

BAYESIAN MULTIVARIATE SPATIAL MODELS AND THEIR APPLICATIONS

A Dissertation

by

JOON JIN SONG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2004

Major Subject: Statistics

BAYESIAN MULTIVARIATE SPATIAL MODELS AND THEIR APPLICATIONS

A Dissertation

by

JOON JIN SONG

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Bani K. Mallick
(Chair of Committee)

Jeffrey D. Hart
(Member)

Marina Vannucci
(Member)

Yalchin Efendiev
(Member)

Shaw-Pin Miaou
(Member)

Michael T. Longnecker
(Head of Department)

August 2004

Major Subject: Statistics

ABSTRACT

Bayesian Multivariate Spatial Models and Their Applications. (August 2004)

Joon Jin Song, B.S., Yeungnam University

M.S., Kyungpook National University

Chair of Advisory Committee: Dr. Bani K. Mallick

Univariate hierarchical Bayes models are being vigorously researched for use in disease mapping, engineering, geology, and ecology. This dissertation shows how the models can also be used to build model-based risk maps for area-based roadway traffic crashes. County-level vehicle crash records and roadway data from Texas are used to illustrate the method. A potential extension that uses univariate hierarchical models to develop network-based risk maps is also discussed.

Several Bayesian multivariate spatial models for estimating the traffic crash rates from different types of crashes simultaneously are then developed. The specific class of spatial models considered is conditional autoregressive (CAR) model. The univariate CAR model is generalized for several multivariate cases. A general theorem for each case is provided to ensure that the posterior distribution is proper under improper and flat prior. The performance of various multivariate spatial models is compared using a Bayesian information criterion. The Markov chain Monte Carlo (MCMC) computational techniques are used for the model parameter estimation and statistical inference. These models are illustrated and compared again with the Texas crash data.

There are many directions in which this study can be extended. This dissertation concludes with a short summary of this research and recommends several promising extensions.

*To My Parents,
Parents-in-Law,
and My Lovely Wife, Wonjeong*

ACKNOWLEDGEMENTS

There are many people who through their generosity and knowledge have made important contributions to this dissertation. It would be impossible to list everyone who contributed or to adequately list the extent of the contributions for those who are mentioned.

First and foremost, I am extremely grateful to my advisor, Dr. Bani K. Mallick, for his guidance and support throughout my graduate study. I especially thank him for giving me the opportunity to participate in several of his research projects which deal with many challenging statistical issues.

I wish to thank my committee members, Dr. Jeffrey D. Hart, Dr. Marina Vannucci, and Dr. Yalchin Efendiev for their insightful suggestions and help.

I would like to extend a special thank you to Dr. Shaw-Pin Miaou, who has been a committee member and a great mentor and friend to me. I cannot express my thanks in writing for his endless encouragement, concern, and support in my study and life.

I also wish to acknowledge my friends, Ho-Jin Lee, Jeesun Jung, and Deukwoo Kwon, with whom I shared my joy, complaints, and laughter through these past years.

Finally, I would like to thank my parents, Jong-Chool Song and Nam-Soon Lee, and my parents-in-law, Soo-Hwan Kim and Gyo-Sook Jeong, for their constant encouragement and support. I also thank my sister, Won-Suk Song, and brother-in-law, Wonjo Kim, for their emotional support during the difficult time of this long journey.

To all people, I am most thankful to my wife, Wonjeong, who is definitely the main contributor to all my significant accomplishments, including this dissertation. I thank God for having her as my wife, best friend, and soul companion.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
DEDICATION	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	viii
LIST OF TABLES	x
CHAPTER	
I INTRODUCTION	1
1.1 Spatial Data	1
1.2 Research Problems, Objectives and Scope	4
1.3 Research Contributions	4
1.4 Organization of Dissertation	5
II LITERATURE REVIEW	7
2.1 Conditional Autoregressive (CAR) Models	7
2.2 Multivariate Conditional Autoregressive (MCAR) Models	9
2.3 Posterior Propriety	11
III UNIVARIATE HIERARCHICAL SPATIAL MODELS	14
3.1 Introduction	14
3.2 Description of Data	20
3.3 Bayesian Hierarchical Models	24
3.4 Deviance Information Criterion and Variants	37
3.5 Results	40
3.6 Discussion	47
IV MULTIVARIATE HIERARCHICAL SPATIAL MODELS	51
4.1 Introduction	51
4.2 Univariate Hierarchical Model	53
4.3 Multivariate Hierarchical Model	54

CHAPTER	Page
4.4 Data Analysis	62
V CONCLUSIONS	73
REFERENCES	75
APPENDIX A	84
APPENDIX B	88
APPENDIX C	90
APPENDIX D	92
VITA	95

LIST OF FIGURES

FIGURE	Page
1 Geographic Districts, Counties, and Urbanized Areas in Texas.	21
2 The Number of KAB Crashes on Rural, 2-Lane, Low-Volume, On-System Roads in Each Texas County: 1999.	22
3 Vehicle-Miles Traveled on Rural, 2-Lane, Low-Volume, On-System Roads in Each Texas County: 1999.	24
4 "Raw" Annual KAB Crash Rates in Crashes per MVMT by County: 1992-1999.	25
5 Proportion of KAB Crashes That Occurred on Sharp Horizontal Curves in Each County: 1992-1999.	26
6 Proportion of KAB Crashes That Occurred under Wet Pavement Conditions for Each County: 1999.	27
7 Proportion of KAB Crashes That Were Intersection, Intersection Related, or Driveway Access Related for Each County: 1999.	28
8 Proportion of KAB Crashes Involving Vehicles That Ran Off Roads and Hit Fixed Objects on the Roadside for Each County: 1999.	29
9 Estimated KAB Crash Rates in Crashes per MVMT by County from Model 27: 1999.	48
10 Locations of KAB Crashes on the State-Maintained Highway Network of a Texas County in 1999.	49
11 Plot of the Posterior Distributions of the Covariates for Interaction Crash.	67
12 Plot of the Posterior Distributions of the Covariates for Interaction-Related Crash.	68
13 Plot of the Posterior Distributions of the Covariates for Driveway Crash.	69

FIGURE	Page
14 Plot of the Posterior Distributions of the Covariates for Non-Interaction Crash.	70
15 Plot of the Posterior Distributions of the Correlation Coefficients between the Responses.	71
16 Predicted Map for Different Types of Crash.	72

LIST OF TABLES

TABLE		Page
1	Deviance Information Criterion and Related Performance Measures for Models of Various Complexities.	43
2	Example MCMC Simulation Output for Model 27.	46
3	DIC and p_D Values for Various Multivariate Spatial Models.	65

CHAPTER I

INTRODUCTION

Statistical spatial models have been used in diverse applications, such as engineering, geology, ecology, and public health, for analyzing geographically referenced data. Advances in computing power, Geographic information system (GIS), and computational techniques, such as the Markov Chain Monte Carlo (MCMC), allow sophisticated spatial models to be developed. In biostatistical field, spatial models have particularly been increasingly employed to analyze disease rates and develop disease maps.

This chapter is organized as follows. First, a brief review of spatial data is given. The research objectives and contributions are then presented. Finally, the organization of this dissertation is outlined.

1.1 Spatial Data

Spatial data can be viewed as realizations as a spatial stochastic process

$$\{Y(\mathbf{s}) : \mathbf{s} \in D\},$$

where \mathbf{s} is the location from which the data is observed and D is a random set in d -dimensional Euclidean space. A realization of an underlying spatial stochastic process is denoted by $\{y(\mathbf{s}) : \mathbf{s} \in D\}$.

Spatial data are generally categorized into three types; (1) geostatistical or point-referenced data, (2) lattice or areal data, and (3) point pattern data. A brief review

The format and style follow that of *Journal of the American Statistical Association*.

for these types of data and the associated spatial models is provided in following subsections.

1.1.1 Geostatistical Data

The basic principle for geostatistical data analysis is based on a stochastic process $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$, where D is assumed to be a fixed subset of R^d . Let $Y(\mathbf{s}) = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))^T$ denote n observations at sites $\mathbf{s}_1, \dots, \mathbf{s}_n$ in a region of interest $D \in R^d$. The observed data are used to predict some unknown observations at unobserved sites. Kriging is the most popular method for spatial prediction which is a optimal least squares interpolation. This prediction method depends on the second-order properties of the spatial process $Y(\mathbf{s})$.

The spatial process $Y(\mathbf{s})$ is often assumed to follow Gaussian process. The basic model is developed by means of the following model

$$Y(\mathbf{s}) = m(\mathbf{s}) + \epsilon(\mathbf{s}) \tag{1.1}$$

where $m(\mathbf{s})$ is the mean function of the process and $\epsilon(\mathbf{s})$ is a random process. The total variation in (1.1) is divided into large-scale variation (the mean function $m(\mathbf{s})$) and small-scale variation (the residual random process $\epsilon(\mathbf{s})$). Residual process is associated with a covariance function, which expresses covariance of two values of $\epsilon(\mathbf{s}_i)$ and $\epsilon(\mathbf{s}_j)$.

The several approaches for the analysis have been developed, such as plug-in approach under a Gaussian process (Kitanidis 1983; Mardia and Marshall 1984; Kitanidis and Lane 1985), nonparametric regression methods for spatial prediction (Laslett 1994; Journel 1983), median polishing kriging for nonstationary mean (Cressie 1986, 1993), and nonparametric estimation of nonstationary spatial covariance (Sampson and Guttorp 1992). For further methods in detail, see Cressie (1993).

1.1.2 Lattice Data

Analogous to geostatistical data, denote that $Y_{\mathbf{s}_1}, \dots, Y_{\mathbf{s}_n}$ are lattice data observed at n sites. The apparent difference between geostatistical data and lattice data is that the latter are observed at every site. In addition, D is a fixed subset of R^d and it is partitioned into a finite number of lattices (or areal units), while site index \mathbf{s} in geostatistical data varies continuously over D .

In practice, the lattices (or areal units) are irregular such as zip codes or counties and the data are regularly sums or average of quantities of interest over these lattices. Spatial association over the lattices are introduced by a neighborhood structure and there are two popular models, the simultaneously autoregressive (SAR) and the conditional autoregressive (CAR) models, that incorporate such neighborhood structure. Whittle (1954) proposed the SAR model which has the advantage in computation when likelihood methods are used, while the CAR model is developed by Besag (1974) and this model is computationally convenient for Gibbs sampling in Bayesian framework.

The general approach in Bayesian hierarchical spatio-temporal models is to embed spatial random effect with the CAR prior and time effect in generalized linear model. This approach is commonly used in disease mapping studies and is also utilized in Chapter III and IV.

1.1.3 Point Pattern Data

Recall that $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ is a set of the locations in a region of interest. A quantity of interest to be analyzed in point pattern data is the locations of n events in a region D . The objective of the analysis is to investigate whether the pattern of data shows complete spatial randomness, clustering, or regularity. For example, consider residences of persons with a particular disease or locations of a certain species

of tree in a forest.

In contrast to previous two types of data, a quantity of interest, occurrence of events, is usually fixed and the locations \mathbf{s}_i are random. The studies on spatial point pattern can be found in Pielou (1959, 1977), Getis and Boots (1978), Marquiss et al. (1978), Ripley (1981), Diggle (1983), and Upton and Fingleton (1985).

1.2 Research Problems, Objectives and Scope

Spatial data is frequently multivariate. For example, incidences of several diseases, such as leukemia, pediatric asthma, and lung cancer, would be generally collected at county or census tract level in public health. In such case, we expect not only the dependence between incidences of different diseases at a given areal unit, but also spatial association between the incidences across areal units.

The main objective of this dissertation is to explore multivariate spatial modeling for multivariate measurements over areal units. For areal-based data, we propose several types of multivariate extension based on general univariate conditional autoregressive (CAR) model. The second goal is to ensure posterior propriety for proposed models with improper prior, vague flat prior. Finally, the multivariate spatial models are applied to real data, Texas crash data and risk maps are generated based on estimated crash risk rate.

1.3 Research Contributions

Statistical methodologies in transportation safety community play an important role to investigate traffic crash rate and to improve roadway safety. However, research on spatial models for roadway traffic crashes has not conducted much. One of contribution of this research is to introduce model-based statistical approach to the community and to illustrate this by vehicle crash records and roadway inventory data

at county level in Texas.

A variety of spatial models for univariate data have been developed in the past decade. As indicated earlier, it is often necessary to have spatial models to analyze multivariate data. Recently, Carlin and Banerjee (2003) proposed multivariate conditional autoregressive (MCAR) model for spatially and temporally correlated survival data. We also propose several spatial models for multivariate data based on general univariate conditional autoregressive (CAR) model.

Vague flat prior is usually adopted for fixed effect, such as regression parameters. Though this prior is a simple and convenient choice, it can lead a improper joint posterior distribution, so that the resulting posterior distribution make Bayesian inference impossible. Therefore, it is imperative to ensure that the joint posterior is proper under vague flat prior. The studies to obtain sufficient condition on posterior propriety were found in Chen et al. (2002), Hobert and Casella (1996), and Chen et al. (2003). Ghosh et al. (1998) provided sufficient conditions to gain a proper joint posterior with a univariate CAR prior for spatial random effect. Since vague flat prior is assigned to regression parameters in multivariate model setup proposed here, theorems to obtain proper posteriors corresponding to each spatial prior are provided. These multivariate models are applied to Texas crash data and inferential results and crash risk map are shown.

1.4 Organization of Dissertation

This dissertation is composed of five chapters. In Bayesian framework, all models and methodologies proposed in this dissertation are developed and the analyses of real data are carried out.

In Chapter II, we briefly review selective literature and topics concerning about univariate CAR model, multivariate CAR model, and posterior propriety.

Chapter III shows how hierarchical Bayes models, which are being vigorously researched for use in disease mapping, can also be used to build model-based risk maps for area-based traffic crashes. Country-level vehicle crash records and roadway data from Texas are used to illustrate the method. A potential extension that uses hierarchical models to develop network-based risk maps is also discussed.

We consider several Bayesian multivariate spatial models for estimating the crash rates from different types of crashes in Chapter IV. Conditional autoregressive (CAR) model is considered for the spatial effect model and is generalized for the multivariate case. A general theorem for each case is provided to ensure that posterior is proper under vague flat prior. The different models are compared according to some Bayesian criterion. Markov chain Monte Carlo (MCMC) is used for computation. We illustrate these methods with Texas crash data.

Finally, Chapter V concludes this dissertation with a short summary and some suggestions for future study. The Appendices include proofs for the theorems proposed in Chapter IV.

CHAPTER II

LITERATURE REVIEW

2.1 Conditional Autoregressive (CAR) Models

Conditional autoregressive (CAR) models are introduced by Besag (1974). In recent, the models have been increasingly used in broad application for spatial data analysis because these models allow to model fitting using a Gibbs sampler. CAR models have been implemented by two ways to model spatial association with areal data. Firstly, Geman and Geman (1984) showed direct spatial modeling of observations and the second approach is hierarchical modeling. Spatial association in areal data is commonly described by spatial random effect in hierarchical model and CAR models are employed as the priors of the random effect in Bayesian framework.

The full conditional distributions of CAR models are defined as

$$p(\eta_i | \eta_{j \neq i}) \sim N\left(\sum_j c_{ij} \eta_j, \tau_i^2\right), \quad i = 1, \dots, n. \quad (2.1)$$

From Brook's Lemma, the joint distribution is uniquely determined by

$$f(\boldsymbol{\eta}) \sim N(\mathbf{0}, (\mathbf{I} - \mathbf{C})^{-1} \mathbf{D}), \quad (2.2)$$

where \mathbf{I} is identity matrix, $\mathbf{C} = \{c_{ij}\}$, and $\mathbf{D} = \text{Diag}(\tau_1^2, \dots, \tau_n^2)$. The covariance matrix in (2.2) must be symmetric, and the conditions are obtained,

$$\frac{c_{ij}}{\tau_i^2} = \frac{c_{ji}}{\tau_j^2}, \quad (2.3)$$

for all i and j .

In practice, a proximity matrix \mathbf{W} is usually constructed to describe neighborhood relationship between areal unit. Suppose we set $c_{ij} = w_{ij}/w_{i+}$ and $\tau_i^2 = \sigma_\eta^2/w_{i+}$,

where $\mathbf{W} = \{w_{ij}\}$ and $w_{i+} = \sum_j w_{ij}$. Then the full conditionals (2.1) and the joint distribution (2.2) are rewritten by

$$p(\eta_i | \eta_{j \neq i}) \sim N\left(\sum_j w_{ij} \eta_j / w_{i+}, \sigma_\eta^2 / w_{i+}\right) \quad (2.4)$$

and

$$f(\boldsymbol{\eta}) \propto \exp\left\{-\frac{1}{2\sigma_\eta^2} \boldsymbol{\eta}^T (\mathbf{D}_W - \mathbf{W}) \boldsymbol{\eta}\right\}, \quad (2.5)$$

where \mathbf{D}_W is a diagonal matrix with entries w_{i+} . Note that the distribution in (2.5) is improper because of $(\mathbf{D}_W - \mathbf{W})\mathbf{1} = \mathbf{0}$. To remedy this problem, Sun et al. (2000) introduced a propriety parameter ρ into mean specification in (2.3). The parameter ρ can be interpreted as a measure spatial association over areal unit. If $\lambda_1^{-1} < \rho < \lambda_k^{-1}$, where $\lambda_1 < \dots < \lambda_n$ are the eigenvalues of $\mathbf{D}_W^{-1/2} \mathbf{W} \mathbf{D}_W^{-1/2}$, $\mathbf{D}_W - \rho \mathbf{W}$ becomes nonsingular. Let $\mathbf{W}^* = \text{Diag}(1/w_{i+}) \mathbf{W}$ denote the scaled adjacency matrix. This matrix allows the propriety parameter to be $|\alpha| < 1$. In Chapter IV, the propriety parameter is denoted by α to distinguish from ρ . Carlin and Banerjee (2003) proved that the precision matrix with the propriety parameter α corresponding to the scaled adjacency matrix is symmetric and diagonally dominant. It indicates that the precision matrix is nonsingular because a symmetric and diagonally dominant matrix is positive definite.

Although the introduction of the propriety parameter obviously overcomes impropriety of CAR specification, improper CAR model is still often implemented in spatial modeling. Banerjee et al. (2004) discussed a few reasons why improper CAR models are often used instead of proper CAR models. Firstly, the mean of η_i is intended to be an average of its neighbors in original concept of CAR models, but the mean becomes some proportion of the average of its neighbors by adding the

propriety parameter. They also conducted some simulations in order to examine the performance of calibration of ρ . It is found that a descriptive spatial association measure is not enough to indicate strong spatial correlation even though the propriety parameter ρ is almost equal to 1. This suggests that the parameter which is usually interpreted as "strength of spatial association" can misinform about the strength of association. Finally, proper CAR models can make the range of spatial pattern restricted. Therefore, the choice between two types of CAR models is ambiguous and can be determined by data or researchers.

Besag et al. (1991) proposed the pairwise difference specification which is the most popular formulation in CAR models,

$$f(\boldsymbol{\eta}) \propto \exp \left\{ -\frac{1}{2\sigma_{\eta}^2} \sum_{i \neq j} w_{ij} (\eta_i - \eta_j)^2 \right\}. \quad (2.6)$$

This specification is referred to the intrinsic autoregressive (IAR) model.

The general approach in Bayesian hierarchical spatio-temporal models is to embed spatial random effect and time effect in generalized linear model. At the first level of hierarchy, conditional mean of observations, measurement of interest are assumed to be mutually independent and the spatial modeling is accomplished in the second level of hierarchy.

2.2 Multivariate Conditional Autoregressive (MCAR) Models

Mardia (1988) developed the fundamental theory for multivariate Gaussian Markov random field (GMRF). Based on his work, Carlin and Banerjee (2003) formulated multivariate conditional autoregressive (MCAR) models.

Consider $\boldsymbol{\eta}^T = (\boldsymbol{\eta}_1^T, \dots, \boldsymbol{\eta}_p^T)$, where $\boldsymbol{\eta}_i$ is a $n \times 1$ vector, and a multivariate normal distribution,

$$\boldsymbol{\eta} \sim N(\mathbf{0}, \mathbf{B}^{-1}), \quad (2.7)$$

where \mathbf{B} is a precision matrix with blocks B_{ij} . Similar to univariate CAR models, the full conditionals of $\boldsymbol{\eta}_i$ given $\boldsymbol{\eta}_j, j \neq i$ are obtained by

$$p(\boldsymbol{\eta}_i | \boldsymbol{\eta}_{j \neq i}) \sim \text{N} \left(\sum_j C_{ij} \boldsymbol{\eta}_j, \Sigma_i \right), \quad (2.8)$$

where $C_{ij} = -B_{ii}^{-1} B_{ij}$ and $\Sigma_i = B_{ii}^{-1}$ are $n \times n$ matrices. The joint distribution is also uniquely determined by Brook's Lemma,

$$f(\boldsymbol{\eta}) \sim \text{N}(\mathbf{0}, (\mathbf{I} - \mathbf{C})^{-1} \boldsymbol{\Sigma}), \quad (2.9)$$

where \mathbf{C} and $\boldsymbol{\Sigma}$ are block diagonal matrices with entries C_{ij} and Σ_i , respectively. The propriety parameter could be included in (2.9) to avoid the impropriety problem.

Kim et al. (2001) proposed twofold conditional autoregressive model for bivariate data analysis. The model allows different diseases to share information each other. But, the model is limited to bivariate data and it is infeasible to generalize the model for a number of diseases.

Gelfand and Vounatsou (2003) provided a class of multivariate proper conditional autoregressive models. A new parametric linear transformation is also proposed for an extension which gives fascinating interpretation.

In recent, Jin, et al. (2004) point out the difficulty to specify covariance matrix in multivariate spatial models in areal data. The above studies (Carlin and Banerjee 2003; Gelfand and Vounatsou 2003) are concerned about the precision matrix, instead of the covariance matrix. The key drawback of precision matrix specification approach in multivariate areal models results in obscure interpretation. To overcome this difficulty, they proposed a new class of generalized multivariate conditional autoregressive (GM-CAR) models for areal data. The joint distribution for the multivariate spatial process is defined through simple conditional and marginal forms.

These multivariate conditional autoregressive models are also often used as spatial prior of spatial random effect in the generalized linear mixed model framework. Although several approaches have been proposed, there is no attempt to check posterior propriety with improper prior. As indicated earlier, impropriety of CAR prior could be resolved by plugging in the propriety parameter, but it is still necessary to check posterior propriety unless proper priors for all parameter in the models are specified.

2.3 Posterior Propriety

Once there is no faithful information about parameter θ or an inference only based on data is desired, a noninformative prior which has no information about parameter θ is an appropriate choice. For example, if a parameter space is a bounded continuous, $\Theta = [a, b]$, the uniform distribution is often selected as noninformative prior for θ ,

$$p(\theta) = \frac{1}{(b - a)}, \quad a < \theta < b.$$

Suppose that a parameter space is unbounded, $\Theta = (-\infty, \infty)$. Then a suitable prior could be

$$p(\theta) = c,$$

where $c > 0$ is any constant. However, this prior is improper, $\int p(\theta)d\theta = \infty$, so that it seems that the prior is not acceptable in Bayesian inference. Nevertheless, the prior make Bayesian inference possible if the integration of the likelihood function $f(\mathbf{x}|\theta)$ with respect to θ results in a finite value K . It indicates that there exists some finite normalizing constant and Bayesian inference could be carried out,

$$p(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta) \cdot c}{\int f(\mathbf{x}|\theta) \cdot cd\theta} = \frac{f(\mathbf{x}|\theta)}{K}.$$

However, we need to pay attention to ensure whether the resulting posterior is proper under improper prior because the prior does not always lead proper posterior. For example, in high-dimensional models, the data do not contain sufficient information in order to identify all parameters in models, so that it is required that some priors for parameters should be informative.

Hobert and Casella (1996) warned the users of hierarchical linear mixed models with improper priors not to implement MCMC without ensuring that the resulting posterior is proper. Generally, improper priors are elicited for variance components in hierarchical linear mixed models because of the reasons for the choice in the beginning of this chapter. By dealing with conjugate priors in the prior specification, the full conditionals required for the Gibbs sampling are easily derived and it seems that there is no problem in general Bayesian inference. However, the Gibbs sampler itself could not point out whether the posterior is proper or not. Although the resulting posterior is improper, the output from a Gibbs sampler can behave perfectly. But, the posterior inference based on the output is worthless because it is from a non-existent posterior distribution. They showed this situation with real data analysis and provided the theorems to give sufficient conditions for posterior propriety under improper priors.

Ibrahim and Laud (1991) provide two theorems that allow Jeffrey's priors to be used in generalized linear models with fixed scale parameters. The theorems also give sufficient and necessary conditions for the propriety of the posterior and prior.

Ghosh et al. (1998) present the theorem to ensure posterior propriety in hierarchical Bayes generalized linear mixed models with spatial random effect.

Sun et al. (2001) examined necessary and sufficient conditions for posterior propriety in hierarchical linear mixed models with the improper priors for the fixed effects and variance components.

Chen et al. (2002) investigated the posterior propriety for generalized linear mixed model when an improper prior is placed on the regression parameters. The propriety is considered under a general link function and a general covariance structure for random effects.

CHAPTER III

UNIVARIATE HIERARCHICAL SPATIAL MODELS

3.1 Introduction

Transportation-related deaths and injuries constitute a major public health problem in the United States. Injuries and fatalities occur in all transportation modes, but crashes involving motor vehicles account for almost 95% of all transportation fatalities and most injuries. Despite the progress made in roadway safety in the past several decades, tens of thousands of people are still killed and millions of people are injured in motor vehicle crashes each year. For example, in 1999 nearly 42,000 people were killed in traffic crashes and over 3.2 million more were injured.

Motor vehicle fatalities are the leading cause of unintentional injury deaths, followed by falls, poisonings, and drownings (about 16,000, 10,000, and 4,400 deaths per year, respectively) (NSC 2002). They are also responsible for as many pre-retirement years of life lost as cancer and heart disease, about 1.2 million years annually. In fact, motor vehicle crashes are the leading cause of death for people aged 1 to 33. Societal economic losses from these crashes are huge, estimated by the National Highway Traffic Safety Administration to exceed \$230 billion in 2000. Thus, much work remains to be done to develop a better understanding of the causes of vehicle crashes-their chains of events and operating environments-and to develop countermeasures to reduce the frequency and severity of these crashes (USDOT 1996-1999).

Safety is one of the U.S. Department of Transportation's (USDOT's) five current strategic goals, and Rodney Slater, a former Transportation Secretary stated: "Safety is a promise we keep together." Indeed, roadway safety intersects with all five core functional areas within conventional highway engineering (planning, design, construc-

tion, operation, and maintenance) and crosscuts the boundaries of other engineering (vehicle and material) and nonengineering areas (human factors, public health, law enforcement, education, and other social sciences). Thus, research in roadway safety requires interdisciplinary skills and essential cooperation from various engineering and social science fields.

