 1998; Conlon and Louis, 1999).

Some statistical literatures have indicated that ranking by the posterior mean can perform poorly if ranks of the site-specific parameters are of interest (e.g., Shen and Louis, 1998; Goldstein and Spiegelhalter, 1996). In the same token, if the rank order is of prime interest, ranking by EB means could also perform poorly when an EB approach is taken (as in most of the recent road safety studies). Furthermore, if the ranks are the main interest, it has been shown that ranking by posterior expected rank is optimal under a well-accepted statistical criterion (i.e., the squared error loss criterion) and has thus been recommended for use in applications (Shen and Louis, 1998).

6. Considering crash severities using multivariate spatial models

As part of the illustrations in the next section, we will show how multivariate spatial GLMM can be used to model crashes by injury severity type simultaneously and how the model can be used to rank the site by crash cost rate (in, e.g., US$ 1000 per million VMT or per MVMT). Also, ranking results from the multivariate model using the crash cost rate as decision parameter are compared with those from the univariate model using the crash frequency rate as the decision parameter. In the multivariate model, the response variable for each individual site is a column vector containing number of crashes of different severity types as its elements, e.g., for crash counts of three severity types, we have a column vector $\mu_{it} = (\mu_{1,it}, \mu_{2,it}, \mu_{3,it})'$, where $(\cdots)'$ indicates a transpose of the row vector $(\cdots)$. The statistical theory, on which the adopted multivariate spatial GLMM model is based, is reported in a separate research paper (Song et al., 2003). This research paper extended a univariate CAR spatial model to a multivariate setting following a Bayesian framework. Such models are necessary to analyze more than one type of traffic crash counts simultaneously in that a number of crash or injury severity types may share the same set of risk factors.

7. Data sets and illustrations

Two data sets were employed in this study: Toronto 4-legged signalized intersection data and Texas county data for rural two-lane roads. Both data sets were used to compare the ranking results using the decision parameters (mentioned earlier) under different statistical ranking criteria. To illustrate how crash severities and associated costs could be included in a decision parameter under a multivariate spatial GLMM, Texas data set was used.

7.1. Toronto 4-legged signalized intersection data set

This data set was collected in Toronto, Canada, for years 1990–1995. It has been analyzed as part of a network data set in Lord (2000), and later used in Lord and Persaud (2000) and Miaou and Lord (2003). It includes 868 intersections, which have a total of 54,989 reported crashes for the 6-year period, with approximately a 30–70% split of injury and non-injury crashes, respectively. Individual intersections experienced crashes from 0 to 63 crashes per year. Traffic volumes vary widely from intersections to intersections: from about 5300 to 72,300 vehicles/day for main approaches (summing over both directions) and from 52 to about 42,600 vehicles/day for minor approaches. Flow ratios, calculated as minor approach volume over major approach volume, range from 0.2 to almost 100%, i.e., from “link-like” (or “segment-like”) intersections to intersections with equal flows from the two approaches.

The speed limits on main approaches vary from 50 km/h (30 mph) to 70 km/h (43 mph). In this study, we employed 3 years (1993–1995) of this data set in modeling and ranking.

7.2. Texas rural two-lane low-volume roads data set

This data set is at the county level and includes crashes occurred on rural two-lane roads that, on average, have less than 2000 vehicles/day. In Texas, over 63% of the centerline-miles are these types of roads. They carry only about 8% of the total vehicle miles on state-maintained (or on-system) highways and have less than 7% of the total reported on-system vehicle crashes. Due to the low volume and relatively low crash frequency on these roads, it is often not cost-effective to upgrade these roads to the preferred design standards. However, vehicles traveling on these roadways generally have high-speed and, thus, tend to have relatively more severe injuries when vehicle crashes do occur on these roads. For example, about 26% of the Texas on-system crashes are fatal, incapacitating injury, and non-incapacitating injury (or Type KAB) crashes in 1999. While over 40% of the crashes on rural, two-lane, low-volume on-system roads in 1999 are KAB crashes. A similar data set was previously used and described in Miaou et al. (2003). In this study, 3 years of annual crash frequency data by injury severity type from 1998 to 2000 were used for modeling and ranking analyses. Three explanatory covariates and their interaction terms were considered in Miaou et al. (2003) and were adopted in this study. For the time period studied, 5 of the counties (out of a total of 254 counties), either did not have roads classified as rural, two-lane, low volume roads or had only a couple of centerline miles of such roads, were excluded in this study. (The five excluded counties are Dallas, Kennedy, Galveston, Harris, and Tarrant Counties, and, except for the Kennedy County, the other four counties are highly urbanized counties.)

The average costs of crashes in Texas by crash severity were estimated by Texas Department of Transportation every year. For example, for year 2000, the estimates were: fatal or Type K crashes = US$ 1,191,887, incapacitating injury or Type A crashes = US$ 69,199, non-incapacitating injury or Type B crashes = US$ 25,218, possible injury or Type C crashes = US$ 14,198, and property damage only or Type O crashes = US$ 1969. Note that these crash costs are considered “calculable economic costs” and are lower than the comprehensive or willingness-to-pay costs advocated by National Safety Council. For the KAB crashes the cost ratios are roughly 47:3:1, a clear disparity that gives considerably more weight to the occurrence of fatal crashes than the occurrences of the other two severity types if the crash cost is of main interest.

Table 1
Ranking results compared and figures in which the results are presented

Types of comparisons	Toronto intersection data set	Texas county rural two-lane data set	
	Decision parameter: crash frequency per MVE Spatial sub-model: univariate	Decision parameter: crash frequency per MVMT Spatial sub-model: univariate	Decision parameter: crash cost in US$ 1000 per MVMT Spatial sub-model: multivariate
Comparing among four ranking criteria: RAW, PER, PM, PrWORST	PER vs. PM[a] PER vs. RAW[b] RAW vs. PM vs. PER[c] PrWORST[d]	PER vs. PM[f] PER vs. RAW[g] RAW vs. PM vs. PER[h] PrWORST[i]	PER vs. PM[k] PER vs. RAW[l] RAW vs. PM vs. PER[m] PrWORST[n]
Comparing between two decision parameters		PER of crash frequency rate vs. PER of crash cost rate[j]	
Comparing models with and without a spatial sub-model	PER with vs. PER without a spatial sub-model[e]		

Abbreviations: MVE, million vehicle entries; MVMT, million vehicle miles traveled; RAW, rank based on observed crash frequency rate or cost rate; PER, posterior expected rank; PM, posterior mean; PrWORST, posterior probability for a site being the worst among all sites considered.

[a] See Fig. 2(a and b).
[b] See Fig. 3(a and b).
[c] See Fig. 4(a).
[d] See Fig. 4(b).
[e] See Fig. 12(a and b).
[f] See Fig. 5.
[g] See Fig. 6.
[h] See Fig. 7(a and b).
[i] See Fig. 7(c).
[j] See Fig. 11.
[k] See Fig. 8.
[l] See Fig. 9.
[m] See Fig. 10(a).
[n] See Fig. 10(b).

7.3. Results by decision parameter and ranking criterion

The following decision parameter–data set–spatial component model combinations are used in this study to compare ranking results:

- number of crashes (with all severities combined) per million vehicle entries (MVE) from both approaches: used for the Toronto intersection data in models with a *univariate* spatial component model;
- number of KAB (combined) crashes per MVMT: used for the Texas rural two-lane roads data in models with a *univariate* spatial component model;
- KAB crash cost per MVMT: used for the Texas rural two-lane roads data in models with a *multivariate* spatial component model, in which the number of Type K, Type A, and Type B crashes form a 3×1 vector dependent variable in the Poisson model.

