Bayesian Semiparametric Spatial and Joint Spatio-Temporal Smoothing

by

Gentry White

A dissertation submitted to The University of Missouri-Columbia in conformity with the requirements for the degree of Doctor of Philosophy

Columbia, Missouri

August 4, 2006

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled.

BAYESIAN SEMIPARAMETRIC SPATIAL AND JOINT SPATIO-TEMPORAL SMOOTHING

presented by GENTRY WHITE

A candidate for the degree of Doctor of Philosophy

And hereby certify that in their opinion it is worthy of acceptance.

Dr. Dongchu Sun

Dr. Paul Speckman

Dr. Zhuoqiong He

Dr. Christopher Wikle

Dr. Jeanette Jackson-Thompson

Acknowledgements

Here I must acknowledge a sincere and deep debt of gratitude to my advisers and mentors, Drs. Dongchu Sun and Paul Speckman. Without their guidance, encouragement, patience, enthusiasm and inspiration, this work would have never come to fruition. I also thank my committee members Drs. Christopher Wikle and Zhuoqiong He, whose excellent comments and suggestions helped guide the course of my research.

I am also very grateful to the other member of my advisory committee, Dr. Jackson-Thompson and the staff of the Missouri Cancer Registry, who have all provided a wealth of information and support in my research, and without whom this project would have been much more difficult.

I must also thank Steven L. Sheriff and the Missouri Department of Conservation for sharing not only their data, but their expertise, experience and guidance. The data from the Missouri Turkey Hunting Survey (MTHS) has provided an excellent source of data and inspired much of the course of the research presented here. This progress could not have been made without the input, guidance and experience of the scientists at the Missouri Department of Conservation.

I would like to thank other Statistics Department faculty members and my fellow graduate students for providing a strong and nurturing environment in which to learn.

I am also forever indebted to my family for their understanding and moral support throughout my time at school.

Most importantly, I wish to acknowledge the love and patience of my wife Tanya. It is to her that I dedicate this work.

Abstract

Smoothing is the practice of modeling data in order to eliminate random variation from the observed data and provide estimates of the underlying process. Models are developed here beginning with an additive model that incorporates spatio-temporal smoothing of the observed mortality rates for female breast cancer in Missouri from 1969 through 2001. The next model developed uses an intrinsic auto regressive (IAR) prior to smooth the temporal trends in the data and a conditional auto-regressive (CAR) prior for the spatial effects. These two are combined in a single joint prior for spatio-temporal effects. The third model is a joint spatio-temporal model, using the IAR prior for the temporal trends and a spatial prior based on the thin-plate spline solution. These results open the door for further exploration including an alternate parameterization of the thin-plate splines prior to allow the computation of Bayes factors comparing the CAR prior and the thin-plate splines prior. This example is illustrated using a data set of responses to the Missouri Turkey Hunting Survey of 1996, conducted by the Missouri Department of Conservation. Additional strategies for dimension reduction of large scale problems are explored by reducing the number of basis functions in the thin-plate spline prior, results are compared for various degrees of dimension reduction. The example in this case involves the analysis of data for U.S. mortality due to colorectal cancer among men during the period 1999-2003.

Contents

Acknowledgements ii

Abstract iii

List of Tables vii

List of Figures viii

1 Introduction 1

2 An Additive Hierarchical Model for Mortality Rates 6

2.1 Prior Distributions of z_i and w_i 7

2.2 Summary and Completion of the Hierarchical Model 8

2.3 Propriety of Posterior Distribution 9

2.4 Estimation Via MCMC 9

2.5 Available Conditional Distributions 9

2.6 Results 11

2.6.1 Noninformative and Data Dependent Priors 11

2.6.2 Age effects θ_k and μ_k 13

2.6.3 Variance components δ_0, δ_1 and δ_2 13

2.6.4 Spatial Correlation Parameters ρ_1 and ρ_2 14

2.7 Disease Mapping 14

2.8 Conclusions 15

3 A Joint Model for Spatial and Temporal Effects 23

3.1 The Likelihood of the Data and First Stage Prior 24

3.2 The Joint Spatio-Temporal Prior 24

3.3 Other Priors 28

3.4 Computation 28

3.5 Results for the Joint Model . . . 29
3.5.1 Computational Improvements . . . 30
3.5.2 Noninformative and Data-dependent Priors . . . 31
3.5.3 Interpreting ρ . . . 31
3.5.4 The Variance Components δ_0 and δ_1 . . . 32
3.6 Comparison between the Additive model and the Joint Model . . . 33
3.6.1 Maps . . . 33
3.6.2 Model Selection Criteria . . . 34
3.7 Conclusions . . . 36

4 A Semiparametric Spatio-Temporal Model Using the Thin-Plate Spline Prior 46
4.1 Thin-Plate Splines . . . 49
4.1.1 Derivation . . . 50
4.1.2 Bayesian Thin-Plate Splines . . . 53
4.2 The Joint Spatio-Temporal Model . . . 54
4.2.1 Full-Conditional Distributions . . . 56
4.2.2 Computational Improvements . . . 57
4.3 Results . . . 58
4.3.1 Noninformative and Data-dependent Priors . . . 58
4.3.2 CAR Model vs Thin-Plate Spline Model . . . 59
4.4 Conclusion . . . 62

5 Using a g-prior for Model Comparison and Refinement 73
5.1 The Problem . . . 73
5.2 The Data and Likelihood . . . 74
5.3 The Hierarchical Models . . . 75
5.3.1 Model M_1 for $\boldsymbol{z}$ . . . 76
5.3.2 Model M_2 for $\boldsymbol{z}$ . . . 76
5.4 Bayesian Computation . . . 79
5.4.1 Bayes Factor Computation and the Null Model M_0 . . . 79
5.4.2 Full Conditional Distributions . . . 80
5.4.3 Choices for the Prior Median for η and η_2 . . . 83
5.4.4 Decorrelation to Improve Mixing of the Posterior Sampling Chain 84
5.4.5 Bayes Factor Comparison to the DIC . . . 86
5.5 Results . . . 87
5.6 Discussion and Further Work . . . 88

6 Dimensional Reduction for a Smoothed National Cancer Map 99
6.1 Introduction . . . 100
6.2 Data and Likelihood . . . 100
6.2.1 Data . . . 100
6.2.2 Likelihood . . . 101
6.2.3 Model . . . 101
6.3 Computation . . . 102
6.3.1 Dimension Reduction . . . 103
6.4 Results . . . 104
6.4.1 Data Reduction . . . 105
6.4.2 Calculating the Standardized Mortality Ratios . . . 105
6.4.3 Data Mapping . . . 106
6.4.4 Standardized Mortality Ratios . . . 107
6.5 Discussion . . . 108

List of Tables

2.1 Quantiles of δ_0, δ_1, δ_2, ρ_1 and ρ_2 for Noninformative and Data-dependent Priors 12

3.1 Quantiles of δ_0, δ_1 and ρ for Noninformative and Data-dependent priors 32
3.2 DIC Values for the Additive and Joint Models 35
3.3 $D(m)$ Values for the Additive and Joint Models 35

4.1 Quantiles of δ_0, δ_1 and λ for Noninformative and Data-dependent Priors 59
4.2 DIC Values for the CAR and Thin-plate Spline Models 60
4.3 $D(m)$ Values for the CAR and Thin-plate Spline Models 61

5.1 DIC Values for M_0 M_1 and M_2 86

6.1 DIC values for $q = 3082, 1600, 900$ and 100 107

List of Figures

2.1 Posterior Densities of θ_k, μ_k, δ_0, δ_1, δ_2, ρ_1 and ρ_2 from Additive Model for Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1969-2000. 16

2.2 Maps of Frequency and Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 for $(j,k) = (3,1)$ and $(j,k) = (5,2)$. 17

2.3 Maps of Frequency and Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 for $(j,k) = (7,3)$ and $(j,k) = (11,4)$. 18

2.4 Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 for $j = 1,\ldots,11$ and $k = 1$. 19

2.5 Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 for $j = 1,\ldots,11$ and $k = 2$. 20

2.6 Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 for $j = 1,\ldots,11$ and $k = 3$. 21

2.7 Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 for $j = 1,\ldots,11$ and $k = 4$. 22

3.1 Scatterplots Comparing the Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 from the Additive and Joint CAR Models. 37

3.2 Trace plots of δ_0, δ_1, ρ and θ_k, $k = 2,3,4$ from the Joint CAR Model for Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1969-2000. 38

3.3 Posterior Densities of δ_0, δ_1, ρ and θ_k, $k = 2,3,4$ from the Joint CAR Model for Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1969-2000. 39

3.4 Maps of Frequency and Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 from the Additive and Joint CAR Models for $(j,k) = (3,1)$ and $(j,k) = (5,2)$. 40

3.5 Maps of Frequency and Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 from the Additive and Joint CAR Models for $(j,k) = (7,3)$ and $(j,k) = (11,4)$. 41
3.6 Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 from the Joint CAR Model for $j = 1,\ldots,11$ and $k = 1$. 42
3.7 Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 from the Joint CAR Model for $j = 1,\ldots,11$ and $k = 2$. 43
3.8 Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 from the Joint CAR Model for $j = 1,\ldots,11$ and $k = 3$. 44
3.9 Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 from the Joint CAR Model for $j = 1,\ldots,11$ and $k = 4$. 45

4.1 Scatterplots Comparing the Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 from the CAR and Thin-plate Spline Joint Models. 64
4.2 Trace Plots of δ_0, δ_1, λ and θ_k, $k = 2,3,4$ from the Joint Thin-plate Splines Model for Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1969-2000. 65
4.3 Posterior Densities for δ_0, $\log(\delta_1)$, $\log(\lambda)$, and θ_k, $k = 2,3,4$ from the Joint Thin-plate Splines Model for Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1969-2000. 66
4.4 Maps of Frequency and Bayesian Estimates of Mortality Rates p_{ij} from the Joint Thin-plate Splines Model for Female Breast Cancer in Missouri from 1996-2000 for $(j,k) = (3,1)$ and $(j,k) = (5,2)$. 67
4.5 Maps of Frequency and Bayesian Estimates of Mortality Rates p_{ij} from the Joint Thin-plate Splines Model for Female Breast Cancer in Missouri from 1996-2000 for $(j,k) = (7,3)$ and $(j,k) = (11,4)$. 68
4.6 Maps of Bayesian Estimates of Mortality Rates p_{ij} from the Joint Thin-plate Splines Model for Female Breast Cancer in Missouri from 1996-2000 for $k = 1$ and $j = 1,\ldots,11$. 69
4.7 Maps of Bayesian Estimates of Mortality Rates p_{ij} from the Joint Thin-plate Splines Model for Female Breast Cancer in Missouri from 1996-2000 for $k = 2$ and $j = 1,\ldots,11$. 70
4.8 Maps of Bayesian Estimates of Mortality Rates p_{ij} from the Joint Thin-plate Splines Model for Female Breast Cancer in Missouri from 1996-2000 for $k = 3$ and $j = 1,\ldots,11$. 71

4.9 Maps of Bayesian Estimates of Mortality Rates p_{ij} from the Joint Thinplate Splines Model for Female Breast Cancer in Missouri from 1996-2000 for $k = 4$ and $j = 1, \ldots, 11$. 72

5.1 Trace Plots of (a) $\nu_{11,1}$, (b) ρ, (c) θ_1, (d) η, (e) z_{11} and (f) δ_0 from CAR Model for 1996 MTHS Without Decorrelation Step. 90
5.2 Trace Plots of (a) $\nu_{11,1}$, (b) ρ, (c) θ_1, (d) η, (e) z_{11} and (f) δ_0 from CAR Model for 1996 MTHS Without Decorrelation Step. 91
5.3 Trace Plots of (a) $\nu_{11,1}$, (b) ρ, (c) θ_1, (d) η, (e) z_{11} and (f) δ_0 from g-Prior Model for 1996 MTHS Without Decorrelation Step. 92
5.4 Trace Plots of (a) $\nu_{11,1}$, (b) ρ, (c) θ_1, (d) η, (e) z_{11} and (f) δ_0 from g-Prior Model for 1996 MTHS With Decorrelation Step. 93
5.5 Comparison of Results from CAR and g-Prior Models for 1996 MTHS for θ_1, θ_2 and z_i. 94
5.6 Posterior Densities for δ_0, δ_1, λ, θ_1 and θ_2 from g-Prior Model for 1996 MTHS. 95
5.7 Prior Model Complexity as a Function of b for g-Prior Model of 1996 MTHS. 96
5.8 Prior and Posterior Densities of η_2 for $b = 1$ from g-Prior Model for 1996 MTHS. 97
5.9 Maps of Frequency and Bayesian Estimates of Hunter Success Rates p_{ij} from the 1996 MTHS. 98

6.1 Trace Plots of δ_0, η_1, η_2 and μ from the g-Prior Model for p_i Male Mortality Rates due to Colorectal Cancer 1999-2003. 109
6.2 Posterior Densities for δ_0, η_1, η_2 and μ from the g-Prior Model for p_i Male Mortality Rates due to Colorectal Cancer 1999-2003. 110
6.3 Map of Frequency Estimates for Male Mortality Rates due to Colorectal Cancer p_i per 100,000 from $1999 - 2003$. 111
6.4 Map of Bayesian Estimates for Male Mortality Rates due to Colorectal Cancer p_i per 100,000 from $1999 - 2003$ from the Full Model. 112
6.5 Map of Bayesian Estimates for Male Mortality Rates due to Colorectal Cancer p_i per 100,000 from $1999 - 2003$ from Reduced Model $q = 1600$. 113
6.6 Map of Bayesian Estimates for Male Mortality Rates due to Colorectal Cancer p_i per 100,000 from $1999 - 2003$ from Reduced Model $q = 900$. 114
6.7 Map of Bayesian Estimates for Male Mortality Rates due to Colorectal Cancer p_i per 100,000 from $1999 - 2003$ from Reduced Model $q = 100$. 115
6.8 Significant Standardized Mortality Ratios for Male Mortality Due to Colorectal Cancer during $1999 - 2003$ from Full Model. 116
6.9 Significant Standardized Mortality Ratios for Male Mortality Due to Colorectal Cancer during $1999 - 2003$ from the Reduced Model, $q = 1600$. 117

6.10 Significant Standardized Mortality Ratios for Male Mortality Due to Colorectal Cancer during $1999-2003$ from the Reduced Model, $q = 900.118$

6.11 Significant Standardized Mortality Ratios for Male Mortality Due to Colorectal Cancer during $1999-2003$ from the Reduced Model, $q = 100.119$

Chapter 1

Introduction

Over the past decades a great deal of effort has been expended in the collection and compilation of high quality data on cancer incidence and mortality in the United States. Most of this work is done by governmental agencies such as the National Cancer Institute's Surveillance Epidemiology and End Results (SEER) program (2006*a*) and the Centers for Disease Control and Prevention's (CDC) National Program of Cancer Registries (NPCR) (2006*b*) who jointly prepare the annual U.S. Cancer Statistics (USCS) (2006*c*) report on cancer mortality and incidence in the nation. Other reports are prepared by the North American Association of Central Cancer Registries (NAACCR) (2006*d*) and by state and regional registries, such as the Missouri Cancer Registry (2006*e*), (2006*f*), the Oregon State Cancer Registry (2006*g*) and the Iowa Cancer Registry (2006*h*) and Greater Bay Area Cancer Registry (2006*i*). These data have largely been used in the creation and disbursement of descriptive statistics concerning the state of cancer in the U.S. The information available through these statistics present limited information concerning spatial or temporal trends in the course of cancer in the U.S. Recently, there have been more efforts made to in-

vestigate these trends, such as Jackson-Thompson et al. (2006). National data on mortality due to cancer has been examined using a variety of methods in work by Mungiole & Pickle (1999), Manton, Woodbury, Tallard, Riggan, Creason & Pellom (1989) and Devesa et al. (1999) for example. While others have sought to model national data on cancer incidence (Picle et al. 2003). For public health policy makers, there are two important initial questions about any disease, and cancer specifically, that need to be answered. First, how has the course of the disease changed over time; have incidence or mortality rates increased or decreased? Secondly, are there specific regions where the disease is more or less prevalent than others, and has this changed over time? These are important questions that provide feedback both to assess the effectiveness of public health policy and to provide guidance as to the best allocation of limited resources in preventing the spread of disease.

The use of Bayesian spatial models for disease mapping and smoothing of data dates back to the seminal paper by Clayton & Kaldor (1987), which introduced the use of the conditional autoregressive (CAR) prior from Besag (1974) for spatial effects in an empirical Bayesian model. Other examples of various empirical Bayesian approaches to spatial data include Manton, Woodbury, Stallard, Riggan, Creason & Pellom (1989), Clayton & Bernardinelli (1992), Devine, Halloran & Louis (1994), Devine, Louis & Halloran (1994), and Devine & Louis (1994). Other more recent Bayesian approaches include, Bernardinelli et al. (1995), Ferrándiz et al. (1995), Xia & Carlin (1998), Sun et al. (2000), and Zhang et al. (2006).

This proposal consists of several distinct but related projects, each incorporating or building on existing techniques of spatio-temporal data analysis. These projects demonstrate the evolutionary development of the analysis of the datasets in question.

The first data set considered here consists of the observed number of deaths in each county in Missouri due to female breast cancer from 1969 through 2001. The data are stratified into eleven three year time periods, and into four ten year age-group periods, $40-49$, $50-59$, $60-69$ and $70+$ years of age. One of the goals of the analysis proposed here is to devise a suitable spatio-temporal smoother for this data set. The second datasets include the data collected from the 1996 Missouri Turkey Hunting Survey (MTHS) by the Missouri Department of Conservation. These data consist of responses from a random survey of individuals who purchased hunting permits for the 1996 turkey hunting season. These responses indicate where the individual hunted during each week of the two week season and if they were successful in harvesting a turkey and during which week. This data set has a long history of analysis and spatial modeling, beginning with He & Sun (1998), Woodard et al. (1999) and He & Sun (2000). Later refinements of the spatial models for this data include Woodard et al. (2003), Sheriff et al. (n.d.) and White & Sun (2006). This data set provides an excellent example for the development of new methods of spatial data analysis. The final data set consists of deaths due to colorectal cancer among men across the continental U.S. from 1999-2003.

The first model proposed for the analysis of the breast cancer mortality data is a relatively straightforward additive model with separate terms for the spatial, temporal and age effects. This is similar to a random effects model, but with the distinction that the temporal effects term has a slope that contains both a separate age effect and a separate spatial effect. This model does provide a degree of smoothing, but unfortunately this smoothing is dominated by the age effects and retains little of the regional heterogeneity present in the raw data.

The second model presented for the analysis of the breast cancer mortality data is a semi-parametric model that incorporates a joint spatial and temporal effects term and an age effect term. The joint spatio-temporal effect requires the advent of a joint prior for both the spatial and temporal effects. This is accomplished in this case by using the conditional auto-regressive (CAR) prior for spatial effects introduced in the previous model and implementing an intrinsic auto-regressive (IAR) prior for the temporal effects. This model shows an improvement in smoothing, in terms of retaining the regional heterogeneity of the raw data while reducing the observed noise. However, there are indications that this model does not sufficiently smooth the data and that there is room for improvement.

The third model presented here uses the dataset consisting of the responses from the 1996 Missouri Turkey Hunting Survey, a collection of sample data consisting of the observed number of turkeys harvested and the number of hunters hunting in a given county in Missouri for both weeks of the 1996 turkey season. This data set has previously been analyzed by a variety of methods and is used to demonstrate the development of a model using a spatial effects prior in the model based on the thin-plate spline solution. This model shows excellent results compared to the CAR model for the same data.

The fourth model presented here uses another joint spatio-temporal prior. This model represents the culmination of the previous efforts to devise a suitable spatio-temporal smoother for the data in question. This model uses a similar joint spatio-temporal effect prior as developed in the second model, but instead uses the spatial effects prior based on the thin-plate spline solution demonstrated in the previous model. This model shows promise as a spatio-temporal smoother and provides several

opportunities for future work.

The fifth model presented here is a re-parameterization of the thin-plate splines based prior used in the model presented in Chapter 4. The data used are the results from the 1996 MTHS. The prior used in this model is re-parameterized in order to calculate the Bayes factors comparing the performance of the thin-plate spline prior model to the model using a CAR prior for spatial effects.

The sixth model presented here uses the thin-plate spline prior presented in the previous chapter applied to a national dataset for mortality due to colorectal cancer among men in the continental U.S. during the period $1999 - 2003$. This is used as a platform for exploring a strategy of dimension reduction in order to accelerate the computational process. This dimensional reduction is accomplished by reducing the number of basis functions in the thin-plate spline prior. Several different degrees of dimensional reduction are tried and the results compared with the model evaluated with the full set of basis functions. The capability of these models to detect significant differences between regions is compared using standardized mortality ratios (SMRs).

Chapter 2

An Additive Hierarchical Model for Mortality Rates

In order to begin the description of the additive model, first consider the likelihood of the data. Let y_{ijk} denote the number of cases of a given disease for the i^{th} county, j^{th} time period, and k^{th} age-group. Given the population size n_{ijk} and rate p_{ijk}, we assume that y_{ijk} follows an independent Poisson distribution,

$$(y_{ijk} \mid p_{ijk}) \overset{indep.}{\sim} Poisson(n_{ijk}p_{ijk}). \tag{2.1}$$

We consider the following hierarchical model,

$$\nu_{ijk} \equiv \log(p_{ijk}) = z_i + (\mu_k - w_i)(t_j - \bar{t}) + \theta_k + \epsilon_{ijk}, \tag{2.2}$$

where z_i is the additive effect for the i^{th} county and θ_k is the additive effect for age-group k. The change over time is represented by the rate $(\mu_k + w_i)$ for the i^{th} county and k^{th} age group multiplied by $(t_j - \bar{t})$, where $\bar{t} = J^{-1}\sum_{j=1}^{J} t_j$ and t_j is the midpoint of the j^{th} time period. This allows for each age group and each county to have different temporal slopes. Extra variation due to other sources is included in

the error terms ϵ_{ijk} and is assumed to follow the distribution

$$\epsilon_{ijk} \overset{iid}{\sim} N(0, \delta_0). \tag{2.3}$$

As a result, the prior for ν_{ijk} is

$$(\nu_{ijk} \mid z_i, \theta_k, \mu_k, w_i, \delta_0) \sim N(z_i + (\mu_k - w_i)(t_j - \bar{t}) + \theta_k, \delta_0). \tag{2.4}$$

The priors for θ_k and μ_k must be specified, and z_i and w_i have prior distributions with hyper-parameters that have prior distributions of their own. The extra observed variation ϵ_{ijk} also has a prior distribution with hyper-parameter δ_0, which has its own prior distribution as well. This form of spatio-temporal interaction, in which county slopes are allowed to have spatial correlation, is first suggested in Sun et al. (2000).