In 2002, a series of conferences was hosted by the Bureau of Transportation Statistics under the general title of "Safety in Numbers: Using Statistics to Make the Transportation System Safer." These conferences supported the top strategic safety goal of promoting public health and safety "by working toward the elimination of transportation-related deaths, injuries, and property damage" (USDOT 2002).

3.1.1 Contributing Factors, Countermeasures, and Resources

Motor vehicle crashes are complex events involving the interactions of five major factors: drivers, traffic, roads, vehicles, and the environment (e.g., weather and lighting conditions) (e.g., Miaou 1996). Among these factors, driver error has been identified as the main contributing factor to a great percentage of vehicle crashes, and many research efforts are being undertaken to better understand human and other synergistic factors that cause or facilitate crashes. These factors include operator impairment due to the use of alcohol and drugs, medical conditions, or human fatigue and the operator's interaction with new technologies used on the vehicle.

Countermeasures to reduce the number and severity of vehicle crashes are being sought vigorously through various types of community, education, and law enforcement programs and improved roadway design and vehicle safety technology. However, many of these programs have limited resources and need better tools for risk assessment, prioritization, and resource scheduling and allocation.

Recognizing that "to err is human" and that driver behavior is affected by vir-

tually all elements of the roadway environment, highway engineers are constantly redesigning and rebuilding roadways to meet higher safety standards. This includes designing and building roadways and roadsides that are more "forgiving" when an error is made, more conforming to the physical and operational demands of the vehicle, and that better meet drivers' perceptions and expectations in order to reduce the frequency of human errors (TRB 1987). The relatively low fatality rate on the Interstate Highway System (about half the fatality rate of the remainder of the nation's highways) is evidence of the impact of good design on highway safety (Evans 1991).

Many impediments keep highway engineers from achieving their design and operational goals, including a lack of resources and a vast highway system that needs to be built, operated, maintained, audited, and improved. They must make incremental improvements over time and make difficult decisions on the tradeoffs among cost, safety, and other operational objectives. Consequently, knowing where to improve and how to prioritize and schedule improvements is as important as knowing which roadway and roadside features and elements to add or improve. Tools for identifying, auditing, ranking, and clinically evaluating problem sites; developing countermeasures; and allocating resources are essential for highway engineers who make these decisions.

3.1.2 Disease Mapping and Methods Using Spatial Models

In recent years, a multiplicity of the studies for disease mapping and ecological analysis has been conducted using spatial(-temporal) models in Bayesian framework. This model-based approach has yielded a dramatic gain in the number and scope of applications in public health studies of risks from disease such as leukemia, pediatric, asthma, and lung cancer (Carlin and Louis 1996; Knorr-Held and Besag 1998; Xia et al. 1997; Ghosh et al. 1999; Lawson et al. 1999; Zhu and Carlin 1999; Dey et al.

2000; Sun et al. 2000; Lawson 2001; Green and Richardson 2001). A special issue of "Statistics in Medicine" entitled "Disease Mapping with a Focus on Evaluation" was also published to report the development of this approach (vol. 19, Issues 17-18, 2000). Among other applications, disease mapping have been used to:

- describe the spatial variation in disease incidence for the formulation and validation of etiological hypotheses;
- identify and rank areas with potentially elevated risk and time trends so that action may be taken;
- provide a quantitatively informative map of disease risk in a region to allow better risk assessment, prioritization, and resource allocation in public health.

Clearly, roadway traffic safety planning has similar requirements and can potentially benefit from these kinds of maps.

Studies have shown that risk estimation using hierarchical Bayes models has several advantages over estimation using classical methods. One important point that has been stressed by almost all of these studies is that individual incidences of diseases of concern are relatively rare for a typical analysis unit such as census tract or county. As a result, estimates based on simple aggregation techniques may be unreliable because of large variability from one analysis unit to another. This variability makes it difficult to distinguish chance variability from genuine differences in the estimates and is sometimes misleading for analysis units with a small population size. Hierarchical Bayes models, however, especially those Poisson-based generalized linear models with spatial random effects, have been shown to have the ability to account for the high variance of estimates in low population areas and at the same time clarify overall geographic trends and patterns (Ghosh et al. 1999; Sun et al. 2000).

Note that in the context of sample surveys the type of problem described above is commonly referred to as a small area, local area, or small domain estimation problem. Ghosh and Rao (1994) conducted a comprehensive review of hierarchical Bayes estimations and found them favorable for dealing with small area estimation problems when compared with other statistical methods. Hierarchical models are also gaining enormous popularity in fields such as education and sociology, in which data are often gathered in a nested or hierarchical fashion: for example, as students within classrooms within schools (Goldstein 1999). In these fields, hierarchical models are often called multilevel models, variance component models, or random coefficients models.

The overall strength of the Bayesian approach is its ability to structure complicated models, inferential goals, and analyses. Among the hierarchical Bayes methods, three are most popular in disease mapping studies: empirical Bayes (EB), linear Bayes (LB), and full Bayes methods. These methods offer different levels of flexibility in specifying model structures and complexity in computations. As suggested by Lawson (2001): "While EB and LB methods can be implemented more easily, the use of full Bayesian methods has many advantages, not least of which is the ability to specify a variety of components and prior distributions in the model set-up."

To many statistical practitioners, it is fair to say that the challenges they face dealing with real-world problems come more often from the difficulties of handling nonsampling errors and unobserved heterogeneity (because of the multitude of factors that can produce them) than from handling sampling errors and heterogeneity due to observed covariates. One potential advantage of using the full Bayes model is the flexibility that it can provide in dealing with and adjusting for the unobserved heterogeneity in space and time, whether it is structured or unstructured.

3.1.3 Objectives and Significance of Work

Mapping transforms spatial data into a visual form, enhancing the ability of users to observe, conceptualize, validate, and communicate information. Research efforts in the visualization of traffic safety data, which are usually stored in large and complex databases, are quite limited at this time because of data and methodological constraints (Smith et al. 2001). As a result, it is common for engineers and other traffic safety officials to analyze roadway safety data and make recommendations without actually "seeing" the spatial distribution of the data. This is not an optimal situation.

To the best of our knowledge, unlike the public health community, which has developed models for disease mapping, the roadway safety research community has not done much to develop model-based maps for traffic crash data. One of the objectives of the study presented here was to initiate development of model-based mapping for roadway traffic crashes. Vehicle crash records and roadway inventory data from Texas were used to illustrate the nature of the data, the structure of models, and results from the modeling.

Overall, TxDOT maintains nearly 80,000 centerline-miles of paved roadways, serving about 400 million vehicle-miles per day. Over 63% of the centerline-miles are rural two-lane roads that, on average, carry fewer than 2,000 vehicles per day. These low volume rural roadways 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 deemed cost-effective to upgrade these roads to the preferred design standards. However, vehicles on these roadways generally travel at high speeds and thus tend to have relatively more severe injuries when vehicle crashes

occur. For example, in 1999, about 26% of the Texas on-system crashes were fatal (K), incapacitating injury (A), and nonincapacitating injury (B) (or KAB) crashes, compared with over 40% of the crashes on rural, two-lane, low volume on-system roads (Fitzpatrick et al. 2001). As a result, we have chosen to focus this study on crashes occurring on rural, two-lane, low-volume, on-system roads.

This paper is organized as follows: the next section briefly describes the sources and nature of the data analyzed in this study, followed by a quick review of modeling and computational techniques and a discussion of Poisson-based hierarchical Bayes model with space-time effects and possible variants. Results from models of various levels of complexities are then presented and compared, and we conclude with a discussion of future work.

3.2 Description of Data

The Texas Department of Transportation (TxDOT) maintains highway development with 25 geographic districts and each of them includes 6 to 17 counties. District offices divide their work into area offices and area offices into local maintenance offices. Design and maintenance, right-of-way acquisition, construction oversight, and transportation planning are mainly administrated and accomplished locally due to the diversity of climates and soil conditions in Texas. Figure 1 is a map to show geographic districts, counties, and urbanized areas in Texas.

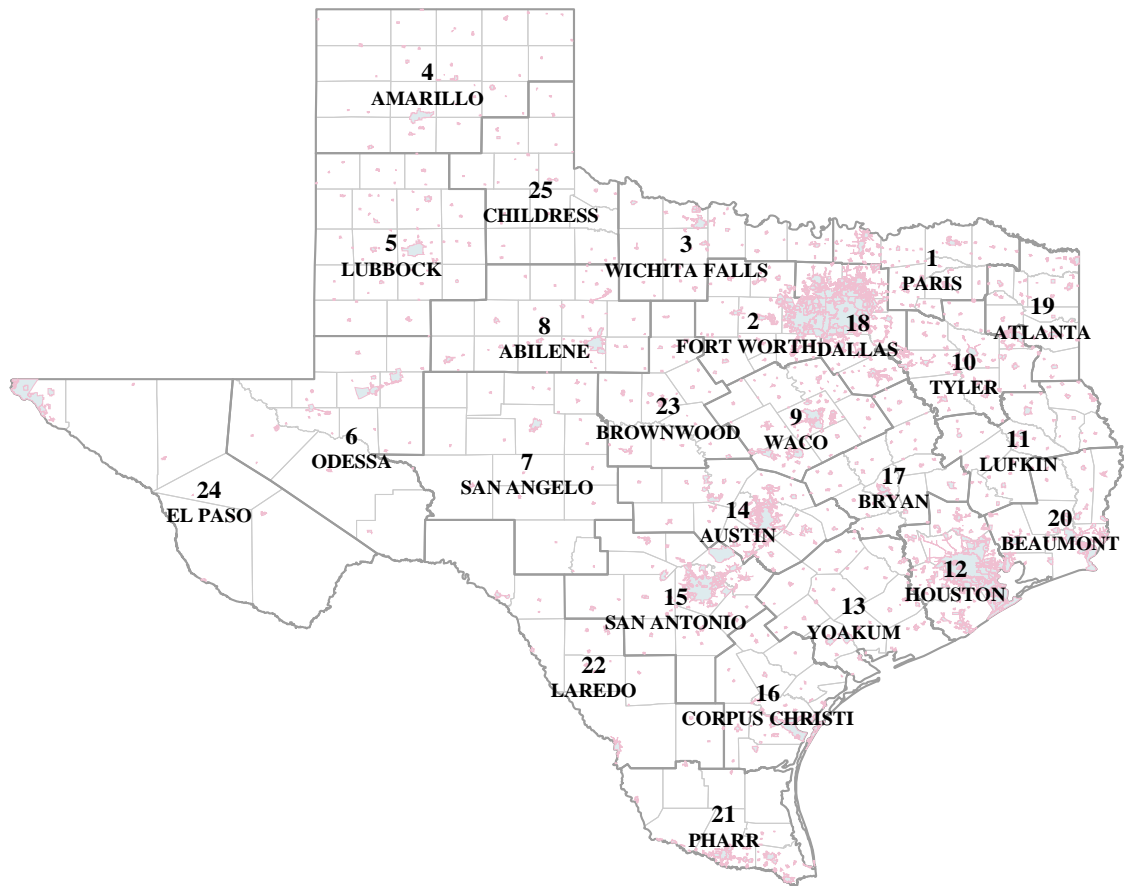


Figure 1: Geographic Districts, Counties, and Urbanized Areas in Texas.

The measurements of interest in this study are Annual KAB crash frequencies for rural, two-lane, low volume on-system roads at the county level from 1992 to 1999. The number of reported KAB crashes by county in 1999 is shown in Figure 2. Low-volume roads refer to road segments carrying fewer than 2,000 vehicles per day and 4,824 KAB crashes were occurred on the roads of interest in 1999. Figure 3 shows total vehicle-miles for the same year (in millions of vehicle-miles traveled, or MVMT). The highest, lowest, and average of the "raw" annual KAB crash rates by county are displayed by a bubble plot in Figure 4 and the rate represents in number of crashes per MVMT. In Figure 4, raw crash rate is expressed in terms of the diameter of the ball. The three balls on the lowest left corner indicate 1.0, 0.5, and 0.25 crashes per

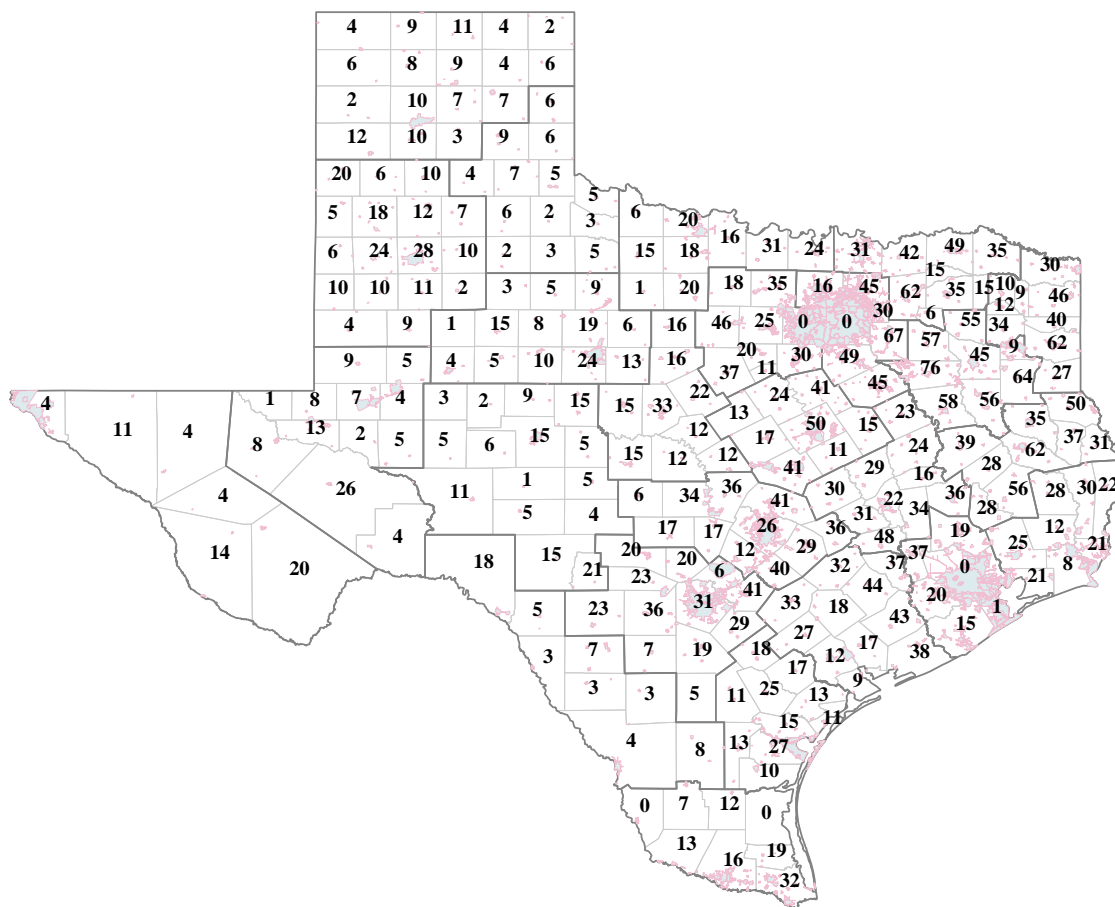


Figure 2: The Number of KAB Crashes on Rural, 2-Lane, Low-Volume, On-System Roads in Each Texas County: 1999.

MVMT, respectively. The rate of county average over 8 years is 0.45 crashes/MVMT. Note that two of the urban counties and one rural county are excluded from this study because these counties have almost no rural two-lane roads with the level of traffic volumes of interest.

Figure 3 shows total vehicle-miles for the same year (in millions of vehicle-miles traveled, or MVMT). The highest, lowest, and average of the "raw" annual KAB crash rates by county are displayed by a bubble plot in Figure 4 and the rate represents in number of crashes per MVMT. In Figure 4, raw crash rate is expressed in terms of the diameter of the ball. The three balls on the lowest left corner indicate 1.0, 0.5,

and 0.25 crashes per MVMT, respectively. The rate of county average over 8 years is 0.45 crashes/MVMT. Note that two of the urban counties and one rural county are excluded from this study because these counties have almost no rural two-lane roads with the level of traffic volumes of interest.

Figure 4 shows that crash rates in most counties over the eight-year period are stable, whereas remarkable differences between the highest and the lowest rates are found in several counties. It is clear that eastern counties have considerable higher rates and east-west is divided in terms of the KAB crash rates. Rural roadways in the eastern counties are limited by the rolling terrain and tend to have less driver-friendly characteristics, with more horizontal and vertical curves (Figure 5), restricted sight distance, and less forgiving roadside development (e.g., trees closer to the travelway and steeper side slopes). Besides, rural roads in more and larger urbanized areas in the east tend to have higher roadside development scores, higher access density, and narrower lanes and/or shoulders (Fitzpatrick et al. 2001). Figure 6 shows that the proportions related to wet-weather crashes are generally higher in northern and eastern counties. In addition, it is found in Figure 7 that eastern counties have more crashes at intersections than western counties.

The National Highway System Designation Act of 1995 repealed the national maximum speed limit and returned authority to set speed limits to the states. Speed limits for daylight on many highways in Texas were increased from 55 mph to 70 for passenger vehicles and to 60 for trucks in early 1996. Griffin et al. (1998) investigated relationship between speed limit raising and the number of KAB crashes increased using monthly time series data from January 1991 to March 1997. The study indicated that the number of KAB crashes on the roads whose speed limits were raised increased in five out of the six highway categories considered during the post-invention periods. Furthermore, the speed limit raising resulted in increases in both the number of

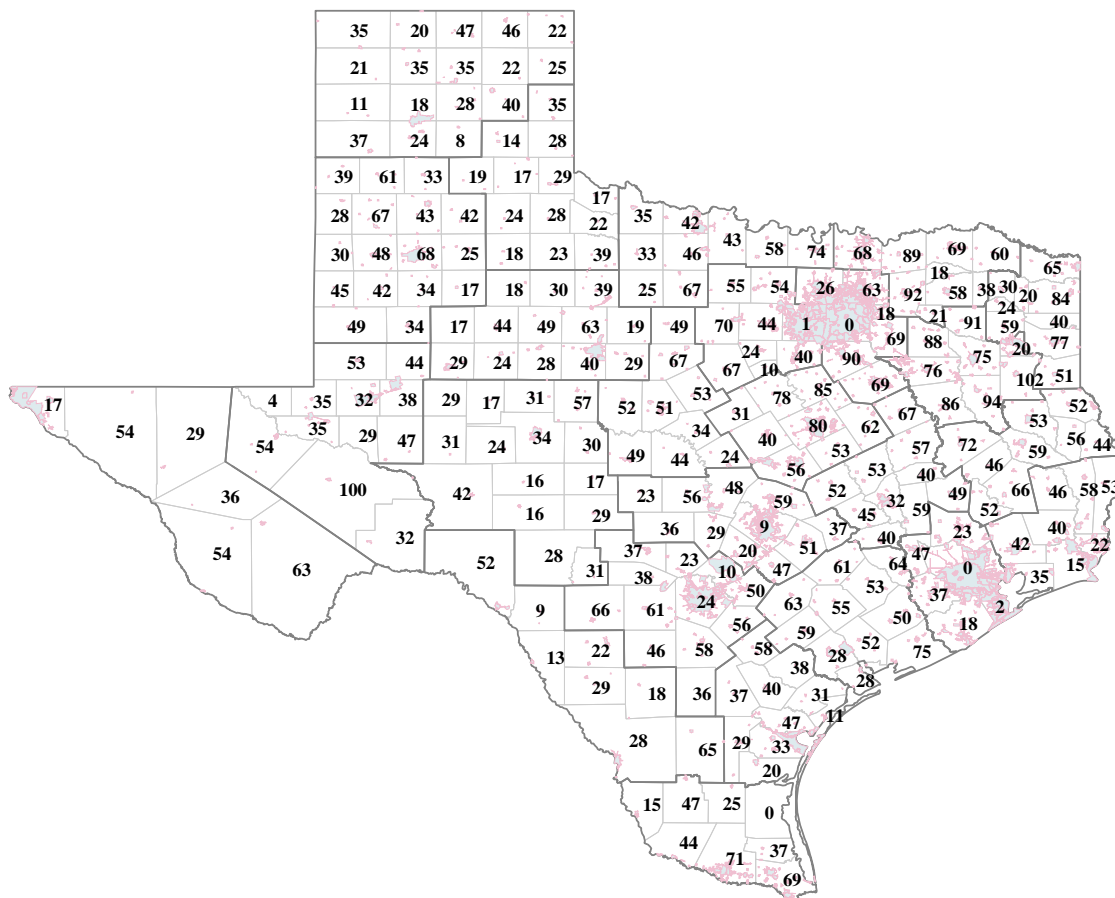


Figure 3: Vehicle-Miles Traveled on Rural, 2-Lane, Low-Volume, On-System Roads in Each Texas County: 1999.

injuries and fatalities related to speed, 3.3% for incapacitating injuries, 7.0% for non-incapacitating injuries, and 14% for fatalities from 1995 to 1996. Hence, a change in KAB crash rates in 1996 is expected in this study.

3.3 Bayesian Hierarchical Models

As part of our modeling efforts, we developed a Poisson hierarchical Bayes model for traffic crash risk mapping at the county level for state-maintained rural, two-lane, low volume roads (fewer than 2,000 vehicles per day) in Texas. In general, the model consists of six components:

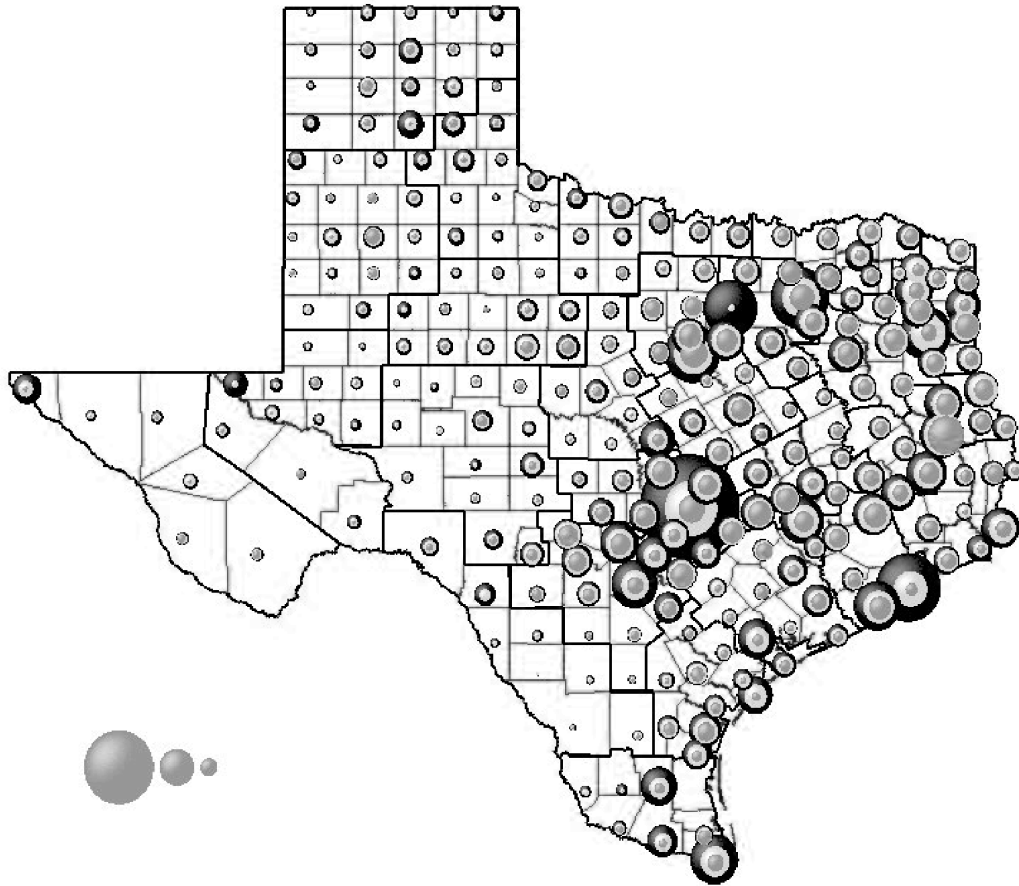


Figure 4: "Raw" Annual KAB Crash Rates in Crashes per MVMT by County: 1992-1999 (Highest, Average, and Lowest in the 8-Year Period). The Diameter of the Dark Outer Circle Represents the Highest Crash Rate; the Light Gray Intermediate Band Represents the Average Crash Rate; and the Medium Gray Inner Circle Represents the Lowest Crash Rate.