The ranking results are compared under the three statistical criteria described earlier, i.e., PM, PER, and PrWORST, and, for reference purposes, results from a fourth criterion that ranks sites based on raw crash frequency or cost rate (RAW) are also provided. The matrix, which details the types of comparisons presented in this paper, is shown in Table 1. The specific figures in which the comparison results are shown are also indicated in the table. As mentioned earlier, if the ranks are the main interest, it has been indicated in the statistical literature that ranking by PER is optimal and has been recommended for use in applications. Thus, in the following comparisons, ranking results from PER are used as baselines.

Comparisons of ranking results for the 868 Toronto intersections between PER and PM and between PER and RAW, are shown in Figs. 2(a) and 3(a), respectively. To get a closer view of how the top-ranked sites compared, Figs. 2(b) and 3(b) zoom-in the comparison on the top 100 intersections

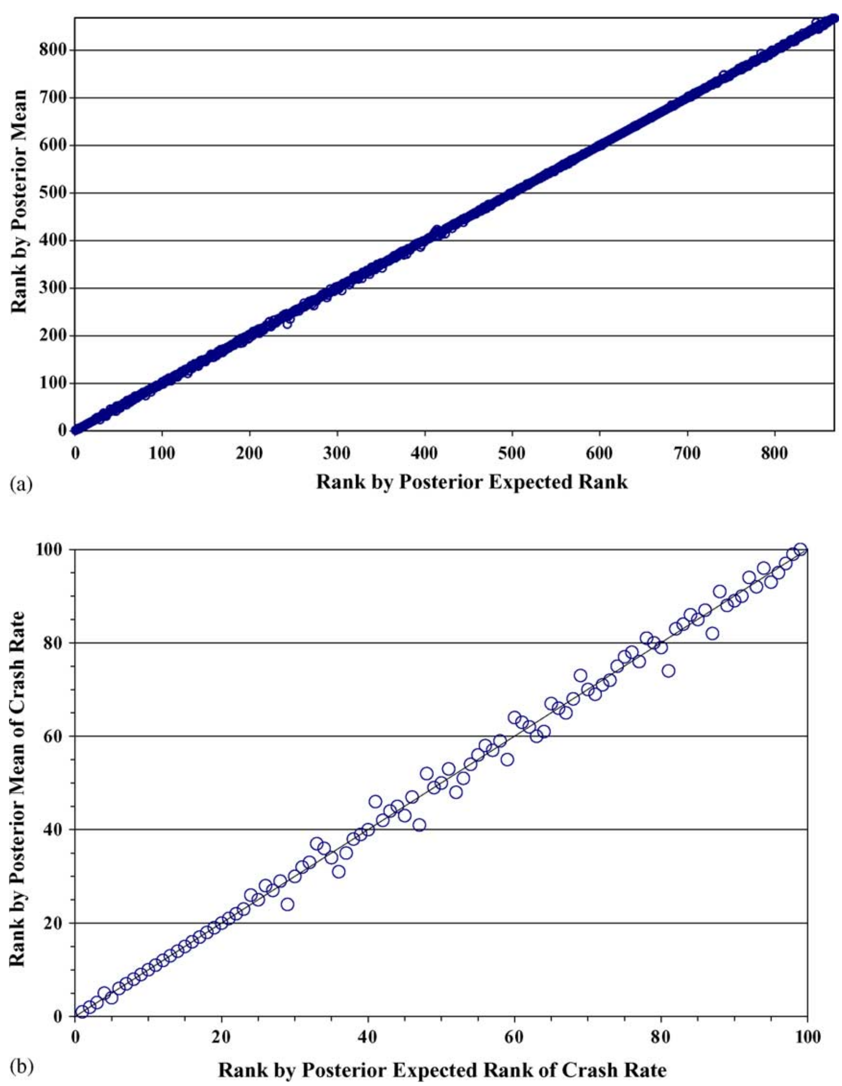


Fig. 2. Comparison of crash rate ranking results for Toronto data: (a) all 868 sites—PER vs. PM and (b) top 100 sites under PER—PER vs. PM.