2.1 Prior Distributions of z_i and w_i

The prior distribution of z_i is given by the conditional autoregressive (CAR) prior proposed in Besag (1974) and Clayton & Kaldor (1987). This prior is defined in part using the $I \times I$ adjacency matrix $\boldsymbol{C}$ with elements C_{uv} defined as

$$C_{uv} = \begin{cases} 1, & \text{if counties } u \text{ and } v \text{ are adjacent,} \\ 0, & \text{otherwise, with } C_{uu} = 0 \end{cases} \tag{2.5}$$

The CAR prior is then defined by the conditional density

$$[\boldsymbol{z} \mid \rho_1, \delta_1] = \frac{|\boldsymbol{I}_I - \rho_1 \boldsymbol{C}|^{1/2}}{(2\pi\delta_1)^{I/2}} \exp\left\{ -\frac{\boldsymbol{z}'(\boldsymbol{I}_I - \rho_1 \boldsymbol{C})\boldsymbol{z}}{2\delta_1} \right\}, \tag{2.6}$$

where $\delta_1 > 0$. In order for this to be a proper prior, the values for ρ_1 are constrained such that $\rho_1 \in (\lambda_1^{-1}, \lambda_I^{-1})$, where λ_1, λ_I are respectively, the maximum and the minimum eigenvalues of the adjacency matrix $\boldsymbol{C}$. Note that this interval also contains

0. In the case $\rho_1 = 0$, the z_i are independent. As in the case of z_i, $\boldsymbol{w} = (w_1, .., w_I)'$ follows the same CAR prior as $\boldsymbol{z} = (z_1, \ldots, z_I)'$

$$[\boldsymbol{w} \mid \rho_2, \delta_2] = \frac{|\boldsymbol{I}_I - \rho_2 \boldsymbol{C}|^{1/2}}{(2\pi\delta_2)^{I/2}} \exp\left\{-\frac{\boldsymbol{w}'(\boldsymbol{I}_I - \rho_2 \boldsymbol{C})\boldsymbol{w}}{2\delta_2}\right\}. \tag{2.7}$$

where $\delta_2 > 0$ and $\rho_2 \in (\lambda_1^{-1}, \lambda_I^{-1})$. Note that the correlation coefficients ρ_1 and ρ_2 are assumed to be independent of age-group and time period. The CAR priors (2.6) and (2.7) chosen for both $\boldsymbol{z}$ and $\boldsymbol{w}$ as in He & Sun (2000) have the benefits of additional correlation parameters not in other priors such as the prior proposed by Besag et al. (1991).

2.2 Summary and Completion of the Hierarchical Model

Evaluation of the model requires the likelihood and additional priors given here. The likelihood in (2.1) can be written in terms of $\nu_{ijk} = \log(p_{ijk})$,

$$[y_{ijk} \mid \nu_{ijk}] \propto \exp(\nu_{ijk} y_{ijk} - n_{ijk} e^{\nu_{ijk}}). \tag{2.8}$$

In order to complete the hierarchical model, the following priors are needed

$$\begin{aligned}
\theta_k &\sim N(\xi_{mk}, \delta_{mk}), && (2.9)\\
\mu_k &\sim N(\xi_{sk}, \delta_{sk}), && (2.10)\\
[\delta_l] &\propto \frac{1}{\delta_l^{a_l - 1}} e^{(-b_l/\delta_l)},\ l = 0, 1, 2, && (2.11)\\
\rho_r &\sim U(\lambda_1^{-1}, \lambda_I^{-1}),\ r = 1, 2. && (2.12)
\end{aligned}$$

The hyper-parameters (ξ_{mk}, δ_{mk}), (ξ_{sk}, δ_{sk}) and (a_l, b_l) are fixed constants. When $a_l > 0$ and $b_l > 0$, δ_l has a proper distribution.

2.3 Propriety of Posterior Distribution

To complete the hierarchical model, the hyper-parameters (ξ_{mj}, δ_{mj}), (ξ_{sj}, δ_{sj}) and (a_l, b_l) need to be specified. The commonly used non-informative prior for θ and μ are flat or constant priors. This can be a limiting case when δ_{mj} and $\delta_{sj} \to \infty$. Flat priors are naturally used for ρ_1 and ρ_2. Noninformative priors for δ_l can present problems. Traditionally the prior $1/\delta_l$ can be used for δ_l. The problem with this prior is that Sun et al. (2000) shows that the resulting posterior obtained will be improper. In this model constant priors for the variance components are used; again this gives rise to the possibility that the resulting posterior distributions may be improper. Sun et al. (2000) gives an estimation procedure based on a theorem on the existence of the posterior.

2.4 Estimation Via MCMC

In order to evaluate this model, Gibbs sampling as proposed in Gelfand & Smith (1990) is used to evaluate the resulting posterior distributions. In order to implement the Gibbs sampler the full conditional distributions need to be sampled; most of these are known densities, while a few others are sampled by proving the log-concavity of the distributions and using the ARS algorithm from Gilks & Wild (1992).

2.5 Available Conditional Distributions

If we define $\boldsymbol{\nu} = (\nu_{111}, \ldots \nu_{11K}, \nu_{121}, \ldots, \nu_{IJK})'$, then the full conditional distribution for the joint posterior of the parameters of interest is

Lemma 1

(a) For given $(z_i, \mu_k, w_i, \theta_k, \delta_0; y_{ijk})$, $(\nu_{1,1,1}, \ldots, \nu_{IJK})$ *are independent, and each* ν_{ijk} *depends only on* y_{ijk},

$$[\nu_{ijk} \mid z_i, \mu_k, w_i, \theta_k, \delta_0; y_{ijk}] \propto \exp\left\{ y_{ijk}\nu_{ijk} - n_{ijk}e^{\nu_{ijk}} - \frac{1}{2\delta_0}(\nu_{ijk} - a_{ijk})^2 \right\},$$

where $a_{ijk} = \theta_k + z_i + (\mu_k + w_i)(t_j - \bar{t})$.

(b) The conditional posterior distribution of ν_{ijk} *in part (a) is log-concave.*

(c) $(\theta_k \mid \nu_{ijk}, z_i, \delta_0; y_{ijk}) \sim N\left(\frac{\frac{\xi_{mk}}{\delta_{mk}} + \frac{1}{\delta_0}\sum_{i,j}(\nu_{ijk} - z_i)}{(IJ/\delta_0) + (1/\delta_{mk})}, \frac{1}{(IJ/\delta_0) + (1/\delta_{mk})} \right)$

(d) $(\boldsymbol{z} \mid \boldsymbol{\nu}, \boldsymbol{\theta}, \rho_1, \delta_0, \delta_1; \boldsymbol{y}) \sim N_I\left(\boldsymbol{c}_1, \boldsymbol{G}_1^{-1}\right)$, *where* $\boldsymbol{G}_1 = \frac{JK}{\delta_0}\boldsymbol{I}_I + \frac{(\boldsymbol{I}_I - \rho_1 C)}{\delta_1}$, $\boldsymbol{c}_1 = \boldsymbol{G}_1^{-1}(d_{11}, d_{12}, \ldots, d_{1I})'$, *and* $d_{1i} = \frac{\sum_{j,k}(\nu_{ijk} - \theta_k)}{\delta_0}$.

(e) $(\boldsymbol{w} \mid \boldsymbol{\nu}, \boldsymbol{\mu}, \rho_2, \delta_0, \delta_2; \boldsymbol{y}) \sim N_I\left(\boldsymbol{c}_3, \boldsymbol{G}_2^{-1}\right)$, *where* $\boldsymbol{G}_2 = \frac{Kc_2}{\delta_0}\boldsymbol{I}_I + \frac{(\boldsymbol{I}_I - \rho_2 C)}{\delta_2}$, $d_2 = \sum_j (t_j - \bar{t})^2$, *and* $\boldsymbol{c}_2 = \frac{1}{\delta_0}\boldsymbol{G}_2^{-1}\sum_{j,k}(\nu_{\cdot jk}(t_j - \bar{t}) - d_2\sum_k \mu_k)$.

(f) $(\mu_k \mid \nu_{ijk}, w_i, \delta_0; y_{ijk}) \sim N\left(c_{3_k}, G_{3_k}^{-1}\right)$, *where* $G_{3_k} = \left(\frac{Id_2}{\delta_0} + \frac{1}{\delta_{sk}}\right)$ *and* $c_{3_k} = G_{3_k}^{-1}\left(\frac{\xi_{sk}}{\delta_{sk}} + \frac{1}{\delta_0}\left[\sum_{i,j}\nu_{ijk}(t_j - \bar{t}) - d_2\sum_i w_i\right]\right)$.

(g) $(\delta_0 \mid \boldsymbol{\nu}, \boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\theta}; \boldsymbol{y}) \sim IG\left(a_0 + \frac{IJK}{2}, b_0 + \frac{1}{2}\sum_i\sum_j\sum_k(\nu_{ijk} - a_{ijk})^2\right)$

(h) $(\delta_1 \mid \boldsymbol{z}, \rho_1; \boldsymbol{y}) \sim IG\left(a_1 + \frac{I}{2}, b_1 + \frac{1}{2}\boldsymbol{z}'(\boldsymbol{I}_I - \rho_1\boldsymbol{C})\boldsymbol{z}\right)$

(i) $(\delta_2 \mid \boldsymbol{w}, \rho_2; \boldsymbol{y}) \sim IG\left(a_2 + \frac{I}{2}, b_2 + \frac{1}{2}\boldsymbol{w}'(\boldsymbol{I}_I - \rho_2\boldsymbol{C})\boldsymbol{w}\right)$.

(j) $[\rho_1 \mid \boldsymbol{z}, \delta_1; \boldsymbol{y}] \propto |\boldsymbol{I}_I - \rho_1\boldsymbol{C}|^{1/2}\exp\left(\frac{\rho_1}{\delta_1}\boldsymbol{z}'\boldsymbol{C}\boldsymbol{z}\right)$.

(k) The conditional density of ρ_1 *in part (a) is log-concave.*

(l) For given $(\boldsymbol{w}, \delta_2; \boldsymbol{y})$ *the conditional posterior density of* ρ_2 *is*

$$[\rho_2 \mid \boldsymbol{w}, \delta_2; \boldsymbol{y}] \propto |\boldsymbol{I}_I - \rho_2 \boldsymbol{C}|^{1/2} \exp\left(\frac{\rho_2}{\delta_2} \boldsymbol{w}' \boldsymbol{C} \boldsymbol{w}\right).$$

(m) The conditional distribution of ρ_2 *in part (l) is log-concave.*

Proof. The proof for part (b) is as follows

$$\frac{\partial^2}{\partial \nu_{ijk}^2} \log[\nu_{ijk} \mid z_i, \mu_k, w_i, \theta_k, \delta_0; y_{ijk}] = -(n_{ijk} e^{\nu_{ijk}} + 1/\delta_0) < 0, \, \forall \, \nu_{ijk}.$$

For part (k),

$$\frac{\partial^2}{\partial \rho_1^2} \log[\rho_1 \mid \boldsymbol{z}, \delta_1; \boldsymbol{y}] = -\frac{1}{2} \sum_{i=1}^{I} \left(\frac{\lambda_i}{1 - \rho_1 \lambda_i}\right)^2 < 0, \, \forall \, \rho_1.$$

Part (l) follows as for part (k). □

2.6 Results

Most of the above conditional distributions are easily sampled, the exceptions being for ν_{ijk}, ρ_1 and ρ_2. Those are shown to have log-concave densities and are then evaluated using the ARS algorithm at each step in the Gibbs sampler. Implementation is done in FORTRAN, with the compiled code running in 100,000 iterations in approximately 100 minutes, with 50,000 iterations for burn-in.

2.6.1 Noninformative and Data Dependent Priors

The model is initially run using non-informative priors for δ_0, δ_1 and δ_2,

$$\pi(\delta_l) \propto \frac{1}{\sqrt{\delta_l}}, \; l = 0, 1, 2.$$

The resulting posterior means and variance are used to calculate a set of data-dependent priors for δ_0, δ_1 and δ_2 by inflating the mean and variances by some factors IF. Typically $IF = (2, 200)$ for the mean and variance respectively. Then we use the inflated mean and variance to calculate new values for the hyper-parameters of the priors. These new data-dependent priors are used to test the model for robustness in prior selection. Results in Table 2.1 for the non-informative priors (NI) and the inflated priors show that the model does indeed appear to be robust in terms of prior selection.

Table 2.1: Quantiles of δ_0, δ_1, δ_2, ρ_1 and ρ_2 for Noninformative (NI) and Inflation Factor (IF) Data-dependent Priors

Summary of Posterior Distributions								
	Prior	Min.	1st Qt.	Median	Mean	3rd Qt.	Max.	Std. Dev.
δ_0	NI	.00095	.00341	.00459	.00467	.00578	.01268	.00166
	IF 2, 200	.00019	.00304	.00418	.00431	.00548	.01282	.00181
	IF 1.5, 25	.00019	.00304	.00418	.00431	.00548	.01282	.00181
	IF 2.5, 500	.00123	.00346	.00432	.00456	.00541	.01183	.00152
δ_1	NI	.00274	.00797	.00952	.00979	.01134	.02295	.00249
	IF 2, 200	.00274	.00794	.00946	.00976	.01126	.03006	.00255
	IF 1.5, 25	.00274	.00794	.00946	.00976	.01126	.03006	.00255
	IF 2.5, 500	.00337	.00830	.00980	.01005	.01151	.02927	.00245
δ_2	NI	.00001	.00019	.00027	.00028	.00035	.00119	.00013
	IF 2, 200	.00001	.00019	.00027	.00030	.00036	.00045	.00019
	IF 1.5, 25	.00001	.00019	.00027	.00030	.00036	.00450	.00019
	IF 2.5, 500	.00005	.00023	.00030	.00032	.00037	.00470	.00016
ρ_1	NI	-.2480	.0787	.1185	.1046	.1442	.1700	.0531
	IF 2,200	-.2480	.0838	.1222	.1084	.1471	.1700	.0519
	IF 1.5, 25	-.2480	.0838	.1222	.1084	.1471	.1700	.0519
	IF 2.5, 500	-.2504	.0803	.1198	.1051	.1453	.1700	.0544
ρ_2	NI	-.3399	-.1188	.0103	-.0188	.0947	.1700	.1334
	IF 2,200	-.3399	-.1019	.02375	-.0076	.1054	.1700	.1333
	IF 1.5,25	-.3399	-.1019	.02375	-.0076	.1054	.1700	.1333
	IF 2.5,500	-.3399	-.0835	.0249	.0006	.1017	.1700	.1208

2.6.2 Age effects θ_k and μ_k

The range of posterior means of θ_k indicates a steady increase in mortality due to female breast cancer with respect to age. Figure 2.1 shows that the rates for each age group appear to be increasing as age increases; the mortality rates are higher for older age-groups, regardless of location in space or time.

The posterior means for μ_k shown in Figure 2.1 demonstrate the change in rates with respect to age over time. The negative values for the two youngest age groups indicate a decrease in mortality rates over the time period for the those age groups, and the positive values for the two oldest age groups represent an increase in the mortality rates over time for those age groups. Looking at both plots indicates that the rates for the youngest age groups are less than the older age groups but also that there is a discrepancy in the rates over time; they are increasing for the older age groups but decreasing for the younger age groups. This interaction between age and time is indication of a possible cohort effect.

2.6.3 Variance components δ_0, δ_1 and δ_2

The relative importance of $\boldsymbol{z}$ and $\boldsymbol{w}$ and can be seen in their respective variances, δ_1, δ_2 and in δ_0, whose posterior distributions are shown in Figure 2.1. The mean of the posterior distribution of δ_2 is smaller than that of δ_0 and δ_1. In addition the posterior density of ρ_2 is quite diffuse and centered about 0. These results indicate that the $\boldsymbol{w}$ components are superfluous to the model, the small variance and density of ρ_2 indicates that there is little contribution by these terms.

2.6.4 Spatial Correlation Parameters ρ_1 and ρ_2

The plots of the posterior distributions of ρ_1 and ρ_2 in Figure 2.1 show that the spatial correlation ρ_1 for $\boldsymbol{z}$ is clearly non-zero, but the distribution for ρ_2, the spatial parameter for the distribution of $\boldsymbol{w}$, is widely spread about 0. The implication in this is that the spatial structure between rates over time is not significant, even though the spatial effect overall is. In the CAR prior for $\boldsymbol{w}$ when ρ_2 is equal to zero, the w_i are in fact i.i.d and would have the effect of adding random noise to the age group component of the temporal slope.

2.7 Disease Mapping

The maps in Figures 2.2 and 2.3 compare the results of the additive model to the raw estimates of rates. As can be seen from these maps, the estimates from the additive model greatly smooth the raw data. Little if any spatial pattern evident in the data is visible from these maps. The smoothing that is taking place is due to the θ_k terms dominating the model estimates. As a result, the rate estimates are being smoothed toward a mean age-group effect. This result is at odds with the maps of the raw rates, which show some indication of possible spatial patterns in the rates.

The second four sets of maps in Figures 2.4–2.7 show the rates for each age group through each time period. These clearly show the trend over time for the two youngest age groups ($k = 1, 2$) to be decreasing. The rates for the third age group ($k = 3$) are flat, and the rates for the oldest age group ($k = 4$) are increasing. These results verify what can be seen in the previous figures showing the posterior densities for μ_k.

2.8 Conclusions

In all, the results for this model show that the dominant term is θ_k, indicating that the age effects dominate the model. While the spatial effects appear significant, they are small relative to the age effects. There appear to be significant temporal trends, though the spatial correlation between the temporal slopes appears insignificant and they are again dominated by the age terms. There is also a clear difference between the oldest and youngest age groups in terms of temporal trends.

This model demonstrates satisfactory results in terms of detecting age group differences in both mean rate and the temporal slope of the mean rate. The model does not detect any apparent spatial trends in the data that could be due to over or under-smoothing of spatial trends.

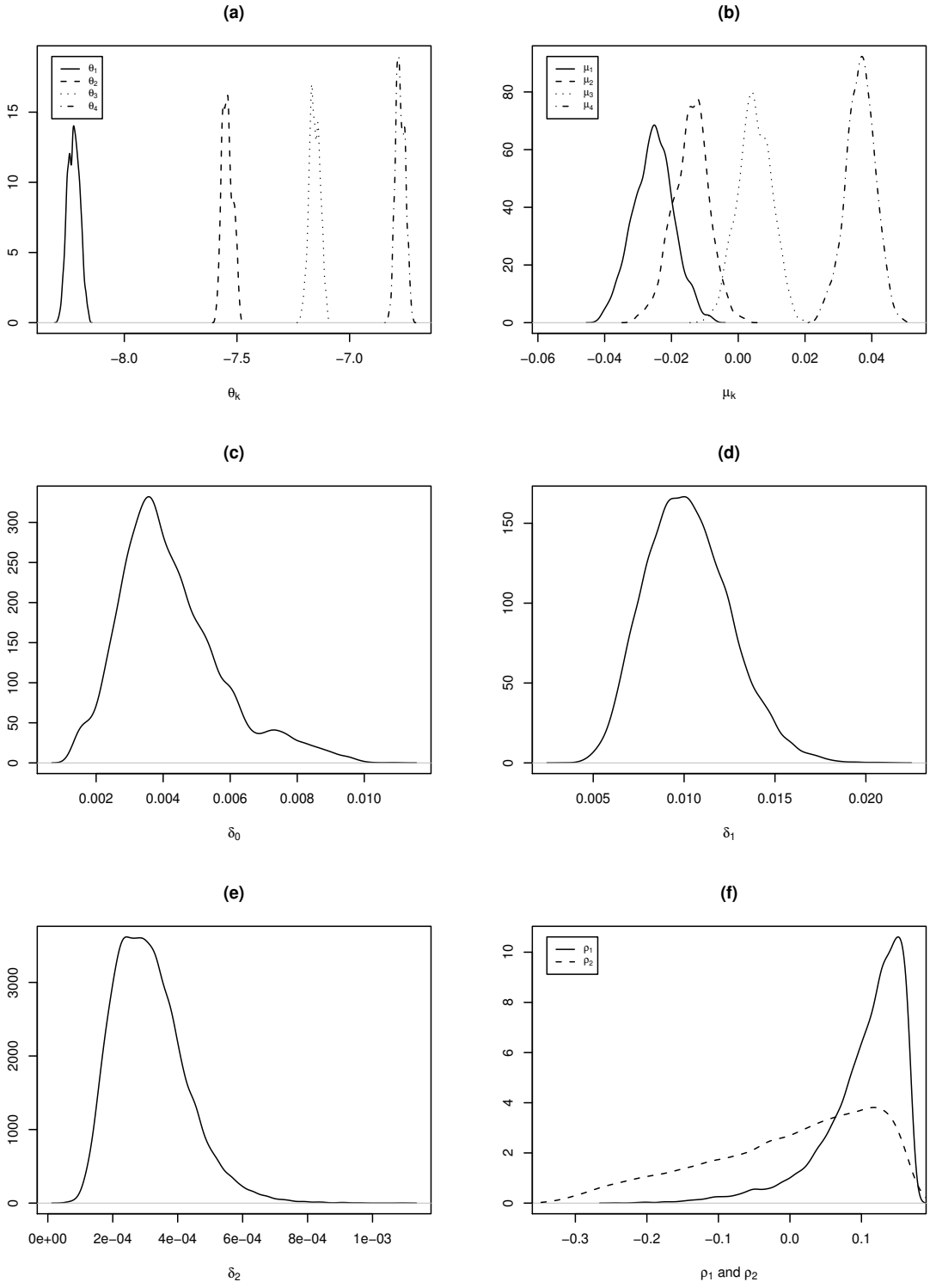

Figure 2.1: Posterior Densities of (a) θ_k, (b) μ_k, (c) δ_0, (d) δ_1, (e) δ_2, (f) ρ_1 and ρ_2 from Additive Model for Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1969-2000.