- an offset term: the amount of travel occurring on state-maintained rural, two-lane, low volume roads (fewer than 2,000 vehicles per day)
- a fixed TxDOT district effect
- a fixed or random covariate effect term
- a random spatial effect component using conditional autoregressive prior in which the inverse of the Great Circle distance between the centroid of counties

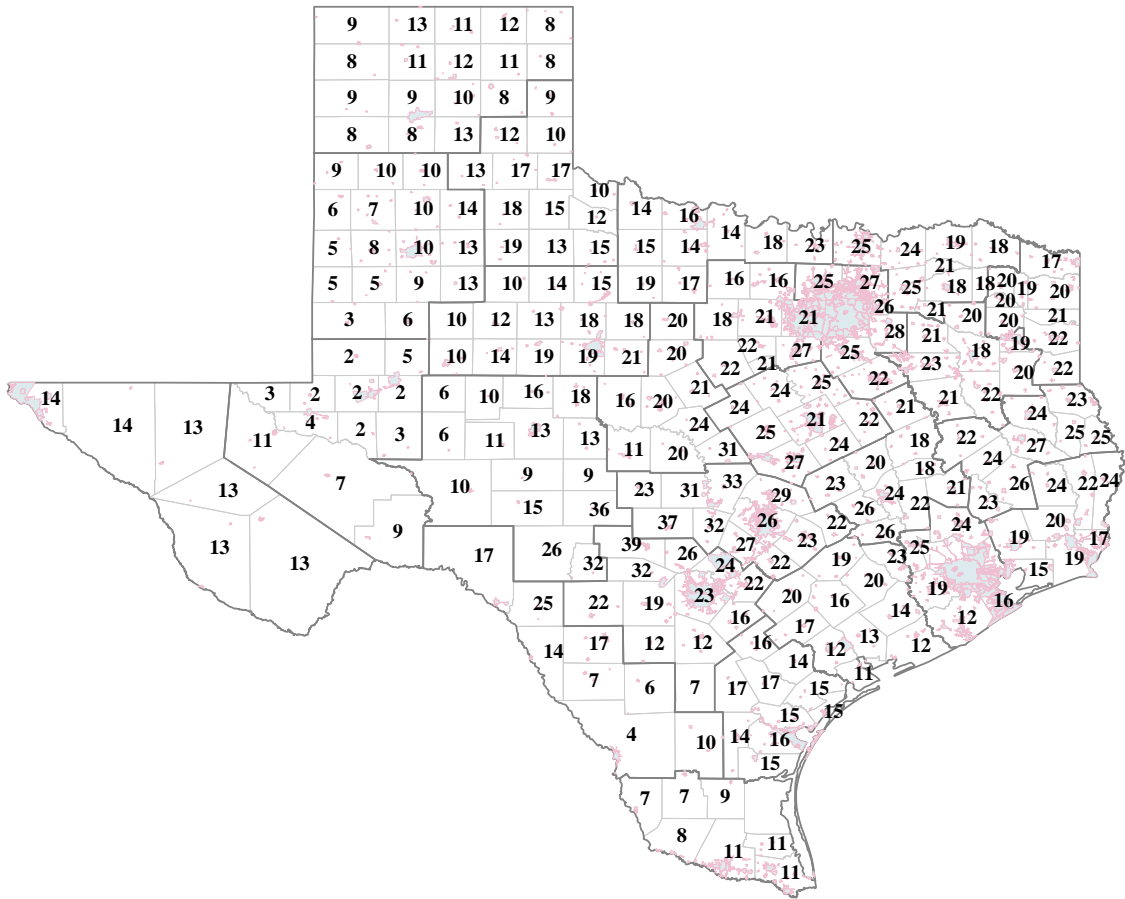


Figure 5: Proportion of KAB Crashes That Occurred on Sharp Horizontal Curves in Each County (In Percent; Averaged over the 1992-1999 Period and 6 Neighboring Counties).

is employed as the weights for structuring spatial association

- a fixed or random time effect term to represent year-to-year changes
- an exchangeable random effect component representing a pure independent random local spatio-temporal variation that is independent of all other components in the model

In this chapter, we consider a fixed effect as an effect that is subject only to the uncertainty associated with an unstructured noninformative prior distribution with no unknown parameters and the sampling variation.¹ A fixed effect can, however, vary

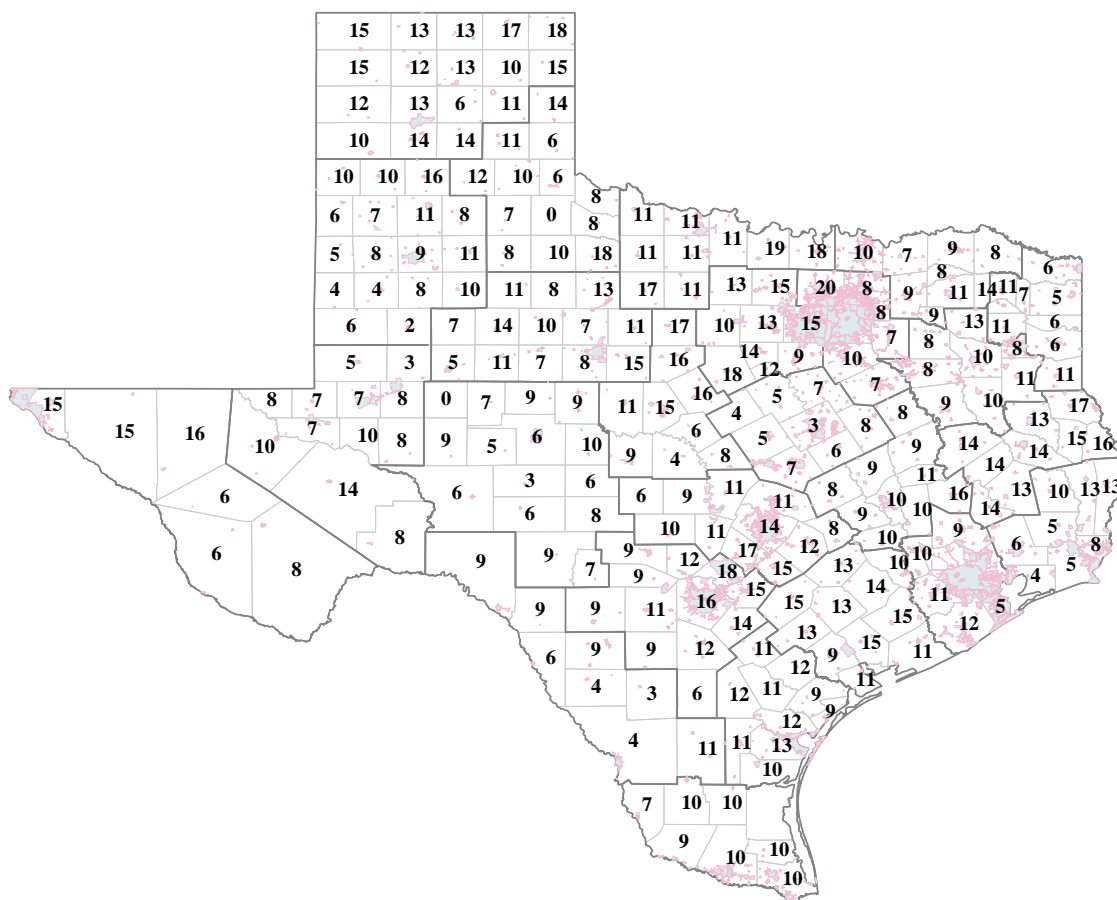


Figure 6: Proportion of KAB Crashes That Occurred under Wet Pavement Conditions for Each County: 1999 (In Percent; Averaged over 6 Neighboring Counties).

by individual districts, counties, and time periods (see the discussion of model hierarchy). Note also that unlike the traditional traffic crash prediction models (Maher and Summersgill 1996; Miaou 1996; and Hauer 1997), which were concerned principally with modeling the fixed effects for individual sites (e.g., road segments or intersections), this study focuses more on exploring the structure of the random component of the model for area-based data. The rediscovery by statisticians in the last 15+ years of the Markov chain Monte Carlo (MCMC) methods and new developments, including convergence diagnostic statistics, are revolutionizing the entire statistical field (Besag et al. 1995; Gilks et al. 1996; Carlin and Louis 1996; Roberts and Rosenthal

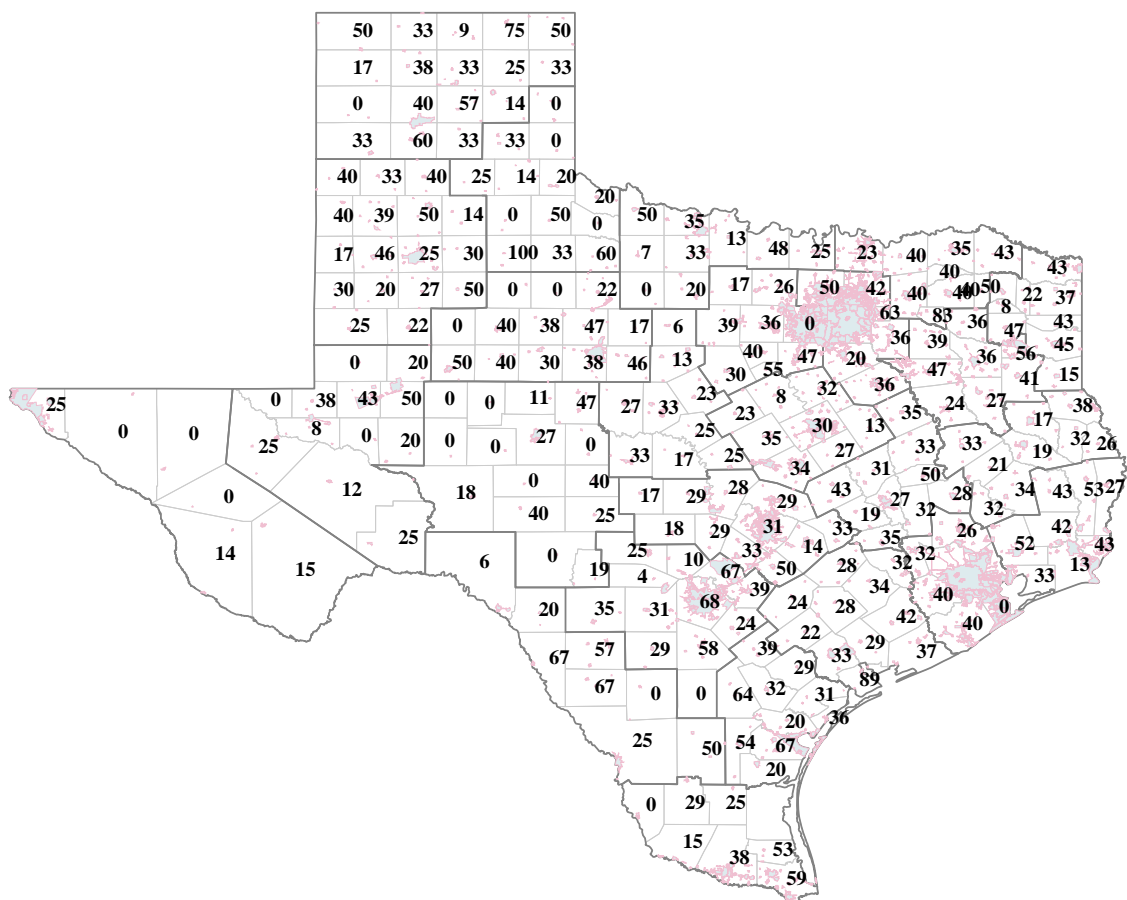


Figure 7: Proportion of KAB Crashes That Were Intersection, Intersection Related, or Driveway Access Related for Each County: 1999 (In Percent).

1998; Robert and Casella 1999). At the same time, improved computer processing speed and lower data-collection and storage costs are allowing more complex statistical models to be put into practice. These complex models are often hierarchical and high dimensional in their probabilistic and functional structures. Furthermore, many models also need to include dynamics of unobserved and unobservable (or latent) variables; deal with data distributions that are heavily tailed, highly overdispersed, or multimodal; and work with datasets with missing data points. MCMC provides a unified framework within which model identification and specification, parameter estimation, performance evaluation, inference, prediction, and communication of

ods for hierarchical models are also available, e.g., iterative generalized least squares (IGLS), expected generalized least squares (EGLS), and generalized estimating equations (GEE). These estimation procedures tend to focus on obtaining a consistent estimate of the fixed effect rather than exploring the structure of the random component of the model (Goldstein 1999).

For some problems, existing software packages such as WinBUGS (Spiegelhalter et al. 2000) and MLwiN (Yang et al. 1999) can provide Gibbs and other MCMC sampling for a variety of hierarchical Bayes models. For the models presented in this paper, we relied solely on the WinBUGS codes. At present, however, the type of spatial and temporal models available in WinBUGS is somewhat limited and will be discussed later.

3.3.1 Notations

We let the indices i , j , and t represent county, TxDOT district, and time period, respectively, where $i = 1, 2, \dots, I$; $j = 1, 2, \dots, J$; and $t = 1, 2, \dots, T$.

For the data analyzed, we have 251 counties, divided among 25 districts, and 8 years of annual data (i.e., $I = 251$, $J = 25$, and $T = 8$). As indicated earlier, each district may include 6 to 17 counties, which will be represented by county set D_j , where $j = 1, 2, \dots, 25$. That is, D_j is a set of indices representing counties administered by TxDOT district j .

We define variable Y_{it} as the total number of reported KAB crashes on the rural road of interest in county i and year t . We also define ν_{it} as the observed total vehicle-miles traveled (VMT) in county i and year t for the roads in discussion, representing the size of the population at risk. In addition, we define x_{itk} as the k th covariate associated with county i and year t . Three covariates were considered.

3.3.2 Covariates

The first covariate x_{it1} is a surrogate variable intended to represent the percentage of time that the road surface is wet due to rain, snow, etc. Not having detailed weather data, we chose to use the proportion of KAB crashes that occurred under wet pavement conditions as a surrogate variable. In addition, we do not expect general weather characteristics to vary much between neighboring counties. Therefore, the proportion for each county is computed as the average of this and six other neighboring counties that are close to the county in terms of their Great Circle distances. We do, however, expect weather conditions to vary significantly from year to year. Thus, for each county i , we have x_{it1} change with t .

The second covariate x_{it2} is intended to represent spatial differences in the number of sharp horizontal curves in different counties. The actual inventory of horizontal curves on the highway network is not currently available. However, when a traffic crash occurs, site characteristics including the horizontal curvature are coded in the traffic crash database. We chose to use the proportion of KAB crashes that occurred on sharp horizontal curves in each county as a surrogate variable, and we define a sharp horizontal curve as any road segment having a horizontal curvature of 4 or higher degrees per 100-foot arc. Given that this roadway characteristic is mainly driven by terrain variations, we do not expect this characteristic to vary much between neighboring counties. Therefore, as in the first covariate, the proportion for each county is computed as the average of this and six other neighboring counties that are close to the county in terms of their Great Circle distances. Furthermore, for this type of road, we did not expect the proportion to vary in any significant way over the eight-year period in consideration. Thus, the average proportion from 1992 to 1999 was actually used for all t . In other words, for each county i , x_{it2} are the same for all

t .

The third covariate x_{it3} is a surrogate variable intended to represent degrees of roadside hazards. As in the second covariate, the actual inventory of hazards (ditches, trees, and utility poles), available clear zones, and geometry and surface type of roadsides are not available. Similar to the first covariate, a surrogate variable was devised to indicate the proportion of KAB crashes that ran off roads and hit fixed objects on the roadside. We also do not expect this characteristic to vary much between neighboring counties over the eight-year period in consideration. Again, the average proportion from 1992 to 1999 was used for all t , i.e., for each county i , x_{it3} are the same for all t . Figure 8 shows the spatial distribution of this variable.

The use of these surrogate variables is purely data driven (as opposed to theory driven) and empirical in nature. We use the proportion of wet crashes (x_{it1}) as an example to explain the use and limitation of such surrogate measures in practice. First, variables such as "percentage of wet crashes" and "wet crashes to dry crashes ratio" are commonly used in wet-weather accident studies. Examples in the literature include Coster (1987), Ivey and Griffin (1990), and Henry (2000). These authors reviewed various wet-weather accident studies, and the relationships between 1) skid numbers (or friction values) of pavement and percentage of wet weather accidents, and 2) skid numbers and wet/dry pavement surfaces were quite well documented. Although they were conducted with limited data, these wet weather accident studies also suggest that crash rates are higher during wet surface conditions than under dry surface conditions, and some indicate that traffic volumes are reduced by about 10% to 20% during wet weather in rural areas (no significant reduction was found in urban areas).

Second, the use of percentage of wet crashes as a surrogate variable in this study to explain the variation of crash rates by county mixes several possible relationships

and has limited explanatory power. A positive correlation of percentage of wet crashes and crash rate mixes has at least two possible relationships: 1) the effect of wet surface conditions on crash rates, and 2) the effect of rainfall (or other precipitation) on traffic volumes. Everything else being equal, if the wet surface crash rate is the same as the dry surface crash rate, then we do not expect this positive correlation to be statistically significant in the model regardless of the relative traffic volumes during wet or dry surface conditions. We interpret a positive correlation as an indication that a higher crash rate is indeed experienced during wet surface conditions than during dry conditions. However, because of the lack of data on traffic volumes by wet and dry surface conditions, we are not able to quantify the difference in crash rates under the two surface conditions. This is the main limitation in using such a surrogate measure.

3.3.3 Probabilistic and Functional Structures

The space-time models considered in this study are similar to the hierarchical Bayes generalized linear model used in several disease mapping studies cited earlier. At the first level of hierarchy, conditional on mean μ_{it} , Y_{it} values are assumed to be mutually independent and Poisson distributed as

$$Y_{it} \sim \text{Poisson}(\mu_{it}). \quad (3.1)$$

The mean of the Poisson is modeled as

$$\mu_{it} = \nu_{it} \lambda_{it}, \quad (3.2)$$

where total VMT ν_{it} is treated as an offset and λ_{it} is the KAB crash rate. The rate, which has to be non-negative, is further structured as

$$\log(\lambda_{it}) = \sum_{t=1}^T \sum_{j=1}^J \alpha_{jt} I(i \in D_j) + \sum_k \beta_k x_{itk} + \delta_t + \eta_i + e_{it}, \quad (3.3)$$

where $I(S)$ is the indicator function of the set S . This makes the first term on the right hand side of equation (3.3) the intercept representing district effects at different years; x_{itk} are covariates discussed earlier and their interactions; δ_t represents year-to-year time effects due, e.g., to speed limit, weather, and socioeconomic changes; η_i is a random spatial effect; e_{it} is an exchangeable, unstructured, space-time random effect; and α_{jt} and β_k are regression parameters to be estimated from the data. As defined earlier, D_j is a set of indices representing counties administered by TxDOT district j .

Many possible variations of equation (3.3) were and could potentially be considered in this study. For each component that was assumed to have a fixed effect, the second level of hierarchy was chosen to be an appropriate noninformative prior. On the other hand, for each component that was assumed to have a random effect, the second level of hierarchy was a prior with certain probabilistic structure that contained unknown parameters. The priors for these unknown parameters (called hyperpriors) constitute the third level of the hierarchy. What follows are discussions of the variation of models considered by this study, some limitations of the WinBUGS software, and possible extensions of the models considered.

The intercept term, which represents the district effect over time, was assumed to have fixed effects with noninformative normal priors. For the covariates x_{itk} , we considered both fixed and random effects. That is, β_k was assumed to be either a fixed value or random variable. The three covariates discussed earlier and three of their interactive terms, $x_{it4} = x_{it1}x_{it2}$, $x_{it5} = x_{it1}x_{it3}$, and $x_{it6} = x_{it2}x_{it3}$, were included in the model. It is important to note that the values of these covariates were *centered* for better numerical performance. Noninformative normal priors were also assumed for fixed-effect models. For the random-effect model, β_k , $k = 1, 2, \dots, 6$, are assumed to be independent and normally distributed with mean μ_{β_k} and variance $\sigma_{\beta_k}^2$, expressed

as $N(\mu_{\beta_k}, \sigma_{\beta_k}^2)$. Noninformative normal and inverse gamma priors (or more precisely hyperpriors) were assumed for μ_{β_k} and $\sigma_{\beta_k}^2$, respectively.

With 251 counties and 8 years of data, the data are considered to be quite rich spatially but rather limited temporally, as are data in many disease mapping studies. Because of this limitation, we only considered two simple temporal effects for δ_t : fixed effects varying by t (or a year-wise fixed-effect model) and an order-one autoregressive model (AR(1)) with the same coefficient for all t . Again, noninformative priors were used for both models. 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. From the fixed effect, we expected to see a change in δ_t at $t = 5$ (1996), due in part to the speed limit increase in that year.

Recent disease mapping research has focused on developing more flexible, yet parsimonious, spatial models that have attractive statistical properties. Based on the Markov random field (MRF) theory, Besag's conditional autoregressive (CAR) model (Besag 1974, 1975) and its variants are by far the most popular ones adopted in disease mapping. We considered several Gaussian CAR models, all of which have the following general form

$$p(\eta_i | \eta_{-i}) \propto r_\eta^{1/2} \exp \left\{ -\frac{r_\eta}{2} \sum_{i^* \in C_i} w_{ii^*} (\eta_i - \eta_{i^*})^2 \right\}, \quad (3.4)$$

where η_{-i} represents all η except η_i , C_i is a set of counties representing "neighbors" of county i , w_{ii^*} is a positive weighting factor associated with the county pair (i, i^*) . This equation is shown to be equivalent to

$$p(\eta_i | \eta_{-i}) \sim N(\mu_{\eta_i}, \sigma_{\eta_i}^2),$$

where $\mu_{\eta_i} = \sum_{i^* \in C_{i^*}} (w_{ii^*}/w_{i+}) \eta_{i^*}$, $\sigma_{\eta_i}^2 = 1/(r_\eta w_{i+})$, and $w_{i+} = \sum_{i^* \in C_{i^*}} w_{ii^*}$. In our study, we had $w_{ii^*} = 1/d_{ii^*}^c$, where $d_{ii^*}^c$ is the Great Circle distance between the

centroid of county i and i^* , and c is a constant parameter equal to 1 or 2 (note that d_{ii^*} ranges roughly from 30 to 700 miles.) With regard to the number of neighbors, we adopted a more generous definition by allowing every other county $i^*(\neq i)$ to be a neighbor of county i .

In theory, we could treat the constant c as an unknown parameter and estimate it from the data. However, in the current version of WinBUGS, the weights of the built-in CAR spatial model do not allow unknown parameters (Spiegelhalter et al. 2000), which we found to be a limitation for our application. In a separate attempt to find a good range of the decay constant for the inverse distance weight in the CAR model, we adopted a simpler model that included only the offset, the yearwise time effect, and the Gaussian CAR components. We estimated the same model with different c values between 0 and 4 and found that model performance was best achieved when the decay constant was set between 1 and 2 (based on the deviance information criterion to be discussed shortly). Weights with an exponential form $w_{ii^*} = \exp(-cd_{ii^*})$ were also examined but are not reported in this paper.

We also explored the L-1 CAR models of the following form:

$$p(\eta_i | \eta_{-i}) \propto r_\eta \exp \left\{ -r_\eta \sum_{i^* \in C_i} w_{ii^*} |\eta_i - \eta_{i^*}| \right\}, \quad (3.5)$$

where r_η is a fixed-effect parameter the same for all i . Weights with the same c as in the Gaussian CAR models were considered. WinBUGS constrains the sum of η_i to zero to make both the Gaussian CAR and L-1 CAR spatial models identifiable. A non-informative gamma distribution was used as hyperpriors for r_η in equations (3.4) and (3.5).

The spatial correlation structure represented by equations (3.4) and (3.5) is considered global in the sense that the distribution functions and associated parameters (c and r_η) do not change by i . More sophisticated models allowing spatial correla-

tion structure to be adaptive or location specific are being actively researched (e.g., Lawson 2000; Green and Richardson 2001). Still, computational challenges seem to be keeping researchers from exploring more flexible, yet parsimonious, space-time interactive effects, and more research in this area needs to be encouraged (Sun et al. 2000).

For the exchangeable random effects, we considered two commonly used distributions. One distribution assumed e_{it} to be independent and identically distributed (*iid*) as

$$e_{it} \sim N(0, \sigma_e^2). \quad (3.6)$$

Another distribution assumed an *iid* one-parameter gamma distribution as

$$\exp(e_{it}) \sim G(\psi, \psi), \quad (3.7)$$

which has a mean equal to 1 and a variance $1/\psi$. The use of a one-parameter gamma distribution (instead of a two-parameter gamma) ensures that all model parameters are identifiable. Again, non-informative inverse gamma and gamma distributions were used as hyperpriors for σ_e^2 and ψ , respectively.

3.4 Deviance Information Criterion and Variants

The deviance information criterion (DIC) has been proposed to compare the fit and complexity (measured by the effective number of parameters) of hierarchical models in which the number of parameters is not clearly defined (Spiegelhalter et al. 1998; Spiegelhalter et al. 2002). DIC is a generalization of the well-known Akaike Information Criterion (AIC) and is based on the posterior distribution of the deviance statistic

$$D(\theta) = -2\log(p(y|\theta)) + \log(f(y)),$$

where $p(y|\theta)$ is the likelihood function for the observed data vector y given the parameter vector θ , and $f(y)$ is some standardizing function of the data alone. For the Poisson model, $f(y)$ is usually set as the saturated likelihood, i.e., $f(y) = p(y|\mu = y)$ where μ is a vector of the statistical means of vector y .

DIC is defined as a classical estimate of fit plus twice the effective number of parameters, which gives

$$DIC = D(\bar{\theta}) + 2p_D = \bar{D} + p_D, \quad (3.8)$$

where $D(\bar{\theta})$ is the deviance evaluated at $\bar{\theta}$, the posterior means of the parameters of interest; p_D is the effective number of parameters for the model; and \bar{D} is the posterior mean of the deviance statistics $D(\theta)$.

As with AIC, models with lower DIC values are preferred. From equation (3.8), we can see that the effective number of parameters p_D is defined as the difference between the posterior mean of the deviance \bar{D} and the deviance at the posterior means of the parameters of interest $D(\bar{\theta})$

$$p_D = \bar{D} - D(\bar{\theta}).$$

It was shown that in nonhierarchical models (or models with negligible prior information) DIC is approximately equivalent to AIC. It has also been emphasized that the quantity of p_D can be trivially obtained from an MCMC analysis by monitoring both θ and $D(\theta)$ during the simulation. For the random-effect model considered in equations (3.1) through (3.3), the parameter vector θ should include α_{jt} , β_k , δ_t , η_i and e_{it} for all i , j , k , and t .

In addition to DIC values and associated quantities \bar{D} , $D(\bar{\theta})$, and p_D , we also used some goodness-of-fit measures that attempted to standardize DIC in some fashion.

This includes DIC divided by sample size n and R_{DIC}^2 , which defined as

$$R_{DIC}^2 = 1 - \frac{DIC_{model} - DIC_{ref}}{DIC_{max} - DIC_{ref}} \quad (3.9)$$

where DIC_{model} is the DIC value for the model under evaluation, DIC_{max} is the maximum DIC value under fixed one-parameter model, DIC_{ref} is a DIC values from a referenced model that, ideally, represents some expected lower bound of the Poisson hierarchical model for a given dataset.

Clearly, R_{DIC}^2 is devised in the spirit of the traditional r^2 goodness-of-fit measure for regression models. Through simulations, Miaou (1996) evaluated several similar measures using AIC for overdispersed Poisson models. Since DIC is known to be non-invariant with respect to the scale of the data (Spiegelhalter et al. 1998; Spiegelhalter et al. 2002), an analytical development of DIC_{ref} is difficult. However, we know that for a model with a good fit, \bar{D} should be close to sample size n (Spiegelhalter et al. 2002). We, therefore, chose $DIC_{ref} = n$ as a conservative measure for computing R_{DIC}^2 ; that is, the effective number of parameters was essentially ignored.

Another goodness-of-fit indicator considered is $1/\psi$, which is the variance of $\exp(e_{it})$ under the gamma model, indicating the extent of overdispersion due to exchangeable random effect. In theory, this value could go to zero when such effects vanish. Thus, similar to R_{DIC}^2 , we can devise the following measure:

$$R_{\psi}^2 = 1 - \frac{(1/\psi)_{model}}{(1/\psi)_{max}},$$

where $(1/\psi)_{model}$ is the variance of $\exp(e_{it})$ for the model under consideration, and $(1/\psi)_{max}$ is the amount of overdispersion under the simplest model. In essence, $(1/\psi)_{ref}$, the expected lower bound, is set to zero.