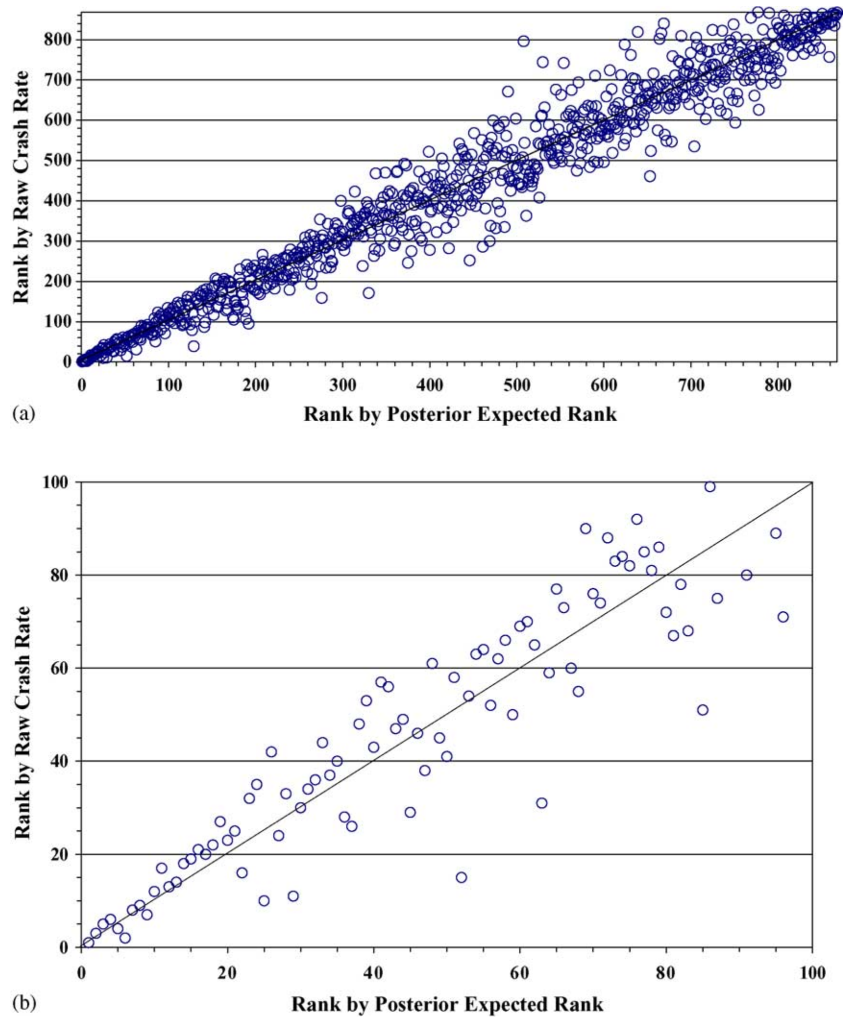


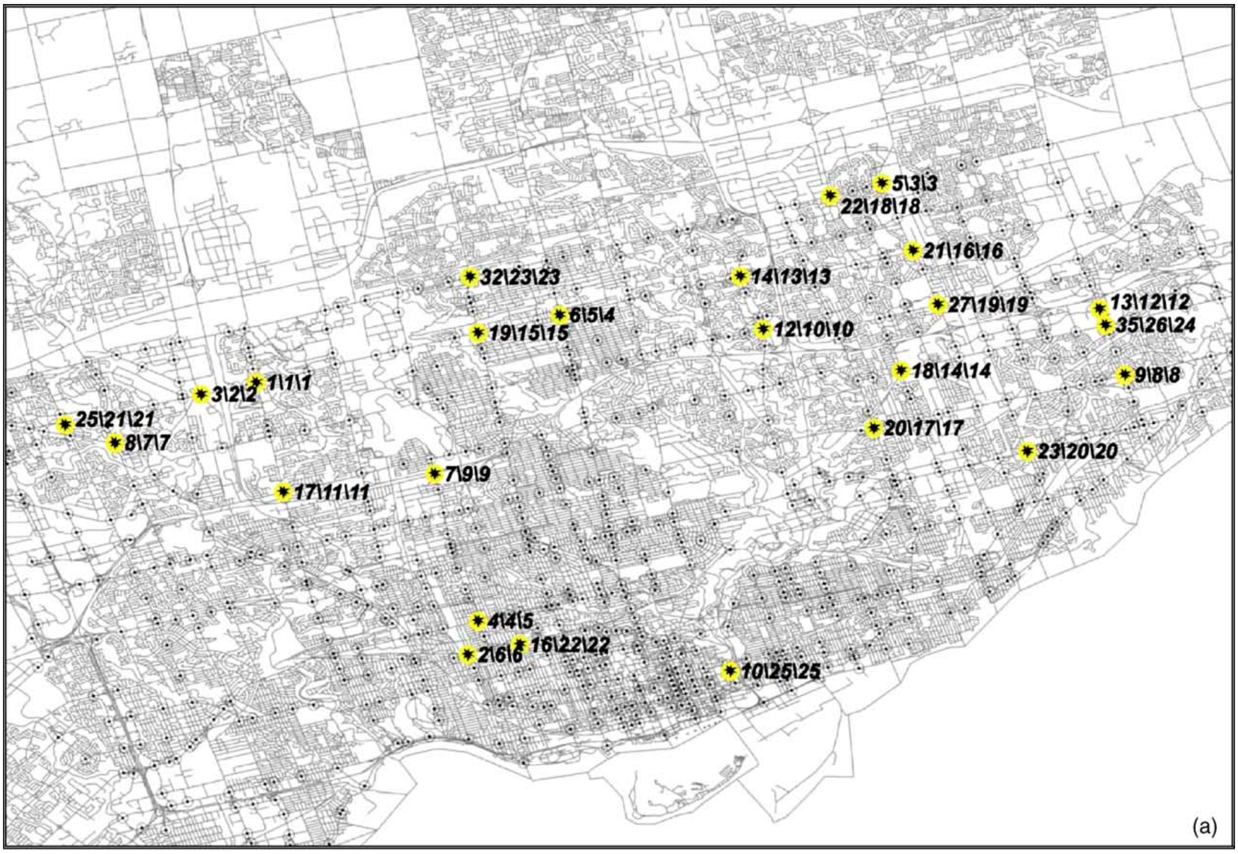
Fig. 3. Comparison of crash rate ranking results for Toronto data: (a) all 868 sites—PER vs. RAW and (b) top 100 sites under PER—PER vs. RAW.

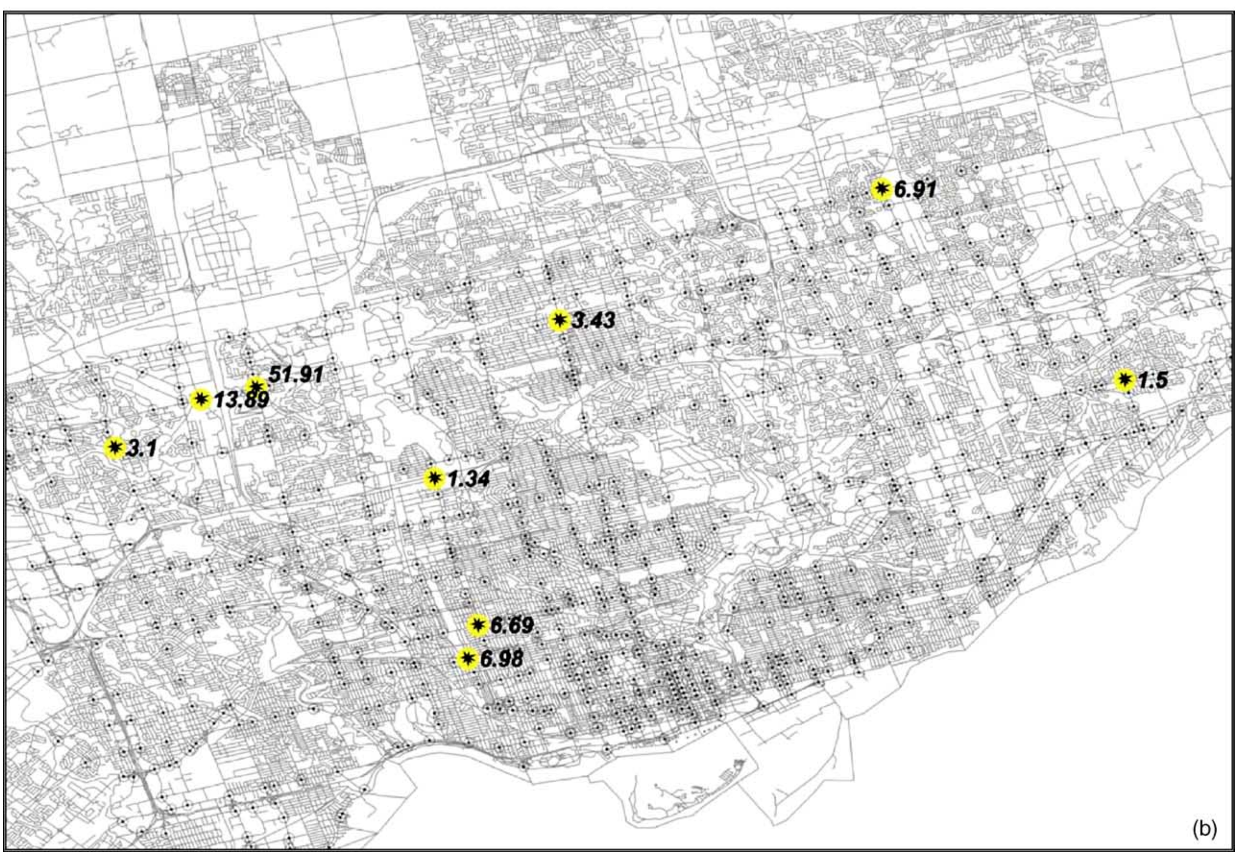
(under the PER criterion). The locations of the top-ranked 25 intersections under the PER criterion are highlighted in Fig. 4(a). Ranks of each of these locations under RAW and PM criteria are presented as well for comparison. For each of these 25 intersections, the three ranks are shown in the following sequence: RAW\PM\PER of crash frequency rate. For example, the second highest ranked intersection under the PER criterion (located on the left side of the map) is ranked as 3, 2, and 2, respectively, under the three criteria. Fig. 4(b) shows the probability (in percent) that the highlighted intersection has the worst crash frequency rate among all intersections. Note that only the intersections with 1% or higher probabilities are displayed. The probabilities of the two intersections that have the highest probabilities of being the worst are 51.91 and 13.89%, respectively.

Some observations can be made from these figures for the Toronto data:

- One can observe that the ranks produced under the PM criterion are quite consistent with those produced from the PER criterion, especially for the top 25 sites.
- In general, ranks generated under the RAW criterion do differ quite significantly from these two criteria. For example, in Fig. 4(a), the highlighted intersection located at the south-central side of the city is shown to rank 10 under RAW, but ranked 25 under both the PM and PER criteria. Interestingly, despite this general observation, the top-ranked 20 sites from RAW do seem to correspond reasonably well with those from the PM and PER criteria.
- As seen from Fig. 4(b), the PrWORST criterion does point to a very small number of intersections as ones that are most likely to have the worst crash frequency rate. Also, as seen from Fig. 4(a and b), the two intersections that have the highest probabilities of being the worst sites do have the highest ranks under both the PM and PER criteria.

Fig. 4. (a) Toronto map for the top-ranked 25 intersections under the posterior expected rank of crash rate criterion (For comparisons, the three ranks shown are based on criteria in the following sequence: RAW\PM\PER). (b) Probability (in percent) that the highlighted intersection in Toronto has the worst crash rate among the intersections considered by this study.