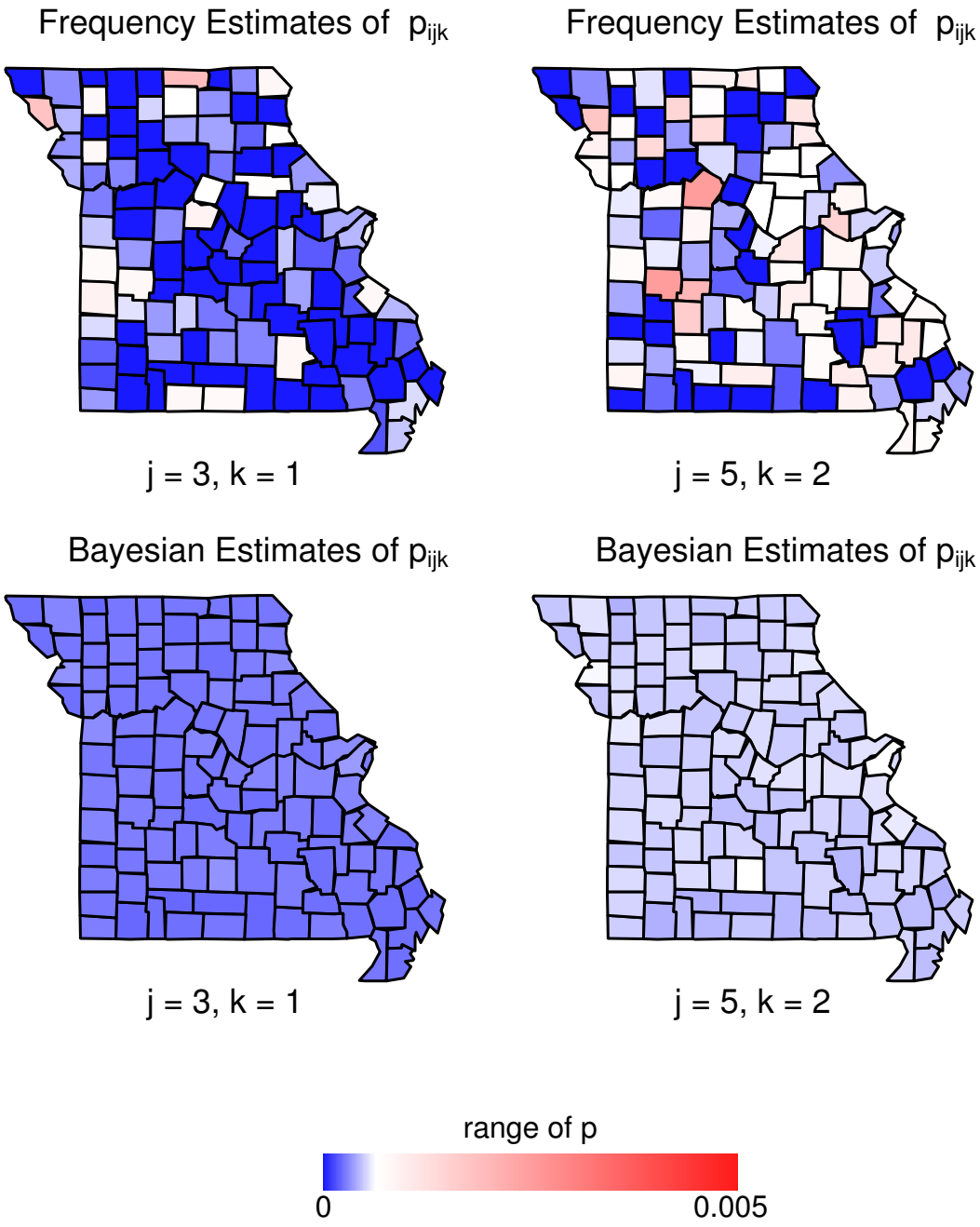

Figure 2.2: Maps of Frequency and Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 for $(j, k) = (3, 1)$ and $(j, k) = (5, 2)$.

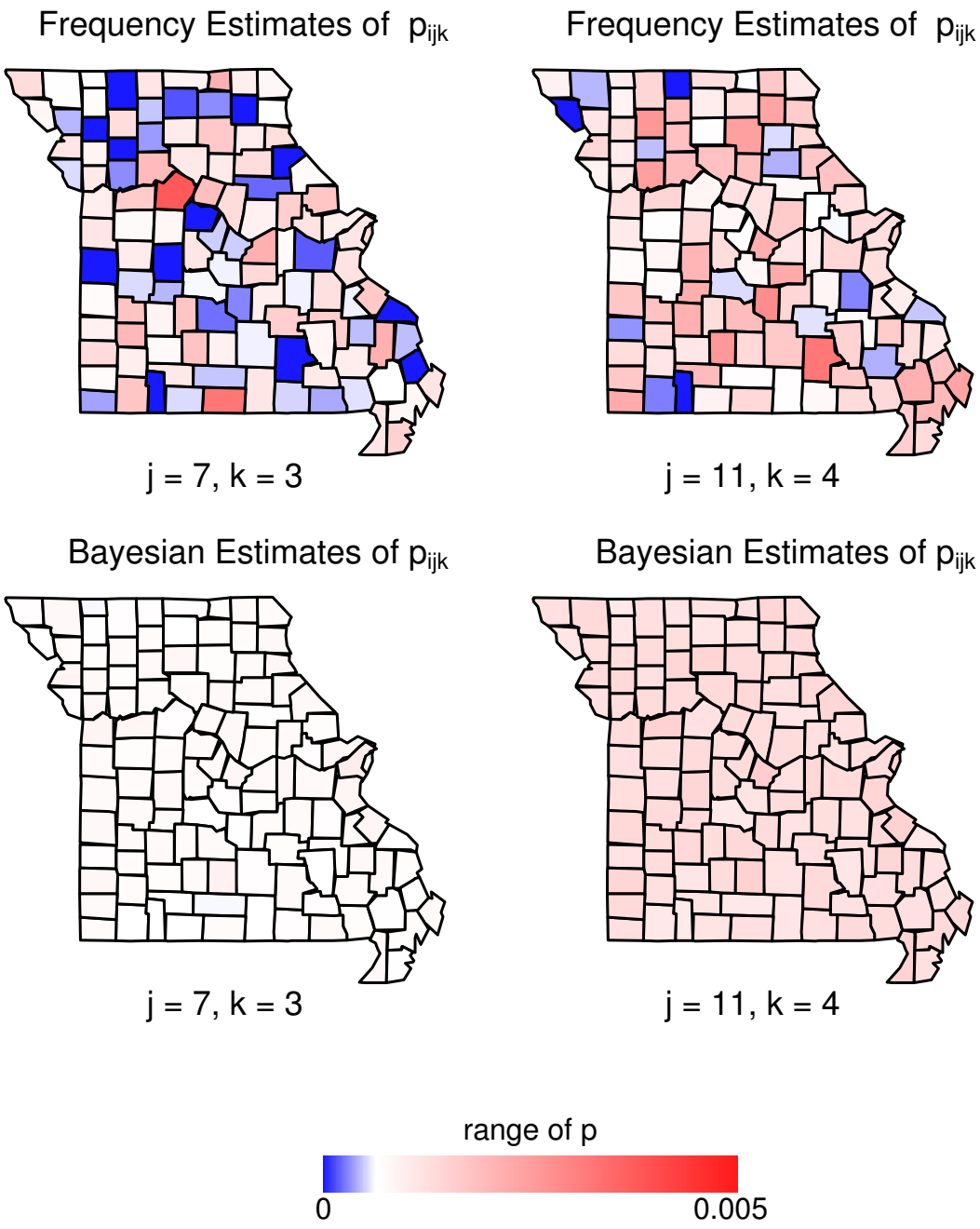

Figure 2.3: Maps of Frequency and Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 for $(j,k) = (7,3)$ and $(j,k) = (11,4)$.

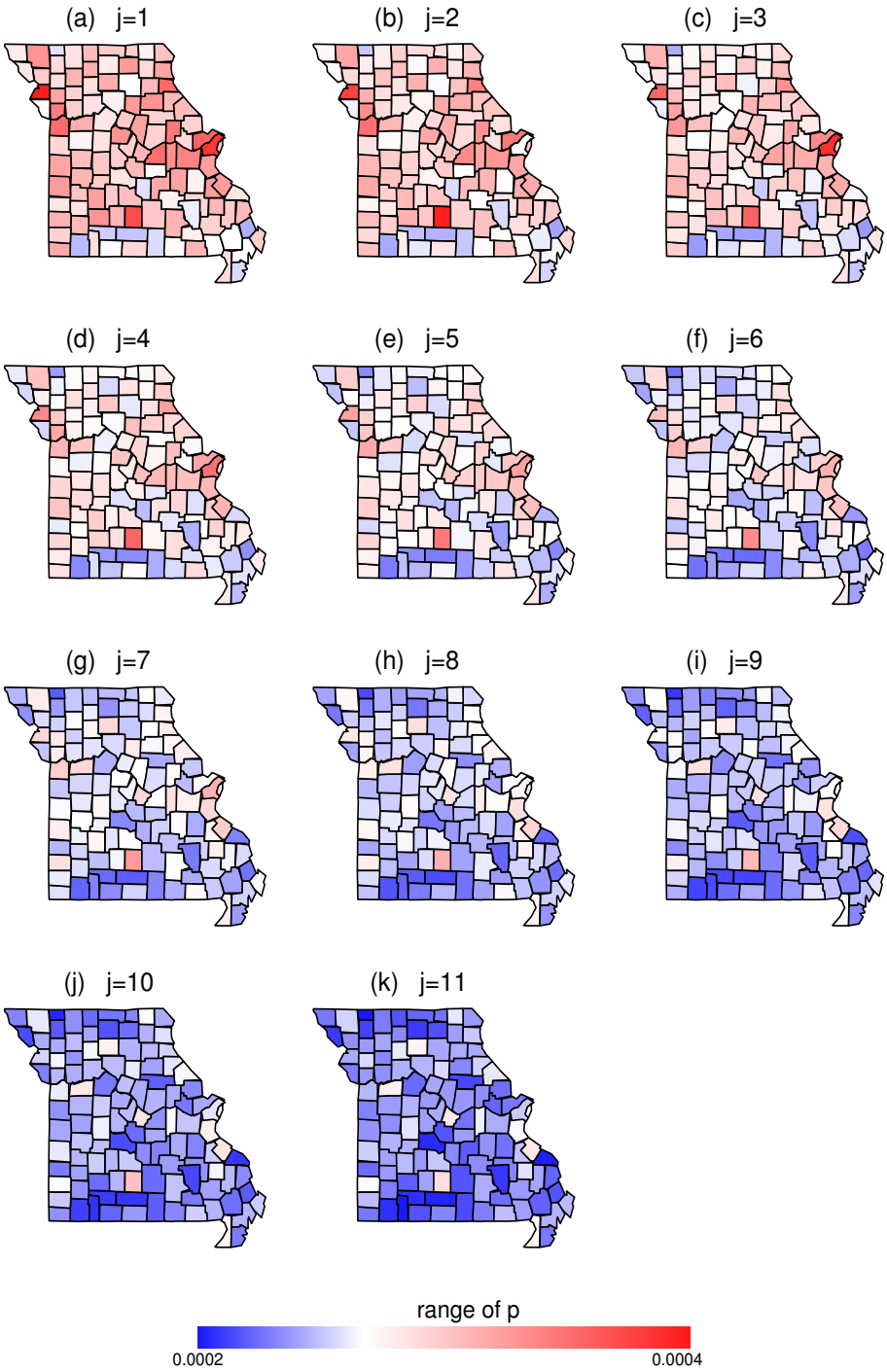

Figure 2.4: Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 for $j = 1, \ldots, 11$ and $k = 1$.

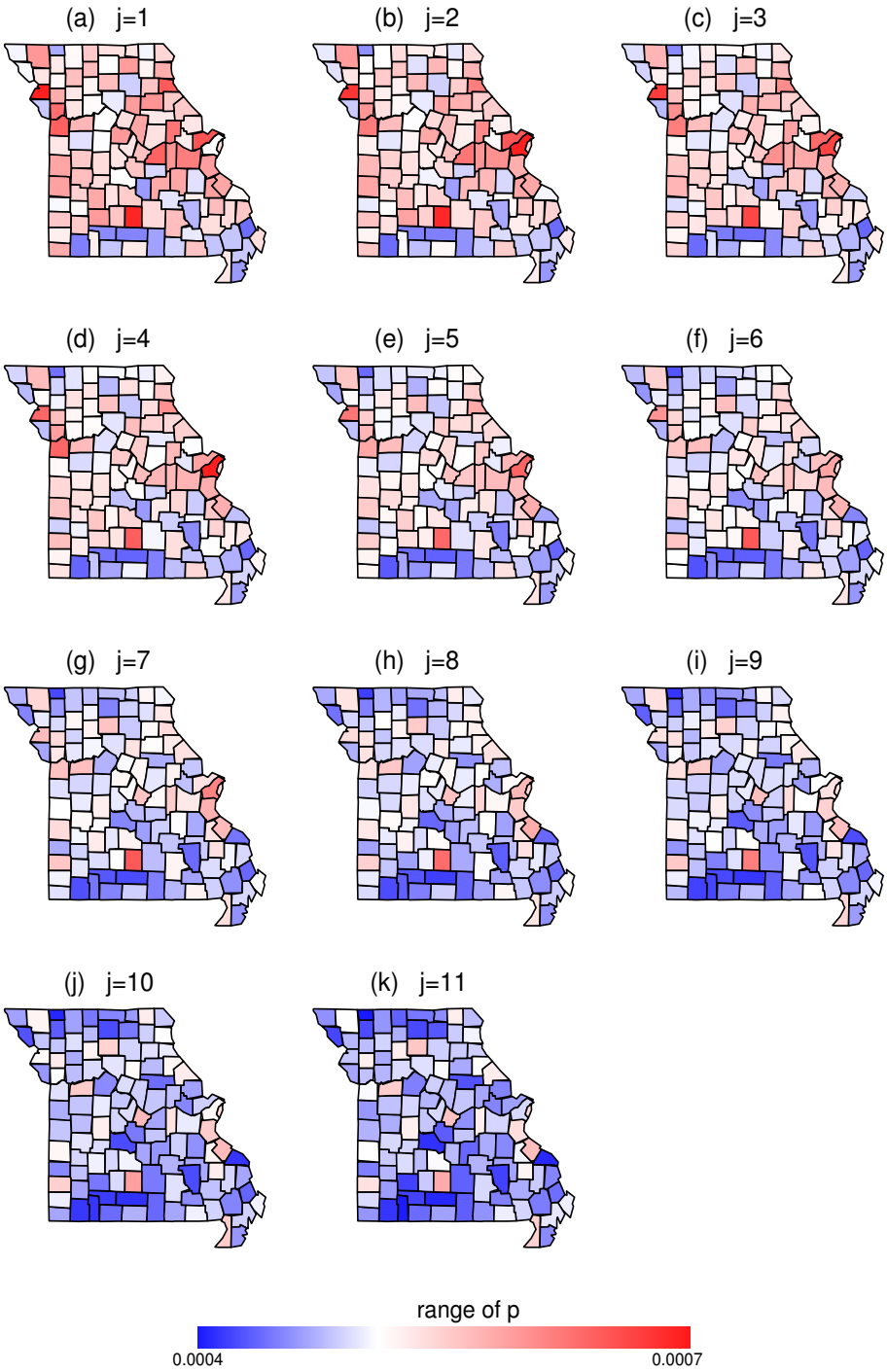

Figure 2.5: Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 for $j = 1, \ldots, 11$ and $k = 2$.

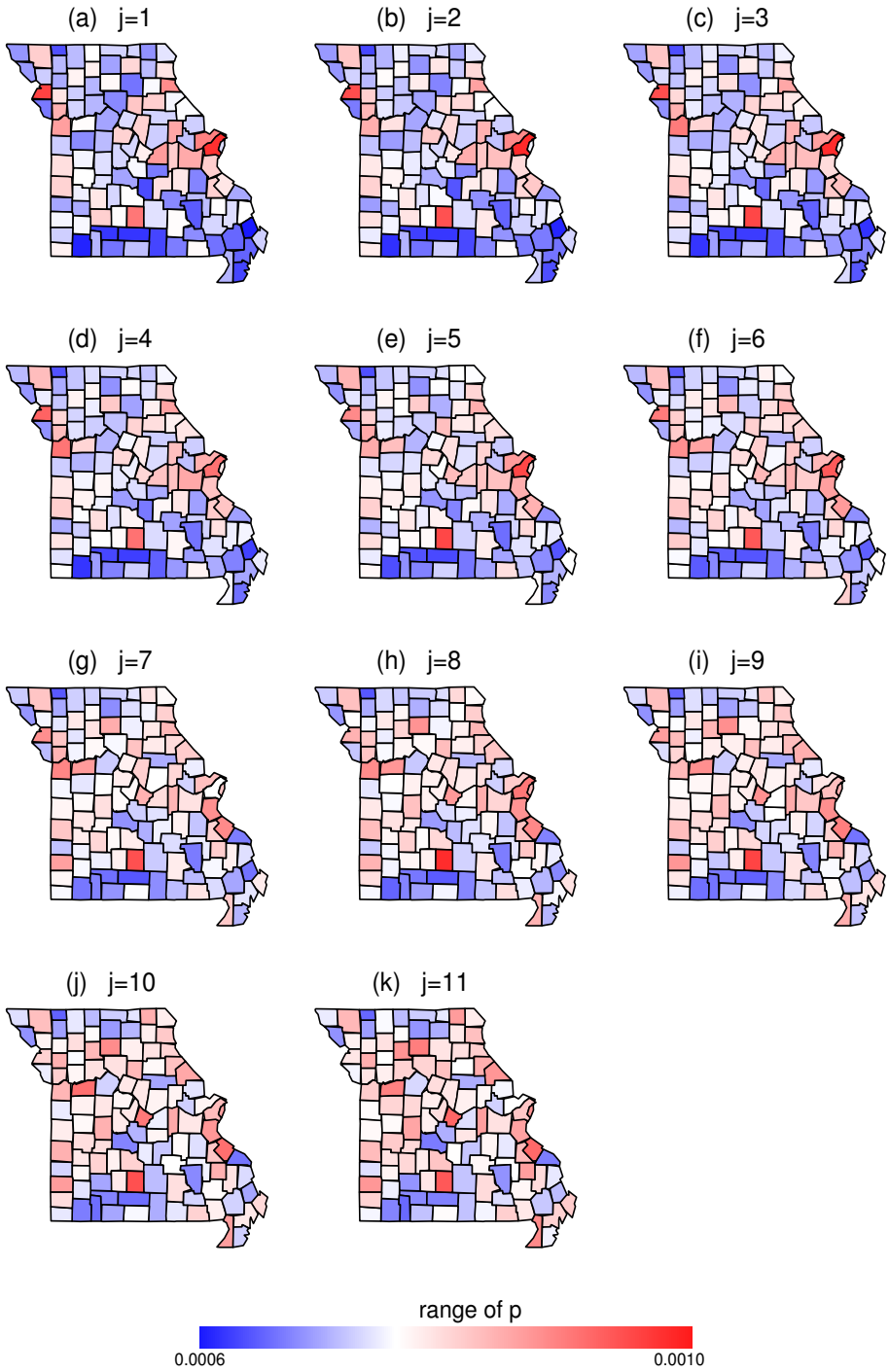

Figure 2.6: Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 for $j = 1, \ldots, 11$ and $k = 3$.

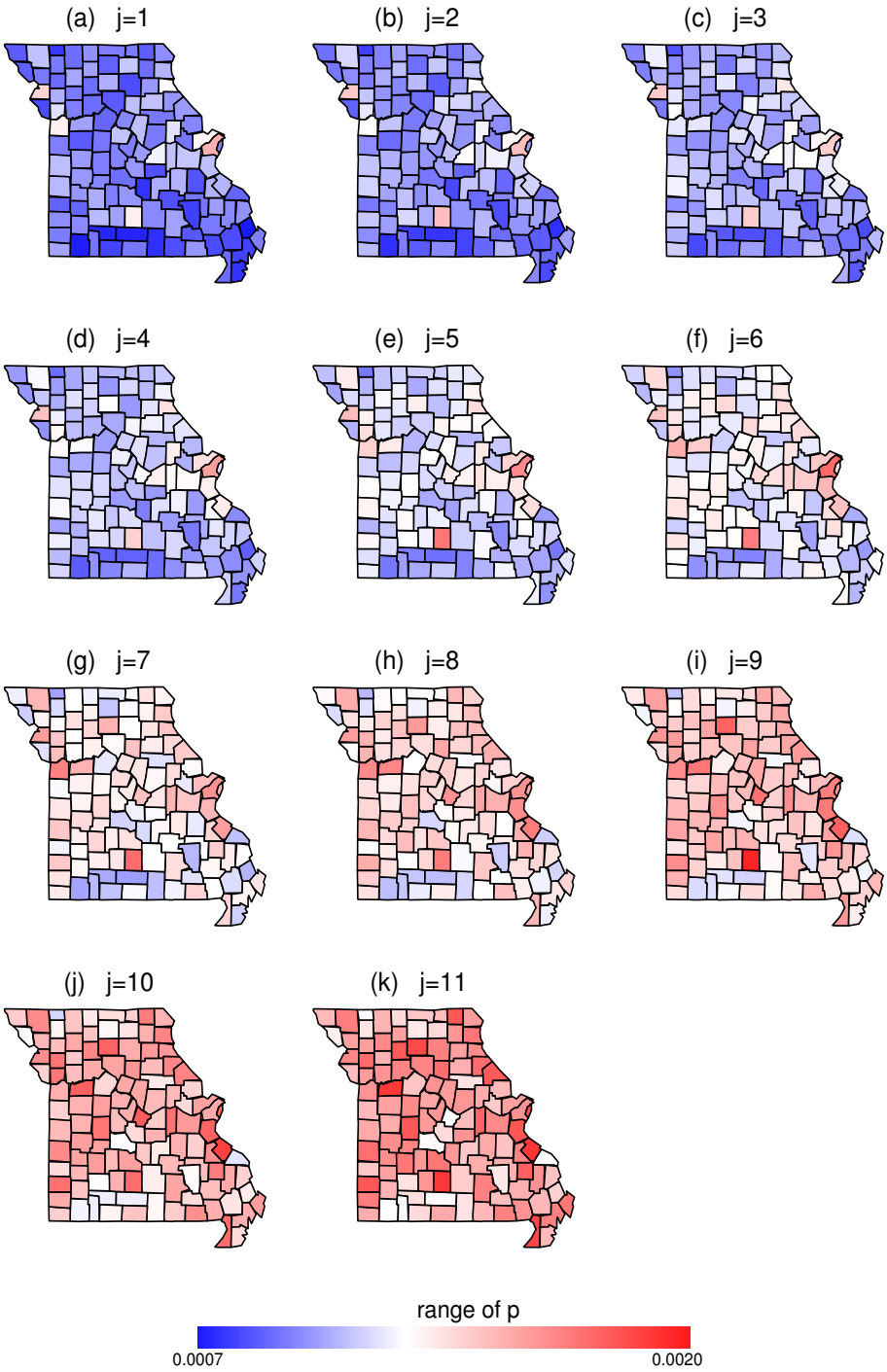

Figure 2.7: Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 for $j = 1, \ldots, 11$ and $k = 4$.