3.5 Results

Table 1 lists 42 models of various complexities examined by this study. These models include simplified versions of the general model presented in equations (3.2) and (3.3), as well as models for reference purposes, e.g., models 1 to 3. Model 1 is a saturated model, in which the estimates of the Poisson means $\hat{\mu}_{it}$ are equal to y_{it} . Model 2, expressed as α_0^* , is a one-parameter Poisson model without the offset, and model 3 is another one-parameter model with the offset. Essentially, model 2 focuses on traffic crash frequency and model 3 on traffic crash rate.

In Table 1, the following symbols are used:

- α_j stands for fixed district effects.
- β_F and β_N respectively represent fixed covariate effects and random covariate effects with independent normal priors.
- δ_F and δ_{AR} respectively stand for fixed time and AR(1) time effects.
- For the random spatial effects, η_{N1} and η_{L1} , represent the Gaussian and L-1 CAR models shown in equations (3.4) and (3.5), respectively, and both have a decay constant c equal to 1.
- η_{N2} and η_{L2} represent similar spatial models with a decay constant c equal to 2.
- The components e_N and e_G represent exchangeable random effects as presented in equations (3.6) and (3.7), respectively.

We experienced some computational difficulties for the models that included the β_N component when we tried to include all six main and interactive effects. Therefore, for all models with the β_N component, we only included the three main effects.

In computing R_{DIC}^2 , DIC_{max} is defined as the maximum DIC value under a fixed one-parameter model, which is model 2 in the table when crash frequency is the focus and model 3 when crash rate is the focus. Similarly, in computing R_{ψ}^2 , $(1/\psi)_{max}$ is set as the amount of overdispersion under the simplest model with an e_G error component, which is model 11 for models focusing on the crash rate.

As a rule, in our development we started with simpler models, and the posterior means of the estimated parameters of these simple models were then used to produce initial values for the MCMC runs of more complex models.

In general, the models presented in the table are ordered by increasing complexity: intercepts only, intercepts + covariate effect, intercepts + covariate effect + exchangeable effect, intercepts + covariate effect + exchangeable effect + spatial/temporal effects, and so on. Models 7 to 9 and the last eight models include a more complex fixed-effect intercept term. The models are presented in the table in line with the order in which they were estimated with the WinBUGS codes.

The MCMC simulations usually reached convergence quite quickly. Depending on the complexity of the models, for typical runs, we performed 10,000 to 20,000 iterations of simulations and removed the first 2,000 to 5,000 iterations as burn ins. As in other iterative parameter estimation approaches, good initial estimates are always the key to convergence. For some of the models, we have hundreds of parameters and MCMC monitoring plots based on the Gelman-Rubin statistics (which are part of the output from the WinBUGS codes). Because estimated parameters usually converge rather quickly, their convergence plots, which are not particularly interesting to show, are not presented here.

From DIC and other performance measures in Table 1, several observations can be made:

- For the exchangeable random effect, models with a gamma assumption (equation 3.6) are preferred over those with a normal assumption (equation 3.5). This is observed by comparing the performance of, e.g., model 15 with model 14, model 18 with model 17, and model 27 with model 26.
- Models with fixed covariate effects are favored over their random-effect counterparts. This is seen by comparing, e.g., model 25 with model 24 and model 33 with model 34.
- Models with fixed time effects (e.g., model 23) performed better than those with AR(1) time effects (e.g., model 22).
- Models with separate district and time effects (α_j and δ_t) are preferred over those with joint district time effects (α_j). For example, we can compare the performance of model 27 with model 42 and model 40 with model 24.
- For comparable model structures, adding a spatial component decreases the DIC value quite significantly, which indicates the importance of the spatial component in the model. As an example, we can compare model 17 with model 20. Except for the spatial component, these two models have the same structures (in intercept terms, covariate effects, and the error component). Model 17 does not have any spatial component, while model 20 includes a normal CAR model. The DIC value drops from 3,287 for model 17 to 2,755 for model 20, a very significant reduction when compared with the differences in DIC values for various models presented in Table 1. Other comparisons that would give the same conclusion include model 19 vs. model 22 or model 38 with models 40 and 42.
- No particular spatial CAR models considered by this study, i.e., η_{N1} , η_{L1} , η_{N2} , or η_{L2} , were clearly favored over other CAR models.

Table 1: Deviance Information Criterion and Related Performance Measures for Models of Various Complexities:
 Notes i : County, j : District, t : Time Period, and $n = 2,008$: Sample Size. Model 1 is the Saturated Model,
 Model 2 is the Constant Frequency Model, and Model 3 is the Constant Rate Model.

No.	Model	D	$D(\theta)$	p_D	DIC	DIC/n	R_{DIC}^2	R_{DIC}^{2*}	$1/\psi$	R_{ψ}^2
1	α_{it}^*	2026	61	1965	3991	1.99	0.91	0.77		
2	α_0^*	23416	23414	1	23417	11.66	0	-		
3	$\nu + \alpha_0$	10706	10701	5	10710	5.33	0.59	0		
4	$\nu + \alpha_0 + \beta_F$	6716	6713	2	6718	3.35	0.78	0.46		
5	$\nu + \alpha_0 + \beta_F + \delta_F$	6686	6676	10	6695	3.33	0.78	0.46		
6	$\nu + \alpha_j + \beta_F$	5126	5090	36	5161	2.57	0.85	0.64		
7	$\nu + \alpha_{j,t}$	5316	5113	202	5518	2.75	0.84	0.6		
8	$\nu + \alpha_{j,t} + \beta_F$	4816	4608	208	5024	2.5	0.86	0.65		
9	$\nu + \alpha_{j,t} + \beta_F + \delta_F$	4816	4623	193	5009	2.49	0.86	0.66		
10	$\alpha_0 + e_N$	2116	350	1765	3881	1.93	0.91	0.78		
11	$\nu + \alpha_0 + e_G$	2086	582	1504	3590	1.79	0.93	0.82	0.263	0
12	$\nu + \alpha_0 + e_N$	2146	611	1534	3680	1.83	0.92	0.81		
13	$\nu + \alpha_0 + \eta_{N1}$	2846	2473	373	3218	1.6	0.94	0.86		
14	$\nu + \alpha_j + e_N$	2156	988	1168	3324	1.66	0.94	0.85		
15	$\nu + \alpha_j + e_G$	2136	978	1158	3294	1.64	0.94	0.85	0.103	0.61
16	$\nu + \alpha_j + \eta_{N1}$	2846	2483	362	3208	1.6	0.94	0.86		
17	$\nu + \alpha_j + \beta_F + e_N$	2186	1085	1101	3287	1.64	0.94	0.85		
18	$\nu + \alpha_j + \beta_F + e_G$	2156	1046	1110	3266	1.63	0.94	0.86	0.089	0.66
19	$\nu + \alpha_j + \beta_F + \delta_{AR} + e_N$	2176	1080	1096	3272	1.63	0.94	0.85		
20	$\nu + \alpha_j + \beta_F + \eta_{N1} + e_N$	2066	1376	689	2755	1.37	0.97	0.91		
21	$\nu + \alpha_j + \beta_F + \eta_{L1} + e_N$	2066	1375	691	2757	1.37	0.97	0.91		
22	$\nu + \alpha_j + \beta_F + \eta_{N1} + \delta_{AR} + e_N$	2066	1380	686	2751	1.37	0.97	0.91		
23	$\nu + \alpha_j + \beta_F + \eta_{N1} + \delta_F + e_N$	2056	1377	679	2735	1.36	0.97	0.92		
24	$\nu + \alpha_j + \beta_F + \eta_{N1} + \delta_F + e_G$	2026	1342	684	2709	1.35	0.97	0.92	0.022	0.92

Continued on next page

Table 1 continued

No.	Model	D	$D(\theta)$	p_D	DIC	DIC/n	R_{DIC}^2	R_{DIC}^{2*}	$1/\psi$	R_{ψ}^2
25	$\nu + \alpha_j + \beta_N + \eta_{N1} + \delta_F + e_G$	2036	1327	709	2744	1.37	0.97	0.92	0.02	0.92
26	$\nu + \alpha_j + \beta_F + \eta_{N2} + \delta_F + e_N$	2046	1360	686	2731	1.36	0.97	0.92		
27	$\nu + \alpha_j + \beta_F + \eta_{N2} + \delta_F + e_G$	2016	1329	687	2703	1.35	0.97	0.92	0.021	0.92
28	$\nu + \alpha_j + \beta_N + \eta_{N2} + \delta_F + e_G$	2026	1318	708	2733	1.36	0.97	0.92	0.02	0.92
29	$\nu + \alpha_j + \beta_F + \eta_{L1} + \delta_F + e_N$	2056	1375	681	2736	1.36	0.97	0.92		
30	$\nu + \alpha_j + \beta_F + \eta_{L1} + \delta_F + e_G$	2026	1339	687	2712	1.35	0.97	0.92	0.022	0.92
31	$\nu + \alpha_j + \beta_N + \eta_{L1} + \delta_F + e_G$	2036	1330	706	2742	1.37	0.97	0.92	0.02	0.92
32	$\nu + \alpha_j + \beta_F + \eta_{L2} + \delta_F + e_N$	1966	1181	785	2751	1.37	0.97	0.91		
33	$\nu + \alpha_j + \beta_F + \eta_{L2} + \delta_F + e_G$	2016	1327	689	2704	1.35	0.97	0.92		
34	$\nu + \alpha_j + \beta_N + \eta_{L2} + \delta_F + e_G$	2026	1317	709	2735	1.36	0.97	0.92	0.02	0.92
35	$\nu + \alpha_{jt} + e_N$	2166	917	1249	3415	1.7	0.93	0.84		
36	$\nu + \alpha_{jt} + e_G$	2126	886	1240	3365	1.68	0.94	0.84	0.108	0.59
37	$\nu + \alpha_{jt} + \beta_F + e_N$	2156	965	1191	3347	1.67	0.94	0.85		
38	$\nu + \alpha_{jt} + \beta_F + e_G$	2136	962	1174	3310	1.65	0.94	0.85	0.089	0.66
39	$\nu + \alpha_{jt} + \beta_F + \eta_{N1} + e_N$	1956	1103	853	2808	1.4	0.96	0.91		
40	$\nu + \alpha_{jt} + \beta_F + \eta_{N1} + e_G$	1936	1091	845	2781	1.38	0.96	0.91	0.027	0.9
41	$\nu + \alpha_{jt} + \beta_F + \eta_{N2} + e_N$	1946	1086	860	2806	1.4	0.96	0.91		
42	$\nu + \alpha_{jt} + \beta_F + \eta_{N2} + e_G$	1926	1072	854	2780	1.38	0.96	0.91	0.027	0.9

- Despite the empirical nature of the two goodness-of-fit measures R_{DIC}^2 and R_{ψ}^2 , seeing some of the better models that have values exceeding 0.9 provides some comfort as to the general explanatory capability of these models.

Table 2 shows some statistics of the estimated posterior density of a selected number of parameters for model 27, which was one of the best models in terms of the DIC value and other performance measures discussed above. Also, Figure 9 presents estimated posterior mean crash rates, as well as their 2.5 and 97.5 percentiles, in a bubble plot for 1999 by county.

From Table 2, one can see that the fixed-time effect δ_t jumps from about 0 in previous years to about 0.05 in $t = 4$ (1995) and has another increase to about 0.09 at $t = 5$ (1996). The value comes down somewhat (about 0.06) in 1998 ($t = 7$) and 1999 ($t = 8$) but is still significantly higher than those in the preintervention periods. It has been suggested that the jump in 1995 was perhaps due to higher driving speeds by drivers in anticipation of a speed limit increase, and higher crash rates in 1996 were due in part to the speed limit increase and less favorable winter weather (Griffin *et al.* 1998). Lower δ_t values in 1998 and 1999 may suggest that drivers had adjusted themselves and become more adapted to driving at higher speeds.

From the same model (model 27), estimates of α_j , i.e., district effects, range from about -0.5 to -1.5, indicating significant district-level variations in crash risk. The covariate effects β_k indicate that the horizontal curve variable is the most influential and statistically significant variable in explaining the crash rate variations over space. Wet pavement condition is the second-most significant variable. The ran-off-road fixed-object variable is not a statistically significant variable, which suggests that ran-off-road fixed-object crash risk is correlated with and perhaps exacerbated by the presence of sharp horizontal curves and wet pavement conditions.

Table 2: Example MCMC Simulation Output for Model 27: Some Statistics of the Estimated Posterior Density for a Selected Number of Parameters. Set δ_1 to 0 as Baseline.

Parameter	Mean	Standard error	2.5%	Median	97.5%
α_1	-0.963	0.154	-1.269	-0.964	-0.662
α_2	-0.639	0.148	-0.929	-0.635	-0.356
α_3	-1.131	0.162	-1.450	-1.128	-0.823
α_4	-1.240	0.183	-1.595	-1.237	-0.882
α_5	-1.288	0.155	-1.595	-1.283	-0.993
α_6	-1.427	0.182	-1.768	-1.429	-1.066
α_7	-1.376	0.128	-1.629	-1.375	-1.127
α_8	-1.218	0.130	-1.479	-1.215	-0.978
α_9	-0.984	0.158	-1.283	-0.986	-0.666
α_{10}	-0.582	0.162	-0.889	-0.584	-0.260
α_{11}	-0.610	0.156	-0.924	-0.602	-0.321
α_{12}	-0.498	0.208	-0.918	-0.489	-0.097
α_{13}	-0.919	0.149	-1.232	-0.914	-0.634
α_{14}	-0.668	0.137	-0.943	-0.668	-0.398
α_{15}	-0.770	0.139	-1.045	-0.772	-0.503
α_{16}	-0.893	0.165	-1.216	-0.891	-0.551
α_{17}	-0.754	0.139	-1.030	-0.756	-0.495
α_{18}	-0.630	0.170	-0.966	-0.621	-0.294
α_{19}	-0.649	0.171	-0.975	-0.645	-0.326
α_{20}	-0.877	0.208	-1.282	-0.880	-0.459
α_{21}	-1.005	0.308	-1.656	-0.981	-0.442
α_{22}	-1.561	0.224	-1.980	-1.566	-1.114
α_{23}	-1.189	0.147	-1.483	-1.187	-0.901
α_{24}	-1.114	0.378	-1.831	-1.127	-0.379
α_{25}	-1.401	0.156	-1.712	-1.396	-1.094
δ_1	0	0	0	0	0
δ_2	0.0132	0.026	-0.0380	0.0129	0.0645
δ_3	-0.0156	0.027	-0.0677	-0.0156	0.0376
δ_4	0.0508	0.027	-0.0009	0.0508	0.1034
δ_5	0.0929	0.027	0.0418	0.0926	0.1453
δ_6	0.0886	0.027	0.0365	0.0886	0.1408
δ_7	0.0632	0.027	0.0111	0.0631	0.1155
δ_8	0.0603	0.026	0.0089	0.0601	0.1123
β_1	0.00286	0.0018	-0.00079	0.0029	0.00648
β_2	0.00723	0.0019	0.0035	0.00721	0.01103
β_3	-0.00057	0.0014	-0.00346	-0.00057	0.00229
β_4	-0.00004	0.0002	-0.00050	-0.00004	0.0004
β_5	0.00009	0.0002	-0.00028	-0.00010	0.00048
β_6	-0.00015	0.0002	-0.00045	-0.00015	0.00014
ψ	46.52	5.04	37.83	46.18	57.41
$1/\psi$	0.0023	0.0002	0.0019	0.0023	0.0028

3.6 Discussion

Most of the methodologies developed in disease mapping were intended for area-based data, e.g., number of cancer cases in a county or census tract during a study period. While we demonstrate the use of some of these methodologies for roadway traffic crashes at the county level, we recognize that, fundamentally, traffic crashes are network-based data, whether they are intersection, intersection-related, driveway access-related, or nonintersection crashes. Figure 10 gives an example of the locations of KAB crashes on the state-maintained highway network of a Texas county in 1999.

Thus, an obvious extension of the current study is to develop risk maps for traffic crashes on road networks. The problem is essentially one of developing hierarchical models for Poisson events on a network (or a graph). We expect that, in different applications, these maps may need to be developed by roadway functional classes, vehicle configurations, types of crashes (e.g., those involving drunk drivers), and crash severity types (e.g., fatal, injury, and noninjury crashes). We also expect these network-based maps to be useful for roadway safety planners and engineers to 1) estimate the cost and benefit of improving or upgrading various design and operational features of the roadway, 2) identify and rank potential problem roadway locations (or hotspots) that require immediate inspection and remedial action, and 3) monitor and evaluate the safety performance of improvement projects after the construction is completed. Such maps need to be constructed from quality accident-, traffic-, and roadway-related databases and with scientifically grounded data visualization and modeling tools.

Modeling and mapping of traffic crash risk need to face all the challenges just as in the field of disease mapping, i.e., multilevel data and functional structures, small areas of occurrence of studied events at each analysis unit, and strong unobserved

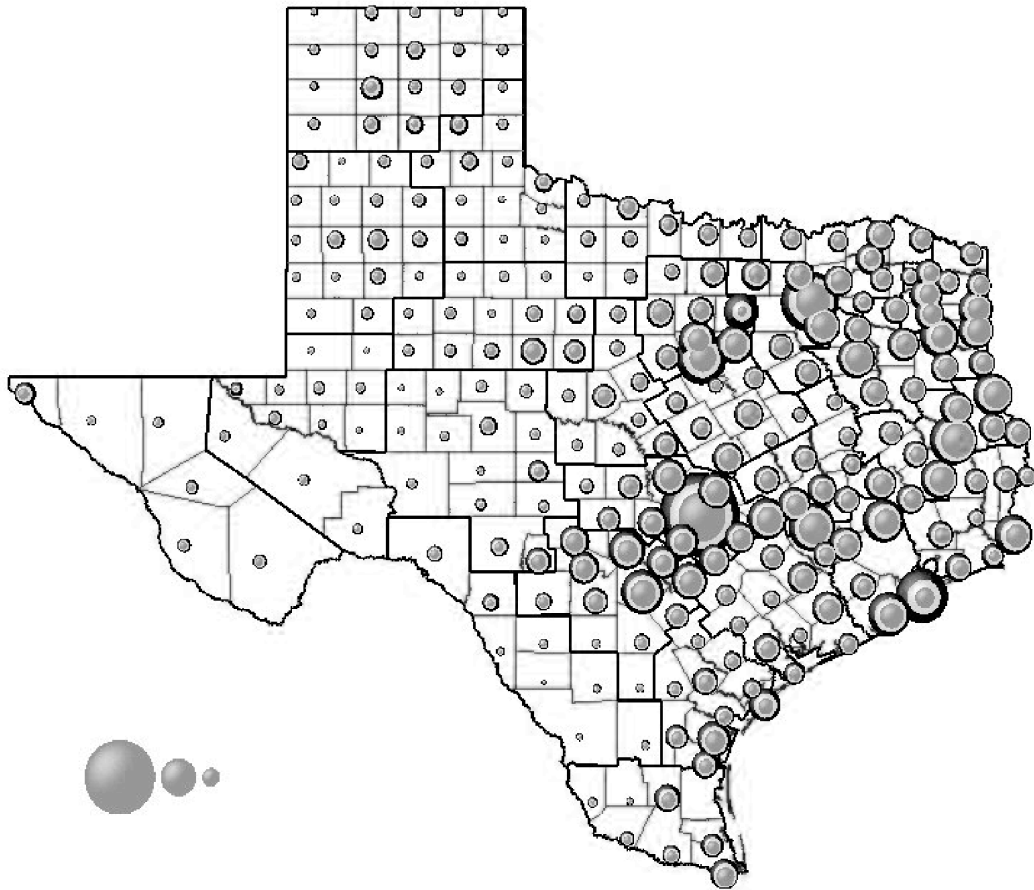


Figure 9: Estimated KAB Crash Rates in Crashes per MVMT by County from Model 27: 1999 (97.5 Percentile Mean, and 2.5 Percentile of the Posterior Density). The Diameter of the Dark Outer Circle Represents the 97.5 Percentile Estimates; the Light Gray Intermediate Band Represents the Mean; and the Medium Gray Inner Circle Represents the 2.5 Percentile Estimates.

heterogeneity. The hierarchical nature of the data can be described as follows: In a typical roadway network, other than the fact that roadway networks are connected or configured in specific ways, individual road entities are classified by key geometric characteristics (e.g., segments, intersections, and ramps), nested within roadway functional or design classifications, further nested within operational and geographical units, and subsequently nested within various administrative and planning organizations. Strong unobserved heterogeneity is expected because of the unobserved driver

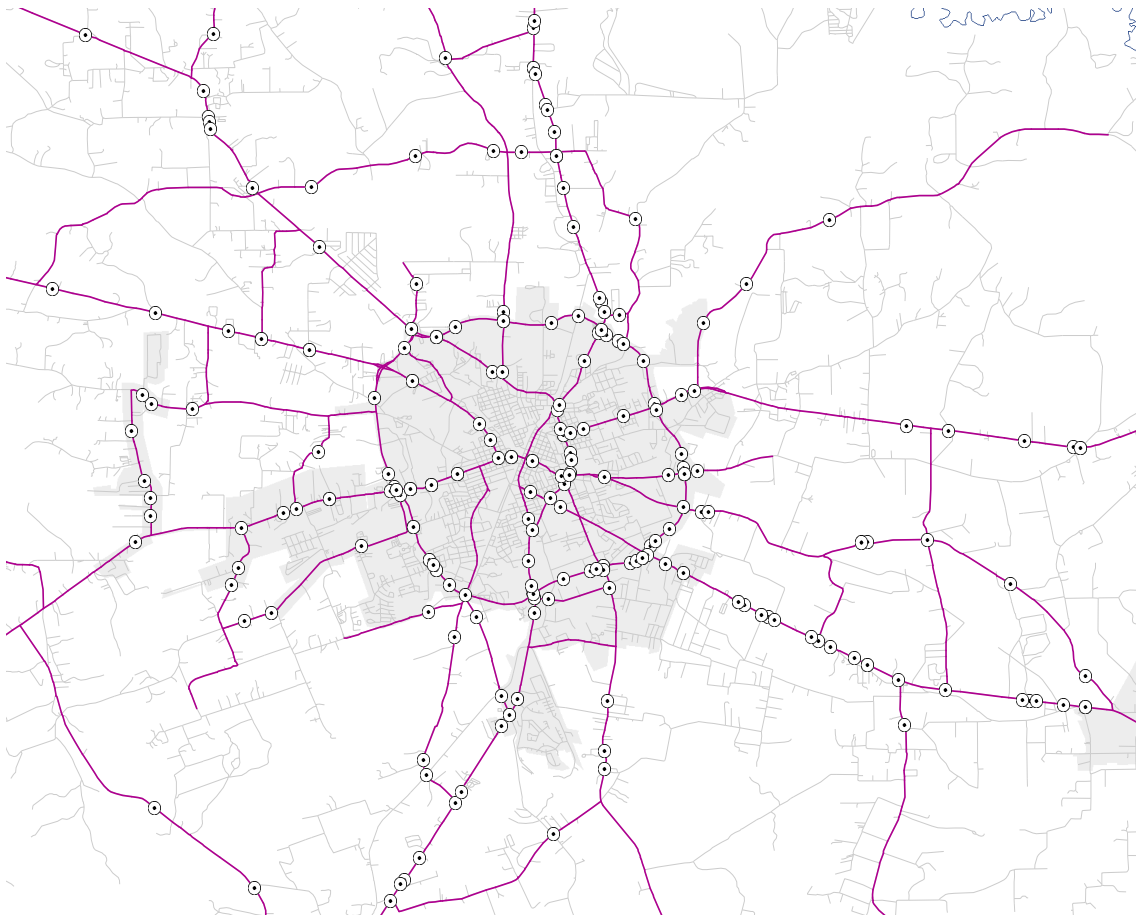


Figure 10: Locations of KAB Crashes on the State-Maintained Highway Network of a Texas County in 1999.

behaviors at individual roadway entities that are responsible for a large percentage of crash events.

Every state maintains databases on vehicle crash records and roadway inventory data. We hope that the results of our study using Texas data will motivate the development of similar studies in other states. We also envision that the network-based hierarchical models we propose can potentially be utilized in other transportation modes and in computer and communication network studies to further the exploration and interpretation of incidence data. Furthermore, the hierarchical Bayes models with spatial random effects described in this paper can be used to develop

more efficient sampling surveys in transportation that alleviate multilevel and small-area problems. Finally, the models have been shown to have the ability to account for the high variance of estimates in low-population areas and at the same time clarify overall geographic trends and patterns, which make them good tools for addressing some of the equity issues required by the Transportation Equity Act for the 21st Century.

CHAPTER IV

MULTIVARIATE HIERARCHICAL SPATIAL MODELS

4.1 Introduction

Though highway safety community is a latecomer in the application of generalized linear models (GLM) in data analysis, recently there a surge of applications of GLM in highway safety research. The use of overdispersed Poisson, including the negative-binomial regression models and their variations has become very popular. Examples include Morris et al. (1991), Hauer (1992), Miaou et al. (1992), Miaou and Lum (1993), Miaou (1994), Miaou (1996), Bonneson and McCoy (1996), Mather and Summersgill (1996), Shankar et al. (1997), and Vogt and Bared (1998). Adjusting for the regression-to-the-mean and local effect has been an important problem surrounding many “before-after” safety evaluation and problem site identification studies using empirical-Bayes estimators (Hauer 1992; Christiansen et al. 1992; Flowers and Griffin 1992). Also, the use of logistic and ordered probit regression models has now become fairly common in studying the factors that affect the crash severity (Duncan et al. 1998; McGinnis et al. 1998).

Most of the above mentioned papers ignore the spatial dependence among the crash data. A very recent exception is Miaou et al. (2003) who studied the geographical pattern of crashes in the state of Texas. The analysis of spatially referenced data has been an increasingly active area of both methodological and applied statistical research. Sophisticated computer programs known as geographic information system (GIS) have revolutionized the analysis and display of such data sets, through their ability to “layer” multiple data sources over a common study area. Finally, Markov chain Monte Carlo (MCMC) algorithms enable the fitting of complex hierarchical

models in a full Bayesian framework, permitting full posterior inference for underlying parameters in complex model settings. That way we can avoid the naive empirical Bayes analysis which usually underestimates uncertainties related to the model.

In this chapter, we will explore the extension of spatial models in a multivariate setup. Such models are necessary to analyze more than one type of crashes simultaneously, since a number of different crashes may share the same set of risk factors. As an example, for different types of crashes, the risk factor could be the excessive curvature or the bad condition of the road. The main purpose of this work is to borrow strength or share information from similar sources, as well as the most directly available sources, to improve crash risk estimates. Estimation of crash risk for a particular crash type may be improved by using information from other types of crash.