5\3\3
22\18\18
21\16\16
32\23\23
14\13\13
27\19\19
13\12\12
35\26\24
6\5\4
19\15\15
12\10\10
18\14\14
9\8\8
1\1\1
3\2\2
25\21\21
8\7\7
20\17\17
23\20\20
7\9\9
17\11\11
4\4\5
2\6\6
16\22\22
10\25\25
(a)


6.91
3.43
51.91
13.89
1.5
3.1
1.34
6.69
6.98
(b)

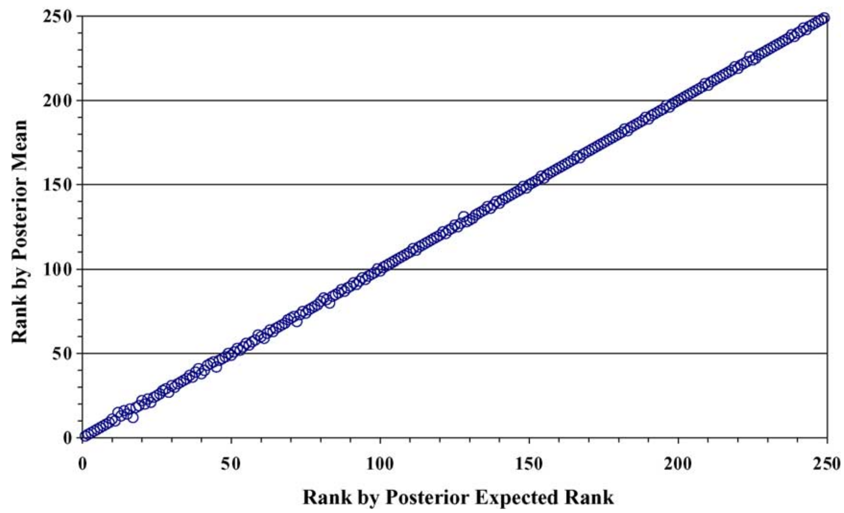


Fig. 5. Comparison of KAB crash rate ranking results for Texas data: PER vs. PM.

The same set of comparisons as in the Toronto data are prepared for the Texas data and are presented in Figs. 5–7(a–c). Similar observations can be made from these figures as in the Toronto data. In general, ranking results from RAW are slightly closer to the PER and PM criteria when compared to the results for the Toronto data. As seen from Fig. 7(a and b), there is a clear regional pattern indicating that these top-ranked counties are all located in the central and eastern parts of Texas.

Using a multivariate spatial GLMM model, the same set of comparisons was developed for the KAB crash cost rate for the Texas data and presented in Figs. 8–11. Again, the ranking results between PM and PER are very close (Fig. 8). The ranks generated from the RAW criterion are, however, clearly out-of-sync with those from the PM and PER criteria (Fig. 9). For the top-ranked counties, the discrepancy can be more closely examined in Fig. 10(a). The large disparity in cost ratios that give considerably more weights to fatal crashes than the other two types of crashes in the decision parameter, as mentioned earlier, in conjunction with the highly stochastic and sporadic nature of this type of crashes over counties are the main reason for this to occur. Putting it in another way, the "small area" nature of the traffic crash problem becomes more conspicuous as smaller domains are considered (in this case, by crash severity type) and as considerably higher weights are given to the rarest of the event in the decision parameter.

Fig. 10(b) marks counties that have some probabilities of being the worst in crash cost rate among all the counties considered. Note that only those counties with 1% or higher probabilities are shown. The three counties that are indicated to have a higher than 10% probability of being the worst in crash cost rate are very consistent with the ranks provided by the PER and PM criteria under the same decision parameter. Under the same PER criterion, Fig. 11 compares the ranks generated by two decision parameters: crash frequency rate and crash cost rate. The results can be seen to be quite differ-

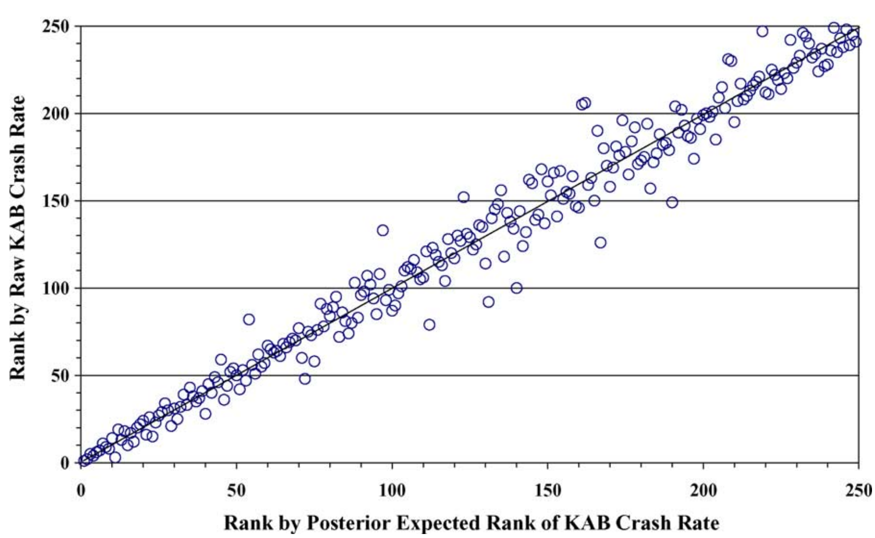


Fig. 6. Comparison of KAB crash rate ranking results for Texas data: PER vs. RAW.

ent. This suggests the importance of considering crash cost by injury severity explicitly in ranking when injury severity and associated cost are the primary concern.

8. Spatial dependence

This study provided several compelling reasons for us to suggest that spatial random effects are very important to consider in developing crash prediction models and in using the model to rank problem sites. First of all, Table 2 shows the overall goodness-of-fit statistics, including the standardized DIC and inverse dispersion parameter discussed earlier, for models with various complexities in their mean functions, from the simplest to the full functional form as presented in Equation (2). The simplest models in Columns (2) and (3), i.e., ‘constant frequency model’ and ‘constant rate model’, respectively, are presented for reference purposes. The