Chapter 3

A Joint Model for Spatial and Temporal Effects

The previous model considered was an additive model that contained separate spatial effects, age effects and a temporal slope term itself containing spatial and age effects. The results of this model show that the mean log rate for a given age group dominated the model. All the observations were essentially shrunk to their respective age group mean rates. This had the effect of obscuring any spatial patterns in the rates. In addition, the temporal slope term was similarly dominated by the mean age group temporal slope, and little spatial effect on the temporal slope was observed. While the posterior density of ρ_1 seemed to indicate the existence of a separate spatial effect, the overall effect in the estimated rates from the model is not easily observed.

In order to overcome these shortcomings in this model, we consider a joint spatio-temporal prior using an intrinsic auto-regressive prior on the temporal trends in the data, and a CAR prior for the spatial effects. This provides for a non-parametric temporal smoothing of the data. The resulting semi-parametric model is a joint

spatio-temporal smoother.

3.1 The Likelihood of the Data and First Stage Prior

The model introduced here is a semi-parametric model that jointly models the spatial and temporal effects, using a non-parametric intrinsic auto-regressive prior as a temporal smoother and a conditional auto regressive prior for the spatial effects. Here we again consider the breast cancer mortality data from Chapter 2. The likelihood of the data is given in (2.1) and (2.8). The difference in this model is in the hierarchical model for p_{ijk}, which in this case is written

$$\log(p_{ijk}) \equiv \nu_{ijk} = z_{ij} + \theta_k + \epsilon_{ijk}, \tag{3.1}$$

where z_{ij} is mean log-rate for the i^{th} county, j^{th} time period and θ_k is the difference between the log-rate of the first age-group and the k^{th} age group. The error term accounting for any other extra variation is then assumed to follow the distribution

$$\epsilon_{ijk} \overset{iid}{\sim} N(0, \delta_0). \tag{3.2}$$

All the effects are additive. This gives rise to the prior for ν_{ijk},

$$(\nu_{ijk} \mid z_{ij}, \theta_k, \delta_0) \sim N(z_{ij} + \theta_k, \delta_0). \tag{3.3}$$

3.2 The Joint Spatio-Temporal Prior

Now in order to define the prior distribution of the z_{ij}, note that $(z_{i1}, z_{i2} \ldots, z_{iJ})$ can be considered a time series for a given county i. To insure the stationarity of a

time series it is often differenced, that is, a new series is defined with elements

$$\nabla x_t = x_t - x_{t+1}. \tag{3.4}$$

In order to generalize this, we define a forwardshift operator D, similar to the backshift operator used in time-series data analysis, as in Shumway & Stoffer (2000)

$$Dx_t = x_{t+1}.$$

This can be extended to powers, $D^2 x_t = D(Dx_{t+1}) = x_{t+2}$, which in the general form gives $D^d x_t = x_{t+d}$. Then (3.4) can be written

$$\nabla^d x_t = (1 - D)^d x_t, \;\; for\, d = 1. \tag{3.5}$$

The higher order difference terms can be found by algebraically expanding (3.5). As an example, the second difference becomes

$$\begin{aligned} \nabla^2 x_t &= (1 - D)^2 x_t \\ &= (1 - 2D + D^2) x_t \\ &= x_t - 2x_{t+1} + x_{t+2}. \end{aligned}$$

For a given vector of time series observations $\boldsymbol{x} = (x_1, x_2, \ldots, x_J)'$, the second order differences can be found by multiplying by the matrix

$$\boldsymbol{B} = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & -2 & 1 \end{pmatrix}_{(J-2)\times J}.$$

The new differenced vector is then

$$\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x}.$$

In the case of our data, if we define $\boldsymbol{Z} = (z_{ij})$, we can define the matrix $\boldsymbol{U} = \boldsymbol{B}\boldsymbol{Z}'$ with the elements.

$$u_{ih} = z_{ih} - 2z_{ih+1} + z_{ih+2}\ ,\ h = 1, 2, \ldots, J-2$$

and

$$\boldsymbol{u}_i' = (u_{i1}, u_{i2}, \ldots, u_{iJ-2})',$$

the i^{th} row vector of $\boldsymbol{U}$. Note that $\boldsymbol{u}_i$ is a stationary temporal process for the i^{th} county, and the vector $\boldsymbol{u}_h^* = (u_{1h}, u_{2h}, \ldots, u_{Ih})'$ is a spatial process for the h^{th} new time period, with a distribution as in (2.6) for any h

$$\boldsymbol{u}_h^* \sim N\left(\boldsymbol{0}, \delta_1(\boldsymbol{I}_I - \rho\boldsymbol{C})^{-1}\right). \tag{3.6}$$

If we define the following

$$\boldsymbol{u} = vec(\boldsymbol{U}) = (u_{11}, u_{12}, \ldots, u_{1J-2}, u_{21}, \ldots, u_{IJ-2})',$$

then the density of $\boldsymbol{u}$ is written as

$$[\mathbf{u} \mid \delta_1, \rho] = \frac{|\mathbf{I_I} - \rho\mathbf{C}|^{\mathbf{1/2}}}{(\mathbf{2}\pi\delta_\mathbf{1})^{\mathbf{I(J-2)/2}}} \exp\left(-\frac{\mathbf{1}}{\mathbf{2}\delta_\mathbf{1}}\mathbf{u}'[\mathbf{I_I} - \rho\mathbf{C}) \otimes \mathbf{I_{J-2}}]\mathbf{u}\right). \tag{3.7}$$

If we define $\boldsymbol{z} = (z_{11}, \ldots, z_{1J}, z_{21}, \ldots, z_{IJ})'$, it can be shown that

$$\boldsymbol{u}'[(\boldsymbol{I}_I - \rho\boldsymbol{C}) \otimes \boldsymbol{I}_{J-2}]\boldsymbol{u} = \boldsymbol{z}'[(\boldsymbol{I}_I - \rho\boldsymbol{C}) \otimes \boldsymbol{A}]\boldsymbol{z}, \tag{3.8}$$

where

$$\boldsymbol{A} = \boldsymbol{B}'\boldsymbol{B} = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ -2 & 5 & -4 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 1 & -4 & 6 & -4 & 1 & \cdots & 0 & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \ddots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & -4 & 6 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -4 & 5 & -2 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & -2 & 1 \end{pmatrix}_{J \times J} \tag{3.9}$$

is a singular matrix of rank $J - 2$. The joint prior density of $\boldsymbol{z}$ is then

$$[\boldsymbol{z}|\delta_1, \rho] \propto \frac{|\boldsymbol{I}_I - \rho\boldsymbol{C}|^{1/2}}{(2\pi\delta_1)^{I(J-2)/2}} \exp\left(-\frac{1}{2\delta_1}\boldsymbol{z}'[(\boldsymbol{I}_I - \rho\boldsymbol{C}) \otimes \boldsymbol{A}]\boldsymbol{z}\right). \tag{3.10}$$

The Kronecker product of $\boldsymbol{I}_I - \rho\boldsymbol{C}$ and $\boldsymbol{A}$ yields a type of precision matrix that is made up of two components, the $\boldsymbol{I}_I - \rho\boldsymbol{C}$ matrix which accounts for the spatial process, and the $\boldsymbol{A}$ matrix which makes up the smoothing of the temporal process; the Kronecker product replicates the spatial function across the temporal processes. The result is that $\boldsymbol{z}$ can be considered a set of spatially correlated temporal processes.

The temporal component of this matrix can be seen as a second order intrinsic auto-regressive (IAR(2)) prior applied to the temporal trend of the data, as related in Besag & Kooperberg (1995) and Künsch (1987). The IAR(2) prior is specifically demonstrated in Fharmeir & Wagenpfeil (1996) in terms of smoothing hazard functions.

3.3 Other Priors

The prior distributions of δ_0, δ_1 are assumed to follow inverse gamma distributions as specified below:

$$[\delta_l] \propto \frac{1}{\delta_l^{a_l+1}} \exp\left(-\frac{b_l}{\delta_l}\right),\ l = 0, 1. \tag{3.11}$$

In order for the matrix $(\mathbf{I_I} - \rho\mathbf{C})$ to be positive definite, the value of ρ is constrained by $(\lambda_I^{-1}, \lambda_1^{-1})$, where $\lambda_1 \leq \ldots \leq \lambda_I$ are the eigenvalues of $\mathbf{C}$. The prior for ρ is then as given in (2.12).

3.4 Computation

If we define $\boldsymbol{\nu} = (\nu_{111}, \ldots \nu_{11K}, \nu_{121}, \ldots, \nu_{IJK})'$, then the full conditional posterior distributions needed to implement the Gibbs sampler are readily calculated.

Lemma 2

(a) For given $(\boldsymbol{z}, \boldsymbol{\theta}, \delta_0; \boldsymbol{y})$, the $(\nu_{111}, \ldots, \nu_{IJK})$ are independent, each ν_{ijk} depends only on y_{ijk}, and

$$[\nu_{ijk} \mid z_{ij}, \theta_k, \delta_0; y_{ijk}] \ \propto\ \exp\left\{\nu_{ijk}y_{ijk} - n_{ijk}e^{\nu_{ijk}} - \frac{1}{2\delta_0}(\nu_{ijk} - z_{ij} - \theta_k)^2\right\}.$$

(b) The conditional posterior distribution of ν_{ijk} is log-concave.

(c) $(\boldsymbol{z} \mid \boldsymbol{\nu}, \boldsymbol{\theta}, \delta_0, \delta_1, \rho; \boldsymbol{y}) \sim N_{IJ}(\boldsymbol{c}, \delta_0\boldsymbol{G}^{-1})$ where $\boldsymbol{G} = K\boldsymbol{I}_{IJ} + \frac{\delta_0}{\delta_1}(\boldsymbol{I}_I - \rho\boldsymbol{C}) \otimes \boldsymbol{A}$, $\boldsymbol{c} = \boldsymbol{G}^{-1}\sum_{k=1}^{K}(\boldsymbol{\nu}_k - \theta_k\mathbf{1_{IJ}})$ and $\boldsymbol{\nu}_k = (\nu_{11k}, \nu_{12k}, \ldots, \nu_{IJk})'$.

(d) $(\delta_0 \mid \boldsymbol{\nu}, \boldsymbol{z}, \boldsymbol{\theta}; \boldsymbol{y}) \sim IG\left(a_0 + IJK/2, b_0 + \frac{1}{2}\sum_{i,j,k}(\nu_{ijk} - Z_{ij} - \theta_k)^2 +\right)$.

(e) $(\delta_1 \mid \boldsymbol{z}; \boldsymbol{y}) \sim IG\left(a_1 + I(J-2)/2, b_1 + \frac{1}{2}\boldsymbol{z}'\left[(\boldsymbol{I}_I - \rho\boldsymbol{C}) \otimes \boldsymbol{A}\right]\boldsymbol{z}\right)$.

(f) $[\rho \mid \boldsymbol{z}, \delta_1; \boldsymbol{y}] \propto \prod_{i=1}^{I}(1-\rho\lambda_i)^{1/2} \exp\left(\frac{\rho}{2\delta_1}\boldsymbol{z}'[\boldsymbol{C}\otimes\boldsymbol{A}]\boldsymbol{z}\right)$.

(g) The conditional posterior distribution of ρ *is log-concave.*

Proof. We only prove the results for log-concavity. For part (b)

$$\frac{\partial^2}{\partial\nu_{ijk}^2}\log[\nu_{ijk} \mid Z_{ij}, \theta_k, \delta_0; y_{ijk}] = -n_{ijk}e^{\nu_{ijk}} - \frac{1}{\delta_0} < 0, \ \forall\, \nu_{ijk}.$$

For part (g), we have

$$\frac{\partial^2}{\partial\rho^2}\log[\rho \mid \boldsymbol{z}, \delta_1; \boldsymbol{y}] = -\frac{1}{2}\sum_{i=1}^{I}\left(\frac{\lambda_i}{1-\rho\lambda_i}\right)^2 < 0, \ \forall\, \rho \in \left(\frac{1}{\lambda_1}, \frac{1}{\lambda_I}\right).$$

The results hold. □

The full conditional posterior distributions of $\boldsymbol{z}, \delta_0, \delta_{1_j}$ are standard forms and can be sampled directly using a Gibbs sampler (Gelfand & Smith 1990). The distributions for the ν_{ijk} and the ρ are shown to be log-concave and can be sampled as in the previous model using the ARS algorithm from Gilks & Wild (1992).

3.5 Results for the Joint Model

The joint model differs from the additive model in several ways. First the use of the IAR(2) prior on the temporal gradient of the data, creates a new set of observations that are all independent and identically distributed. This joint model also allows for the use of a single prior for spatio-temporal effects. This results in a simpler model that only includes a parameter for the age-effects in addition to the spatio-temporal effect. This parameterization avoids the cumbersome form of the additive model, where there are spatial effects and then temporal slopes that are spatially correlated, age group effects and age group specific temporal slopes. In the additive

model updating the full conditional distributions of $\boldsymbol{z}$ and $\boldsymbol{W}$ requires the inversion of two $I \times I$ matrices for each iteration of the Gibbs sampler. In the joint model, there is only one matrix to invert. Unfortunately, the dimension of the matrix is $IJ \times IJ$, which increases the computational difficulty of updating the conditional posterior distribution of $\boldsymbol{z}$.

3.5.1 Computational Improvements

Updating the $\boldsymbol{z}$ using the conditional posterior distribution based on the current value of ρ requires the inversion of an $IJ \times IJ$ matrix. In order to speed up the process, it is desirable to diagonalize this matrix.

- Begin by letting $\boldsymbol{Q}_C$ be an orthogonal matrix such that $\boldsymbol{C} = \boldsymbol{Q}_C \boldsymbol{\Lambda}_C \boldsymbol{Q}_C^{'}$, where $\boldsymbol{\Lambda}_C$ is the diagonal matrix of eigenvalues of $\boldsymbol{C}$, and let $\boldsymbol{Q}_A$ be an orthogonal matrix such that $\boldsymbol{A} = \boldsymbol{Q}_A \boldsymbol{\Lambda}_A \boldsymbol{Q}_A^{'}$, where $\boldsymbol{\Lambda}_A$ is the diagonal matrix of eigenvalues of $\boldsymbol{A}$.

- Then define $\tilde{\boldsymbol{Q}} = \boldsymbol{Q}_C \otimes \boldsymbol{Q}_A$, $\tilde{\boldsymbol{Q}}^{'} = \boldsymbol{Q}_C^{'} \otimes \boldsymbol{Q}_A^{'}$, and

$$\boldsymbol{\Phi}_\rho = K\boldsymbol{I}_{IJ} + ((\boldsymbol{I}_{IJ} - \rho\boldsymbol{\Lambda}_C) \otimes \boldsymbol{\Lambda}_A). \tag{3.12}$$

  Note that $\tilde{\boldsymbol{Q}}$ is orthogonal, and $\boldsymbol{\Phi}$ is diagonal and now easily invertible. It can then be shown that

$$K\boldsymbol{I}_{IJ} + ((\boldsymbol{I}_I - \rho\boldsymbol{C}) \otimes \boldsymbol{A}) = \tilde{\boldsymbol{Q}}\boldsymbol{\Phi}_\rho\tilde{\boldsymbol{Q}}', \tag{3.13}$$

$$(K\boldsymbol{I}_{IJ} + ((\boldsymbol{I}_I - \rho\boldsymbol{C}) \otimes \boldsymbol{A}))^{-1} = \tilde{\boldsymbol{Q}}\boldsymbol{\Phi}_\rho^{-1}\tilde{\boldsymbol{Q}}'. \tag{3.14}$$

We prove the first equality only:

$$\begin{aligned}
&(\boldsymbol{Q}_C \otimes \boldsymbol{Q}_A)(K\boldsymbol{I}_{IJ} + ((\boldsymbol{I}_I - \rho\boldsymbol{\Lambda}_C) \otimes \boldsymbol{\Lambda}_A))(\boldsymbol{Q}'_C \otimes \boldsymbol{Q}'_A) \\
&= K\boldsymbol{I}_{IJ} + \boldsymbol{I}_I - \rho(\boldsymbol{Q}_C\boldsymbol{\Lambda}_C \otimes \boldsymbol{Q}_A\boldsymbol{\Lambda}_A)(\boldsymbol{Q}'_M \otimes \boldsymbol{Q}'_A) \\
&= K\boldsymbol{I}_{IJ} + \boldsymbol{I}_I - \rho(\boldsymbol{Q}_C\boldsymbol{\Lambda}_C\boldsymbol{Q}'_C \otimes \boldsymbol{Q}_A\boldsymbol{\Lambda}_A\boldsymbol{Q}'_A) \\
&= K\boldsymbol{I}_{IJ} + (\boldsymbol{I}_I - \rho\boldsymbol{C}) \otimes \boldsymbol{A}.
\end{aligned}$$

Now to update $\boldsymbol{z}$ the following algorithm is implemented.

1. Update $\boldsymbol{\Phi}_\rho$.
2. Simulate $\boldsymbol{s} \sim N_{IJ}(\boldsymbol{0}, \boldsymbol{I}_{IJ})$.
3. Compute $\boldsymbol{r}_\lambda = \boldsymbol{\Phi}_\lambda^{-\frac{1}{2}}\boldsymbol{s} + \boldsymbol{\Phi}_\lambda^{-1}\tilde{\boldsymbol{Q}}' \sum_{k=1}^{K}(\boldsymbol{\nu}_k - \theta_k \mathbf{1}_{\mathbf{IJ}})$.
4. Let $\boldsymbol{z} = \tilde{\boldsymbol{Q}}\boldsymbol{r}_\lambda$, which has the same distribution as in Lemma 6(a).

3.5.2 Noninformative and Data-dependent Priors

The model is initially run using noninformative priors for δ_0 and δ_1

$$\pi(\delta_l) \propto \frac{1}{\sqrt{\delta_l}},\ l = 0, 1.$$

The resulting statistics from the samples of their posterior distributions are shown in Table 3.1. The posterior means and variances are used to calculate a set of data-dependent priors for δ_0 and δ_1, as discussed in Section 2.6.1.

3.5.3 Interpreting ρ

The trace and density plot of the parameter ρ both indicate that most of the mass of the density is located at the positive limit for the value of ρ. While strictly

Table 3.1: Quantiles of δ_0, δ_1 and ρ for Noninformative (NI) and Inflation Factor (IF) Data-dependent Priors

Summary of Posterior Distributions								
	Prior	Min.	1st Qt.	Median	Mean	3rd Qt.	Max.	Std. Dev.
δ_0	NI	.00233	.00675	.00821	.00835	.00974	.01927	.00223
	IF 2, 200	.00278	.00676	.00811	.00830	.00961	.01706	.00216
	IF 1.5, 25	.00237	.00635	.00771	.00787	.00929	.01724	.00210
	IF 2.5, 500	.00343	.00701	.00825	.00844	.00965	.01729	.00200
δ_1	NI	.00003	.00032	.00047	.00052	.00067	.00222	.00028
	IF 2, 200	.00009	.00032	.00044	.00048	.00058	.0020	.00023
	IF 1.5, 25	.00011	.00034	.00047	.00050	.00061	.00210	.00022
	IF 2.5, 500	.00010	.00034	.00046	.00049	.00060	.00225	.00022
ρ_1	NI	.08411	.1734	.1745	.1736	.1751	.1756	.00342
	IF 2,200	.1083	.1736	.1746	.1739	.1751	.1756	.00250
	IF 1.5, 25	.0691	.1735	.1745	.1737	.1750	.1756	.0035
	IF 2.5, 500	.1313	.1735	.1745	.1738	.1750	.1756	.00225

speaking ρ is not a measure of spatial correlation, if $\rho = 0$ then there is no spatial correlation in the prior and the $\boldsymbol{z}$ would be independent, and unnecessary. On the contrary, these results for the posterior density of ρ seems to indicate that the model is not sufficiently smooth.

3.5.4 The Variance Components δ_0 and δ_1

The two variance components δ_0 and δ_1 show good behavior under the various priors and appear to be robust in the selection of hyperparameters. While a direct comparison with the variance components of the additive model is not possible, it is possible to compare in general their relative magnitudes. In the additive model the relative magnitudes or the ratio of δ_0/δ_1 is the inverse of in the joint model. This would seem to indicate that the relative importance of $\boldsymbol{z}$ in the joint model is greater than in the additive model. This impression is reinforced in looking at the results of

the smoothed estimates of p_{ijk} from each model.

3.6 Comparison between the Additive model and the Joint Model

The results for the joint model provide an interesting comparison to the additive model. As previously mentioned, the additive model seems to shrink the estimates toward an age group mean, while failing to preserve any spatial pattern in the data. This is clear in the sample scatter plots of the additive model estimates versus the joint model estimates. The slope of the resulting scatter plots is almost vertical, showing that the estimates from the additive model vary much less than the results from the joint model. This reveals that the joint model is actually allowing more of the original spatial heterogeneity to remain in the model.

3.6.1 Maps

This result can further be seen in maps comparing the two models and the raw data. Both sets of maps show that the joint model estimates retain more of the original data's spatial distribution. Regions that have a distinctly lower or higher incidence rates still show up on the joint model maps. This is evident in the southeast portion of the state, Reynolds and Shannon counties in particular. This regional heterogeneity is also better preserved in the maps for each age group. Again the same pattern manifests itself in the southeastern portion of the state, particularly in Reynold and Shannon counties, though over time the trends are not as strong.

3.6.2 Model Selection Criteria

Comparison between the two models is also aided by the use of model selection criteria. The first of these criteria is the DIC proposed by Spiegelhalter et al. (2002), is easily implemented in the MCMC code. The DIC measure is based on the deviance D defined as

$$-2log(p(y \mid \theta)).$$

The expected value of the deviance for the model is a measure of the goodness-of-fit, the smaller this value the better the model fits the data. The expectation of the deviance can be estimated from the MCMC output as

$$\bar{D} = E_\theta(D(\theta)) \tag{3.15}$$

Like other model selection criteria the DIC is the sum of a measure of model complexity, and a measure of goodness of fit. The lower this sum the more desirable the model. The measure of model complexity for the DIC is P_D, which is defined as

$$E_\theta(D(\theta)) - D(E_\theta(\theta)).$$

This quantity can be estimated easily from the MCMC output as

$$P_D = \bar{D} - D(\bar{\theta})$$

As a result the DIC is easily estimated from the output of an MCMC sampler,

$$DIC = P_D + \bar{D} \equiv 2\bar{D} - D(\bar{\theta}). \tag{3.16}$$

Selecting the model with the lowest DIC is similar to other criteria and the notion of model selection where it is desirable to select the simplest best fitting model. The results for the additive and joint model are shown in Table 3.2.