We will propose four multivariate models to improve crash risk estimates. In the first two models, the correlation among the regions is induced by a random error term and this is a spatial analog of “shared component” models proposed by Knorr-Held and Best (2000). The third model will be based on the correlated conditional autoregressive (CAR) structure where the correlation is induced through the scale parameters of the CAR model. The final model is a multivariate CAR model following a suggestion of Mardia (1988). A Bayesian criterion is used to choose the best fitted model for our data.

The improper prior is usually used in prior specification of Bayesian hierarchical modeling, which makes it imperative to check that the joint posterior is proper. Ghosh et al. (1998) provided sufficient conditions to obtain a proper posterior for the univariate CAR prior. We will extend the results in the multivariate set up.

The outline of the remaining sections is as follows. In Section 2 of this Chapter, we review briefly the univariate hierarchical Bayesian model. Several versions of

multivariate hierarchical Bayesian models are introduced in Section 3. Data analysis based on the multivariate models is carried out and some concluding remarks are made in Section 4. The proofs of some of the technical results are deferred to the Appendices.

4.2 Univariate Hierarchical Model

Let y_1, y_2, \dots, y_n denote measurements in a given period of time for the n regions. Conditional on $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^T$, y_1, \dots, y_n are assumed to be independent with pdf's

$$p(y_i|\theta_i) = \exp(y_i\theta_i - \Psi(\theta_i))h(y_i).$$

This is the one-parameter exponential family model.

Ghosh et al. (1999) developed a hierarchical model as $\theta_i = q_i + \mathbf{x}_i^T \boldsymbol{\beta} + \eta_i + e_i$ for $i = 1, \dots, n$, where q_i is a known parameter. The \mathbf{x}_i are region-level covariates, having parameter coefficient $\boldsymbol{\beta}$. The e_i capture region-wide heterogeneity via an exchangeable normal prior. Finally, the η_i are the parameters that make this a truly spatial model by capturing regional clustering. They assumed that the spatial random effects η_i and random errors e_i were mutually independent, Also the η_i have a *pairwise difference prior* with joint pdf

$$p(\boldsymbol{\eta}) \propto (\sigma_\eta^2)^{-1/2} \exp \left\{ -\frac{1}{2\sigma_\eta^2} \sum_{i \neq j} w_{ij} (\eta_i - \eta_j)^2 \right\}, \quad (4.1)$$

where $w_{ij} = w_{ji}$. This is the so-called *pairwise difference* prior considered quite extensively in Besag et al. (1995). The errors e_i were assumed to be iid with 0 mean and variance σ_e^2 . Finally, $\boldsymbol{\beta}$, σ_e^2 , and σ_η^2 were mutually independent and $\boldsymbol{\beta} \sim \text{Uniform}(R^p)$, $(\sigma_e^2)^{-1} \sim G(a/2, b/2)$, and $(\sigma_\eta^2)^{-1} \sim G(c/2, d/2)$. Throughout this chapter, a random variable Z is said to have a $G(\alpha, p)$ distribution if it has a pdf of

the form $f(z) \propto \exp(-\alpha z)z^{p-1}$. The joint posterior under the given prior is

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \mathbf{e}, r_\eta, r_e | \mathbf{y}) &\propto \prod_i p(y_i | \theta_i) \\ &\times r_e^{n/2} \exp \left\{ -\frac{r_e}{2} \sum_{i=1}^n (\theta_i - q_i - \mathbf{x}_i^T \boldsymbol{\beta} - \eta_i)^2 \right\} \\ &\times r_\eta^{n/2} \exp \left\{ -\frac{r_\eta}{2} \sum_{1 \leq i < l \leq n} w_{il} (\eta_i - \eta_l)^2 \right\} \\ &\times r_\eta^{(d/2)-1} \exp \left(-\frac{cr_\eta}{2} \right) r_e^{(b/2)-1} \exp \left(-\frac{ar_e}{2} \right), \end{aligned} \quad (4.2)$$

where $r_\eta = \sigma_\eta^{-2}$ and $r_e = \sigma_e^{-2}$.

Ghosh et al. (1999) provided sufficient conditions to ensure that the posterior is proper. The Bayesian analysis was implemented by the Markov chain Monte Carlo (MCMC) numerical integration technique. The full conditionals needed for such implementation are available in Ghosh et al. (1999).

4.3 Multivariate Hierarchical Model

4.3.1 Introduction

In this section, we propose four multivariate hierarchical Bayesian spatial models. Let $\mathbf{y}_i = (y_{i1}, \dots, y_{iq})^T$, $i = 1, \dots, n$ denote the n response vectors. For our specific example, the responses are the numbers of crashes at n regions due to q different causes. Analogous to the previous section, we begin with the one-parameter exponential family model

$$p(y_{ij} | \theta_{ij}) = \exp[\theta_{ij} y_{ij} - \psi(\theta_{ij})] h(y_{ij}), \quad (4.3)$$

$j = 1, \dots, q; i = 1, \dots, n$. In the next stage, we model the θ_{ij} as

$$\theta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \eta_{ij} + e_{ij} \quad (j = 1, \dots, q; i = 1, \dots, n), \quad (4.4)$$

where the \mathbf{x}_{ij} are p -component column vectors ($p < q$). Writing $\boldsymbol{\theta}_i = (\theta_{i1}, \dots, \theta_{iq})^T$, $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{iq})^T$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iq})$ and $\mathbf{e}_i = (e_{i1}, \dots, e_{iq})^T$, we can rewrite (4.4) as

$$\boldsymbol{\theta}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\eta}_i + \mathbf{e}_i, \quad i = 1, \dots, n. \quad (4.5)$$

In the above, the errors \mathbf{e}_i and the spatial effects $\boldsymbol{\eta}_i$ are assumed to be mutually independent. Throughout this chapter, we assume that $\mathbf{e}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_e)$ and $\text{rank}(\mathbf{X}_i) = p$. We will introduce various spatial priors for the $\boldsymbol{\eta}_i$ in the next four subsections. In particular, we will consider various CAR priors for the $\boldsymbol{\eta}_i$. We will label these priors as CAR priors I-IV.

4.3.2 CAR Prior I

We first consider the case when $\boldsymbol{\eta}_i = \eta_i \mathbf{1}_q$, $i = 1, \dots, n$. This amounts to the assumption that all the components of the spatial vector $\boldsymbol{\eta}_i$ in a given region are equal, i.e. the spatial influence is not cause-specific. For η_1, \dots, η_n , we consider the pairwise difference prior as given in (4.1). At the final stage of the hierarchical model, it is assumed that $\boldsymbol{\beta}$, r_η and $\boldsymbol{\Sigma}_e$ are mutually independent with $\boldsymbol{\beta} \sim \text{uniform}(R^p)$, $r_\eta \sim G(a/2, b/2)$, and $\boldsymbol{\Sigma}_e$ has an inverse Wishart distribution with pdf

$$\pi(\boldsymbol{\Sigma}_e) \propto |\boldsymbol{\Sigma}_e|^{-(\gamma+q+1)} \exp[-(1/2)\text{tr}(\boldsymbol{\Sigma}_e^{-1} \mathbf{A})].$$

This distribution will be written symbolically as $\text{IW}(\mathbf{A}, \gamma)$. Now writing

$$\mathbf{y} = (y_{11}, \dots, y_{1q}, \dots, y_{n1}, \dots, y_{nq})^T, \quad \boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T \quad \text{and} \quad \boldsymbol{\theta}^T = (\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_n^T),$$

the joint posterior is given by

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_e, r_\eta | \mathbf{y}) &\propto \prod_{i,j} p(y_{ij} | \theta_{ij}) \\ &\times |\boldsymbol{\Sigma}_e|^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\boldsymbol{\theta}_i - \mathbf{K}_i^1)^T \boldsymbol{\Sigma}_e^{-1} (\boldsymbol{\theta}_i - \mathbf{K}_i^1) \right\} \end{aligned}$$

$$\begin{aligned}
& \times r_\eta^{n/2} \exp \left\{ -\frac{r_\eta}{2} \sum_{1 \leq i < l \leq n} w_{il} (\eta_i - \eta_l)^2 \right\} \\
& \times |\Sigma_e|^{-(\gamma+q+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \Sigma_e^{-1} \mathbf{A} \right\} \\
& \times r_\eta^{(b/2)-1} \exp \left(-\frac{ar_\eta}{2} \right), \tag{4.6}
\end{aligned}$$

where $\mathbf{K}_i^{-1} = \eta_i \mathbf{1}_q + \mathbf{X}_i \boldsymbol{\beta}$. The prior for $\boldsymbol{\beta}$ is improper. We present a general theorem ensuring that the posterior is proper.

Theorem 4.3.1 *Assume $a > 0$, $n + b > 0$, and $n > p + q$. Then, if*

$$\int_{-\infty}^{\infty} \exp\{y_{ij}\theta - \psi(\theta)\} d\theta < \infty$$

for all y_{ij} , the joint posterior pdf of the θ_{ij} given \mathbf{y} is proper.

The proof of the theorem is deferred to Appendix A.

Direct evaluation of the posterior of the θ_{ij} given \mathbf{y} involves high-dimensional numerical integration and is not computationally feasible. Instead the Gibbs sampler is used requiring generation of samples from the full conditional distributions of the parameters. These conditionals are given by

$$r_\eta | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \Sigma_e, \mathbf{y} \sim \text{Gamma} \left(\frac{1}{2} \left(\sum_{1 \leq i < l \leq n} w_{il} (\eta_i - \eta_l)^2 + a \right), \frac{n+b}{2} \right);$$

$$\Sigma_e | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, r_\eta, \mathbf{y} \sim \text{IW} \left(\mathbf{A} + \sum_{i=1}^n (\boldsymbol{\theta}_i - \boldsymbol{\eta}_i - \mathbf{X}_i \boldsymbol{\beta})(\boldsymbol{\theta}_i - \boldsymbol{\eta}_i - \mathbf{X}_i \boldsymbol{\beta})^T, n + \gamma \right);$$

$$\boldsymbol{\beta} | \boldsymbol{\theta}, \boldsymbol{\eta}, \Sigma_e, r_\eta, \mathbf{y} \sim N_p(\boldsymbol{\mu}_\beta, \Sigma_\beta);$$

$$\eta_i | \boldsymbol{\theta}, \boldsymbol{\beta}, \eta_l (l \neq i), \boldsymbol{\Sigma}_e, r_\eta, \mathbf{y} \sim N \left(\frac{(\boldsymbol{\theta}_i - \mathbf{X}_i \boldsymbol{\beta})^T \boldsymbol{\Sigma}_e^{-1} \mathbf{1}_q + r_\eta w_{i+} \bar{\eta}_i}{\mathbf{1}_q^T \boldsymbol{\Sigma}_e^{-1} \mathbf{1}_q + r_\eta w_{i+}}, \frac{1}{\mathbf{1}_q^T \boldsymbol{\Sigma}_e^{-1} \mathbf{1}_q + r_\eta w_{i+}} \right);$$

$$\pi(\boldsymbol{\theta}_i | \boldsymbol{\theta}_l (l \neq i), \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_e, r_\eta, \mathbf{y}) \propto \prod_j p(y_{ij} | \theta_{ij}) \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta}_i - \mathbf{K}_i^1)^T \boldsymbol{\Sigma}_e^{-1} (\boldsymbol{\theta}_i - \mathbf{K}_i^1) \right\},$$

where $\boldsymbol{\mu}_\beta = (\sum_{i=1}^n \mathbf{X}_i^T \boldsymbol{\Sigma}_e^{-1} \mathbf{X}_i)^{-1} (\sum_{i=1}^n \mathbf{X}_i^T \boldsymbol{\Sigma}_e^{-1} (\boldsymbol{\theta}_i - \eta_i \mathbf{1}_q))$, $\boldsymbol{\Sigma}_\beta = (\sum_{i=1}^n \mathbf{X}_i^T \boldsymbol{\Sigma}_e^{-1} \mathbf{X}_i)^{-1}$, $w_{i+} = \sum_{l \neq i} w_{li}$ and $\bar{\eta}_i = \sum_{l \neq i} w_{li} \eta_l / w_{i+}$. The full conditionals for r_η , $\boldsymbol{\Sigma}_e$ and $\boldsymbol{\beta}$ are standard, and it is easy to generate samples from them. Also, the conditionals of the $\boldsymbol{\theta}_i$ are log-concave, so that one can use the adaptive rejection sampling (Gilks and Wild 1992) to generate samples from them.

4.3.3 CAR Prior II

The model considered in the previous subsection is based on the assumption that all the components of $\boldsymbol{\eta}_i$, the i th the spatial effect vector are the same ($i = 1, \dots, n$). In this subsection, we consider the situation when the vectors $(\eta_{1j}, \dots, \eta_{nj})$ ($j = 1, \dots, q$) are mutually independent, and $\eta_{1j}, \dots, \eta_{nj}$ have the joint prior

$$\pi(\eta_{1j}, \dots, \eta_{nj} | r_{\eta_j}) \propto r_{\eta_j}^{n/2} \exp \left\{ -\frac{r_{\eta_j}}{2} \sum_{1 \leq i < l \leq n} w_{il} (\eta_{ij} - \eta_{lj})^2 \right\}. \quad (4.7)$$

Also, we assign the same prior distributions for all the other parameters as in the previous subsection. Then the joint posterior is given by

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_e, r_\eta | \mathbf{y}) &\propto \prod_{i,j} p(y_{ij} | \theta_{ij}) \\ &\times |\boldsymbol{\Sigma}_e|^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\boldsymbol{\theta}_i - \mathbf{K}_i^2)^T \boldsymbol{\Sigma}_e^{-1} (\boldsymbol{\theta}_i - \mathbf{K}_i^2) \right\} \\ &\times \prod_{j=1}^q \left[r_{\eta_j}^{n/2} \exp \left\{ -\frac{r_{\eta_j}}{2} \sum_{1 \leq i < l \leq n} w_{il} (\eta_{ij} - \eta_{lj})^2 \right\} \right] \\ &\times |\boldsymbol{\Sigma}_e|^{-(\gamma+q+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_e^{-1} \mathbf{A} \right\} \end{aligned}$$

$$(4.8) \quad \times \prod_{j=1}^q r_{\eta_j}^{(b_j/2)-1} \exp\left(-\frac{a_j r_{\eta_j}}{2}\right),$$

where $\mathbf{K}_i^2 = \boldsymbol{\eta}_i + \mathbf{X}_i \boldsymbol{\beta}$ and $\mathbf{r}_\eta = (r_{\eta_1}, \dots, r_{\eta_q})^T$. The following theorem is provided to ensure that the posterior is proper under vague flat prior for $\boldsymbol{\beta}$.

Theorem 4.3.2 *Assume $a_j > 0$, $n + b_j > 0$, $j = 1, \dots, q$, and $n > p + q$. Then, if*

$$\int_{-\infty}^{\infty} \exp\{y_{ij}\theta - \psi(\theta)\} d\theta < \infty$$

for all y_{ij} , the joint posterior probability density function of the θ_{ij} given \mathbf{y} is proper.

The proof is provided in Appendix B.

The full conditionals required for Gibbs sampling are given by

$$r_{\eta_j} | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_e, \mathbf{y} \sim \text{Gamma}\left(\frac{1}{2} \left(\sum_{1 \leq i < l \leq n} w_{il} (\eta_{ij} - \eta_{lj})^2 + a_j \right), \frac{n + b_j}{2}\right);$$

$$\boldsymbol{\Sigma}_e | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, r_\eta, \mathbf{y} \sim \text{IW}\left(\mathbf{A} + \sum_{i=1}^n (\boldsymbol{\theta}_i - \boldsymbol{\eta}_i - \mathbf{X}_i \boldsymbol{\beta})(\boldsymbol{\theta}_i - \boldsymbol{\eta}_i - \mathbf{X}_i \boldsymbol{\beta})^T, n + \gamma\right);$$

$$\boldsymbol{\beta} | \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_e, r_\eta, \mathbf{y} \sim N_p(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta);$$

$$\boldsymbol{\eta}_i | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}_{l(\neq i)}, \boldsymbol{\Sigma}_e, r_\eta, \mathbf{y} \sim N(\boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_\eta);$$

$$\pi(\boldsymbol{\theta}_i | \boldsymbol{\theta}_{j(j \neq i)}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_e, r_\eta, \mathbf{y}) \propto \prod_j p(y_{ij} | \theta_{ij}) \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}_i - \mathbf{K}_i^2)^T \boldsymbol{\Sigma}_e^{-1} (\boldsymbol{\theta}_i - \mathbf{K}_i^2)\right\},$$

where $\boldsymbol{\mu}_\beta = (\sum_{i=1}^n \mathbf{X}_i^T \boldsymbol{\Sigma}_e^{-1} \mathbf{X}_i)^{-1} \sum_{i=1}^n \mathbf{X}_i^T \boldsymbol{\Sigma}_e^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\eta}_i)$, $\boldsymbol{\Sigma}_\beta = (\sum_{i=1}^n \mathbf{X}_i^T \boldsymbol{\Sigma}_e^{-1} \mathbf{X}_i)^{-1}$, $\boldsymbol{\mu}_\eta = (\boldsymbol{\Sigma}_e^{-1} + w_{i+} \mathbf{R})^{-1} (\boldsymbol{\Sigma}_e^{-1} (\boldsymbol{\theta}_i - \mathbf{X}_i \boldsymbol{\beta}) + \frac{\mathbf{R} w_{i+} \boldsymbol{\eta}_{i+}}{2})$, $\boldsymbol{\Sigma}_\eta = (\boldsymbol{\Sigma}_e^{-1} + R w_{i+})^{-1}$, and $\mathbf{R} = \text{Diag}(r_{\eta_1}, \dots, r_{\eta_q})$.

4.3.4 CAR Prior III

The first two spatial models do not induce correlation among the type of crashes directly. In this subsection, we consider correlated CAR(CCAR) priors for spatial random effects where the scale parameters, say, r_{η_j} vary across the different components $j = 1, \dots, q$. Also, we assume that the logarithms of the scale parameters have a joint multivariate normal distribution. Writing $\boldsymbol{\rho} = (\rho_1, \dots, \rho_q)^T = (\log r_{\eta_1}, \dots, \log r_{\eta_q})^T$, we assume that $\boldsymbol{\rho} \sim N_q(\mathbf{0}, \boldsymbol{\Sigma})$. Now the spatial models for different crash types are correlated through the scale parameter and we can measure the strength of the correlation as well. The other components of the model remain the same as in the previous subsection. We first prove the following theorem which provides sufficient conditions for the proper joint posterior.

Theorem 4.3.3 *Assume $n + \gamma > 0$. Then if*

$$\int_{-\infty}^{\infty} \exp\{y_{ij}\theta - \psi(\theta)\}d\theta < \infty$$

for all y_{ij} , the joint posterior probability density function of the θ_{ij} given \mathbf{y} is proper.

The proof is provided in Appendix C.

The full conditionals needed for Gibbs sampler are given by

$$\rho_j | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_e, \boldsymbol{\rho}_l (l \neq j), \mathbf{y} \propto \exp \left\{ \frac{\rho_j n - \exp(\rho_j) \sum_{1 \leq i < l \leq n} w_{il} (\eta_{ij} - \eta_{lj})^2 - \boldsymbol{\rho}^T \boldsymbol{\Sigma}_\eta^{-1} \boldsymbol{\rho}}{2} \right\};$$

$$\boldsymbol{\Sigma}_e | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\rho}, \mathbf{y} \sim \text{IW} \left(\mathbf{A} + \sum_{i=1}^n (\boldsymbol{\theta}_i - \boldsymbol{\eta}_i - \mathbf{X}_i \boldsymbol{\beta})(\boldsymbol{\theta}_i - \boldsymbol{\eta}_i - \mathbf{X}_i \boldsymbol{\beta})^T, n + \gamma \right);$$

$$\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_e, \boldsymbol{\rho}, \mathbf{y} \sim MN(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta);$$

$$\boldsymbol{\eta}_i|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}_{-i}, \boldsymbol{\Sigma}_e, \boldsymbol{\rho}, \mathbf{y} \sim N(\boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_\eta);$$

$$\pi(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{j(j \neq i)}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_e, \boldsymbol{\rho}, \mathbf{y}) \propto \prod_j p(y_{ij}|\theta_{ij}) \exp \left\{ -\frac{1}{2}(\boldsymbol{\theta}_i - \mathbf{K}_i^2)^T \boldsymbol{\Sigma}_e^{-1}(\boldsymbol{\theta}_i - \mathbf{K}_i^2) \right\},$$

where $\boldsymbol{\mu}_\beta$, $\boldsymbol{\Sigma}_\beta$, $\boldsymbol{\mu}_\eta$, and $\boldsymbol{\Sigma}_\eta$ are the same as in the previous subsection and $\mathbf{R} = \text{Diag}(\exp(\rho_1), \dots, \exp(\rho_q))$.

4.3.5 CAR Prior IV

In this subsection, we consider a different Bayesian version of a multivariate CAR model first introduced by Mardia (1988). Carlin and Banerjee (2003) considered a special case which is what we consider as well. Under this framework, conditional on \mathbf{V} , the spatial effect is given by $\mathbf{V}^{-1} = (\mathbf{D} - \alpha \mathbf{W}) \otimes \boldsymbol{\Lambda}$. Here \otimes is the Kronecker product, $\mathbf{D} = \text{Diag}(m_1, \dots, m_n)$, m_i being the number of neighbors for the i th region; \mathbf{W} is the adjacency matrix; $\boldsymbol{\Lambda}^{-1}$ describe the relative variability and covariance relationships between the different crashes given the neighboring sites; $\alpha \in (0, 1)$ is the propriety parameters for \mathbf{V} to repair the possible singularities in it. Thus, \mathbf{V}^{-1} may be looked upon as the Kronecker product of two partial precision matrices: $\mathbf{D} - \alpha \mathbf{W}$ for spatial components, and $\boldsymbol{\Lambda}$ for variation across crashes.

We assume a beta (c, d) prior for α and a Wishart (s, \mathbf{B}) prior for $\boldsymbol{\Lambda}$. Other prior specifications remain the same as in the previous section. Then the joint posterior is given by

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_e, \alpha, \boldsymbol{\Lambda}|\mathbf{y}) &\propto \prod_{i,j} p(y_{ij}|\theta_{ij}) \\ &\times |\boldsymbol{\Sigma}_e|^{-n/2} \exp \left[-\frac{1}{2} \sum_{i=1}^n (\boldsymbol{\theta}_i - \mathbf{K}_i^2)^T \boldsymbol{\Sigma}_e^{-1} (\boldsymbol{\theta}_i - \mathbf{K}_i^2) \right] \end{aligned}$$

$$\begin{aligned}
& \times |\mathbf{D} - \alpha \mathbf{W}|^{q/2} |\mathbf{\Lambda}|^{n/2} \exp \left(-\frac{1}{2} \boldsymbol{\eta}^T \mathbf{V}^{-1} \boldsymbol{\eta} \right) \alpha^{c-1} (1 - \alpha)^{d-1} \\
& \times |\boldsymbol{\Sigma}_e|^{-(\gamma+q+1)/2} \exp \left[-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_e^{-1} \mathbf{A}) \right] \\
& \times |\mathbf{\Lambda}|^{(s-q-1)/2} \exp \left[-\frac{1}{2} \text{tr}(\mathbf{\Lambda} \mathbf{B}) \right], \tag{4.9}
\end{aligned}$$

where $\mathbf{K}_i^2 = \boldsymbol{\eta}_i + \mathbf{X}_i \boldsymbol{\beta}$. The following theorem is proved to ensure that the posterior is proper.

Theorem 4.3.4 *Suppose $n + s > q$, $n + \gamma > 0$, and $\int p(y_{ij}|\boldsymbol{\theta})d\boldsymbol{\theta} < \infty$ for all (i, j) . Then the posterior is proper.*

The proof is provided in Appendix D.

For Gibbs sampling, the full conditionals are given by

$$\boldsymbol{\Sigma}_e | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \alpha, \mathbf{\Lambda}, \mathbf{y} \sim \text{IW} \left(\mathbf{A} + \sum_{i=1}^n (\boldsymbol{\theta}_i - \boldsymbol{\eta}_i - \mathbf{X}_i \boldsymbol{\beta})(\boldsymbol{\theta}_i - \boldsymbol{\eta}_i - \mathbf{X}_i \boldsymbol{\beta})^T, n + \gamma \right);$$

$$\boldsymbol{\beta} | \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_e, \alpha, \mathbf{\Lambda}, \mathbf{y} \sim \text{MN}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta);$$

$$\boldsymbol{\eta}_i | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}_{-i}, \boldsymbol{\Sigma}_e, \alpha, \mathbf{\Lambda}, \mathbf{y} \sim N(\boldsymbol{\mu}_\eta^*, \boldsymbol{\Sigma}_\eta^*);$$

$$\pi(\alpha | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_e, \mathbf{\Lambda}, \mathbf{y}) \propto |\mathbf{D} - \alpha \mathbf{W}|^{q/2} \alpha^{c-1} (1 - \alpha)^{d-1};$$

$$\mathbf{\Lambda} | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_e, \alpha, \mathbf{y} \sim \text{Wishat}(\mathbf{B}, n + s);$$

$$\pi(\boldsymbol{\theta}_i | \boldsymbol{\theta}_{j(j \neq i)}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_e, \mathbf{V}, \mathbf{y}) \propto \prod_j p(y_{ij} | \boldsymbol{\theta}_{ij}) \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta}_i - \mathbf{K}_i^2)^t \boldsymbol{\Sigma}_e^{-1} (\boldsymbol{\theta}_i - \mathbf{K}_i^2) \right\},$$

where $\boldsymbol{\mu}_\beta = (\sum_{i=1}^n \mathbf{X}_i^T \boldsymbol{\Sigma}_e^{-1} \mathbf{X}_i)^{-1} (\sum_{i=1}^n \mathbf{X}_i^T \boldsymbol{\Sigma}_e^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\eta}_i))$, $\boldsymbol{\Sigma}_\beta = (\sum_{i=1}^n \mathbf{X}_i^T \boldsymbol{\Sigma}_e^{-1} \mathbf{X}_i)^{-1}$, $\boldsymbol{\Sigma}_\eta^* = [\boldsymbol{\Sigma}_e^{-1} + (m_i - \alpha w_{ii}) \boldsymbol{\Lambda}]^{-1}$, and $\boldsymbol{\mu}_\eta^* = \boldsymbol{\Sigma}_\eta^* [\boldsymbol{\Sigma}_e^{-1} (\boldsymbol{\theta}_i - \mathbf{X}_i \boldsymbol{\beta}) + \frac{1}{2} \sum_{j(\neq i)} (\alpha w_{ij}) \boldsymbol{\Lambda} \eta_{ij}]$.