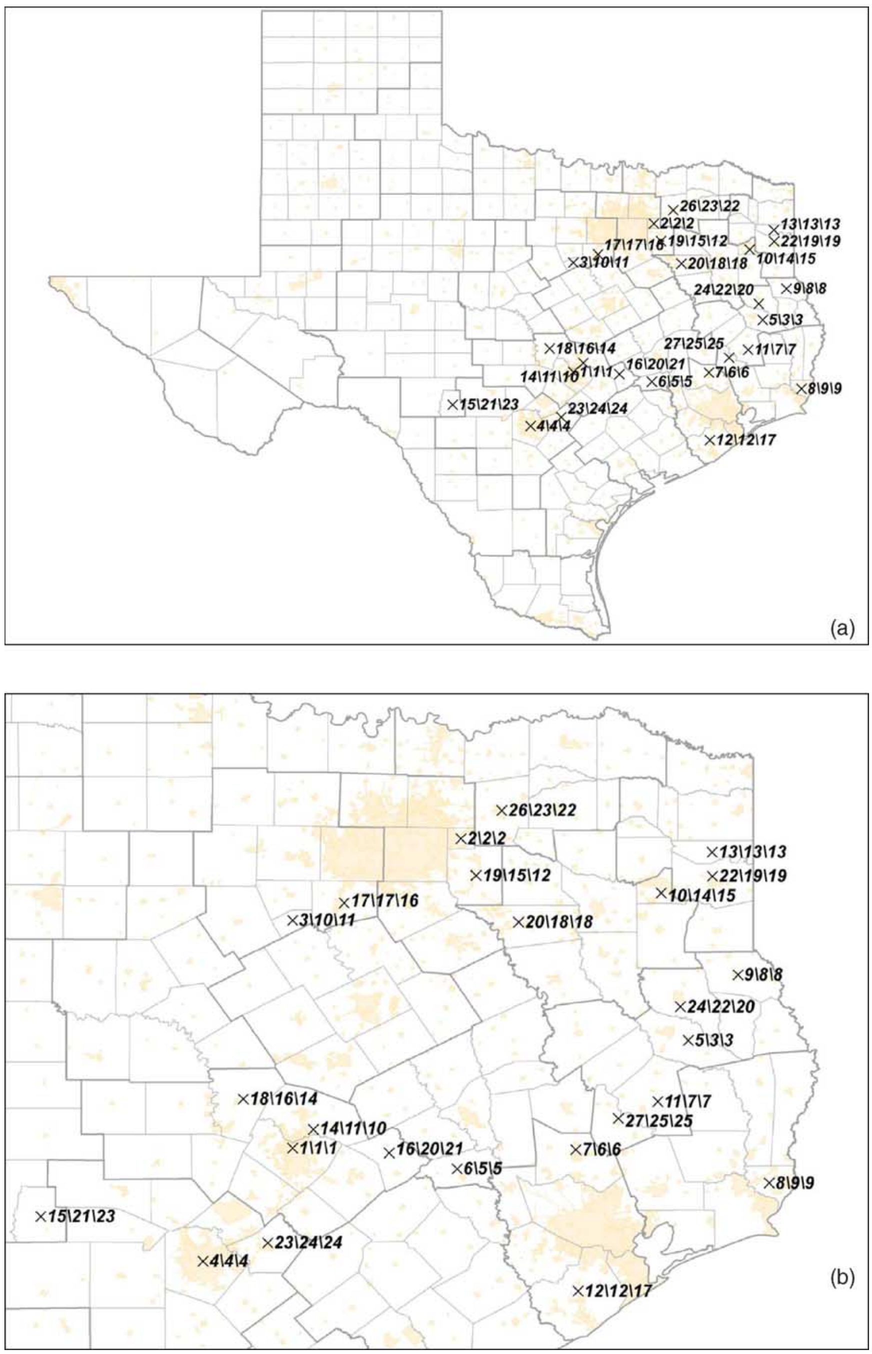


Fig. 7. (a) Map for the top-ranked 25 Texas counties under the posterior expected rank of KAB crash rate criterion (for comparisons, the three ranks shown are based on criteria in the following sequence: RAW\PM\PER). (b) Same as in (a): focusing on the 25 top-ranked counties located in the central and eastern Texas. (c) Probabilities (in percent) that the marked Texas counties have the worst KAB crash rate among the 249 counties considered by this study.

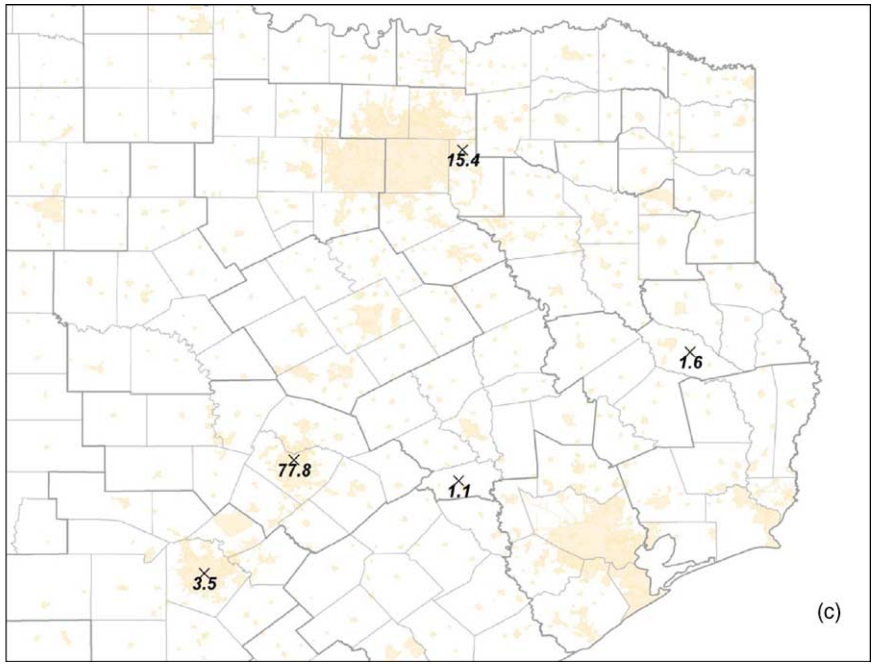


Fig. 7. (a) (*Continued*).

readers are referred to Miaou et al. (2003) on how to utilize these simplest models as reference models to construct R^2-like goodness-of-fit measures for evaluating a model of interest. Models from Columns (4) to (7) include varying numbers of component models and associated parameters in Equation (2):

- Column (4): constants + exchangeable;
- Column (5): constants + VMT/flows as offsets + exchangeable;
- Column (6): constants + VMT/flows + covariates + exchangeables;
- Column (7): constants + VMT/flows + covariates + spatial effects + exchangeables.

Using Toronto data as an example, the DIC/n value decreases from 1.91 in Column (4) to 1.70 in Column (5) when adding VMT/Flows as offsets in the model. It further reduces to 1.59 in Column (6) when explanatory covariates and some "dummy" time effects are considered. Finally, with the in-

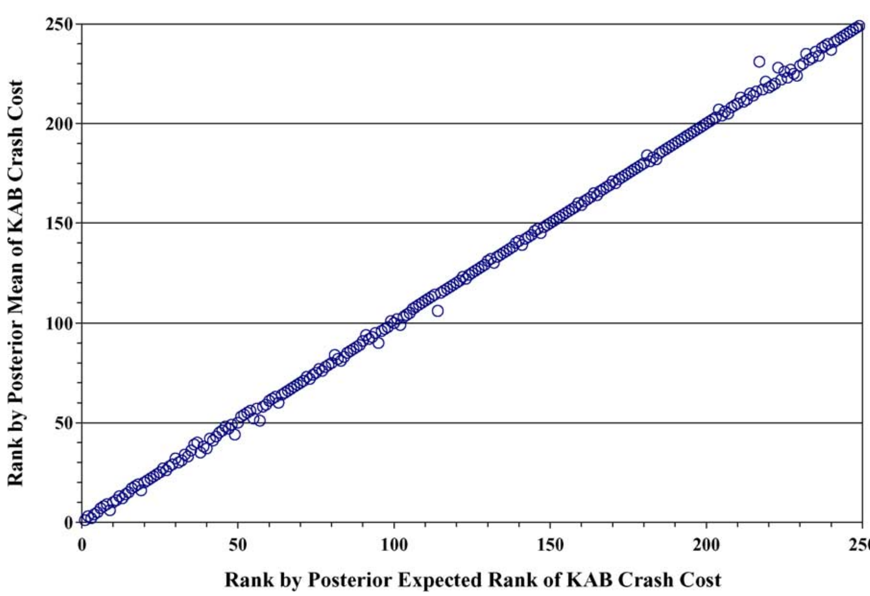


Fig. 8. Comparison of ranking results for Texas data using KAB crash cost (per MVMT) as the decision parameter: PER vs. PM.