Table 3.2: DIC Values for the Additive and Joint Models

Model Comparison Using DIC			
Model	P_D	$\bar{D}$	DIC
Additive	198.57	434911	435109
Joint	443.909	449051	449495

Table 3.3: $D(m)$ Values for the Additive and Joint Models

Model Comparison Using D(m)			
Model	$P(m)$	$G(m)$	$D(m)$
Additive	64.45	18692.64	18757.09
Joint	188.78	30016.74	30205.52

The second criteria, is from Gelfand and Ghosh (1998), selects the model that minimizes the posterior expected loss. This is shown to be the equivalent of minimizing the quantity

$$D(m) = G(m) + P(m),$$

where $G(m)$ is the sum of squares predictive error, and $P(m)$ is the sum of the predictive variances. The term $G(m)$ acts as a goodness-of-fit measure while the $P(m)$ is a penalty term. The interpretation is intuitive. Simple models suffer under both $G(m)$ and $P(m)$, whereas over-fitted models tend to have larger predictive variances. The results for these two models are shown in Table 3.3.

These model selection criteria provide two similar results. The DIC and the $D(m)$ criteria select the additive model, citing both better fit and a less complex model. The joint model however shows a posterior density for ρ that is heavily skewed toward the positive limit of its prior density. This is thought to suggest that there is in fact some degree of spatial heterogeneity that is not being explained by either the additive or the joint model. The suggestion here is that both models are in some sense under-

smoothing, in that they are unable to adequately explain the spatial pattern and in the case of the additive model are shrinking the data estimates toward a mean. In the case of the joint model, the heavily skewed value for the posterior density of ρ suggest that the model is under-smoothing the data. These reasons suggest that the results of the model selection criteria are misleading and leads to the conclusion that another form of spatial prior might be beneficial and provide better results.

3.7 Conclusions

The use of a joint spatial temporal semi-parametric model has shown itself to be potentially beneficial and to possibly provide good smoothing characteristics. The constraints on the smoothing parameter ρ placed by the propriety of the CAR prior, and the resulting posterior distribution for ρ indicate that the data actually may require more nuanced smoothing than is provided by the CAR prior. This leads to the notion that there may be a more suitable prior to use for the spatial effects. This will lead to the next example where we derive a joint spatio-temporal model using a prior based on thin-plate splines for spatial effects, and compare it to the joint model using the CAR prior for spatial effects presented here.

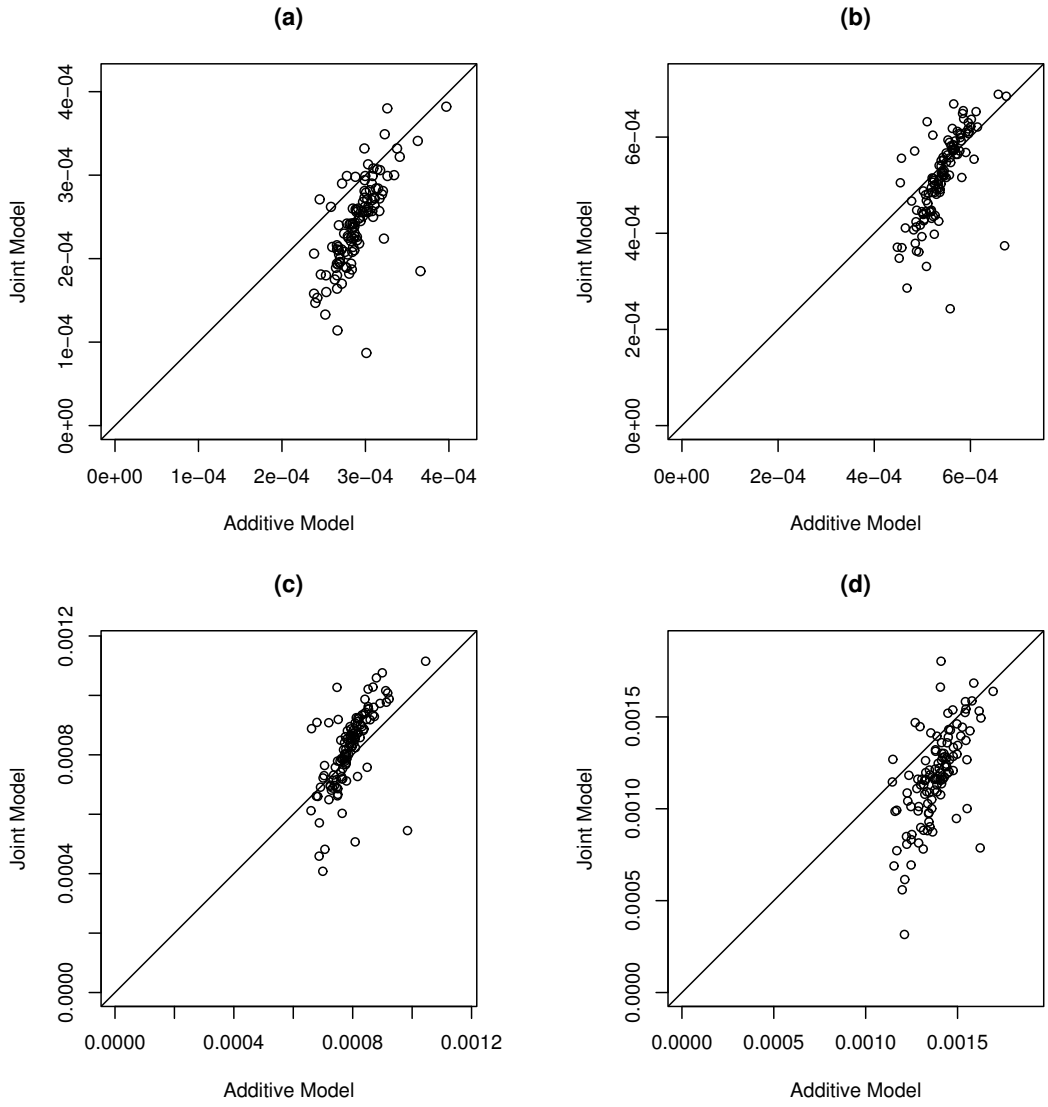

Figure 3.1: Scatterplots Comparing the Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 from the Additive and Joint CAR Models (a) $(j,k) = (3,1)$, (b) $(j,k) = (5,2)$, (c) $(j,k) = (7,3)$, (d) $(j,k) = (11,4)$.

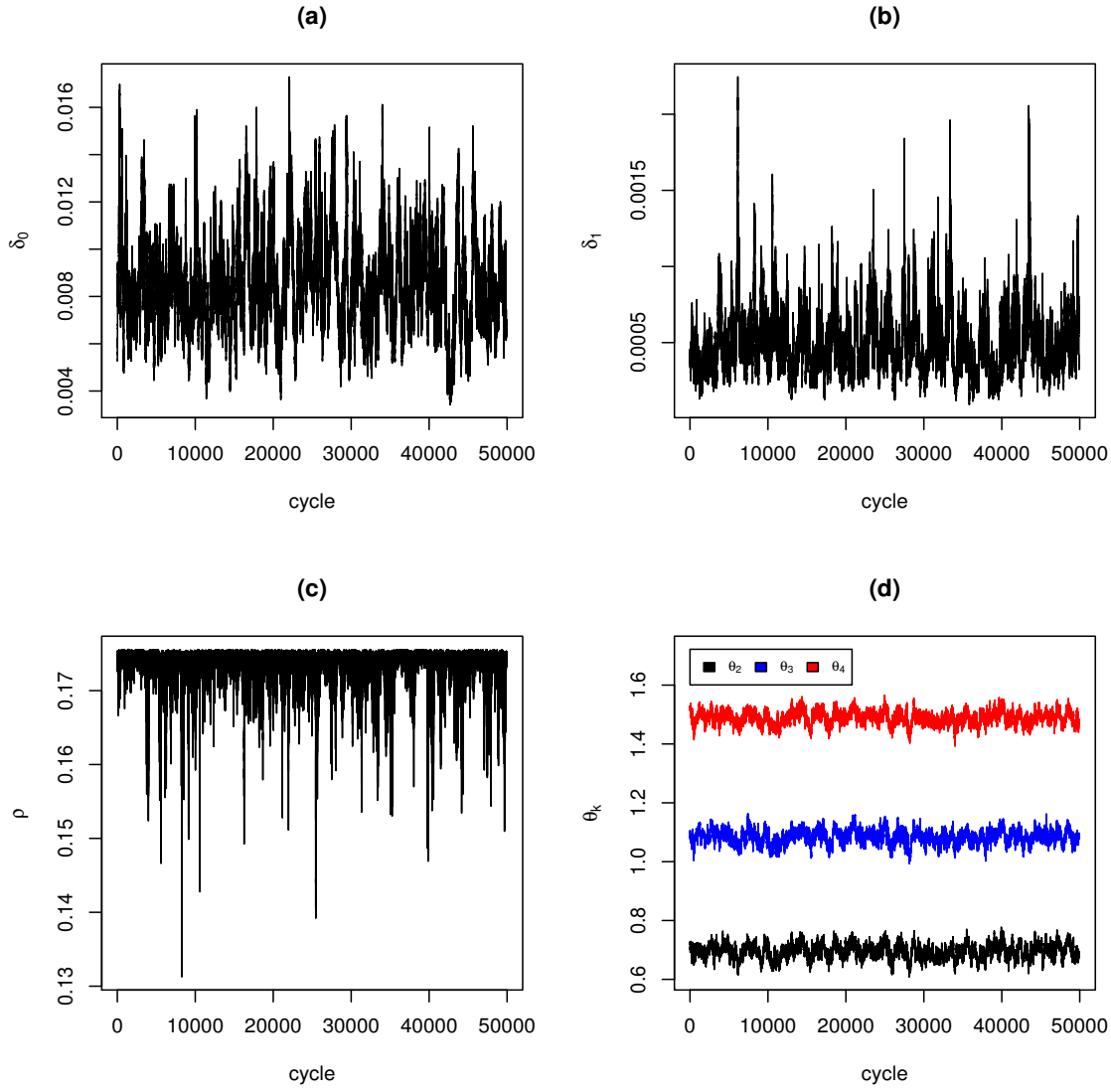

Figure 3.2: Trace plots of (a) δ_0, (b) δ_1, (c) ρ, and (d) θ_k, $k = 2, 3, 4$ from the Joint CAR Model for Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1969-2000.

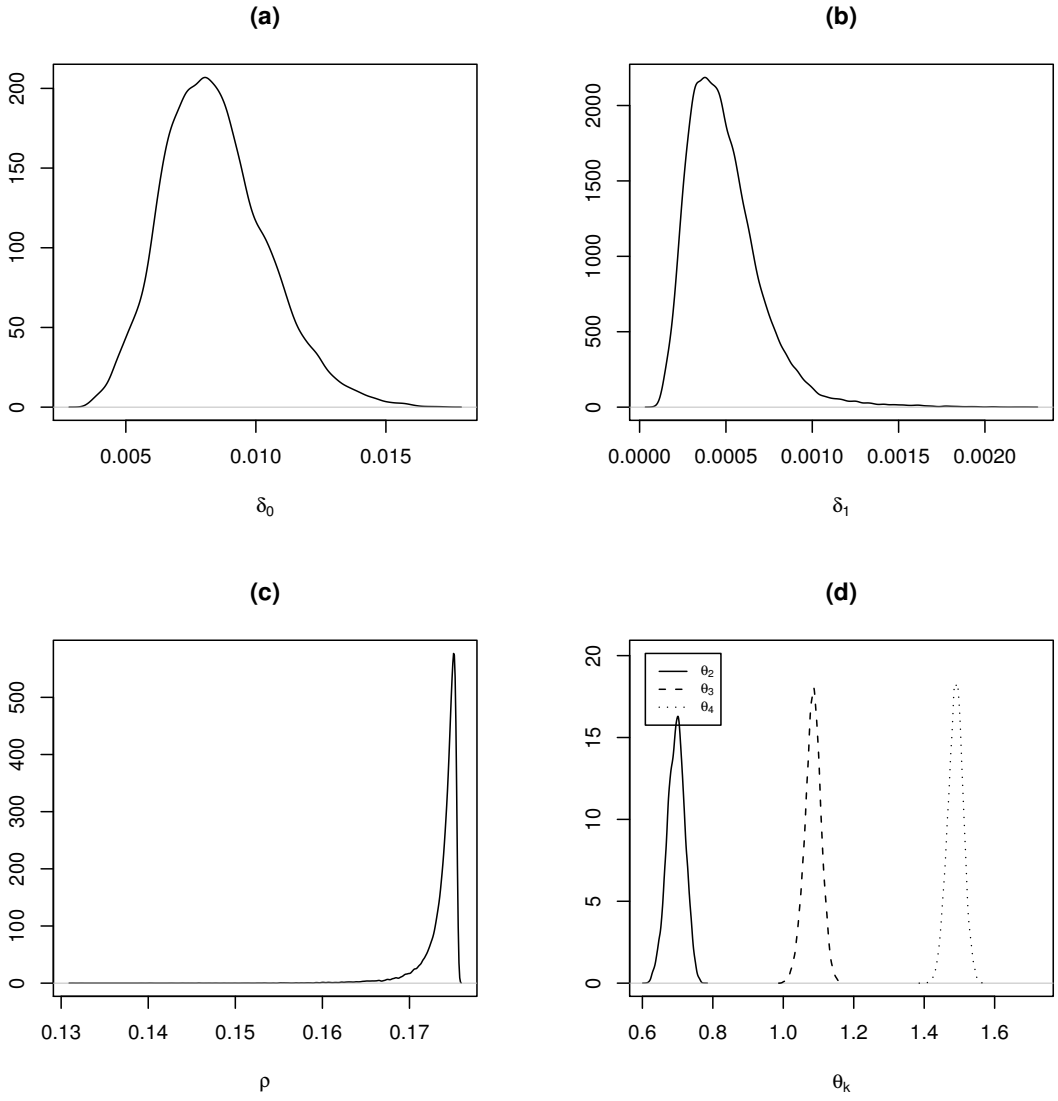

Figure 3.3: Posterior Densities of (a) δ_0, (b) δ_1, (c) ρ, and (d) θ_k, $k = 2, 3, 4$ from the Joint CAR Model for Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1969-2000.

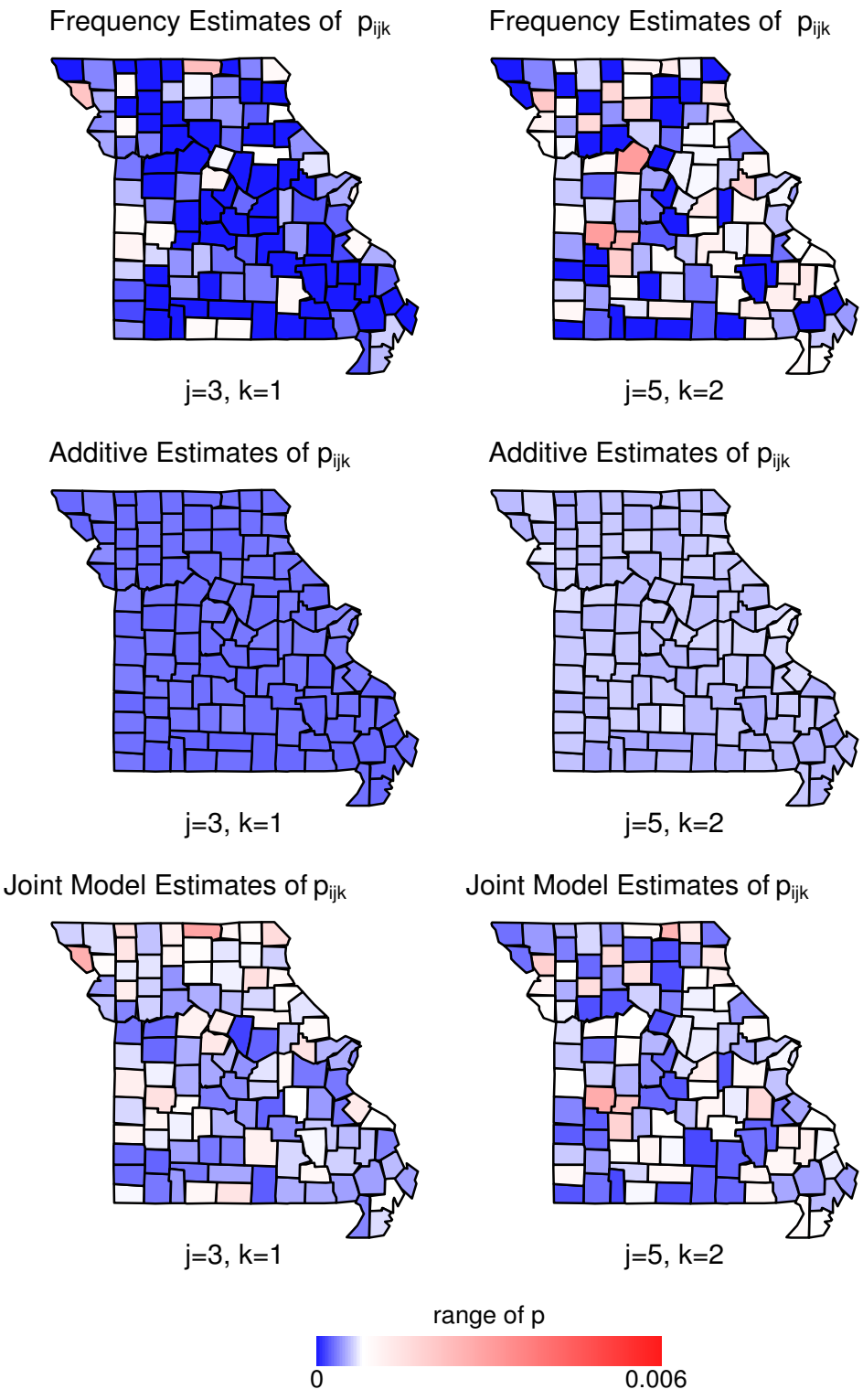

Figure 3.4: Maps of Frequency and Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 from the Additive and Joint CAR Models for $(j,k) = (3,1)$ and $(j,k) = (5,2)$.

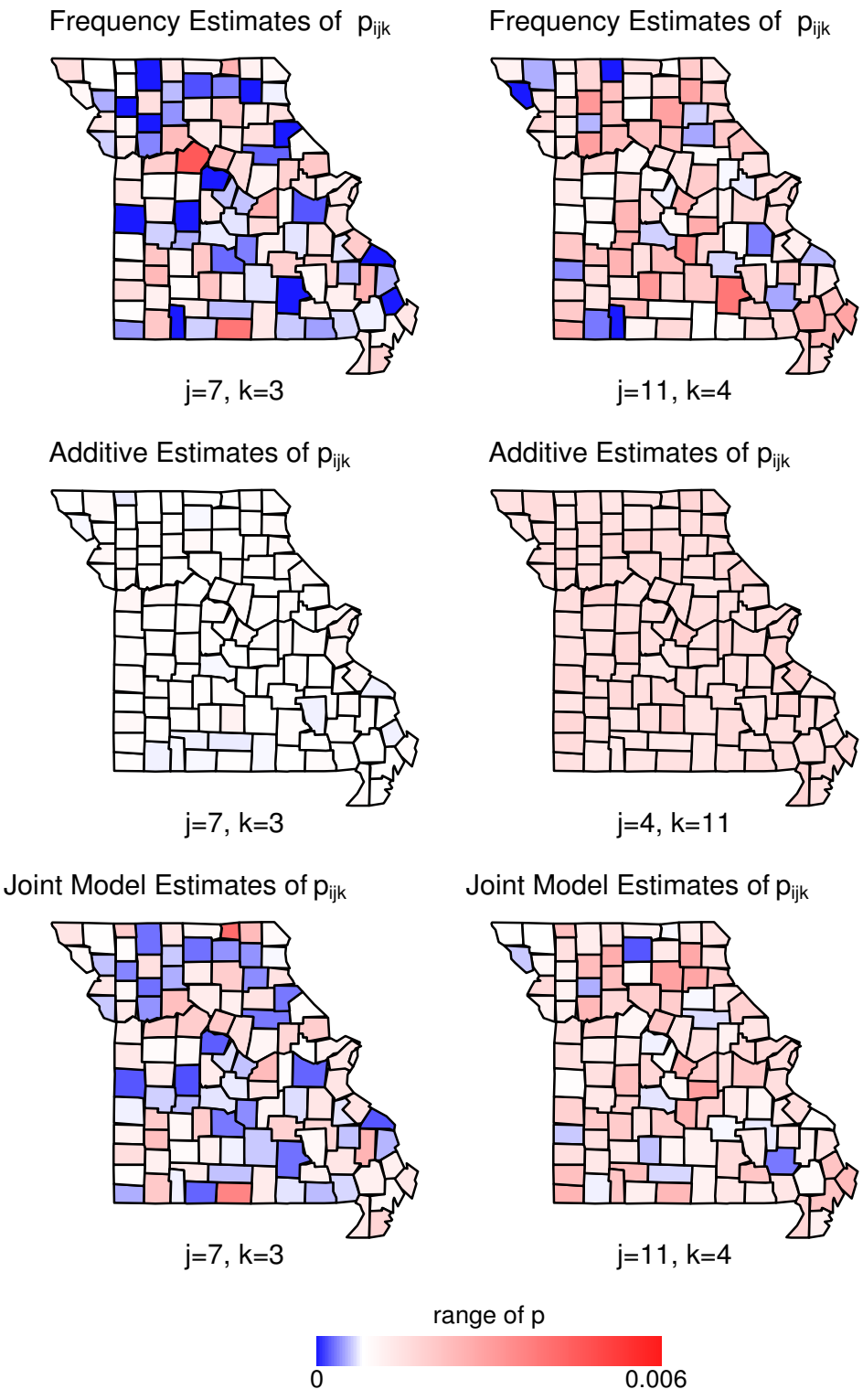

Figure 3.5: Maps of Frequency and Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 from the Additive and Joint CAR Models for $(j,k) = (7,3)$ and $(j,k) = (11,4)$.