4.4 Data Analysis

The data for the illustration for proposed multivariate spatial models also comes from county-level vehicle crash records and roadway data in Texas. The TXDOT has maintained the traffic crash data by separating four types of crash based on a location in which a traffic crash occurs:

- Intersection crash: a traffic crash which occurs within the limits of an intersection.
- Intersection-related crash: a traffic crash which (1) occurs on an approach to or exit from an intersection and (2) result from an activity, behavior or control related to the movement of traffic units through the intersection.
- Driveway access crash: a traffic crash occurs a driveway access or involves a road vehicle entering or leaving another roadway by way of on a driveway access.
- Non-intersection crash: a traffic crash that is not intersection crash, intersection-related crash, and driveway access crash.

The same covariates in Chapter III are considered here:

- Wet: a surrogate variable intended to represent the percentage of time that the road surface is wet due to rain, snow, and so forth. Not having detailed weather data, we chose to use the proportion of KAB crashes that occurred under wet pavement conditions.

- Curve: a surrogate variable to capture spatial variations in the number of sharp horizontal curves in different counties. Since actual inventory of horizontal curves on the highway network is not available, we chose to use the proportion of KAB crashes that occurred on sharp horizontal curves in each county as a surrogate variable.
- Obj: a surrogate variable to represent degree of roadside hazards. The proportion of KAB crashes that ran off roads and hit fixed objects on the roadside is used as a surrogate variable due to similar reason to the first covariate.

Refer to Chapter III or Miaou et al. (2003) for more detail background and description of the data. The interaction terms between covariates are involved in the model. In addition, note that two of the urban counties and one rural county were removed from the analysis for having no rural two-lane roads with the level of traffic volume of interest, i.e., fewer than 2,000 vehicles per day on average.

Let Y_{ij} be the number of j th type of reported KAB crashes in county i , $i = 1, \dots, n (= 251)$, $j = 1, \dots, q (= 4)$. At the first level of hierarchy, conditional on mean μ_{ij} , Y_{ij} are assumed to be mutually independent and Poisson distributed as

$$Y_{ij} \sim \text{Poisson}(\mu_{ij}). \quad (4.10)$$

The mean of the Poisson is modeled

$$\mu_{ij} = \nu_{ij} \lambda_{ij} \quad (4.11)$$

where ν_{ij} is an offset (in million of vehicle-miles traveled, or MVMT) and λ_{ij} is the KAB crash rate. Since the rate has to be nonnegative, it is structured as

$$\theta_{ij} = \log(\mu_{ij}) = \log(\lambda_{ij}) + \log(\nu_{ij}) = \log(\nu_{ij}) + \mathbf{x}_i^T \boldsymbol{\beta}_j + \eta_{ij} + e_{ij}, \quad (4.12)$$

where \mathbf{x}_i is covariates, β_j is regression coefficient vector, η_{ij} is spatial random effect, and e_{ij} is exchangeable random effect. For simplicity of notation, we can rewrite the expression as

$$\boldsymbol{\theta}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\eta}_i + \mathbf{e}_i, \quad (4.13)$$

where $\mathbf{X}_i = \mathbf{I}_q \otimes \mathbf{x}_i^T$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_q^T)^T$, and \mathbf{x}_i^T is a $p \times 1$ row vector for county i . $\boldsymbol{\theta}_i$'s can be expressed as a $N(= n \times q) \times 1$ column matrix $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_n^T)^T$ and the model is given by

$$\boldsymbol{\theta} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\eta} + \mathbf{e}, \quad (4.14)$$

where $\mathbf{X}^T = (\mathbf{X}_1^T, \dots, \mathbf{X}_n^T)^T$, $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^T, \dots, \boldsymbol{\eta}_n^T)^T$, and $\mathbf{e} = (\mathbf{e}_1^T, \dots, \mathbf{e}_n^T)^T$.

Prior distributions of all parameters in the model are specified as those in previous section and four types of spatial priors for multivariate models are considered in this analysis. Posterior propriety for each proposed spatial prior is ensured through the theorems with the integrability of the likelihood. Let $\theta_{ij} = \log(\mu_{ij})$ and $\psi(\theta_{ij}) = \exp(\theta_{ij})$. Then, the integral in the theorems is replaced by

$$\int_0^\infty \xi_{ij}^{y_{ij}-1} \exp(-\xi_{ij}) d\xi_{ij} < \infty,$$

which hold when $y_{ij} = 1, 2, \dots$. Therefore, all proposed theorems hold for poisson models with additional requirement $y_{ij} = 1, 2, \dots$.

As mentioned earlier, posterior inference is carried out by MCMC and Gibb sampler is implemented for most of the parameters whose full conditionals are available in closed form. The rest of them are sampled using Metropolis-Hastings algorithm. It is only necessary to replace exponential family by poisson density in the full conditionals and note that sampling step for $\boldsymbol{\theta}$ is only depend on likelihood function. Hyperparameters which are satisfied with the conditions suggested in the theorems

Table 3: DIC and p_D Values for Various Multivariate Spatial Models: Model 1=Model with Same Spatial Effect. Model 2=Model with Independent CAR. Model 3=Model with Correlated CAR. Model 4=Model with Multivariate CAR and Different Choices of α .

	P_D	DIC
Model 1	516.8	1480.3
Model 2	481.1	1399.6
Model 3	455.0	1385.2
Model 4 (with fixed $\alpha=1$)	462.8	1391.0
Model 4 (with single, unknown α)	459.6755	1391.2
Model 4 (with multiple, unknown α)	457.5247	1385.9

are specified.

We have used all of our models to fit the data and made model comparison based on the DIC values have been presented in the Table 3. It is clear that correlated CAR and multivariate CAR with unknown α is performing well. We will present other results based on the correlated CAR model.

We plot the posterior distribution of the regression parameters corresponding to the covariates and their interactions for each of the responses in Figures 11, 12, 13 and 14. From Figure 11 it is clear that the covariates curve and obj has significant effect on intersection crashes. Also the intersection between wet-curve and curve-obj is significant. From Figures 12 and 13, we reach to the similar conclusion for intersection related crashes and driveway crashes. For the non-intersection crashes the main significant variables remain same except this time the covariate obj assign significant mass towards 0 from Figure 14. Altogether the covariate wet is not significant for all the responses but the other two covariates curve and obj and their interactions have significant effect.

We have also plotted the posterior distributions of the correlation of the spatial scale parameters ρ in Figure 15. All of the parameters have significant positive correlations which is expected. Higher correlation has been seen among intersec-

tion, intersection related and driveway crashes. All these responses have lower but significant positive correlation with non-intersection crash.

The predicted maps based on our model is presented in Figure 16. From the map it is clear that east Texas has higher crash risk than the west. By further investigation we found the high risk sites for each type of crashes are rural areas near to the big cities like Dallas, Austin, San Antonio, and Fort Worth. Limited by the rolling terrain in the eastern counties, roadways in rural area tend to have less driver-friendly characteristics with, e.g., more horizontal and vertical curves, restricted sight-distance, and less forgiving roadside development (e.g. tree closer to the travelway and steeper side-slopes). In addition, with more and larger urbanized areas in the east, rural roads tend to have higher roadside development scores, higher access density, and narrow lanes and/or shoulder (Fitzpatrick et al. 2002).

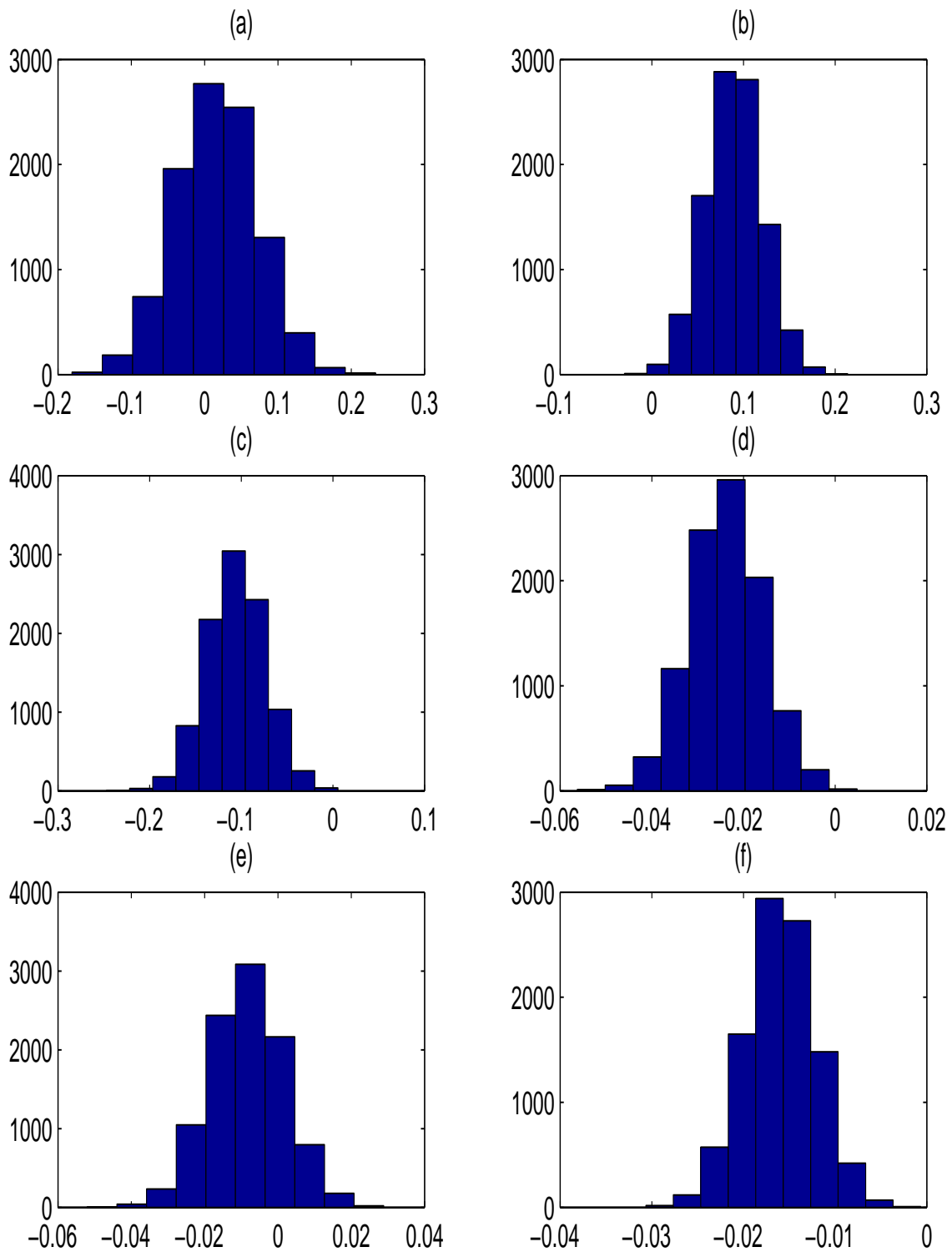


Figure 11: Plot of the Posterior Distributions of the Covariates for Interaction Crash. Regression Parameters Corresponding to (a) Wet, (b) Curve, (c) Obj, and the Interactions (d) Wet.Curve, (e) Wet.Obj, (f) Curve.Obj.

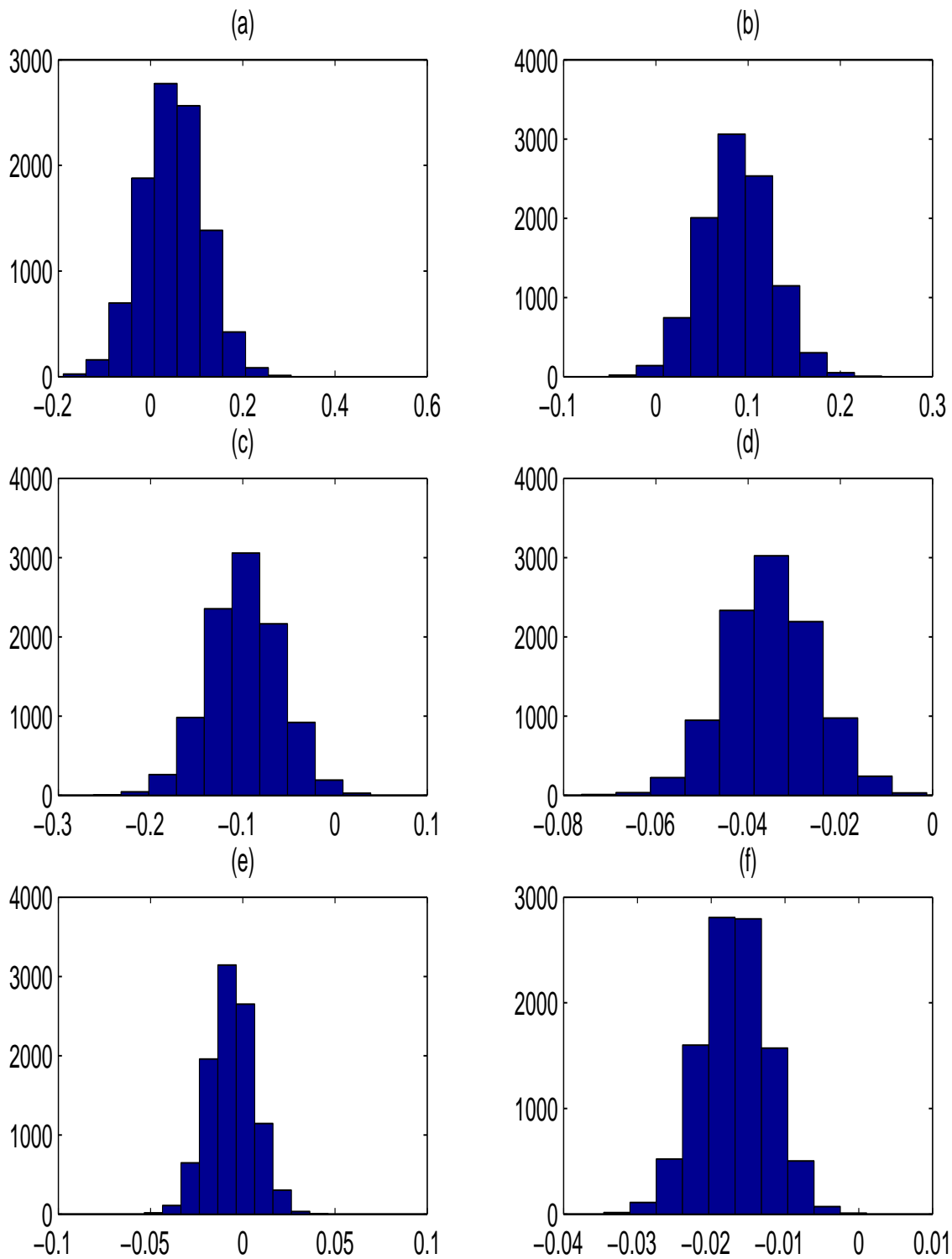


Figure 12: Plot of the Posterior Distributions of the Covariates for Interaction-Related Crash. Regression Parameters Corresponding to (a) Wet, (b) Curve, (c) Obj, and the Interactions (d) Wet.Curve, (e) Wet.Obj, (f) Curve.Obj.

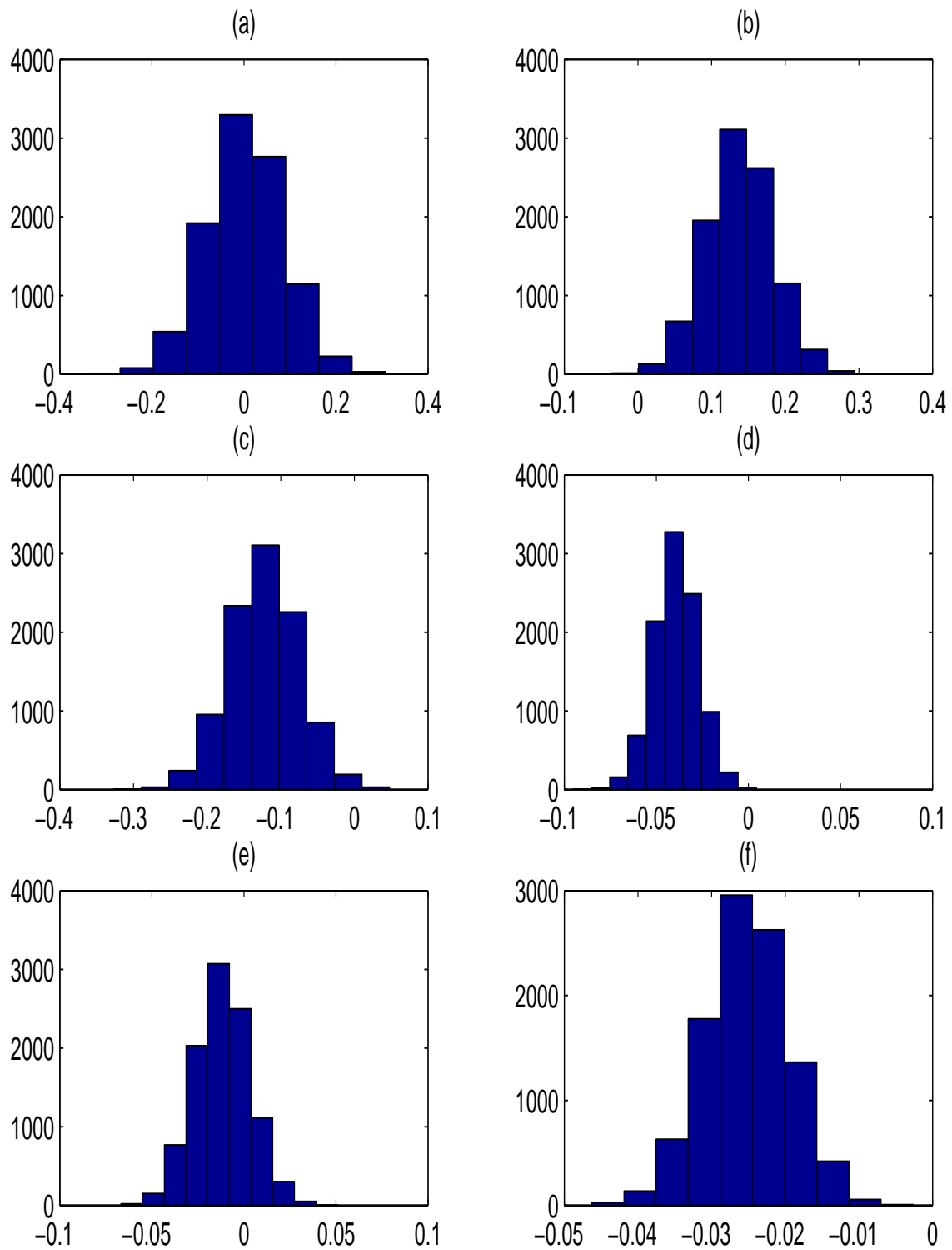


Figure 13: Plot of the Posterior Distributions of the Covariates for Driveway Crash. Regression Parameters Corresponding to (a) Wet, (b) Curve, (c) Obj, and the Interactions d) Wet.Curve, (e) Wet.Obj, (f) Curve.Obj.

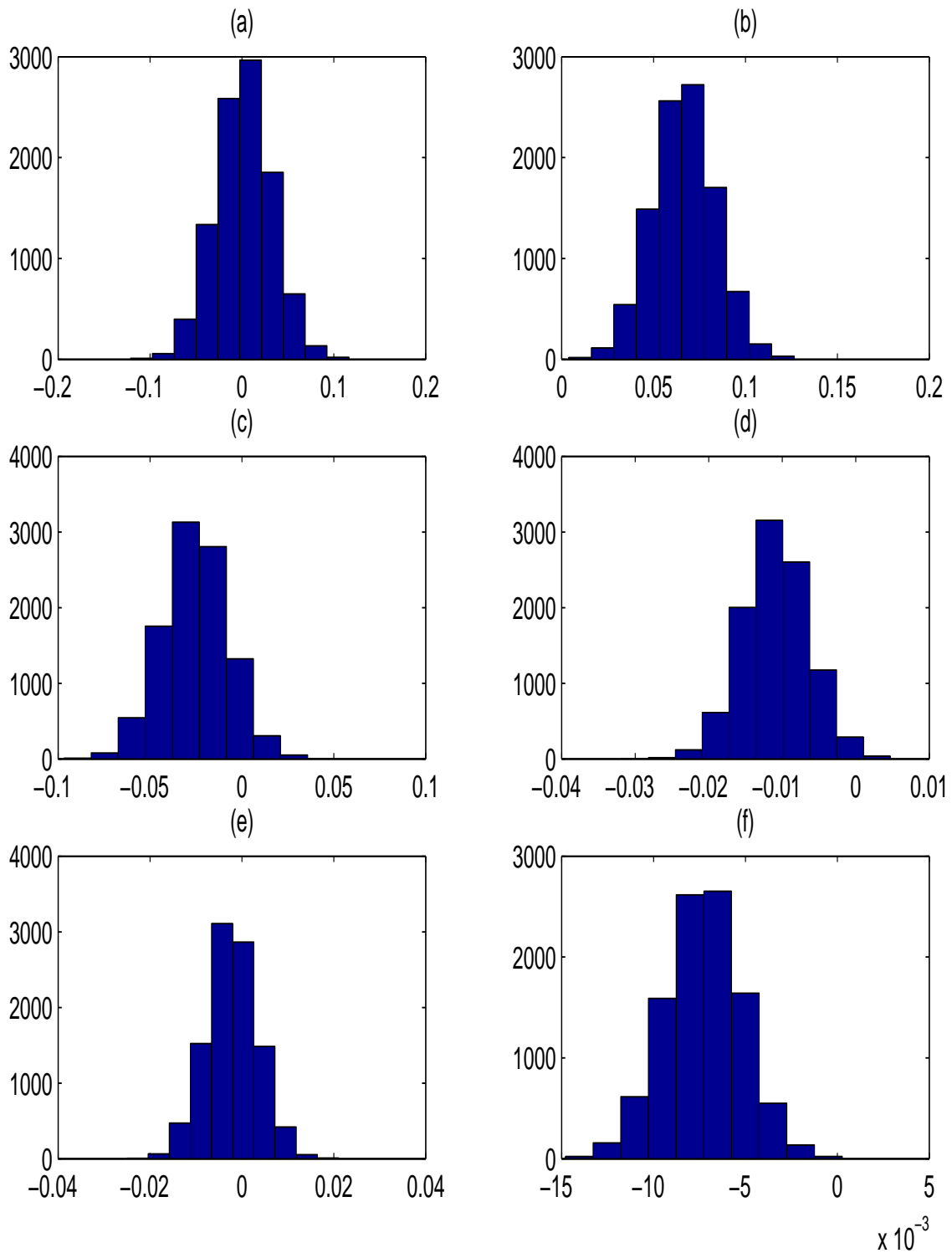


Figure 14: Plot of the Posterior Distributions of the Covariates for Non-Interaction Crash. Regression Parameters Corresponding to (a) Wet, (b) Curve, (c) Obj, and the Interactions d) Wet.Curve, (e) Wet.Obj, (f) Curve.Obj.

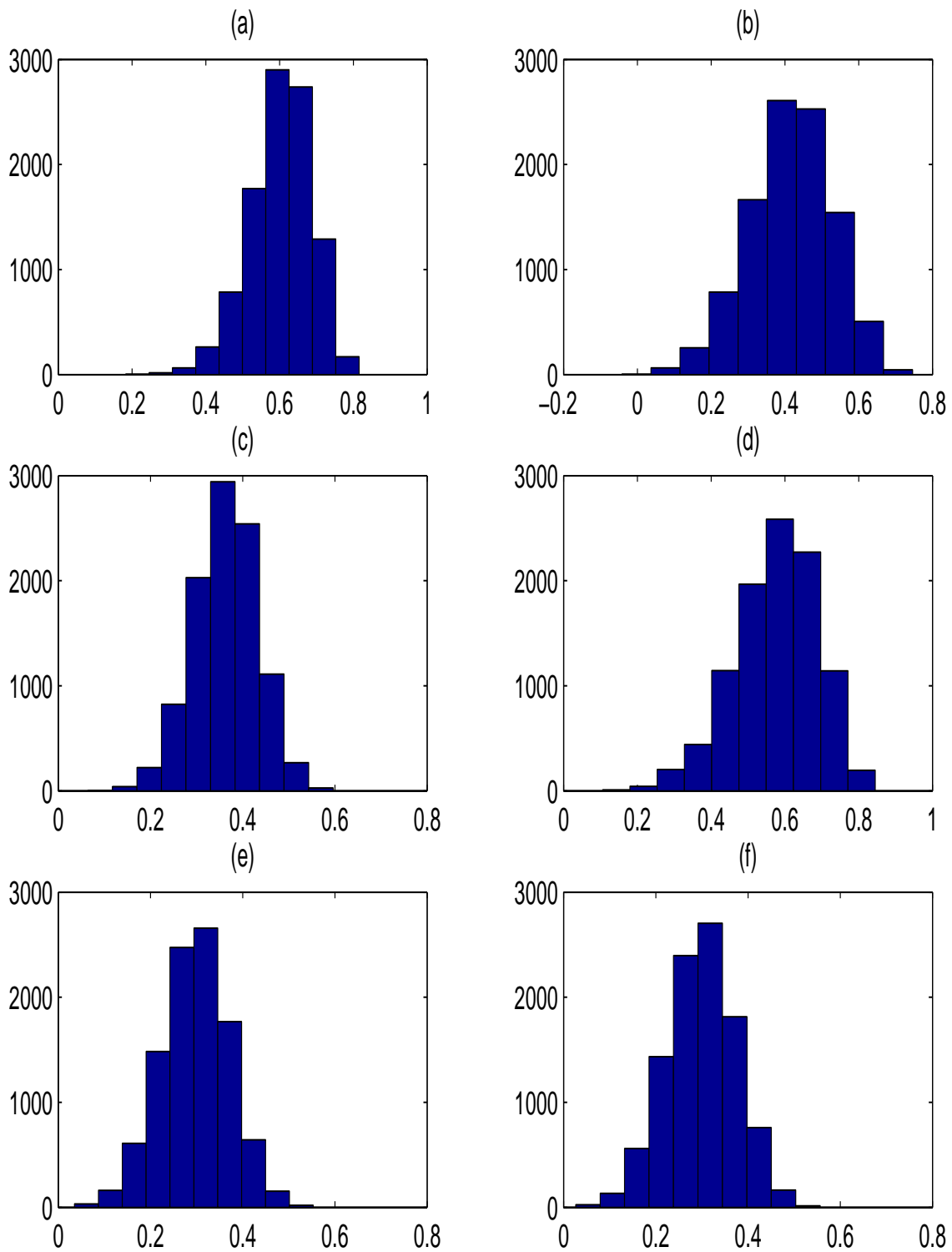


Figure 15: Plot of the Posterior Distributions of the Correlation Coefficients between the Responses. Correlation Coefficients Corresponding to (a) Intersection and Intersection-Related, (b) Intersection vs Driveway Access, (c) Intersection vs Non-Intersection, d) Intersection-Related vs Driveway Access, (e) Intersection-Related vs Non-Intersection, (f) Driveway Access vs Non-Intersection.