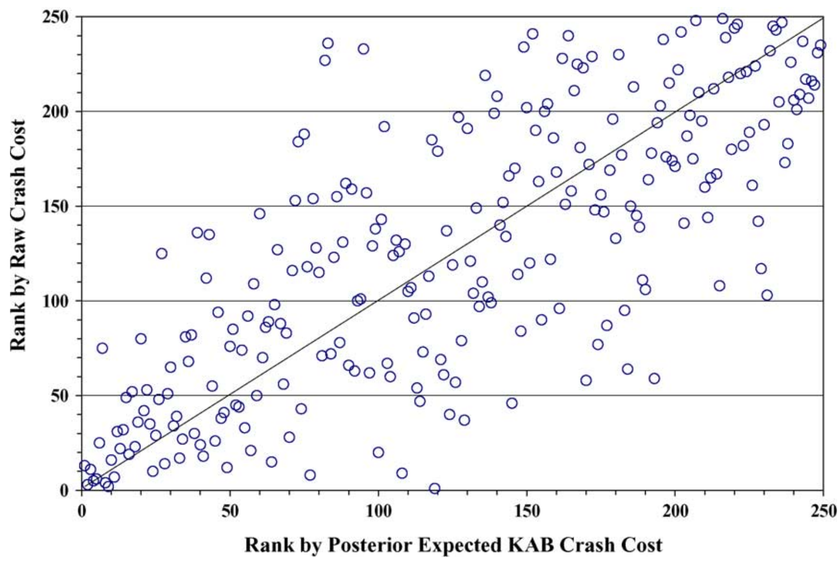


Fig. 9. Comparison of ranking results for Texas data using KAB crash cost (per MVMT) as the decision parameter: PER vs. RAW.

clusion of a CAR spatial component model, the DIC/n value drops to 1.34 in Column (7), a drop of 0.25 in value when compared to that in Column (6). In terms of reducing the overall DIC/n values, the contribution of the spatial component is about 44% (=0.25/0.57). Note that 0.57 is the total reduction in DIC/n value as the complexity of the models increases from Column (4) to Column (7). Similar observations can be made about the Texas data and the contribution from the spatial component in reducing the DIC/n value is about 47%.

Second, the degree of overdispersions in the model, due to un-modeled and isolated site-time specific heterogeneities, can be measured using the estimated mean value of $1/\psi_t$. Again, using the Toronto data as an example, we can see that this overdispersion almost vanishes completely as the CAR spatial sub-model is included into the model. This is indicated by the significant increase in the estimated mean value for ψ_t: from about 6.9 to 96 in 1 year and from 7 to over 100 for the other 2 years. A similar observation about the degree of overdispersions can be made from Table 2 for the Texas data, though the decrease is not as dramatic.

Third, the full model estimated for both Toronto and Texas data sets exhibits very high values in spatial correlation parameter ρ and both are highly statistically significant. The posterior means and standard errors for ρ are: 0.96 ($\pm$0.03) for Toronto data and 0.85 ($\pm$0.12) for Texas data. As indicated earlier, these high ρ values need to be interpreted with some caution, especially for the grid-like urban intersection data of Toronto.

Another evidence that points to the importance of considering spatial effects in modeling crash data is shown in Fig. 12(a and b). Specifically, these two figures compare the ranking results for the Toronto data using the PER criterion for two models: one with and one without the CAR spatial component model (i.e., model in Column (7) versus model in Column (6)). Except for the top-ranked 10–20 sites, the ranks generated from the two models can be seen to be quite different in general.

Taking in all these findings collectively, we conclude that the inclusion of a spatial component in the crash prediction model significantly improved the overall goodness-of-fit performance of the model and affected the ranking results for the two data sets that we examined.

9. Future extensions

There are many directions in which the current study can be extended. Some are presented as follows:

- This study employed 3 years of data from both the Toronto and Texas data sets, for which more years of data are currently available. One possible extension is to study the sensitivity of ranking results when more or less number of years of data is used to develop models and ranks.
- This study touched on 5 of the 10 basic issues regarding the use of ranking methodologies in improving roadway safety as outlined at the outset of this paper. Another possible extension is thus to explore the other five issues, including the appropriate roles of the ranking methodology in engineering practices and other implementation issues.
- Some statistical literatures have indicated that ranking by posterior mean can perform poorly if the ranks are of interest. In addition, if the ranks are the main interest, it has been shown that ranking by posterior expected rank is optimal under well-accepted statistical criteria and has thus been recommended for use in applications. Our comparisons show that the ranking results from the posterior means are actually quite consistent with those generated from the posterior expected rank for both data

Table 2
Comparisons of model goodness-of-fit performances as model complexity increases (for crash frequency rate models only)

Data Set and goodness-of-fit criterion[a]	Increased model complexity (from column 2 to column 7)					
	Constant frequency model (one constant, no VMT or flows; no exchangeable)	Constant rate model (VMT or total flows as offsets; no exchangeable)	With a constant and an exchangeable	With a constant, VMT or total flows as offsets, and an exchangeable	With yearly constants, VMT or flows, other covariates (including flow ratios), and yearly exchangeables	With yearly constants, VMT or flows, other covariates (including flow ratios), yearly exchangeables + spatial effects
Toronto intersection data						
Standardized deviance information criterion (DIC/n)	7.55	3.96	1.91	1.70	1.59	1.34
Inverse dispersion parameter (ψ)			1.90	3.57	6.85–7.19[b]	96–176[b]
Posterior mean			(±0.1)[c]	(±0.1)[c]	(±0.6)–(±0.6)[c]	(±51)–(±94)[c]
Texas rural two-lane road data						
Standardized deviance information criterion (DIC/n)	10.47	4.99	1.86	1.79	1.61	1.39
Inverse dispersion parameter (ψ)			1.85	4.12	10.3–13.7[b]	78–142
Posterior mean			(±0.1)[c]	(±0.3)[c]	(±1.7)–(±2.4)[c]	(±41)–(±72)[b,c]

[a] The smaller the DIC/n value, the better the overall model performance; and the larger the ψ value, the smaller the unexplained site-specific heterogeneities. Note that n is the sample size, 2604 and 747, respectively, for Toronto and Texas data sets.
[b] The inverse dispersion parameter values shown are the lowest and highest estimates for the 3 years considered.
[c] Values in parentheses are estimated standard errors of the parameter values above.