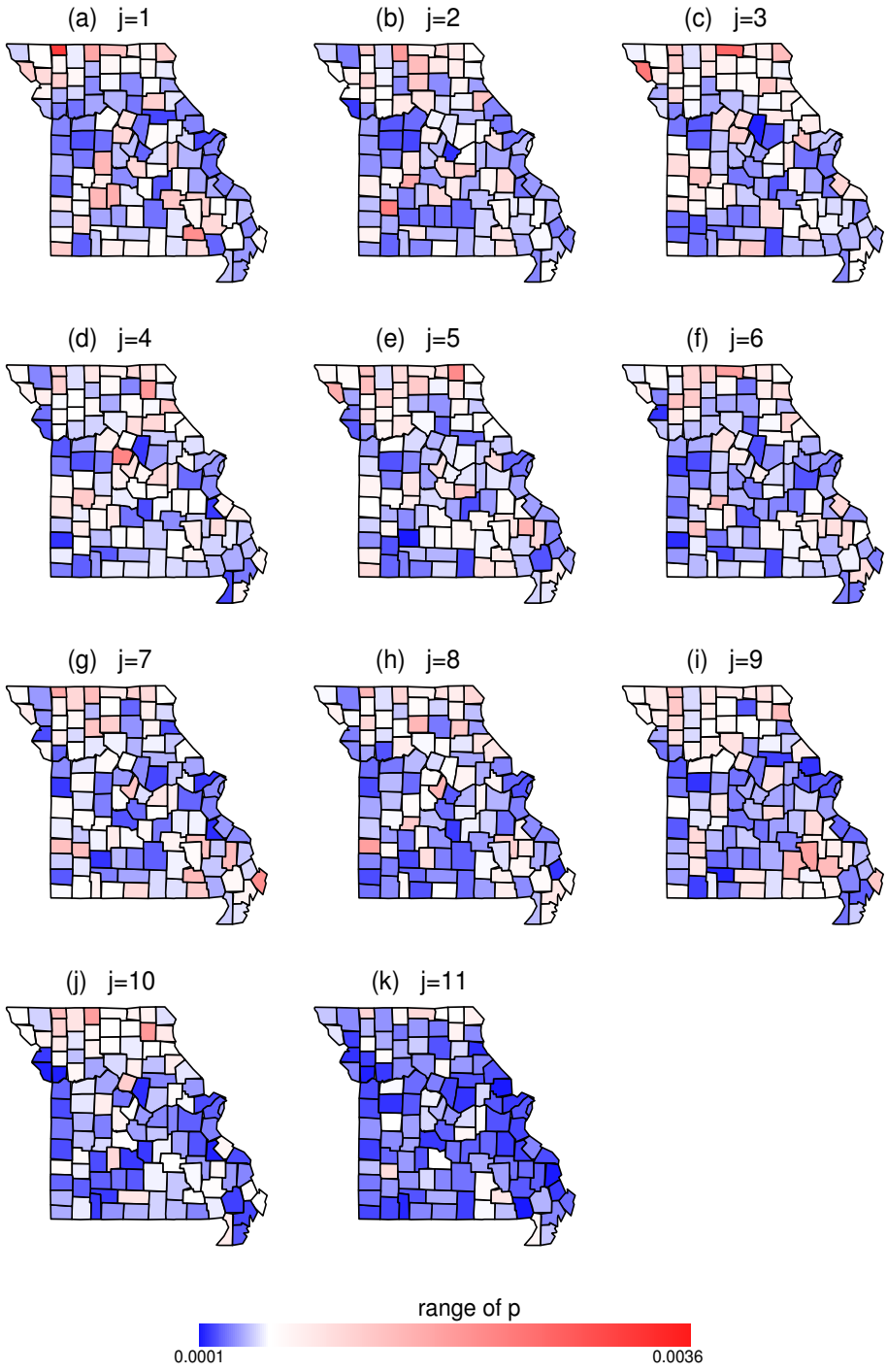

Figure 3.6: Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 from the Joint CAR Model for $j = 1, \ldots, 11$ and $k = 1$.

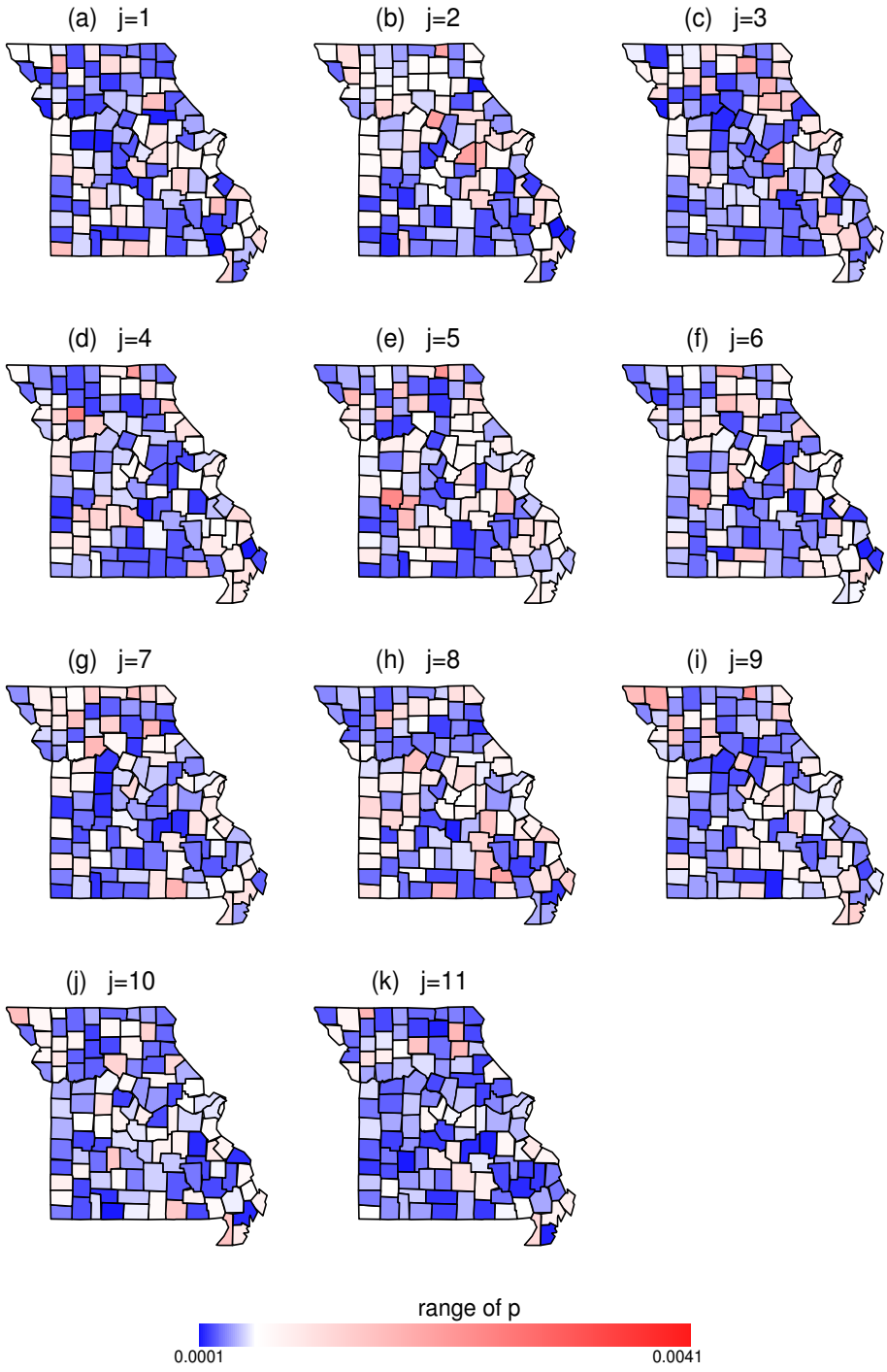

Figure 3.7: Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 p_{ijk} from the Joint CAR Model for $j = 1, \ldots, 11$ and $k = 2$.

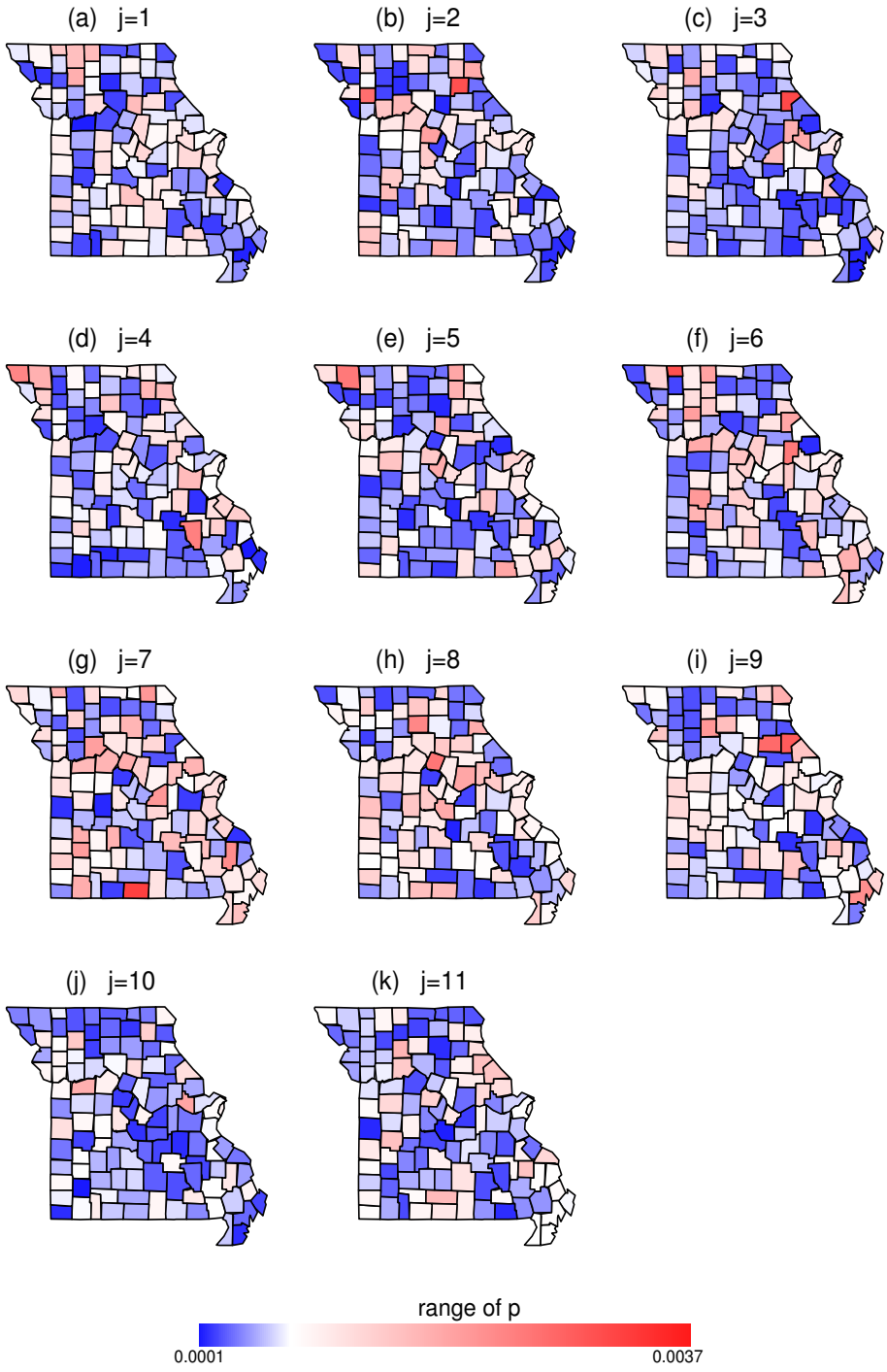

Figure 3.8: Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 from the Joint CAR Model for $j = 1, \ldots, 11$ and $k = 3$.

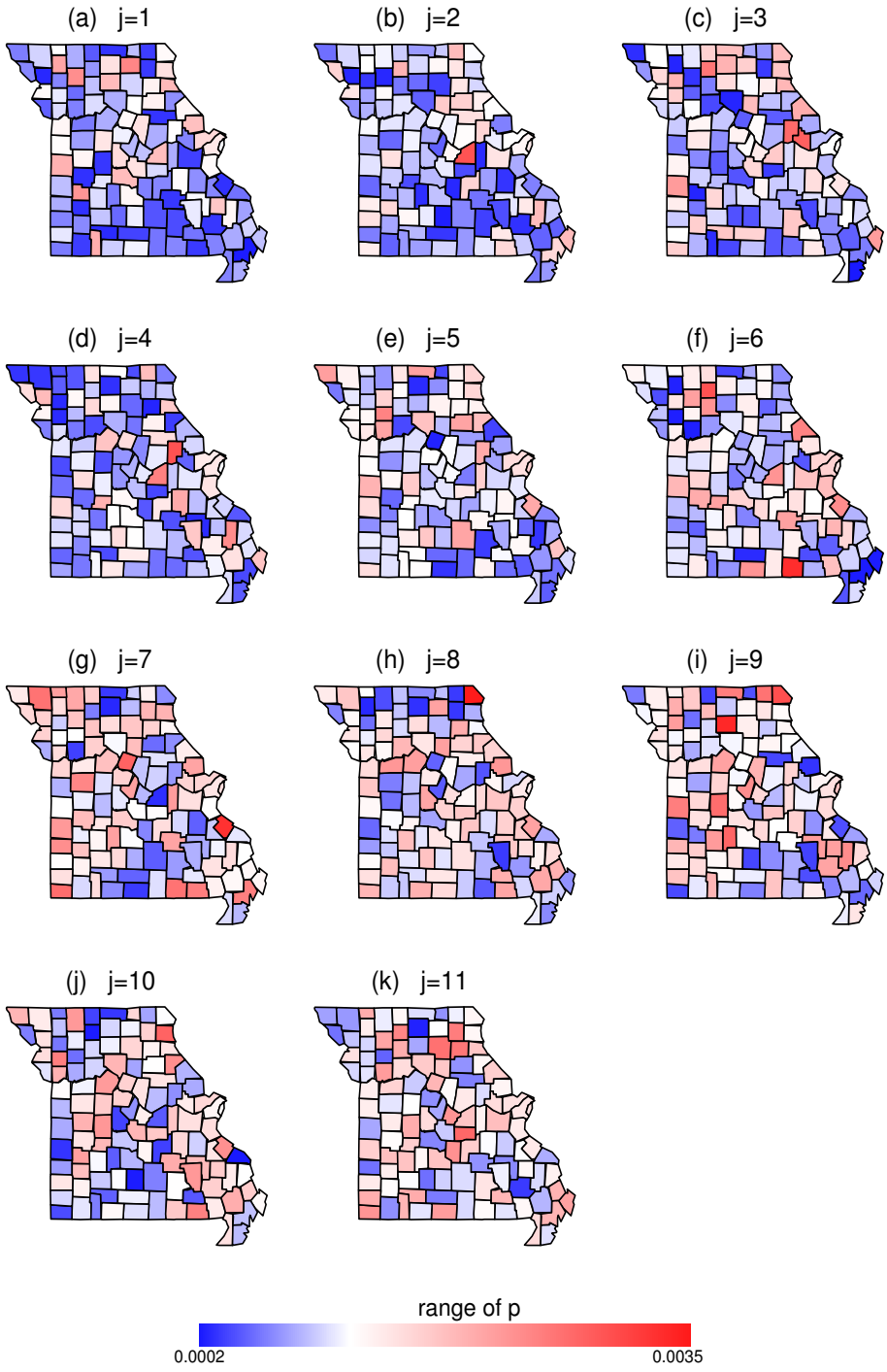

Figure 3.9: Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 from the Joint CAR Model for $j = 1, \ldots, 11$ and $k = 4$.

Chapter 4

A Semiparametric Spatio-Temporal Model Using the Thin-Plate Spline Prior

Spatial data occurs in many contexts, and the detection of trends in or clustering of such data is often the central question of interest in data analysis. Often the data collected has noise associated with it that conceals existing spatial patterns. The use of spatial covariance functions or smoothing functions seeks to remove noise from the observed data by a variety of mechanisms, depending on the method employed. No matter what the method used, the end result is an estimate of the desired quantity preserving the spatial pattern, free from the noise obscuring that pattern in the observed data. A good smoothing function should have the properties of removing noise from the data and providing reasonable estimates of the desired quantity. This should be accomplished without over-smoothing the data, thus preserving existing spatial clustering or trends in the data. For these reasons, thin-plate splines as im-

plemented here appear to be a reasonable method to be used as a spatial smoother. Results show that the smoother is in fact relatively simple to implement and compares favorably with other methods of spatial data smoothing, namely a CAR model applied to areal data. Examples of the specific use of thin-plate splines as spatial smoothers can be found in Wahba et al. (1995) and van der Linde et al. (1995). In addition, the comparison between the use of thin-plate splines and other non-parametric smoothing functions and more traditional geo-statistical techniques such as kriging have been made in Laslett (1994), Hutchinson & Gessler (1994), Laslett & McBratney (1990), as well as Nychka (2000), who provides several examples as well. The methods presented here differ slightly from the traditional representation of thin plate splines seen in texts. The derivation of the solution is identical to those presented in other sources such as Wahba (1990) and Green & Silverman (1994), who provide a thorough technical and historical coverage of smoothing splines. The model here is derived using Bayesian methodology as suggested in Wahba (1978), Wahba (1983), and Kimmeldorf & Wahba (1970) and Kimmeldorf & Wahba (1971), and implemented in a Gibbs sampler.

Typically thin-plate splines are used for point referenced data. In this paper the data used are collected at the areal level and include issues of missing or sparse data for several areas. There exists a well established body of work concerning the use of areal models for modeling spatial data that address these issues. The history of these models is well known, beginning with Besag (1974) who introduced the conditional auto-regressive CAR model, now one of the most used models for disease mapping and other applications. Examples of the CAR model include Clayton & Kaldor (1987), Cressie & Chan (1989), Marshall (1991), Waller et al. (1997), and He & Sun (2000), among many others. This popularity is due in part to the ease of implementation

using a Gibbs sampler as first suggested by Gelfand & Smith (1990). He & Sun (2000) implement a CAR model as a spatial smoother in a generalized linear hierarchical model evaluated in a Bayesian context using Gibbs sampling to analyze the same dataset presented in this paper. There are some slight differences in the construction of the generalized linear models in these two papers and the hierarchical structures, yet the similarities allow for comparison between the two models as spatial smoothers for the data in question.

Thin-plate splines allow for the smoothing of rough multi-dimensional point referenced data by fitting a smoothed surface to the data. This fitting is accomplished by maximizing a penalized likelihood function; the penalty corresponding to certain smoothness conditions on the fitted surface. These conditions provide the smoothing effect of the thin-plate spline. Rather than fitting an overall mean plane to the data as in a least squares solution, the thin-plate spline penalty term allows for variation in the response surface, preserving regional heterogeneity, while enforcing a smooth transition between regions. It is natural then to apply thin-plate splines to point referenced spatial data as a smoothing function. This paper implements a thin-plate spline based model for spatial data in order to evaluate the efficacy and ease of using thin-plate splines as a spatial smoother for areal data in comparison to the use of a conditional auto-regressive (CAR) function for spatial effects. Furthermore, the fitting of the model is performed in a Bayesian context using posterior means as point estimates of the smoothed values of the observed data.

Over the past decades a great deal of effort has gone into the collection and compilation of high quality data on cancer incidence and mortality. However, only recently have efforts been made toward more in-depth analysis of the data, including

the spatial-temporal modeling of incidence and mortality rates. While various models previously presented have used a variety of spatial smoothing techniques in a Bayesian context and have incorporated temporal effects, here we present a spatio-temporal model that is semiparametric and uses a single prior for both spatial and temporal effects. The spatial smoother is based on the thin-plate spline solution while the temporal smoothing is based on an intrinsic auto-regressive (IAR) model. The results show good computational characteristics for this model as well as promising results in terms of comparison with other spatio-temporal models.

The first section of the paper contains a cursory presentation of the thin-plate spline material as relevant to the development of the model as well as the Bayesian interpretation of the thin-plate spline solution. The temporal smoothing by use of an IAR(2) prior is also presented and incorporated into the model. The full conditional distributions are derived as well as a technique for simplifying the MCMC sampling by diagonalizing the covariance matrix used to sample from the full conditionals. The second section tests the thin-plate spline based model for robustness using noninformative and data-dependent priors. The estimates of the mortality rates for the thin-plate spline model, the CAR model, and the raw data are compared and mapped. Model selection and testing are also considered using the DIC and $D(m)$. The final section discusses the implication of these results and comments on further areas of investigation.

4.1 Thin-Plate Splines

The material presented here is intended to highlight the representation of the thin-plate spline solution that is unique to this paper, not as a thorough demonstration of

the derivation of the thin-plate spline solution. For a more thorough demonstration of this material see Nychka (2000). For an elegant derivation of the background material concerning thin-plate splines, see Duchon (1977) or Meinguet (1979)

4.1.1 Derivation

Consider the non-parametric regression problem,

$$y_i = f(\boldsymbol{x}_i) + \epsilon_i, \;\; i = 1, \ldots, n, \tag{4.1}$$

where f is an unknown function on fixed domain $\mathcal{D} \subset \mathbb{R}^d$, $\boldsymbol{x}_i \in \mathcal{D}$ are fixed points, and errors ϵ_i are iid (independently identically distributed) $N(0, \delta_0)$.

To estimate unknown f, we consider the penalized sum of squares

$$S_\eta(f) = \frac{1}{n} \sum_i^n w_i (y_i - f(\boldsymbol{x}_i))^2 + \lambda J_m(f) \tag{4.2}$$

for some $\lambda > 0$, where w_i are some fixed constants. The estimate of $\hat{f}$ at $\boldsymbol{x}_i$ is the minimizer of $S_\eta(f)$. The term $J_m(f)$ can be thought of as a "roughness" penalty, and is defined as

$$J_m(f) = \int_{\mathbb{R}^d} \sum \frac{m!}{\alpha_1! \ldots \alpha_d!} \left(\frac{\partial^m f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}} \right)^2 d\boldsymbol{x}. \tag{4.3}$$

The sum in the integrand is taken over all the non-negative integer vectors $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)'$ such that $\alpha_1 + \ldots + \alpha_d = m$, where $2m > d$. In the case of spatial data, $d = 2$. It is then common to choose $m = 2$. Matheron (1973) and Duchon (1977) show that the solution to (4.2) belongs to the finite dimensional space

$$f(\boldsymbol{x}) = \sum_{j=1}^t \phi_j(\boldsymbol{x}) \beta_j + \sum_{i=1}^n \psi_i(\boldsymbol{x}) \gamma_i, \tag{4.4}$$

where $(\phi_1, \ldots \phi_t)$ is a set of functions that span the space of all d-dimensioned polynomials of degree less than m. For $d = 2$ and $m = 2$, $t = 3$ and $\boldsymbol{x} = (x_1, x_2)$. As a

result

$$\phi_1(\boldsymbol{x}) = 1, \;\; \phi_2(\boldsymbol{x}) = x_1, \;\; \phi_3(\boldsymbol{x}) = x_2.$$

For $d = 2$ and $m = 3$, $t = 6$, we can choose

$$\phi_1(\boldsymbol{x}) = 1, \;\; \phi_2(\boldsymbol{x}) = x_1, \;\; \phi_3(\boldsymbol{x}) = x_2,$$
$$\phi_4(\boldsymbol{x}) = x_1 x_2, \;\; \phi_5(\boldsymbol{x}) = x_1^2, \;\; \phi_6(\boldsymbol{x}) = x_2^2.$$

The functions $(\psi_1 \ldots, \psi_n)$ are a set of n radial basis functions defined as

$$\psi_i(\boldsymbol{x}) = h_{md}(\boldsymbol{x} - \boldsymbol{x}_i), \quad h_{md}(\boldsymbol{r}) = \begin{cases} a_{md}\|\boldsymbol{r}\|^{2m-d} \log \|\boldsymbol{r}\|, & \text{if } d \text{ is even,} \\ a_{md}\|\boldsymbol{r}\|^{2m-d}, & \text{if } d \text{ is odd,} \end{cases} \quad , \tag{4.5}$$

for some constants a_{md}. Here $\|\boldsymbol{r}\| = \sqrt{\boldsymbol{r}'\boldsymbol{r}}$ is the Euclidean norm of $\boldsymbol{r}$.