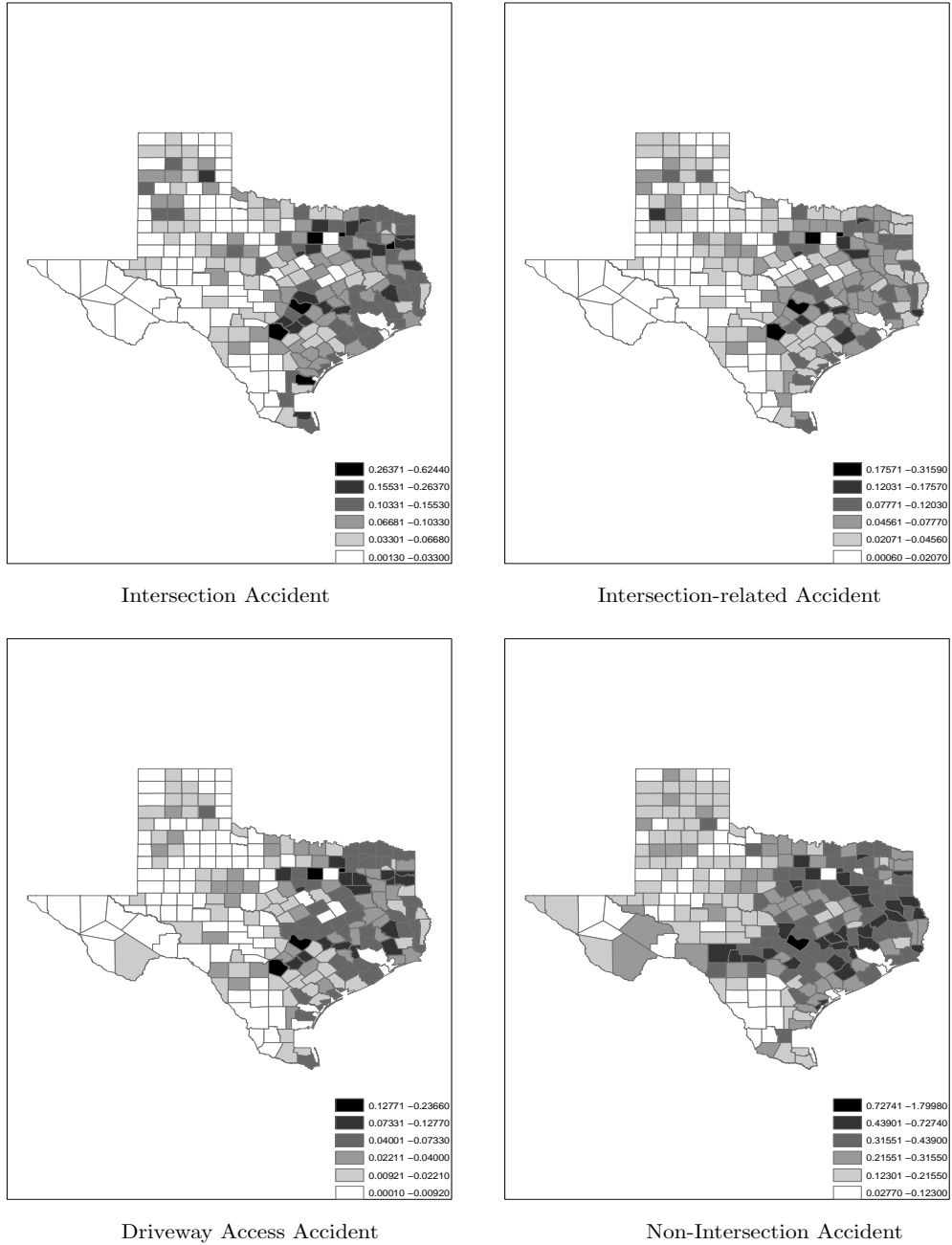


Figure 16: Predicted Map for Different Types of Crash.

CHAPTER V

CONCLUSIONS

We explored possible hierarchical spatial models in multivariate data. As preliminary study, Texas crash data is analyzed with univariate spatial models and it is also considered for the illustration of multivariate spatial models. The sufficient conditions to ensure posterior propriety using vague flat prior on regression parameter are obtained.

The best model in terms of DIC is suggested in univariate spatial model framework and estimated crash risk map is also shown with the selected model. It substantially supports that there is spatial pattern in Texas crash data and spatial effect is significant. The development of models and risk maps for traffic crash on road network is suggested as an extension of the study in discussion of Chapter III.

It is also an interesting topic to apply statistical ranking criteria to identify sites on a road network for further engineering inspection and safety improvement. A future study in transportation application can be to explore some of the issues raised regarding ranking methodology in light of the recent statistical development in spatio-temporal generalized linear mixed models.

We proposed the extension of univariate CAR model to multivariate setup. DIC is also selected for model comparison and it suggests that correlated CAR and multivariate CAR with unknown α outperform than the other models. Based on the estimated crash rate, crash risk maps are generated with four types of crashes.

Some studies on multivariate spatial models have been conducted, but posterior propriety is not considered. As indicated earlier, the propriety is always not guaranteed under improper priors and it should be considered as an important step

in MCMC. The proper joint posterior distribution corresponding to each proposed multivariate spatial prior is ensured by the theorem and the detail proofs are in Appendices.

In contrast to the models in Chapter III, only spatial random effects are included in the proposed models. It is potentially useful to add time effect and to investigate time trend in data when data are observed over certain time period.

REFERENCES

- Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: Chapman & Hall/CRC.
- Besag, J. (1974). “Spatial Interaction and the Statistical Analysis of Lattice Systems (with Discussion).” *Journal of the Royal Statistical Society, Ser. B*, 36, 192–236.
- (1975). “Statistical Analysis of Non-Lattice Data.” *Statistician*, 24, 179–195.
- Bonneson, J. and McCoy, P. (1996). “Effect of Median Treatment on Urban Arterial Safety: An Accident Prediction Model.” In *72nd Annual Meeting of the Transportation Research Board*. National Research Council, Washington, DC.
- Carlin, B. and Banerjee, S. (2003). “Hierarchical Multivariate CAR Models for Spatio-Temporally Correlated Survival Data.” in *Bayesian Statistics 7*, eds. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West. Oxford: Oxford University Press.
- Carlin, B. and Louis, T. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. London: Chapman and Hall/CRC.
- Chen, M.-H., Ibrahim, J., Shao, Q.-M., and Weiss, R. (2003). “Prior Elicitation for Model Selection and Estimation in Generalized Linear Mixed Models.” *Journal of Statistical Planning and Inference*, 111, 57–76.
- Chen, M.-H., Shao, Q.-M., and Xu, D. (2002). “Necessary and Sufficient Conditions on the Property of Posterior Distributions for Generalized Linear Mixed Models.” *Sankhya*, 64, 57–85.

- Christiansen, C., Morris, C., and Pendleton, O. (1992). “Hierarchical Poisson Model with Beta Adjustments for Traffic Accident Analysis.” Technical Report, Center for Statistical Sciences, University of Texas at Austin.
- Coster, J. (1987). “Literature Survey of Investigations Performed to Determine the Skid Resistance/Accident Relationship.” Technical Report RP/37, National Institute for Transport and Road Research, South Africa.
- Cressie, N. (1986). “Kriging Nonstationary Data.” *Journal of the American Statistical Association*, 81, 625–634.
- (1993). *Statistics for Spatial Data*. New York: Wiley.
- Dey, D., Ghosh, S., and Mallick, B. (2000). *Bayesian Generalized Linear Model*. New York: Marcel Dekker.
- Diggle, P. (1993). *Statistical Analysis of Spatial Point Patterns*. New York: Academic Press.
- Duncan, C., Khattak, A., and Council, F. (1998). “Applying the Ordered Probit Model to Injury Severity in Truck-Passenger Car Rear-End Collisions.” *Transportation Research Record*, 1635, 63–71.
- Fitzpatrick, K., Parham, A., Brewer, M., and Miaou, S. (2001). “Characteristics of and Potential Treatments for Crashes on Low-Volume, Rural Two-Lane Highways in Texas.” Technical Report, Texas Transportation Institute, College Station, TX.
- Flowers, R. and Griffin, L. (1992). “Development of a Plan for Identifying Highway Locations That May Be Over Represented in Accident Frequency and/or Severity.” Technical Report, Texas Transportation Institute, College Station, TX.

- Gelfand, A. and Vounatsou, P. (2003). “Proper Multivariate Conditional Autogressive Models for Spatial Data Analysis.” *Biostatistics*, 4, 11–25.
- Geman, S. and Geman, D. (1984). “Stochastic Relaxation Gibbs Distributions and the Bayesian Restoration of Image.” *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Getis, A. and Boots, B. (1978). *Models of Spatial Processes*. Cambridge: Cambridge University Press.
- Ghosh, M., Natarajan, K., Stroud, T., and Carlin, B. (1998). “Generalized Linear Models for Small Area Estimation.” *Journal of the American Statistical Association*, 96, 273–282.
- Ghosh, M., Natarajan, K., Waller, L., and Kim, D. (1999). “Hierarchical Bayes for the Analysis of Spatial Data: An Application to Disease Mapping.” *Journal of Statistical Planning and Inference*, 75, 305–318.
- Ghosh, M. and Rao, J. (1994). “Small Area Estimation: An Appraisal.” *Statistical Science*, 9, 55–76.
- Gilks, W., Richardson, S., and Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- Gilks, W. and Wild, P. (1992). “Adaptive Rejection Sampling for Gibbs Sampling.” *Applied Statistics*, 41, 337–348.
- Goldstein, H. (1999). *Multilevel Statistical Models, 1st Internet edition*. Available at <http://www.ioe.ac.uk/multilevel/>.

- Green, P. and Richardson, S. (2001). “Hidden Markov Models and Disease Mapping.” Working paper, Department of Mathematics, University of Bristol, United Kingdom.
- Griffin, L., Pendleton, O., and Morris, D. (1998). “An Evaluation of the Safety Consequences of Raising the Speed Limit on Texas Highways to 70 Miles per Hour.” Technical Report, Texas Transportation Institute, College Station, TX.
- Hauer, E. (1992). “Empirical Bayes Approach to the Estimation of Unsafety: The Multivariate Regression Approach.” *Accident Analysis & Prevention*, 24, 456–478.
- Henry, J. (2000). “Design and Testing of Pavement Friction Characteristics.” NCHRP Project 20-5, Synthesis of Highway Practice, Topic 30-11. Washington, DC: Transportation Research Board.
- Hobert, J. and Casella, G. (1996). “The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models.” *Journal of the American Statistical Association*, 91, 1461–1473.
- Ibrahim, J. and Laud, W. (1991). “On Bayesian Analysis of Generalized Linear Models Using Jeffrey’s Prior.” *Journal of the American Statistical Association*, 86, 981–986.
- Ivey, D. and Griffin, L. (1990). “Proposed Program to Reduce Skid Initiated Accidents in Texas.” Technical Report, Texas Transportation Institute, College Station, TX.
- Jin, X., Carlin, B., and Banerjee, S. (2004). “Generalized Hierarchical Multivariate CAR Models for Areal Data.” Research Report 04-001, University of Minnesota, Division of Biostatistics,.

- Journel, A. (1983). “Non-Parametric Estimation of Spatial Distributions.” *Journal of the International Association for Mathematical Geology*, 15, 445–468.
- Kim, H., Sun, D., and Tsutakawa, R. (2001). “A Bivariate Bayes Methods for Improving the Estimates of Mortality Rates with a Twofold Conditional Autoregressive Model.” *Journal of the American Statistical Association*, 96, 1506–1521.
- Kitanidis, P. (1983). “Statistical Estimation of Polynomial Generalized Covariance Functions and Hydrologic Applications.” *Water Resources Research*, 19, 909–921.
- Kitanidis, P. and Lane, R. (1985). “Maximum Likelihood Parameter Estimation of Hydrologic Spatial Processes by the Gauss-Newton Method.” *Journal of Hydrology*, 79, 53–71.
- Knorr-Held, L. and Besag, J. (1998). “Modelling Risk from a Disease in Time and Space.” *Statistics in Medicine*, 17, 2045–2060.
- Knorr-Held, L. and Best, N. G. (2000). “A Shared Component Model for Detecting Joint and Selective Clustering of Two Diseases.” *Journal of the Royal Statistical Society, Ser. A*, 164, 73–85.
- Laslett, G. (1994). “Kriging and Splines: An Empirical Comparison of Their Predictive Performance in Some Applications.” *Journal of the American Statistical Association*, 89, 391–409.
- Lawson, A. (2000). “Cluster Modelling of Disease Incidence via RJMCMC Methods: A Comparative Evaluation.” *Statistics in Medicine*, 19, 2361–2375.
- (2001). “Tutorial in Biostatistics: Disease Map Reconstruction.” *Statistics in Medicine*, 20, 2183–2204.

- Lawson, A., Biggeri, A., Bohning, D., Lessafre, E., Viel, J., and R., B. (1999). *Disease Mapping and Risk Assessment for Public Health*. Chichester, UK: Wiley.
- Maher, M. J. and Summersgill, L. (1996). “A Comprehensive Methodology for the Fitting of Predictive Accident Models.” *Accident Analysis & Prevention*, 28, 281–296.
- Mardia, K. and Marshall, R. (1984). “Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression.” *Biometrika*, 71, 135–146.
- Mardia, K. V. (1988). “Multi-Dimensional Multivariate Gaussian Markov Random Fields with Application to Image Processing.” *Journal of Multivariate Analysis*, 24, 265–284.
- Marquiss, M., Newton, I., and Radcliffe, D. (1978). “The Decline of the Raven, *Corvus corax*, in Relation to Afforestation in Southern Scotland and Northern England.” *Journal of Applied Ecology*, 15, 129–144.
- McGinnis, R., Wissinger, L., Kelly, R., and Acuna, C. (1998). “Estimating the Influence of Driver, Highway, and Environmental Factors on Run-off-Road Crashes Using Logistic Regression.” In *Transportation Research Board 78th Annual Meeting*. National Research Council, Washington, DC.
- Miaou, S.-P. (1994). “The Relationship Between Truck Accidents and Geometric Design of Road Sections: Poisson versus Negative Binomial Regressions.” *Accident Analysis & Prevention*, 26, 471–482.
- (1996). “Measuring the Goodness-of-Fit of Accident Prediction Models.” FHWA-RD-96-040, Federal Highway Administration, U.S. Department of Transportation.

- Miaou, S.-P., Hu, P., Wright, T., Rathi, A., and Davis, S. (1992). “Relationship Between Truck Accidents and Geometric Design: A Poisson Regression Approach.” *Transportation Research Record*, 1376, 10–18.
- Miaou, S.-P. and Lum, H. (1993). “Modeling Vehicle Accidents and Highway Geometric Design Relationships.” *Accident Analysis & Prevention*, 28, 689–709.
- Miaou, S.-P., Song, J., and Mallick, B. (2003). “Roadway Traffic Crash Mapping: A Space-Time Modeling Approach.” *Journal of Transportation and Statistics*, 6, 33–58.
- Morris, C., Christiansen, C., and Pendleton, O. (1991). “Application of New Accident Analysis Methodologies, Volume III: Theoretical Development.” FHWA-RD-91-015, Federal Highway Administration, U.S. Department of Transportation.
- National Safety Council (NSC) (2002). *Report on Injuries in America, 2001*. Available at <http://www.nsc.org/library/rept2000.htm>, as of July 2003.
- Pielou, E. (1959). “The Use of Point-to-Plant Distances in the Pattern of Plant Populations.” *Journal of Ecology*, 47, 607–613.
- (1977). *Mathematical Ecology*. New York: Wiley.
- Ripley, B. (1981). *Spatial Statistics*. New York: Wiley.
- Robert, C. and Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer-Verlag.
- Roberts, G. and Rosenthal, J. (1998). “Markov Chain Monte Carlo: Some Practical Implications of Theoretical Results.” *Canadian Journal of Statistics*, 26, 5–31.

- Sampson, P. and Guttorp, P. (1992). “Non-Parametric Estimation of Nonstationary Spatial Covariance Structure.” *Journal of the American Statistical Association*, 87, 108–119.
- Shankar, V., Milton, J., and Mannering, F. (1997). “Modeling Statewide Accident Frequencies as Zero-Inflated Probability Processes: An Empirical Inquiry.” *Accident Analysis & Prevention*, 29, 829–837.
- Smith, R., Harkey, D., and Harris, B. (2001). “Implementation of GIS-Based Highway Safety Analyses: Bridging the Gap.” FHWA-RD-01-039. Prepared for the Federal Highway Administration, U.S. Department of Transportation.
- Spiegelhalter, D., Best, N., and Carlin, B. (1998). “Bayesian Deviance, the Effective Number of Parameters, and the Comparison of Arbitrarily Complex Models.” Tech. 98-009. Division of Biostatistics, University of Minnesota.
- Spiegelhalter, D., Best, N., Carlin, B., and Linde, A. (2002). “Bayesian Measure of Model Complexity and Fit.” *Journal of the Royal Statistical Society, Ser. B*, 64, 1–34.
- Spiegelhalter, D., Thomas, A., and Best, N. (2000). *WinBUGS Version 1.3 User Manual*. Cambridge, UK: MRC Biostatistics Unit.
- Sun, D., Tsutakawa, R., and He, Z. (2001). “Propriety of Posteriors with Improper Priors in Hierarchical Linear Mixed Models.” *Statistica Sinica*, 11, 77–95.
- Sun, D., Tsutakawa, R., Kim, H., and He, Z. (2000). “Spatio-Temporal Interaction with Disease Mapping.” *Statistics in Medicine*, 19, 2015–2035.
- Transportation Research Board (TRB) (1987). “Designing Safer Roads: Practices

for Resurfacing, Restoration, and Rehabilitation.” Special Report 214, National Research Council, Washington, D.C.

Upton, G. and Fingleton, B. (1985). *Spatial Data Analysis by Example*, Vol. 1: Point Pattern and Quantitative Data. Chichester:Wiley.

U.S. Department of Transportation (USDOT) (2002). *Safety in Numbers Conferences*. Available at <http://www.bts.gov/sdi/conferences>, as of August 2003.

U.S. Department of Transportation (USDOT), Bureau of Transportation Statistics (BTS) (1996-1999). “Transportation Statistics Annual Report.” Washington, D.C.

Vogt, A. and Bared, J. (1998). “Accident Models for Two-Lane Rural Roads: Segments and Intersections.” FHWA-RD-98-133, Office of Safety and Traffic Operations R&D, Federal Highway Administration, U.S. Department of Transportation.

Whittle, O. (1954). “On Stationary Processes in the Plane.” *Biometrika*, 41, 434–449.

Xia, H., Carlin, B., and Waller, L. (1997). “Hierarchical Models for Mapping Ohio Lung Cancer Rates.” *Environmetrics*, 8, 107–120.

Yang, M., Rasbash, J., Goldstein, H., and Barbosa, M. (1999). *MLwiN Macros for Advanced Multilevel Modeling, Version 2.0: Multilevel Models Project*. Institute of Education, University of London.

Zhu, L. and Carlin, B. (1999). “Comparing Hierarchical Models for Spatio-Temporally Misaligned Data Using the DIC Criterion.” Research Report 99-006, University of Minnesota, Division of Biostatistics,.

APPENDIX A

PROOF OF THEOREM 1

Let $z_i = \eta_i - \eta_n$ ($i = 1, \dots, n-1$) and $z_n = 0$. Then, the transformed posterior is

$$\begin{aligned}
\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{z}, \eta_n, \boldsymbol{\Sigma}_e, r_\eta | \mathbf{y}) &\propto \prod_{i,j} p(y_{ij} | \theta_{ij}) \\
&\times \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{c}_i - \eta_n \mathbf{1}_q)^T \boldsymbol{\Sigma}_e^{-1} (\mathbf{c}_i - \eta_n \mathbf{1}_q) \right\} \\
&\times \exp \left\{ -\frac{r_\eta}{2} \sum_{1 \leq i < l \leq n} w_{il} (z_i - z_l)^2 \right\} r_\eta^{n/2} \\
&\times |\boldsymbol{\Sigma}_e|^{-(\gamma+n+q+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_e^{-1} \mathbf{A} \right\} \\
&\times r_\eta^{(b/2)-1} \exp \left(-\frac{ar_\eta}{2} \right),
\end{aligned}$$

where $\mathbf{z} = (z_1, \dots, z_{n-1})^T$ and $\mathbf{c}_i = \boldsymbol{\theta}_i - \mathbf{X}_i \boldsymbol{\beta} - z_i \mathbf{1}_q$. Writing $\bar{\mathbf{c}} = n^{-1} \sum_{i=1}^n \mathbf{c}_i$, one has

$$\begin{aligned}
\sum_{i=1}^n (\eta_n \mathbf{1}_q - \mathbf{c}_i)^T \boldsymbol{\Sigma}_e^{-1} (\eta_n \mathbf{1}_q - \mathbf{c}_i) &= n\eta_n^2 (\mathbf{1}_q^T \boldsymbol{\Sigma}_e^{-1} \mathbf{1}_q) - 2n\eta_n (\mathbf{1}_q^T \boldsymbol{\Sigma}_e^{-1} \bar{\mathbf{c}}) + \sum_{i=1}^n \mathbf{c}_i^T \boldsymbol{\Sigma}_e^{-1} \mathbf{c}_i \\
&= n(\mathbf{1}_q^T \boldsymbol{\Sigma}_e^{-1} \mathbf{1}_q) \left(\eta_n - \frac{\mathbf{1}_q^T \boldsymbol{\Sigma}_e^{-1} \bar{\mathbf{c}}}{\mathbf{1}_q^T \boldsymbol{\Sigma}_e^{-1} \mathbf{1}_q} \right)^2 + \sum_{i=1}^n \mathbf{c}_i^T \boldsymbol{\Sigma}_e^{-1} \mathbf{c}_i - \frac{n(\mathbf{1}_q^T \boldsymbol{\Sigma}_e^{-1} \bar{\mathbf{c}})^2}{\mathbf{1}_q^T \boldsymbol{\Sigma}_e^{-1} \mathbf{1}_q}
\end{aligned}$$

Now integrating with respect to η_n ,

$$\begin{aligned}
\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{z}, \boldsymbol{\Sigma}_e, r_\eta | \mathbf{y}) &\propto \prod_{i,j} p(y_{ij} | \theta_{ij}) \\
&\times (\mathbf{1}_q^T \boldsymbol{\Sigma}_e^{-1} \mathbf{1}_q)^{-1/2} \exp \left\{ -\frac{1}{2} \left(\sum_{i=1}^n \mathbf{c}_i^T \boldsymbol{\Sigma}_e^{-1} \mathbf{c}_i - \frac{n(\mathbf{1}_q^T \boldsymbol{\Sigma}_e^{-1} \bar{\mathbf{c}})^2}{\mathbf{1}_q^T \boldsymbol{\Sigma}_e^{-1} \mathbf{1}_q} \right) \right\}
\end{aligned}$$

$$\begin{aligned}
& \times \exp \left[-\frac{r_\eta}{2} \left\{ a + \sum_{1 \leq i < l \leq n} w_{il} (z_i - z_l)^2 \right\} \right] r_\eta^{((n+b)/2)-1} \\
& \times |\Sigma_e|^{-(\gamma+n+q+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \Sigma_e^{-1} \mathbf{A} \right\}.
\end{aligned}$$

Next, by the inequality

$$\begin{aligned}
\sum_{i=1}^n \mathbf{c}_i^T \Sigma_e^{-1} \mathbf{c}_i - \frac{n(\mathbf{1}_q^T \Sigma_e^{-1} \bar{\mathbf{c}})^2}{\mathbf{1}_q^T \Sigma_e^{-1} \mathbf{1}_q} &= \sum_{i=1}^n (\mathbf{c}_i - \bar{\mathbf{c}})^T \Sigma_e^{-1} (\mathbf{c}_i - \bar{\mathbf{c}}) + n \left[\mathbf{c}^T \Sigma_e^{-1} \bar{\mathbf{c}} - \frac{(\mathbf{1}_q^T \Sigma_e^{-1} \bar{\mathbf{c}})^2}{\mathbf{1}_q^T \Sigma_e^{-1} \mathbf{1}_q} \right] \\
&\geq \sum_{i=1}^n (\mathbf{c}_i - \bar{\mathbf{c}})^T \Sigma_e^{-1} (\mathbf{c}_i - \bar{\mathbf{c}}) \\
&= \sum_{i=1}^n [\mathbf{g}_i - (\mathbf{X}_i - \bar{\mathbf{X}}) \boldsymbol{\beta}]^T \Sigma_e^{-1} [\mathbf{g}_i - (\mathbf{X}_i - \bar{\mathbf{X}}) \boldsymbol{\beta}],
\end{aligned}$$

where $\mathbf{g}_i = \boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}} - (z_i - \bar{z}) \mathbf{1}_q$ ($i = 1, \dots, n$), one gets

$$\begin{aligned}
\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{z}, \Sigma_e, r_\eta | \mathbf{y}) &\leq K \prod_{i,j} p(y_{ij} | \theta_{ij}) \times (\mathbf{1}_q^T \Sigma_e^{-1} \mathbf{1}_q)^{-1/2} \\
&\times \exp \left[-\frac{1}{2} \sum_{i=1}^n \{ \mathbf{g}_i - (\mathbf{X}_i - \bar{\mathbf{X}}) \boldsymbol{\beta} \}^T \Sigma_e^{-1} \{ \mathbf{g}_i - (\mathbf{X}_i - \bar{\mathbf{X}}) \boldsymbol{\beta} \} \right] \\
&\times \exp \left[-\frac{r_\eta}{2} \left\{ a + \sum_{1 \leq i < l \leq n} w_{il} (z_i - z_l)^2 \right\} \right] r_\eta^{((n+b)/2)-1} \\
&\times |\Sigma_e|^{-(\gamma+n+q+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \Sigma_e^{-1} \mathbf{A} \right\},
\end{aligned}$$

where in the above and in what follows, $K(> 0)$ is a generic constant.