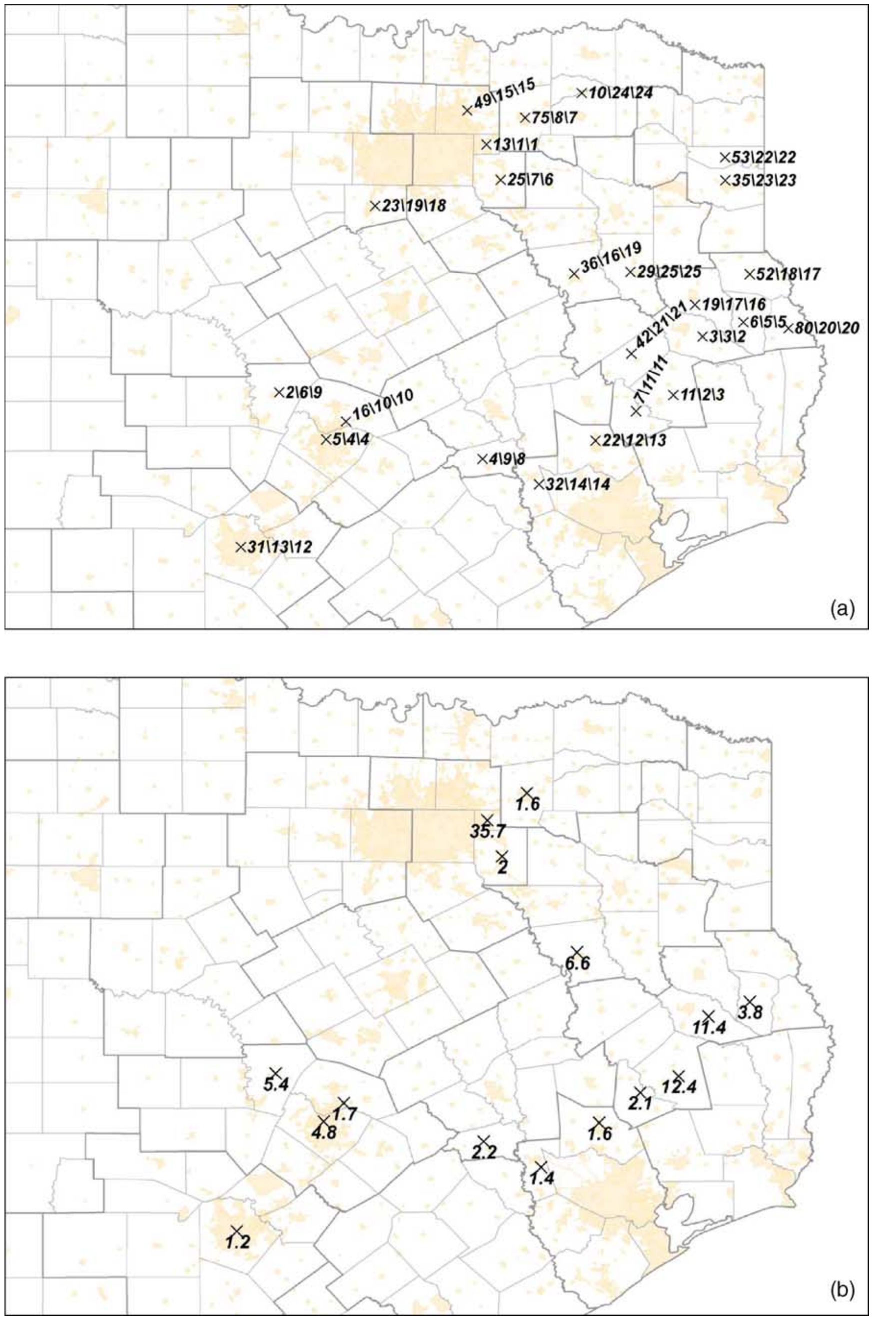


Fig. 10. (a) Map for the top-ranked 25 Texas counties under the posterior expected rank of KAB crash cost per MVMT criterion (ranks shown are in the following sequence: RAW\PM\PER). (b) Probabilities (in percent) that the marked counties have the worst KAB crash cost (per MVMT) among all 249 counties considered.

sets examined in this study. It seems that more data sets need to be studied to get some clues as to when ranking by posterior mean can perform poorly, relative to the ranks produced by the posterior expected rank criterion.

- For the two data sets analyzed in this study, it was found that the inclusion of a spatial effect component in the model could significantly improve the overall goodness-of-fit performance of the model and affect the ranking results. The analysis can be extended to include more data sets with wider ranges of variations in, e.g., traffic volumes and roadway classes, to confirm this finding and to form a better overall understanding of how important spatial correlation effects really are in modeling traffic crash data.
- Another related finding from this study was that the inclusion of a spatial effect component in the model could remove almost all of the overdispersion due to the un-modeled and isolated site-time specific heterogeneities. This suggests that the overdispersion phenomenon observed in the popular Poisson-gamma traffic crash prediction models, which do not consider

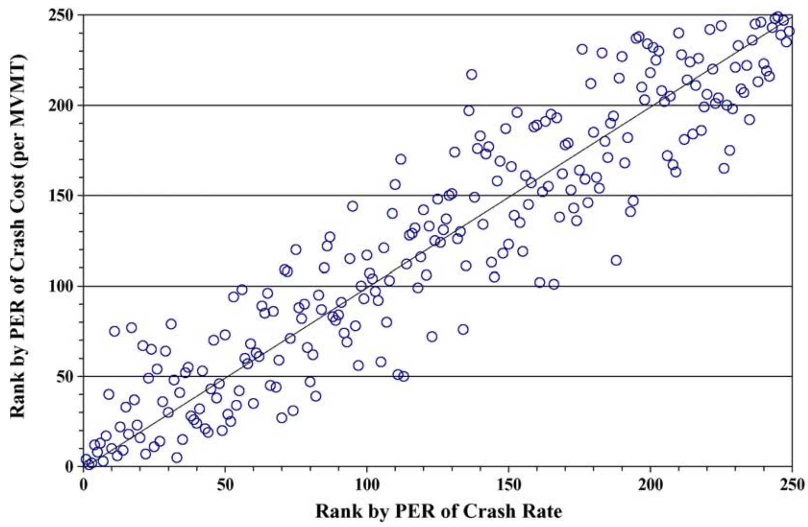
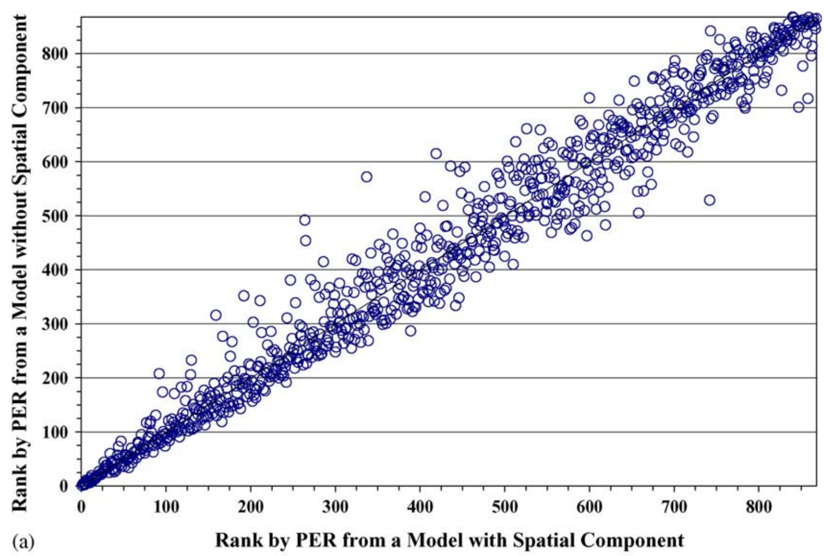


Fig. 11. Comparison of ranking results for Texas data: PER of KAB crash frequency rate vs. PER of KAB crash cost rate (in per MVMT).



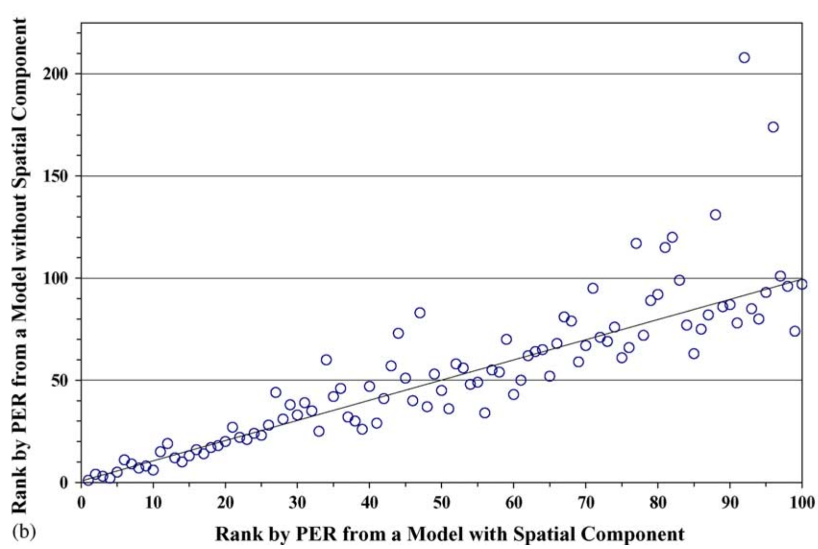


Fig. 12. Comparison of ranking results for Toronto data using PER of crash frequency rates: (a) all 868 sites—model with spatial component vs. model without spatial component and (b) top 100 sites—model with spatial component vs. model without spatial component.

spatial effects, may have come mainly from the spatial correlation induced by omitted variables. Thus, another logical extension of the current study is to understand how and what omitted variables might have contributed to these spatial correlations. From traffic planning and operational point of views, the variables of interest could potentially include those that characterize land uses, roadside development, “business” environment, etc.

Acknowledgements

Support for this study was provided by a research grant (#167145) from the Southwest Region University Transportation Center through the University Transportation Centers Program, US Department of Transportation. The contents of this paper reflect the views of the authors, who are solely responsible for the accuracy of the data, opinions, findings, and conclusions presented herein. Special thanks are due to Mr. Jim Smith, Mr. Blair Lagden, and Mr. Steven Kodama of Transportation Services, City of Toronto, for providing the Toronto data used in this study, and to our colleague Dr. Dominique Lord at Texas Transportation Institute for his assistance with the data. The authors would also like to thank anonymous reviewers for their constructive comments and suggestions on an earlier version of the paper.