In matrix notation, we write

$$\boldsymbol{\beta} = (\beta_1, \cdots, \beta_t)', \tag{4.6}$$
$$\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_n)', \tag{4.7}$$
$$\boldsymbol{T} = (\phi_j(\boldsymbol{x}_i))_{n \times t}, \tag{4.8}$$
$$\boldsymbol{K} = (\psi_i(\boldsymbol{x}_j))_{n \times n}. \tag{4.9}$$

Then (4.4) is expressed as

$$\begin{pmatrix} f(\boldsymbol{x}_1) \\ \vdots \\ f(\boldsymbol{x}_n) \end{pmatrix}_{n \times 1} = \boldsymbol{T}\boldsymbol{\beta} + \boldsymbol{K}\boldsymbol{\gamma}. \tag{4.10}$$

Meinguet (1979) and Duchon (1977) also showed that equation (4.3) can be written as

$$J_m(f) = \boldsymbol{\gamma}'\boldsymbol{K}\boldsymbol{\gamma}, \tag{4.11}$$

subject to the constraint that $\boldsymbol{T}'\boldsymbol{\gamma} = 0$. The problem minimizing (4.2) then becomes a constrained minimization problem with objective function

$$S_\eta(f) = (\boldsymbol{y} - \boldsymbol{T}\boldsymbol{\beta} - \boldsymbol{K}\boldsymbol{\gamma})'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{T}\boldsymbol{\beta} - \boldsymbol{K}\boldsymbol{\gamma}) + \lambda\boldsymbol{\gamma}'\boldsymbol{K}\boldsymbol{\gamma}, \tag{4.12}$$

where $\boldsymbol{y} = (y_1, \cdots, y_n)'$ and $\boldsymbol{W} = diag(w_1, \cdots, w_n)$. Following Wahba (1990), consider the spectral decomposition of $\boldsymbol{T}\boldsymbol{T}'$, i.e.,

$$\boldsymbol{T}\boldsymbol{T}' = \boldsymbol{F}\boldsymbol{\Lambda}\boldsymbol{F}', \tag{4.13}$$

where $\boldsymbol{F}$ is orthogonal and $\boldsymbol{\Lambda}$ is diagonal. We write $\boldsymbol{F}$ as

$$\boldsymbol{F} = (\boldsymbol{F}_1, \boldsymbol{F}_2), \tag{4.14}$$

where $\boldsymbol{F}_1$ is the $n \times t$ matrix of vectors spanning the column space of $\boldsymbol{T}$ and $\boldsymbol{F}_2$ has dimension $n \times (n-t)$. Because $\boldsymbol{F}$ is orthogonal, $\boldsymbol{F}_1'\boldsymbol{F}_2 = 0$. Consequently $\boldsymbol{T}'\boldsymbol{\gamma} = 0$ if and only if $\boldsymbol{\gamma} = \boldsymbol{F}_2\boldsymbol{\eta}$ for some $\boldsymbol{\eta} \in \mathbb{R}^{n-t}$. The minimization problem (4.12) is equivalent to

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^t,\, \boldsymbol{\eta}\in\mathbb{R}^{n-t}} \Big\{(\boldsymbol{y} - \boldsymbol{T}\boldsymbol{\beta} - \boldsymbol{K}\boldsymbol{F}_2\boldsymbol{\eta})'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{T}\boldsymbol{\beta} - \boldsymbol{K}\boldsymbol{F}_2\boldsymbol{\eta}) + \lambda\boldsymbol{\eta}'\boldsymbol{F}_2'\boldsymbol{K}\boldsymbol{F}_2\boldsymbol{\eta}\Big\}. \tag{4.15}$$

We define the following matrices and vector

$$\boldsymbol{G} = (\boldsymbol{T}, \boldsymbol{K}\boldsymbol{F}_2)_{n\times n}, \quad \boldsymbol{H} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{F}_2'\boldsymbol{K}\boldsymbol{F}_2 \end{pmatrix}_{n\times n}, \quad \boldsymbol{\omega} = \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\eta} \end{pmatrix}.$$

Note that $\boldsymbol{F}_2'\boldsymbol{K}\boldsymbol{F}_2$ has dimension $(n-t) \times (n-t)$ and is invertible, so $\boldsymbol{H}$ has rank $n-t$. Then (4.15) can be written as

$$\min_{\boldsymbol{\omega}\in\mathbb{R}^n} \Big\{(\boldsymbol{y} - \boldsymbol{G}\boldsymbol{\omega})'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{G}\boldsymbol{\omega}) + \lambda\boldsymbol{\omega}'\boldsymbol{H}\boldsymbol{\omega}\Big\}. \tag{4.16}$$

Finally we define

$$\boldsymbol{v} = \boldsymbol{G}\boldsymbol{\omega}, \tag{4.17}$$

$$\boldsymbol{M} = (\mathbf{G}^{-\mathbf{1}})'\mathbf{H}\mathbf{G}^{-\mathbf{1}}, \tag{4.18}$$

then (4.16) can be written as

$$\min_{\boldsymbol{v}\in\mathbb{R}^n} \Big\{(\boldsymbol{y}-\boldsymbol{v})'\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{v}) + \lambda \boldsymbol{v}'\boldsymbol{M}\boldsymbol{v}\Big\}. \tag{4.19}$$

Clearly, the solution to (4.19) is

$$\hat{\boldsymbol{v}} \equiv \hat{f}(\boldsymbol{x}) = (\boldsymbol{W}+\lambda\boldsymbol{M})^{-1}\boldsymbol{W}\boldsymbol{y}. \tag{4.20}$$

Note that $\boldsymbol{M}$ has rank $n-t$.

4.1.2 Bayesian Thin-Plate Splines

The minimization problem in (4.19) has a Bayesian interpretation, first suggested by Kimmeldorf & Wahba (1971) and Wahba (1978). Suppose $\boldsymbol{y}$ follows a normal distribution

$$(\boldsymbol{y} \mid \boldsymbol{v}, \delta_0) \sim N(\boldsymbol{v}, \delta_0 \boldsymbol{W}^{-1}). \tag{4.21}$$

Next suppose that $\boldsymbol{v}$ has a prior with density function

$$\pi_1(\boldsymbol{v} \mid \delta_1) = \frac{|\boldsymbol{M}|_+^{1/2}}{(2\pi\delta_1)^{(n-t)/2}} \exp\left(-\frac{1}{2\delta_1}\boldsymbol{v}'\boldsymbol{M}\boldsymbol{v}\right), \tag{4.22}$$

because $\boldsymbol{M}$ has rank $n-t$ it is not invertible the prior given in (4.22) is improper. It is easy to show that the conditional posterior distribution of $\boldsymbol{v}$ given $(\delta_0, \delta_1; \boldsymbol{y})$ is

$$(\boldsymbol{v} \mid \delta_0, \delta_1; \boldsymbol{y}) \sim N((\boldsymbol{W}+\lambda\boldsymbol{M})^{-1}\boldsymbol{W}\boldsymbol{y},\, \delta_0(\boldsymbol{W}+\lambda\boldsymbol{M})^{-1}), \tag{4.23}$$

where $|\boldsymbol{M}|_+$ is the product of the positive eigenvalues of $\boldsymbol{M}$ and $\lambda = \delta_0/\delta_1$. Clearly the conditional posterior mean or mode of $\boldsymbol{v}$ given $(\delta_0, \delta_1; \boldsymbol{y})$ is the same as the solution given in (4.20).

To see the structure of the prior (4.22), recall

$$\boldsymbol{v} = \boldsymbol{G}\boldsymbol{\omega} = \boldsymbol{G}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\eta}\end{pmatrix},$$

and

$$\pi_2(\boldsymbol{\beta}, \boldsymbol{\eta} \mid \delta_1) = \pi_1 \left(\boldsymbol{G} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\eta} \end{pmatrix} \mid \delta_1 \right) \cdot |\boldsymbol{G}|, \tag{4.24}$$

where the first term of the right hand side is given by (4.22). Note that (4.18) implies

$$\boldsymbol{G}'\boldsymbol{M}\boldsymbol{G} = \boldsymbol{H} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F}_2'\boldsymbol{K}\boldsymbol{F}_2 \end{pmatrix}.$$

Thus, we have

$$\begin{aligned} \boldsymbol{v}'\boldsymbol{M}\boldsymbol{v} &= (\boldsymbol{\beta}'\,\boldsymbol{\eta}')\boldsymbol{G}'\boldsymbol{M}\boldsymbol{G}\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\eta} \end{pmatrix} = (\boldsymbol{\beta}'\,\boldsymbol{\eta}')\boldsymbol{H}\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\eta} \end{pmatrix} = \boldsymbol{\eta}'\boldsymbol{F}_2'\boldsymbol{K}\boldsymbol{F}_2, \qquad (4.25) \\ |\boldsymbol{M}|_+^{1/2}|\boldsymbol{G}| &= (|\boldsymbol{G}'||\boldsymbol{M}|_+|\boldsymbol{G}|)^{1/2} = |\boldsymbol{G}'\boldsymbol{M}\boldsymbol{G}|_+^{1/2} \\ &= |\boldsymbol{H}|_+^{1/2} = |\boldsymbol{F}_2'\boldsymbol{K}\boldsymbol{F}_2|^{1/2}. \qquad (4.26) \end{aligned}$$

Substituting (4.25) and (4.26) into (4.24) we get

$$\pi_2(\boldsymbol{\beta}, \boldsymbol{\eta} \mid \delta_1) = \frac{|\boldsymbol{F}_2'\boldsymbol{K}\boldsymbol{F}_2|^{1/2}}{(2\pi\delta_1)^{(n-t)/2}} \exp\left(-\frac{1}{2\delta_1}\boldsymbol{\eta}'\boldsymbol{F}_2'\boldsymbol{K}\boldsymbol{F}_2\boldsymbol{\eta} \right). \tag{4.27}$$

Clearly $\boldsymbol{\beta}$ has a constant prior and $\boldsymbol{\eta}$ has a proper normal prior. Consequently, $\boldsymbol{v} = \boldsymbol{G}\boldsymbol{\omega}$ has a partially informative normal prior from Speckman & Sun (2003). In order to complete a full Bayesian hierarchical model for estimating the unknown $\boldsymbol{v}$, the prior densities for δ_1 and δ_0 could be applied. These are assumed to follow inverse gamma distributions with densities

$$[\delta_l] \propto \frac{1}{\delta_l^{a_l+1}} \exp\left(-\frac{b_l}{\delta_l} \right), \text{ for } l = 0, 1. \tag{4.28}$$

See, for example, Speckman & Sun (2003).

4.2 The Joint Spatio-Temporal Model

The model for our data here begins with the likelihood defined in (2.1) and (2.8). Additionally we refer to the same hierarchical structure on the prior of p_{ijk} as in (3.1).

The difference between this model and the model (3.1) lies in the prior for $\boldsymbol{z}$. This prior is developed in the same manner as (3.10),

$$[\boldsymbol{z} \mid \delta_1] = \frac{|\boldsymbol{M} \otimes \boldsymbol{A}|_+^{1/2}}{(2\pi\delta_1)^{(I-3)(J-2)/2}} \exp\left(-\frac{1}{2\delta_1}\boldsymbol{z}'(\boldsymbol{M} \otimes \boldsymbol{A})\boldsymbol{z}\right), \tag{4.29}$$

where the matrix $\boldsymbol{M}$ is defined in (4.18) and the matrix $\boldsymbol{A}$ is defined in (3.9). The Kronecker product of $\boldsymbol{M}$ and $\boldsymbol{A}$ yields a type of precision matrix that is made up of two components, the $\boldsymbol{M}$ matrix which accounts for the spatial structure of the data, and the $\boldsymbol{A}$ matrix which smooths the temporal process. The Kronecker product applies the spatial function across the temporal processes. The result is that $\boldsymbol{z}$ can be considered a set of spatially correlated temporal processes.

The temporal component of this matrix can be seen as a second order intrinsic auto-regressive, (IAR(2)) prior applied to the temporal trend of the data, as related in Besag and Kooperberg (1995) and Künsch (1987). The IAR(2) prior is used in Fahrmeir and Wagenpfeil (1996) for smoothing hazard functions.

Since θ_k is the difference between the mean for age-group k and age-group 1, $\theta_1 = 0$. We assume the priors for the remaining θ_k are constant, i.e.

$$\pi(\theta_k) \propto 1,\ k = 2, 3, \ldots, K. \tag{4.30}$$

Additionally, we specify the prior distribution of ν_{ijk} as in (3.3),

$$(\nu_{ijk} \mid z_{ij}, \theta_k, \delta_0) \sim N(z_{ij} + \theta_k, \delta_0). \tag{4.31}$$

Let $\boldsymbol{\nu} = (\nu_{111}, \ldots, \nu_{1J1}, \nu_{211} \ldots, \nu_{IJ1}, \nu_{112}, \ldots, \nu_{IJK})'$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)'$. The prior distribution of $\boldsymbol{\nu}$ is written as

$$(\boldsymbol{\nu} \mid \boldsymbol{z}, \boldsymbol{\theta}, \delta_0) \sim N_{IJK}\left(\mathbf{1}_K \otimes \boldsymbol{z} + \boldsymbol{\theta} \otimes \mathbf{1}_{IJ},\ \delta_0 \boldsymbol{I}_{IJK}\right). \tag{4.32}$$

In order to complete this model, we choose the priors for δ_0 and δ_1 as in (3.11).

4.2.1 Full-Conditional Distributions

The full conditional distributions of the model parameters are needed to implement a Gibbs sampler.

Lemma 3

(a) For given $(\boldsymbol{z}, \theta_k, \delta_0; \boldsymbol{y})$, $(\nu_{111}, \ldots, \nu_{IJK})$ *are independent, and each* ν_{ijk} *depends only on* y_{ijk} *with*

$$[\nu_{ijk} \mid z_{ij}, \theta_k, \delta_0; y_{ijk}] \;\propto\; \exp\left\{\nu_{ijk} y_{ijk} - n_{ijk} e^{\nu_{ijk}} - \frac{1}{2\delta_0}(\nu_{ijk} - z_{ij} - \theta_k)^2\right\}.$$

(b) The conditional posterior density of ν_{ijk} *in (a) is log-concave.*

(c) Write $\boldsymbol{\nu}_k = (\nu_{11k}, \nu_{12k}, \ldots, \nu_{IJk})'$. *Then* $(\boldsymbol{z} \mid \boldsymbol{\nu}, \boldsymbol{\theta}, \delta_0, \delta_1; \boldsymbol{y}) \sim N_{IJ}(\boldsymbol{c}, \delta_0 \boldsymbol{G}^{-1})$, *where*

$$\boldsymbol{G} = K\boldsymbol{I}_{IJ} + \lambda(\boldsymbol{M} \otimes \boldsymbol{A}),\ \boldsymbol{c} = \boldsymbol{G}^{-1} \sum_{k=1}^{K} (\boldsymbol{\nu}_k - \theta_k \mathbf{1}_{\mathbf{IJ}}),$$

and $\lambda = \delta_0/\delta_1$.

(d) $(\theta_k \mid \boldsymbol{\nu}, \boldsymbol{z}, \delta_0; \boldsymbol{y}) \sim N\left((IJ)^{-1} \sum_{i=1}^{I} \sum_{j=1}^{J} (\nu_{ijk} - z_{ij}), \delta_0 (IJ)^{-1}\right)$.

(e) $(\delta_0 \mid \boldsymbol{\nu}, \boldsymbol{z}, \boldsymbol{\theta}; \boldsymbol{y}) \sim IG(a_0 + \frac{1}{2} IJK, b_0 + \frac{1}{2} \sum_i^I \sum_j^J \sum_k^K (\nu_{ijk} - z_{ij} - \theta_k)^2)$.

(f) $(\delta_1 \mid \boldsymbol{z}; \boldsymbol{y}) \sim IG(a_1 + \frac{1}{2}(I-3)(J-2), b_1 + \frac{1}{2}\boldsymbol{z}'(\boldsymbol{M} \otimes \boldsymbol{A})\boldsymbol{z})$.

Proof. We only prove part(b). In fact,

$$\frac{\partial^2}{\partial \nu_{ijk}^2} \log[\nu_{ijk} \mid z_{ij}, \theta_k, \delta_0; y_{ijk}] = -n_{ijk} e^{\nu_{ijk}} + \frac{1}{\delta_0} \;<\; 0,\ \forall\, \nu_{ijk}.$$

The result then holds. □

4.2.2 Computational Improvements

Updating $\boldsymbol{z}$ using the conditional distribution based on the current value of λ in Lemma 11(a) requires the inversion of an $IJ \times IJ$ matrix. In order to speed up the process, it is desirable to diagonalize this matrix. This algorithm extends the method of Section 3.5.1.

- Begin by letting $\boldsymbol{Q}_M$ be an orthogonal matrix such that $\boldsymbol{M} = \boldsymbol{Q}_M \boldsymbol{\Lambda}_M \boldsymbol{Q}_M'$, where $\boldsymbol{\Lambda}_M$ is the diagonal matrix of eigenvalues of $\boldsymbol{M}$, and let $\boldsymbol{A} = \boldsymbol{Q}_A \boldsymbol{\Lambda}_A \boldsymbol{Q}_A'$, where $\boldsymbol{\Lambda}_A$ is the diagonal matrix of eigenvalues of $\boldsymbol{A}$.

- Then define $\tilde{\boldsymbol{Q}} = \boldsymbol{Q}_M \otimes \boldsymbol{Q}_A$, $\tilde{\boldsymbol{Q}}' = \boldsymbol{Q}_M' \otimes \boldsymbol{Q}_A'$, and

$$\boldsymbol{\Phi}_\lambda = K\boldsymbol{I}_{IJ} + \lambda(\boldsymbol{\Lambda}_M \otimes \boldsymbol{\Lambda}_A). \tag{4.33}$$

  Note that $\tilde{\boldsymbol{Q}}$ is orthogonal and $\boldsymbol{\Phi}$ is diagonal and now easily invertible. It can then be shown that

$$\begin{aligned} K\boldsymbol{I}_{IJ} + \lambda(\boldsymbol{M} \otimes \boldsymbol{A}) &= \tilde{\boldsymbol{Q}} \boldsymbol{\Phi}_\lambda \tilde{\boldsymbol{Q}}', \\ (K\boldsymbol{I}_{IJ} + \lambda(\boldsymbol{M} \otimes \boldsymbol{A}))^{-1} &= \tilde{\boldsymbol{Q}} \boldsymbol{\Phi}_\lambda^{-1} \tilde{\boldsymbol{Q}}'. \end{aligned}$$

  We prove the first equality only:

$$\begin{aligned} &\tilde{\boldsymbol{Q}}(K\boldsymbol{I}_{IJ} + \lambda(\boldsymbol{\Lambda}_M \otimes \boldsymbol{\Lambda}_A))\tilde{\boldsymbol{Q}}' \\ &= (\boldsymbol{Q}_M \otimes \boldsymbol{Q}_A)(K\boldsymbol{I}_{IJ} + \lambda(\boldsymbol{\Lambda}_M \otimes \boldsymbol{\Lambda}_A))(\boldsymbol{Q}_M' \otimes \boldsymbol{Q}_A') \\ &= K\boldsymbol{I}_{IJ} + \lambda(\boldsymbol{Q}_M \boldsymbol{\Lambda}_M \otimes \boldsymbol{Q}_A \boldsymbol{\Lambda}_A)(\boldsymbol{Q}_M' \otimes \boldsymbol{Q}_A') \\ &= K\boldsymbol{I}_{IJ} + \lambda(\boldsymbol{Q}_M \boldsymbol{\Lambda}_M \boldsymbol{Q}_M' \otimes \boldsymbol{Q}_A \boldsymbol{\Lambda}_A \boldsymbol{Q}_A') \\ &= K\boldsymbol{I}_{IJ} + \lambda(\boldsymbol{M} \otimes \boldsymbol{A}). \end{aligned}$$

Now to update $\boldsymbol{z}$ the following algorithm could be implemented.

1. Update $\boldsymbol{\Phi}_\lambda$.

2. Simulate $\boldsymbol{s} \sim N_{IJ}(\boldsymbol{0}, \boldsymbol{I}_{IJ})$.

3. Compute $\boldsymbol{r}_\lambda = \boldsymbol{\Phi}_\lambda^{-\frac{1}{2}}\boldsymbol{s} + \boldsymbol{\Phi}_\lambda^{-1}\tilde{\boldsymbol{Q}}' \sum_{k=1}^{K}(\boldsymbol{\nu}_k - \theta_k \mathbf{1_{IJ}})$.

4. Let $\boldsymbol{z} = \tilde{\boldsymbol{Q}}\boldsymbol{r}_\lambda$, which has the same distribution as Lemma 9(c).

4.3 Results

Since almost all the full conditionals are of closed form, the implementation of the Gibbs sampler (Gelfand & Smith 1990) is relatively simple. The exception is for ν_{ij}, which can be evaluated using an Adaptive Rejection Sampler (Gilks & Wild 1992) at each step of the Gibbs sampler. Convergence is rapid, and as a result, implementation in FORTRAN takes a little over 100 minutes to produce 100,000 iterations, with the first 50,000 discarded as burn-in.

4.3.1 Noninformative and Data-dependent Priors

The model is initially run using non-informative priors for δ_0 and δ_1,

$$\pi(\delta_l) \quad \propto \quad 1,\, l = 0, 1.$$

Statistics from the posterior distributions for δ_0, δ_1 and λ are shown in Table 4.1. The posterior means and variance are then used to calculate sets of data-dependent priors for both δ_0 and δ_1. These are calculated by inflating the posterior means and variances of the respective parameters and deriving the respective hyperparameters corresponding to the inflated values, as in Section 2.6.1. These new hyperparameters

are then used and the model re-run in order to determine the robustness of the model to perturbations in the variance parameters' priors.