Next integrating with respect to $\boldsymbol{\beta}$,

$$\pi(\boldsymbol{\theta}, \mathbf{z}, \Sigma_e, r_\eta | \mathbf{y}) \leq K \prod_{i,j} p(y_{ij} | \theta_{ij}) (\mathbf{1}_q^T \Sigma_e^{-1} \mathbf{1}_q)^{-1/2} |\Sigma_\beta|^{1/2}$$

$$\begin{aligned}
& \times \exp \left[-\frac{1}{2} \sum_{i=1}^n \{ \mathbf{g}_i - (\mathbf{X}_i - \bar{\mathbf{X}}) \boldsymbol{\beta}_* \}^T \boldsymbol{\Sigma}_e^{-1} \{ \mathbf{g}_i - (\mathbf{X}_i - \bar{\mathbf{X}}) \boldsymbol{\beta}_* \} \right] \\
& \times \exp \left[-\frac{r_\eta}{2} \left\{ a + \sum_{1 \leq i < l \leq n} w_{il} (z_i - z_l)^2 \right\} \right] r_\eta^{((n+b)/2)-1} \\
& \times |\boldsymbol{\Sigma}_e|^{-(\gamma+n+q+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_e^{-1} \mathbf{A} \right\},
\end{aligned}$$

where $\boldsymbol{\Sigma}_\beta^{-1} = \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^T \boldsymbol{\Sigma}_e^{-1} (\mathbf{X}_i - \bar{\mathbf{X}})$ and $\boldsymbol{\beta}_* = \boldsymbol{\Sigma}_\beta^{-1} [\sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^T \boldsymbol{\Sigma}_e^{-1} \mathbf{g}_i]$.

The above is bounded above by

$$\begin{aligned}
\pi(\boldsymbol{\theta}, \mathbf{z}, \boldsymbol{\Sigma}_e, r_\eta | \mathbf{y}) & \leq K \prod_{i,j} p(y_{ij} | \theta_{ij}) (\mathbf{1}_q^T \boldsymbol{\Sigma}_e^{-1} \mathbf{1}_q)^{-1/2} |\boldsymbol{\Sigma}_\beta|^{1/2} \\
& \times \exp \left[-\frac{r_\eta}{2} \left\{ a + \sum_{1 \leq i < l \leq n} w_{il} (z_i - z_l)^2 \right\} \right] r_\eta^{((n+b)/2)-1} \\
& \times |\boldsymbol{\Sigma}_e|^{-(\gamma+n+q+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_e^{-1} \mathbf{A} \right\},
\end{aligned}$$

Next observe that $\sum \sum_{1 \leq i < l \leq n} w_{il} (z_i - z_l)^2 = \mathbf{z}^T \mathbf{W} \mathbf{z}$, where

$$\mathbf{W} = \begin{bmatrix} \sum_{l=1}^n w_{1,l} & -w_{1,2} & \cdots & -w_{1,n-1} \\ -w_{2,1} & \sum_{l=1}^n w_{2,l} & \cdots & -w_{2,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ -w_{n-1,1} & w_{n-1,2} & \cdots & \sum_{l=1}^n w_{n-1,l} \end{bmatrix}.$$

Hence, integrating first with respect to \mathbf{z} and then with respect to r_η , one gets

$$\begin{aligned}
\pi(\boldsymbol{\theta}, \boldsymbol{\Sigma}_e | \mathbf{y}) & \leq K \prod_{i,j} p(y_{ij} | \theta_{ij}) (\mathbf{1}_q^T \boldsymbol{\Sigma}_e^{-1} \mathbf{1}_q)^{-1/2} |\boldsymbol{\Sigma}_\beta|^{1/2} \\
& \times |\boldsymbol{\Sigma}_e|^{-(\gamma+n+q+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_e^{-1} \mathbf{A} \right\}.
\end{aligned}$$

Let $\zeta_1 \leq \zeta_2 \leq \dots \leq \zeta_q$ denote the eigenvalues of Σ_e . Then, ζ_q^{-1} is the smallest eigenvalue of Σ_e^{-1} .

Now by the inequalities $(\mathbf{1}_q^T \Sigma_e^{-1} \mathbf{1}_q)^{-1/2} \leq \zeta_q^{1/2} q^{-1/2}$ and $|\Sigma_\beta|^{1/2} = |\sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^T \Sigma_e^{-1} (\mathbf{X}_i - \bar{\mathbf{X}})|^{-1/2} \leq \zeta_q^{p/2} |\sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^T (\mathbf{X}_i - \bar{\mathbf{X}})|^{-1/2}$,

$$\begin{aligned}
(\mathbf{1}_q^T \Sigma_e^{-1} \mathbf{1}_q)^{-1/2} |\Sigma_\beta|^{1/2} |\Sigma_e|^{-(\gamma+n+q+1)/2} &\leq K \zeta_q^{(p+1)/2} \left(\prod_{j=1}^q \zeta_j \right)^{-(n+q+\gamma+1)/2} \\
&= K \left(\prod_{j=1}^{q-1} \zeta_j^{-1} \right)^{(n+q+\gamma+1)/2} (\zeta_q^{-1})^{(n+q-p)/2} \\
&= K \left(\prod_{j=1}^q \zeta_j^{-1} \right)^{(n+q-p)/2} \left(\prod_{j=1}^{q-1} \zeta_j^{-1} \right)^{(p+\gamma+1)/2} \\
&= K |\Sigma_e|^{-(n+q-p)/2} \left(\frac{1}{q-1} \sum_{j=1}^{q-1} \zeta_j^{-1} \right)^{(q-1)(p+\gamma+1)/2} \\
&\leq K |\Sigma_e|^{-(n+q-p)/2} \left(\sum_{j=1}^q \zeta_j^{-1} \right)^{(q-1)(p+\gamma+1)/2} \\
&= K |\Sigma_e|^{-(n+q-p)/2} [\text{tr}(\Sigma_e^{-1})]^{(q-1)(p+\gamma+1)/2}.
\end{aligned}$$

Now, by the fact

$$\int [\text{tr}(\Sigma_e^{-1})]^{(q-1)(p+\gamma+1)/2} |\Sigma_e|^{-(n+q-p)/2} \exp \left[-\frac{1}{2} \text{tr}(\Sigma_e^{-1} \mathbf{A}) \right] d\Sigma_e < \infty,$$

one gets

$$\pi(\boldsymbol{\theta} | \mathbf{y}) \leq K \prod_{i,j} p(y_{ij} | \theta_{ij}).$$

The propriety of the posterior now follow from the assumption that $\int p(y_{ij} | \theta) d\theta < \infty$ for all i and j .

APPENDIX B

PROOF OF THEOREM 2

By the transformation $\mathbf{z}_i = \boldsymbol{\eta}_i - \boldsymbol{\eta}_n$ ($i = 1, \dots, n-1$) and $\mathbf{z}_n = \mathbf{0}$, writing $\mathbf{z}^T = (\mathbf{z}_1^T, \dots, \mathbf{z}_{n-1}^T)$, the joint posterior is given by

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{z}, \boldsymbol{\eta}_n, \boldsymbol{\Sigma}_e, \mathbf{r}_\eta | \mathbf{y}) &\propto \prod_{i,j} p(y_{ij} | \theta_{ij}) \\ &\times |\boldsymbol{\Sigma}_e|^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{c}_i - \boldsymbol{\eta}_n)^T \boldsymbol{\Sigma}_e^{-1} (\mathbf{c}_i - \boldsymbol{\eta}_n) \right\} \\ &\times |\boldsymbol{\Sigma}_e|^{-(\gamma+q+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_e^{-1} A \right\} \\ &\times \prod_{j=1}^q r_{\eta_j}^{n/2} \exp \left\{ -\frac{r_{\eta_j}}{2} \sum_{1 \leq i < l \leq n} w_{il} (z_{ij} - z_{lj})^2 \right\} \\ &\times \prod_{j=1}^q r_{\eta_j}^{(b_j/2)-1} \exp \left(-\frac{a_j r_{\eta_j}}{2} \right), \end{aligned}$$

where $\mathbf{c}_i = \boldsymbol{\theta}_i - \mathbf{z}_i - \mathbf{X}_i \boldsymbol{\beta}$. Let $\bar{\mathbf{c}} = n^{-1} \sum_{i=1}^n \mathbf{c}_i$. Then integrating with respect to $\boldsymbol{\eta}_n$,

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{z}, \boldsymbol{\Sigma}_e, \mathbf{r}_\eta | \mathbf{y}) &\propto \prod_{i,j} p(y_{ij} | \theta_{ij}) \\ &\times |\boldsymbol{\Sigma}_e|^{-(n-1)/2} \exp \left[-\frac{1}{2} \sum_{i=1}^n (\mathbf{c}_i - \bar{\mathbf{c}})^T \boldsymbol{\Sigma}_e^{-1} (\mathbf{c}_i - \bar{\mathbf{c}}) \right] \\ &\times \prod_{j=1}^q r_{\eta_j}^{n/2} \exp \left\{ -\frac{r_{\eta_j}}{2} \sum_{1 \leq i < l \leq n} w_{il} (z_{ij} - z_{lj})^2 \right\} \\ &\times |\boldsymbol{\Sigma}_e|^{-(\gamma+q+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_e^{-1} A \right\} \\ &\times \prod_{j=1}^q r_{\eta_j}^{(b_j/2)-1} \exp \left(-\frac{a_j r_{\eta_j}}{2} \right). \end{aligned}$$

Now writing $\mathbf{c}_i - \bar{\mathbf{c}} = \mathbf{g}_i - (\mathbf{X}_i - \bar{\mathbf{X}})\boldsymbol{\beta}$, where $\mathbf{g}_i = \boldsymbol{\theta}_i - \mathbf{z}_i$ and $\bar{\mathbf{X}} = n^{-1} \sum_{i=1}^n \mathbf{X}_i$, integration with respect to $\boldsymbol{\beta}$ yields

$$\begin{aligned} \pi(\boldsymbol{\theta}, \mathbf{z}, \boldsymbol{\Sigma}_e, \mathbf{r}_\eta | \mathbf{y}) &\propto \prod_{i,j} p(y_{ij} | \theta_{ij}) \times |\boldsymbol{\Sigma}_\beta|^{1/2} \\ &\times \exp \left[-\frac{1}{2} \sum_{i=1}^n (\mathbf{g}_i - (\mathbf{X}_i - \bar{\mathbf{X}})\boldsymbol{\beta}_*)^T \boldsymbol{\Sigma}_\beta^{-1} (\mathbf{g}_i - (\mathbf{X}_i - \bar{\mathbf{X}})\boldsymbol{\beta}_*) \right] \\ &\times |\boldsymbol{\Sigma}_e|^{-(\gamma+n+q-p)/2} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_e^{-1} \mathbf{A} \right\} \\ &\times \prod_{j=1}^q \left[r_{\eta_j}^{((n+b_j)/2)-1} \exp \left\{ -\frac{r_{\eta_j}}{2} \left(a_j + \sum_{1 \leq i < l \leq n} w_{il} (z_{ij} - z_{lj})^2 \right) \right\} \right], \end{aligned}$$

where as before $\boldsymbol{\Sigma}_\beta^{-1} = \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^T \boldsymbol{\Sigma}_e^{-1} (\mathbf{X}_i - \bar{\mathbf{X}})$ and $\boldsymbol{\beta}_* = \boldsymbol{\Sigma}_\beta^{-1} [\sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^T \boldsymbol{\Sigma}_e^{-1} \mathbf{g}_i]$. Now, writing $K (> 0)$ once again for a generic constant,

$$\begin{aligned} \pi(\boldsymbol{\theta}, \mathbf{z}, \boldsymbol{\Sigma}_e, \mathbf{r}_\eta | \mathbf{y}) &\leq K \prod_{i,j} p(y_{ij} | \theta_{ij}) |\boldsymbol{\Sigma}_\beta|^{1/2} |\boldsymbol{\Sigma}_e|^{-(\gamma+n+q)/2} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_e^{-1} \mathbf{A} \right\} \\ &\times \prod_{j=1}^q \left[r_{\eta_j}^{((n+b_j)/2)-1} \exp \left\{ -\frac{r_{\eta_j}}{2} \left(a_j + \sum_{1 \leq i < l \leq n} w_{il} (z_{ij} - z_{lj})^2 \right) \right\} \right]. \end{aligned}$$

Next, integrating first with respect to \mathbf{z} and then with respect to \mathbf{r}_η , one gets

$$\pi(\boldsymbol{\theta}, \boldsymbol{\Sigma}_e, | \mathbf{y}) \leq K \prod_{i,j} p(y_{ij} | \theta_{ij}) |\boldsymbol{\Sigma}_\beta|^{1/2} |\boldsymbol{\Sigma}_e|^{-(\gamma+n+q)/2} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_e^{-1} \mathbf{A} \right\}$$

Arguing as in the previous section, we get

$$\pi(\boldsymbol{\theta}, \boldsymbol{\Sigma}_e, | \mathbf{y}) \leq K \prod_{i,j} p(y_{ij} | \theta_{ij}) [\text{tr}(\boldsymbol{\Sigma}_e^{-1})]^{(q-1)(p+\gamma)/2} |\boldsymbol{\Sigma}_e|^{-(n+q-p)/2} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_e^{-1} \mathbf{A} \right\}$$

This leads to

$$\pi(\boldsymbol{\theta} | \mathbf{y}) \leq K \prod_{i,j} p(y_{ij} | \theta_{ij}),$$

after integration with respect to $\boldsymbol{\Sigma}_e$. The result follows now from the condition of the theorem.

APPENDIX C

PROOF OF THEOREM 3

The joint posterior is given by

$$\begin{aligned}
\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_e, \boldsymbol{\rho}, \mathbf{y}) &\propto \prod_{i,j} p(y_{ij}|\theta_{ij}) \\
&\times |\boldsymbol{\Sigma}_e|^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\boldsymbol{\theta}_i - \mathbf{K}_i^2)^T \boldsymbol{\Sigma}_e^{-1} (\boldsymbol{\theta}_i - \mathbf{K}_i^2) \right\} \\
&\times \prod_{j=1}^q \left[\exp(\rho_j n/2) \exp \left\{ -\frac{\exp(\rho_j)}{2} \sum_{1 \leq i < l \leq n} w_{il} (\eta_{ij} - \eta_{lj})^2 \right\} \right] \\
&\times |\boldsymbol{\Sigma}_e|^{-(\gamma+q+1)/2} \exp \left\{ -\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_e^{-1} \mathbf{A}) \right\} \\
&\times \exp \left\{ -\frac{1}{2} \boldsymbol{\rho}^T \boldsymbol{\Sigma}_\eta^{-1} \boldsymbol{\rho} \right\},
\end{aligned}$$

where $\mathbf{K}_i^2 = \boldsymbol{\eta}_i + \mathbf{X}_i \boldsymbol{\beta}$. As in the previous section, writing $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^T = \boldsymbol{\eta}_i - \boldsymbol{\eta}_n$ ($i = 1, \dots, n-1$), $\mathbf{z}^T = (z_1^T, \dots, z_{n-1}^T)$, $\mathbf{g}_i = (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}) - (\mathbf{z}_i - \bar{\mathbf{z}})$, integration with respect to $\boldsymbol{\eta}_n$ yields

$$\begin{aligned}
\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{z}, \boldsymbol{\Sigma}_e, \boldsymbol{\rho}, \mathbf{y}) &\propto \prod_{i,j} p(y_{ij}|\theta_{ij}) \times |\boldsymbol{\Sigma}_e|^{-(n-1)/2} \\
&\times \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{g}_i - (\mathbf{X}_i - \bar{\mathbf{X}}) \boldsymbol{\beta})^T \boldsymbol{\Sigma}_e^{-1} (\mathbf{g}_i - (\mathbf{X}_i - \bar{\mathbf{X}}) \boldsymbol{\beta}) \right\} \\
&\times \prod_{j=1}^q \left[\exp(\rho_j n/2) \exp \left\{ -\frac{\exp(\rho_j)}{2} \sum_{1 \leq i < l \leq n} w_{il} (z_{ij} - z_{lj})^2 \right\} \right] \\
&\times |\boldsymbol{\Sigma}_e|^{-(\gamma+q+1)/2} \exp \left\{ -\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_e^{-1} \mathbf{A}) \right\} \\
&\times \exp \left\{ -\frac{1}{2} \boldsymbol{\rho}^T \boldsymbol{\Sigma}_\eta^{-1} \boldsymbol{\rho} \right\}.
\end{aligned}$$

Next integrating with respect to $\boldsymbol{\beta}$ and writing $K(> 0)$ for a generic constant, one gets

$$\begin{aligned} \pi(\boldsymbol{\theta}, \mathbf{z}, \boldsymbol{\Sigma}_e, \boldsymbol{\rho}, \mathbf{y}) &\leq K \prod_{i,j} p(y_{ij}|\theta_{ij}) |\boldsymbol{\Sigma}_\beta|^{1/2} \\ &\times \prod_{j=1}^q \left[\exp(\rho_j n/2) \exp \left\{ -\frac{\exp(\rho_j)}{2} \sum_{1 \leq i < l \leq n} w_{il} (z_{ij} - z_{lj})^2 \right\} \right] \\ &\times |\boldsymbol{\Sigma}_e|^{-(\gamma+n+q)/2} \exp \left\{ -\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_e^{-1} \mathbf{A}) \right\} \\ &\times \left. -\frac{1}{2} \boldsymbol{\rho}^T \boldsymbol{\Sigma}_\eta^{-1} \boldsymbol{\rho} \right\}, \end{aligned}$$

where as before $\boldsymbol{\Sigma}_\beta^{-1} = \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^T \boldsymbol{\Sigma}_e^{-1} (\mathbf{X}_i - \bar{\mathbf{X}})$.

Now integrating with respect to \mathbf{z} , one gets

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\Sigma}_e, \boldsymbol{\rho}, \mathbf{y}) &\leq K \prod_{i,j} p(y_{ij}|\theta_{ij}) |\boldsymbol{\Sigma}_\beta|^{1/2} \prod_{j=1}^q \exp((n-1)\rho_j/2) \\ &\times \exp(-\boldsymbol{\rho}^T \boldsymbol{\Sigma}_\eta^{-1} \boldsymbol{\rho}/2) |\boldsymbol{\Sigma}_e|^{-(\gamma+n+q)/2} \exp \left\{ -\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_e^{-1} \mathbf{A}) \right\} \end{aligned}$$

Next integrating with respect to $\boldsymbol{\rho}$, and using the finiteness of the mgf of a multivariate normal distribution, one gets

$$\pi(\boldsymbol{\theta}, \boldsymbol{\Sigma}_e, \mathbf{y}) \leq K \prod_{i,j} p(y_{ij}|\theta_{ij}) |\boldsymbol{\Sigma}_\beta|^{1/2} |\boldsymbol{\Sigma}_e|^{-(\gamma+n+q)/2} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_e^{-1} \mathbf{A} \right\}$$

The rest of the proof is the same as in the previous sections.

APPENDIX D

PROOF OF THEOREM 4

Let $\mathbf{g}_i = \boldsymbol{\theta}_i - \mathbf{X}_i\boldsymbol{\beta}$ ($i = 1, \dots, n$), and $\mathbf{g}^T = (\mathbf{g}_1^T, \dots, \mathbf{g}_n^T)$. Now we write

$$\begin{aligned}
& \sum_{i=1}^n (\boldsymbol{\theta}_i - \mathbf{X}_i\boldsymbol{\beta} - \boldsymbol{\eta}_i)^T \boldsymbol{\Sigma}_e^{-1} (\boldsymbol{\theta}_i - \mathbf{X}_i\boldsymbol{\beta} - \boldsymbol{\eta}_i) + \boldsymbol{\eta}^T \mathbf{V}^{-1} \boldsymbol{\eta} \\
&= \boldsymbol{\eta}^T (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1} + \mathbf{V}^{-1}) \boldsymbol{\eta} - 2\mathbf{g}^T (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1}) \boldsymbol{\eta} + \mathbf{g}^T (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1}) \mathbf{g} \\
&= [\boldsymbol{\eta} - (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1} + \mathbf{V}^{-1})^{-1} (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1}) \mathbf{g}]^T (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1} + \mathbf{V}^{-1}) \\
&\times [\boldsymbol{\eta} - (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1} + \mathbf{V}^{-1})^{-1} (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1}) \mathbf{g}] + \mathbf{g}^T [(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1}) \\
&- (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1}) (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1} + \mathbf{V}^{-1})^{-1} (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1})] \mathbf{g}
\end{aligned}$$

Noting that $(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1}) - (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1}) (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1} + \mathbf{V}^{-1})^{-1} (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1}) = [(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1})^{-1} + \mathbf{V}]^{-1} = \mathbf{C}$, say, integration with respect to $\boldsymbol{\eta}$ yields

$$\begin{aligned}
\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\Sigma}_e, \alpha, \boldsymbol{\Lambda} | \mathbf{y}) &\propto \prod_{i,j} p(y_{ij} | \theta_{ij}) \\
&\times |\boldsymbol{\Sigma}_e|^{-n/2} |\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1} + \mathbf{V}^{-1}|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{g}^T \mathbf{C} \mathbf{g}\right) \\
&\times |\mathbf{D} - \alpha \mathbf{W}|^{q/2} |\boldsymbol{\Lambda}|^{n/2} \alpha^{c-1} (1 - \alpha)^{d-1} \\
&\times |\boldsymbol{\Sigma}_e|^{-(\gamma+q+1)/2} \exp\left[-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_e^{-1} \mathbf{A})\right] \\
&\times |\boldsymbol{\Lambda}|^{(s-q-1)/2} \exp\left[-\frac{1}{2} \text{tr}(\boldsymbol{\Lambda} \mathbf{B})\right].
\end{aligned}$$

Next writing $\mathbf{X}^T = (\mathbf{X}_1^T, \dots, \mathbf{X}_n^T)$, $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_n^T)$,

$$\begin{aligned} \mathbf{g}^T \mathbf{C} \mathbf{g} &= \boldsymbol{\beta}^T (\mathbf{X}^T \mathbf{C} \mathbf{X}) \boldsymbol{\beta} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{C} \boldsymbol{\theta} + \boldsymbol{\theta}^T \mathbf{C} \boldsymbol{\theta} \\ &= [\boldsymbol{\beta} - (\mathbf{X}^T \mathbf{C} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{C} \boldsymbol{\theta}]^T (\mathbf{X}^T \mathbf{C} \mathbf{X}) [\boldsymbol{\beta} - (\mathbf{X}^T \mathbf{C} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{C} \boldsymbol{\theta}] \\ &\quad + \boldsymbol{\theta}^T [\mathbf{C} - \mathbf{C}^T \mathbf{X} (\mathbf{X}^T \mathbf{C} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{C}] \boldsymbol{\theta}. \end{aligned}$$

Hence, integrating with respect to $\boldsymbol{\beta}$, one gets

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\Sigma}_e, \alpha, \boldsymbol{\Lambda} | \mathbf{y}) &\propto \prod_{i,j} p(y_{ij} | \theta_{ij}) \\ &\quad \times |\boldsymbol{\Sigma}_e|^{-n/2} |\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1} + \mathbf{V}^{-1}|^{-1/2} |\mathbf{X}^T \mathbf{C} \mathbf{X}|^{-1/2} \\ &\quad \times \exp \left[-\frac{1}{2} \boldsymbol{\theta}^T \{ \mathbf{C} - \mathbf{C}^T \mathbf{X} (\mathbf{X}^T \mathbf{C} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{C} \} \boldsymbol{\theta} \right] \\ &\quad \times |\mathbf{D} - \alpha \mathbf{W}|^{q/2} |\boldsymbol{\Lambda}|^{n/2} \alpha^{c-1} (1 - \alpha)^{d-1} \\ &\quad \times |\boldsymbol{\Sigma}_e|^{-(\gamma+q+1)/2} \exp \left[-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_e^{-1} \mathbf{A}) \right] \\ &\quad \times |\boldsymbol{\Lambda}|^{(s-q-1)/2} \exp \left[-\frac{1}{2} \text{tr}(\boldsymbol{\Lambda} \mathbf{B}) \right]. \end{aligned}$$

Hence, writing $K (> 0)$ for a generic constant which does not depend on any unknown parameters,

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\Sigma}_e, \alpha, \boldsymbol{\Lambda} | \mathbf{y}) &\leq K \prod_{i,j} p(y_{ij} | \theta_{ij}) \\ &\quad \times |\boldsymbol{\Sigma}_e|^{-(n+\gamma+q+1)/2} |\mathbf{I}_n \otimes \boldsymbol{\Sigma}_e^{-1} + \mathbf{V}^{-1}|^{-1/2} |\mathbf{X}^T \mathbf{C} \mathbf{X}|^{-1/2} \\ &\quad \times |\mathbf{D} - \alpha \mathbf{W}|^{q/2} \alpha^{c-1} (1 - \alpha)^{d-1} \exp \left[-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_e^{-1} \mathbf{A}) \right] \end{aligned}$$

$$\times |\mathbf{\Lambda}|^{(n+s-q-1)/2} \exp \left[-\frac{1}{2} \text{tr}(\mathbf{\Lambda} \mathbf{B}) \right].$$

But, $|\mathbf{I}_n \otimes \mathbf{\Sigma}_e^{-1} + \mathbf{V}^{-1}|^{-1/2} \leq |\mathbf{V}^{-1}|^{-1/2} = |\mathbf{V}|^{1/2} = |\mathbf{D} - \alpha \mathbf{W}|^{-q/2} |\mathbf{\Lambda}|^{-n/2}$. Thus,

$$\begin{aligned} \pi(\boldsymbol{\theta}, \mathbf{\Sigma}_e, \alpha, \mathbf{\Lambda} | \mathbf{y}) &\leq K \prod_{i,j} p(y_{ij} | \theta_{ij}) \\ &\times |\mathbf{\Sigma}_e|^{-(n+\gamma+q+1)/2} |\mathbf{X}^T \mathbf{C} \mathbf{X}|^{-1/2} \\ &\times \alpha^{c-1} (1-\alpha)^{d-1} \exp \left[-\frac{1}{2} \text{tr}(\mathbf{\Sigma}_e^{-1} \mathbf{A}) \right] \\ &\times |\mathbf{\Lambda}|^{(s-q-1)/2} \exp \left[-\frac{1}{2} \text{tr}(\mathbf{\Lambda} \mathbf{B}) \right]. \end{aligned}$$

The rest of the proof is the same as in previous sections.

VITA

Joon Jin Song, son of Jong-Chool Song and Nam-Soon Lee, was born on June 1, 1974 in Kimchon, Korea. He graduated from Kimchon High School in Kimchon, Korea in 1992. He received a Bachelor of Science degree in statistics from Yeungnam University in Kyungsan, Korea in 1997. He received a Master of Science degree in statistics from Kyungpook National University in Daegu, Korea under the direction of Professor Joong Kweon Sohn in 1999. He continued to study statistics under the direction of Professor Bani K. Mallick, and received a Doctor of Philosophy degree in statistics from Texas A&M University in August 2004.

Joon Jin Song is married to Wonjeong Kim. His permanent address is:

331-4 Bugok-Dong, Kimchon

Kyungpook, Korea