References

Agent, K.R., O’Connell, L., Green, E.R., Kreis, D., Pigman, J.G., Tollner, N., Thompson, E., 2003. Development of Procedures for Identifying High-Crash Locations and Prioritizing Safety Improvements. Research Report KTC-03-15/SPR250-02-1F. Kentucky Transportation Center, University of Kentucky.

Anselin, L., 2003. Spatial externalities, spatial multipliers and spatial econometrics. Int. Reg. Sci. Rev. 26 (2) (April), 153–166.

Anselin, L., Griffith, D., 1988. Do spatial effects really matter in regression analysis? Pap. Reg. Sci. Assoc. 65, 11–34.

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2004. Hierarchical Modeling and Analysis for Spatial Data. CRC Press.

Besag, J., Green, P., Higdon, D., Mengersen, K., 1995. Bayesian computation and stochastic systems. Stat. Sci. 10, 3–66 (with discussion).

Boyle, A.J., Wright, C.-C., 1984. Accident “migration” after remedial treatment at accident blackspots. Traffic Eng. Control 25, 260–267.

Carlin, B.P., Louis, T.A., 1996. Bayes and Empirical Bayes Methods for Data Analysis. Chapman and Hall/CRC Press, London.

Congdon, P., 2001. Bayesian Statistical Modelling. John Wiley & Sons Ltd., New York.

Conlon, E.M., Louis, T.A., 1999. Addressing multiple goals in evaluating region-specific risk using Bayesian methods. In: Lawson, A., Biggeri, A., Bohning, D., Lesaffre, E., Viel, J.-F., Bertollini, R. (Eds.), Disease Mapping and Risk Assessment for Public Health. Wiley, pp. 31–47 (Chapter 3).

Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley, New York.

Dey, D., Ghosh, S., Mallick, B.K., 2000. Bayesian Generalized Linear Model. Marcel Dekker, New York.

Diggle, P.J., 2003. Statistical Analysis of Spatial Point Patterns, second ed. Arnold, London.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley, New York.

Gatrell, A.C., Bailey, T.C., 1996. Interactive Spatial Data Analysis. Prentice Hall, Upper Saddle River, NJ.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. Bayesian Data Analysis. Chapman & Hall/CRC Press.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Markov Chain Monte Carlo in Practice. Chapman and Hall, London.

Goldstein, H., 2003. Multilevel Statistical Models, third ed. Hodder Arnold, London.

Ghosh, M., Natarajan, K., Waller, L.A., Kim, D., 1999. Hierarchical Bayes for the analysis of spatial data: an application to disease mapping. J. Stat. Plan. Inference 75, 305–318.

Goldstein, H., Spiegelhalter, D.J., 1996. League tables and their limitations: statistical issues in comparisons of institutional performance. J. R Stat. Soc. Ser A 159, 385–443 (with discussion).

Griffith, D.A., 1988. Advanced Spatial Statistics. Kluwer Academic Publishers, Boston.

Haining, R., 1990. Spatial Data Analysis in the Social and Environmental Sciences. Cambridge University Press.

Hallmark, S., Basavaraju, R., 2002. Evaluation of the Iowa DOT’s Safety Improvement Candidate List Process. CTRE Project 00-74. Final Report. Center for Transportation Research and Education, Iowa State University.

Hauer, E., Kononov, J., Allery, B., Griffith, M.S., 2002. Screening the road network for sites with promise. Paper No. 02-2182. Transportation Res. Rec. 1784, 27–32.

Heydecker, B.G., Wu, J., 2001. Identification of sites for road accident remedial work by bayesian statistical methods: an example of uncertain inference. Adv. Eng. Software 32, 859–869.

Holloway, G., Shankar, V., Rahman, S., 2002. Bayesian spatial probit estimation: a primer and an application to HYV rice adoption. Agric. Econ. 27, 383–402.

Kononov, J., 2002. Identifying locations with potential for accident reductions: use of direct diagnostics and pattern recognition methodologies. Paper No. 02-2182. Transportation Res. Rec. 1784, 153–158.

Lawson, A.B., 2001. Tutorial in biostatistics: disease map reconstruction. Stat. Med. 20, 2183–2204.

Lawson, A.B., Denison, D.G.T., 2002. Spatial Cluster Modelling. Chapman and Hall/CRC, London.

LeSage, J.P., 2001. A family of geographically weighted regression models. Working Paper. Department of Economics, University of Toledo.

Lord, D., 2000. The prediction of accidents on digital networks: characteristics and issues related to the application of accident prediction models. Ph.D. Dissertation. University of Toronto.

Lord, D., Persaud, B., 2000. Accident prediction models with and without trend: application of the generalized estimating equation. In: Transportation Research Record 1717. Transportation Research Board, National Research Council, Washington, DC, pp. 102–108.

Maher, M.J., 1990. A bivariate negative binomial model to explain traffic accident migration. Accid. Anal. Prev. 22 (5), 487–498.

Miaou, S.-P., Song, J.J., Mallick, B.K., 2003. Roadway traffic-crash mapping: a space–time modeling approach. J. Transportation Stat. 6 (1), 33–57.

Miaou, S.-P., Lord, D., 2003. Modeling traffic crash–flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes methods. Paper No. 03-3363, Transportation Research Record 1840, Journal of Transportation Research Board, National Research Council, 31–40.

Midwest Research Institute, 2002. SafetyAnalyst: Software Tools for Safety Management of Specific Highway Sites: Task K. White Paper for Module 1—Network Screening. Prepared for Federal Highway Administration.

Møller, J., 2003. Spatial Statistics and Computational Methods. Springer, New York.

Morris, C.N., Christiansen, C.L., 1996. Hierarchical models for ranking and for identifying extremes, with applications. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, vol. 5. Oxford University Press, Oxford, pp. 277–296 (with Discussion).

Levine, N., 2004. CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations. Ned Levine & Associates/The National Institute of Justice, Houston, TX/Washington, DC.

Persaud, B., Lyon, C., Nguyen, T., 1999. Empirical Bayes procedure for ranking sites for safety investigation by potential for safety improvement. Transportation Res. Rec. 1665, 7–12.

Rao, J.N.K., 2003. Small Area Estimation. Wiley, New York.

Robert, C.P., Casella, G., 1999. Monte Carlo Statistical Methods. Springer-Verlag, New York.

Schluter, P.J., Deely, J.J., Nicholson, A.J., 1997. Ranking and selecting motor vehicle accident sites by using a hierarchical Bayesian model. Statistician 46 (3), 293–316.

Shen, W., Louis, T.A., 1998. Triple-goal estimates in two-stage hierarchical models. J. R. Stat. Soc. B 60, 455–471.

Song, J.J., Ghosh, M., Miaou, S.-P., Mallick, B., 2003. Bayesian multivariate spatial models for roadway traffic crash mapping. Research Paper. Department of Statistics, Texas A&M University, College Station, TX.

Spiegelhalter, D.J., Best, N., Carlin, B.P., Linde, A., 2002. Bayesian measure of model complexity and fit. J. R. Stat. Soc. 64 (Part 3), 1–34.

Spiegelhalter, D.J., Thomas, A., Best, N.G., 2003. WinBUGS Version 1.4 User Manual. Imperial College and Medical Research Council, UK, http://www.mrc-cam.ac.uk/bugs.

Sun, D., Tsutakawa, R.K., Kim, H., He, Z., 2000. Spatio–temporal interaction with disease mapping. Stat. Med. 19, 2015–2035.

Thomas, A., Best, N., Arnold, R., Spiegelhalter, D., 2002. GeoBUGS User Manual. Version 1.1 Beta.

Tunaru, R., 1999. Hierarchical Bayesian models for road accident data. Traffic Engineering + Control 40 (6) (June), 318–324.