Table 4.1: Quantiles of δ_0, δ_1 and ρ for Noninformative (NI) and Inflation Factor (IF) Data-dependent Priors

Summary of Posterior Distributions								
	Prior	Min.	1st Qt.	Median	Mean	3rd Qt.	Max.	Std. Dev.
δ_0	NI	.0027	.0077	.0092	.0094	.0109	.0219	.0024
	IF 2, 200	.0035	.0074	.0087	.0089	.0101	.0200	.0021
	IF 1.5, 150	.0035	.0067	.0080	.0082	.0096	.0174	.0021
	IF 2.5, 500	.0036	.0079	.0092	.0094	.0107	.0178	.0021
δ_1	NI	.00000	.00001	.00002	.00004	.00005	.00041	.00005
	IF 2, 200	.0026	.0050	.0057	.0058	.0065	.0117	.0011
	IF 1.5, 150	.0023	.0041	.0046	.0047	.0053	.0104	.0009
	IF 2.5, 500	.0031	.0055	.0062	.0064	.0071	.0119	.0011
λ	NI	11.58	170.5	386.5	9743.8	1281.6	406881	29764.1
	IF 2,200	.4376	1.2423	1.5115	1.5845	1.846	5.1474	.4893
	IF 1.5,25	.4670	1.3685	1.7188	1.8129	2.1679	4.9362	.61283
	IF 2.5,500	.4867	1.2060	1.4703	1.5232	17908	4.5884	.4443

As can be seen, the choice of priors seems to have little influence on the posterior distribution of δ_0, though the choice of these informative priors seems to have a great deal of influence over the posterior distributions of δ_1 and λ. This is a bit troubling. We would like to see a more non-informative prior for these parameters. This is an issue which will see further discussion. For now the results from the non-informative priors will be used for further discussion.

4.3.2 CAR Model vs Thin-Plate Spline Model

The results here are compared with a previous model for the same data, which used a conditional autoregressive prior for the spatial effects. The CAR prior density is defined in (3.10).

Scatterplots in Figure 4.1 of the estimates of p_{ijk} for the two models illustrate the similarities between them. The thin-plate spline model tends to smooth the data much like the CAR model. One of the original motivations for using the thin-plate spline model was its demonstrated improvement over the CAR model in Chapter 4. The posterior density of ρ from the CAR model in Figure 3.7, showed that it is heavily skewed to the right. This is thought to be indicative of under smoothing, though this interpretation is open to discussion.

The maps in Figures 4.4-4.9 show the estimates of p_{ijk} based on the raw data, the CAR model and the thin-plate spline model. The results reinforce what is shown in the scatterplots. The results of the two models look similar. Both models appear to have shrunk the estimates toward an overall mean.

One method of model selection between the CAR model and the thin-plate spline model is the DIC as proposed by Spiegelhalter et al. (2002) and discussed in Section 3.6.2. Table 4.2 lists the results for the two models.

Table 4.2: DIC Values for the CAR and Thin-plate Spline Models

Model Comparison Using DIC			
Model	P_D	$\bar{D}$	DIC
CAR	443.909	449051	449495
Thin-Plate Spline	432.868	449067	449500

In this case the DIC indicates that the CAR model is a better model. Spiegelhalter et al. (2002) give only a rough rule of thumb stating that models within one to two of the best score should be considered and models within three to seven should be "less so." Given the variability in the goodness-of-fit term $\bar{D}$ and the smaller complexity term P_D for the thin-plate spline model it is difficult to reject it out of hand based on the DIC alone.

Another model selection criteria is the $D(m)$, based on minimizing the posterior expected loss as proposed by Gelfand & Ghosh (1998). The results for the comparison of the two models using this criteria are shown in Table 4.3.

Table 4.3: $D(m)$ Values for the CAR and Thin-plate Spline Models

Model Comparison Using D(m)			
Model	$P(m)$	$G(m)$	$D(m)$
CAR	188.78	30016.74	30205.52
Thin-Plate Spline	192.56	22988.85	23181.41

These results strongly favor the thin-plate spline model. Despite its larger estimate variances, it appears to fit the data better. The caveat with this interpretation is that the measure of goodness of fit $G(m)$ is based on the squared error loss, which may not be the appropriate loss function.

Viewing the posterior densities of some parameters for both models, we can see in Figure 3.7 that in the case of the CAR model that the posterior distribution of ρ is heavily skewed and the center of mass is near the positive limit for ρ. This is thought to indicate that the model based on the CAR prior for the spatial effects does not sufficiently smooth the data. This may be true. Referring to Table 4.1 and looking at the posterior density of the smoothing parameter in Figure 4.3 λ and δ_0 we see that it has large values for the mean and median, and correspondingly small values for δ_1 in the noninformative case. The trace of these two parameters indicate that the smoothing parameter seems to spike from time to time, indicating that the model is trying to fit a flat surface. Looking at the maps, it appears that this is probably what both models are trying to do. Interpreting the skewed posterior density of ρ suggesting that the model is under smoothing is probably correct in a strict sense. The best scoring model in terms of DIC has been the additive model that tended to

shrink estimates toward a mean and didn't indicate much spatial pattern to the data. What is probably going on is that there is no real need for a spatial component in the model for this data.

4.4 Conclusion

Though the results of the model selection criteria are perhaps unclear, comparison of the maps of the estimated rates from both the CAR and the thin-plate splines models show that the estimates reveal a subtle spatial pattern at best for the estimates. These results suggest that there could be spatial and temporal trends at work and that the nature of these trends appear to interact, in space and time at the county level. There is also the possibility that the temporal trend is not adequately explained using the IAR prior. These trends can be seen in the maps of rates for each age groups and reveal the complex nature of the changes in rates across age groups over time. It appears that age effects and these temporal changes most likely dominate the data.

It is likely an overall mean would fit the data as well as any of these spatial models. The problem arises in looking at the results based on the CAR model, particularly the posterior density of ρ and assuming that the skewed density indicated the presence of a strong spatial pattern, that was not fully explained by the CAR prior. The posterior density of the smoothing parameter λ in the thin-plate spline model was actually more useful as it indicated that the model was trying to fit a flat surface. Perhaps there was not much spatial pattern to the data after all. This was easier to detect in the thin-plate spline model, because the smoothing parameter has a more straightforward interpretation in terms of the presence or absence of spatial

heterogeneity. The mixed performance of the model selection criteria is possibly the result of neither model being appropriate.

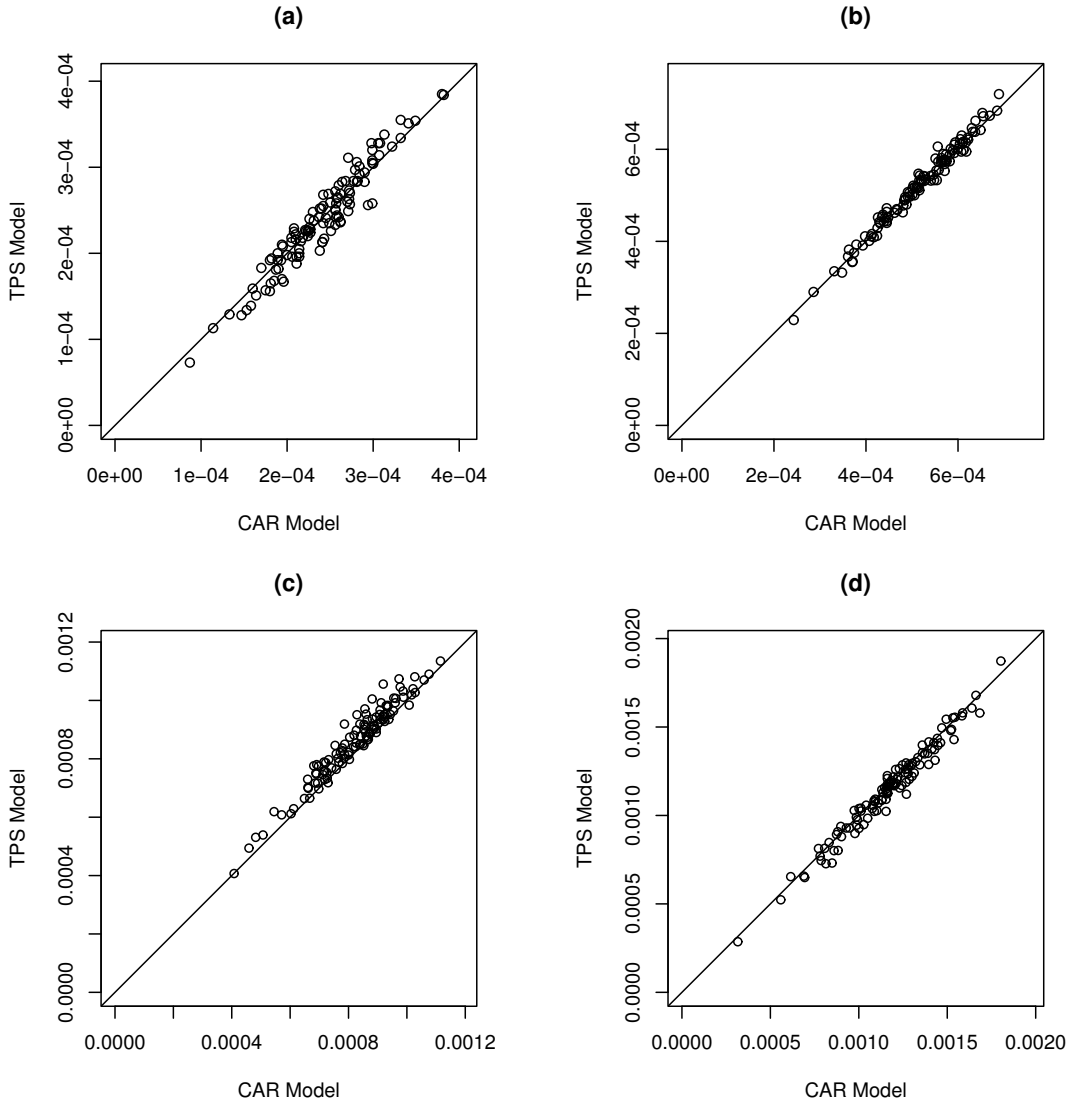

Figure 4.1: Scatterplots Comparing the Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1996-2000 from the CAR and Thin-plate Spline Joint Models: (a) $(j,k) = (3,1)$, (b) $(j,k) = (5,2)$, (c) $(j,k) = (7,3)$, and (d) $(j,k) = (11,4)$.

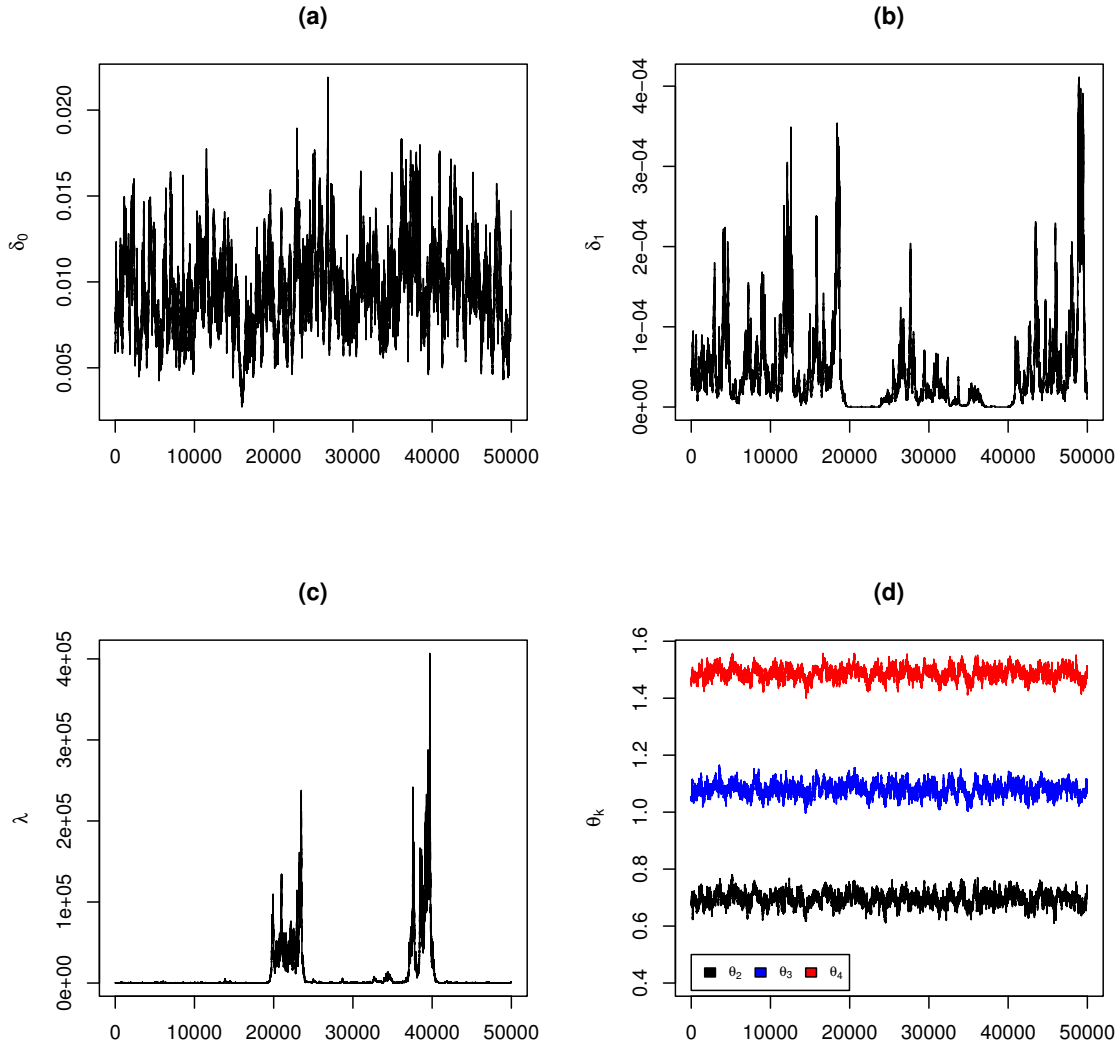

Figure 4.2: Trace Plots of (a) δ_0, (b) δ_1, (c) λ, and (d) θ_k, $k = 2, 3, 4$ from the Joint Thin-plate Splines Model for Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1969-2000.

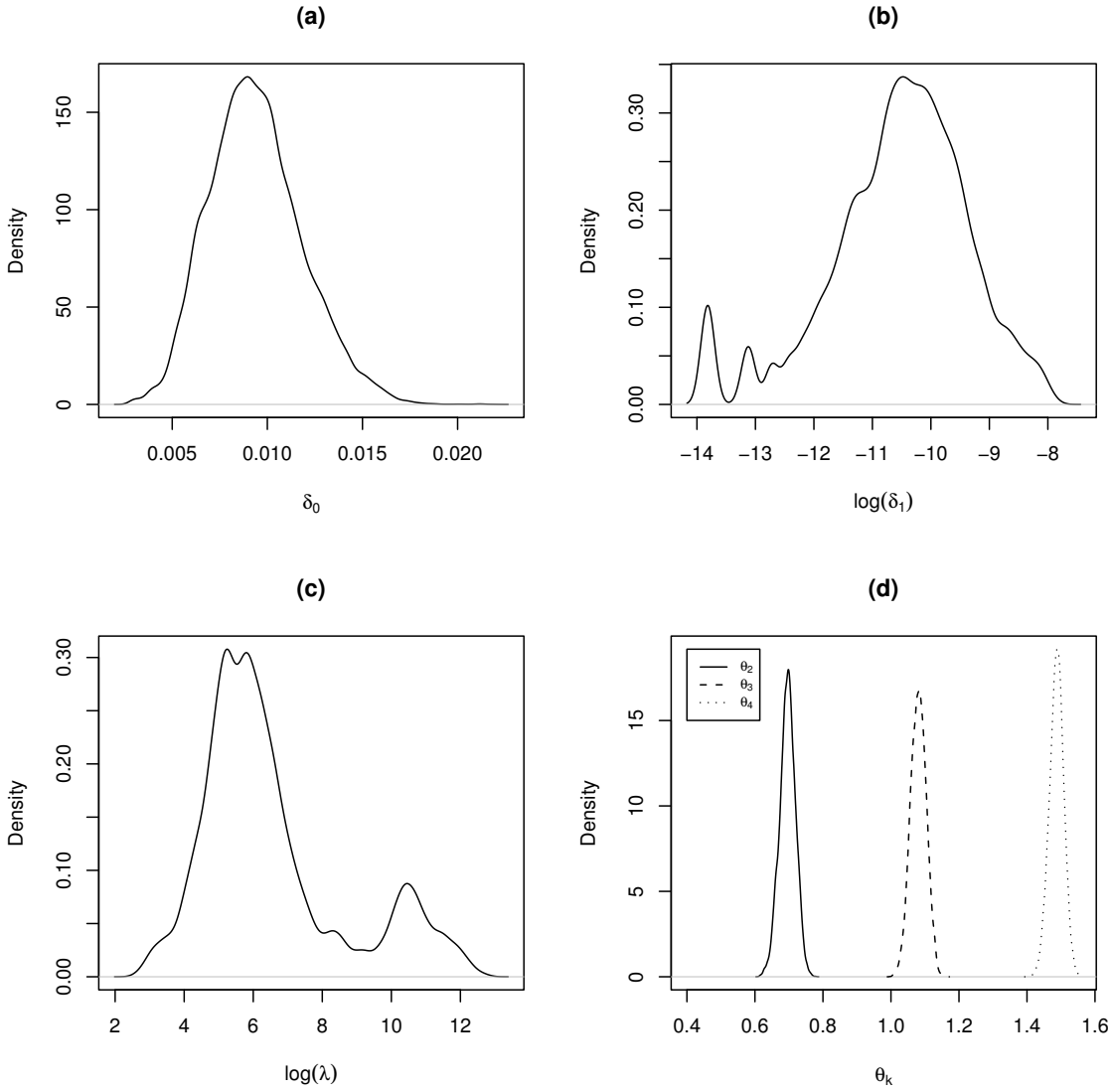

Figure 4.3: Posterior Densities for (a) δ_0, (b) $\log(\delta_1)$, (c) $\log(\lambda)$, and (d) θ_k, $k = 2, 3, 4$ from the Joint Thin-plate Splines Model for Mortality Rates p_{ijk} for Female Breast Cancer in Missouri from 1969-2000.

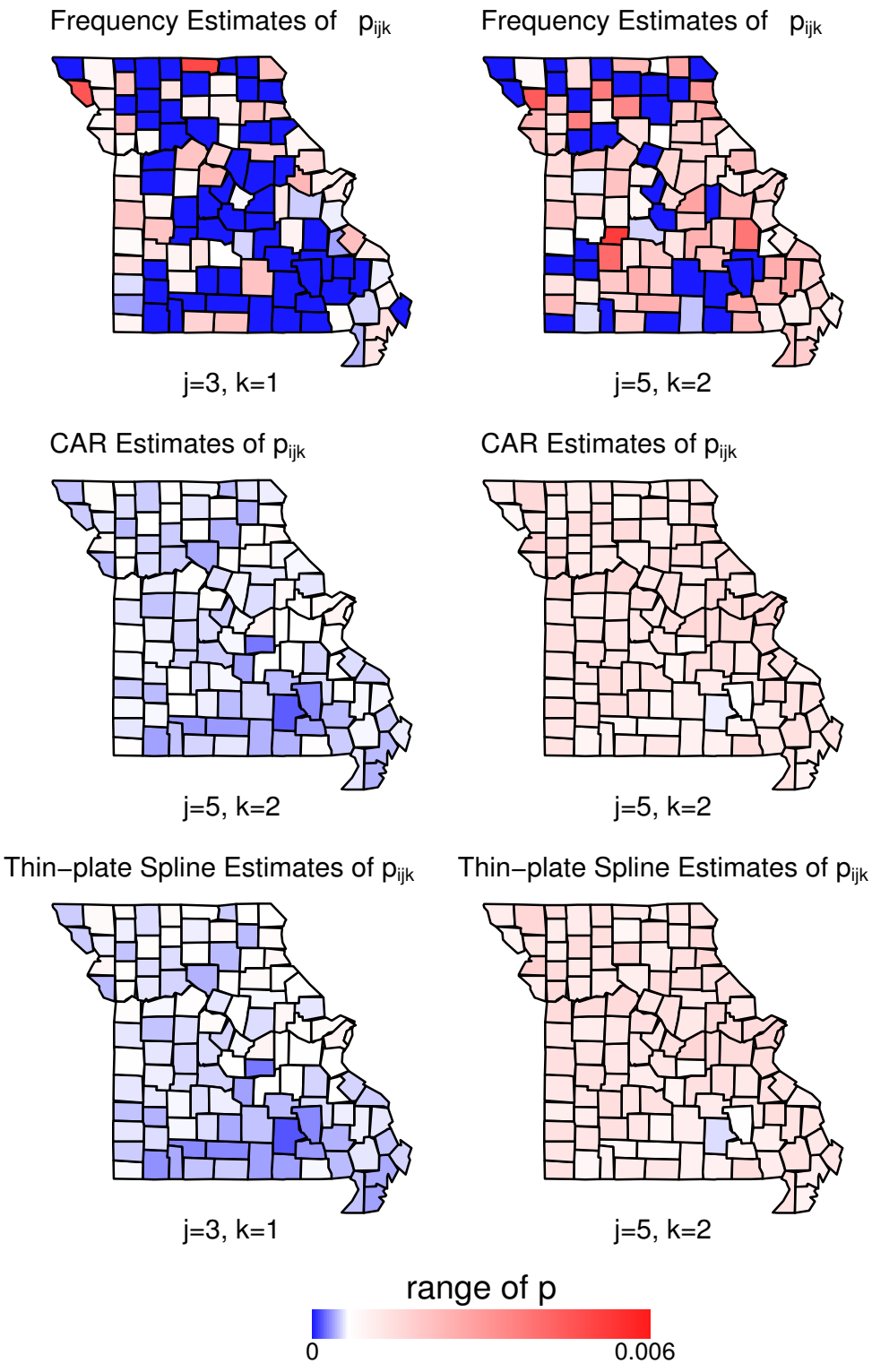

Figure 4.4: Maps of Frequency and Bayesian Estimates of Mortality Rates p_{ij} from the Joint Thin-plate Splines Model for Female Breast Cancer in Missouri from 1996-2000 for $(j,k) = (3,1)$ and $(j,k) = (5,2)$.

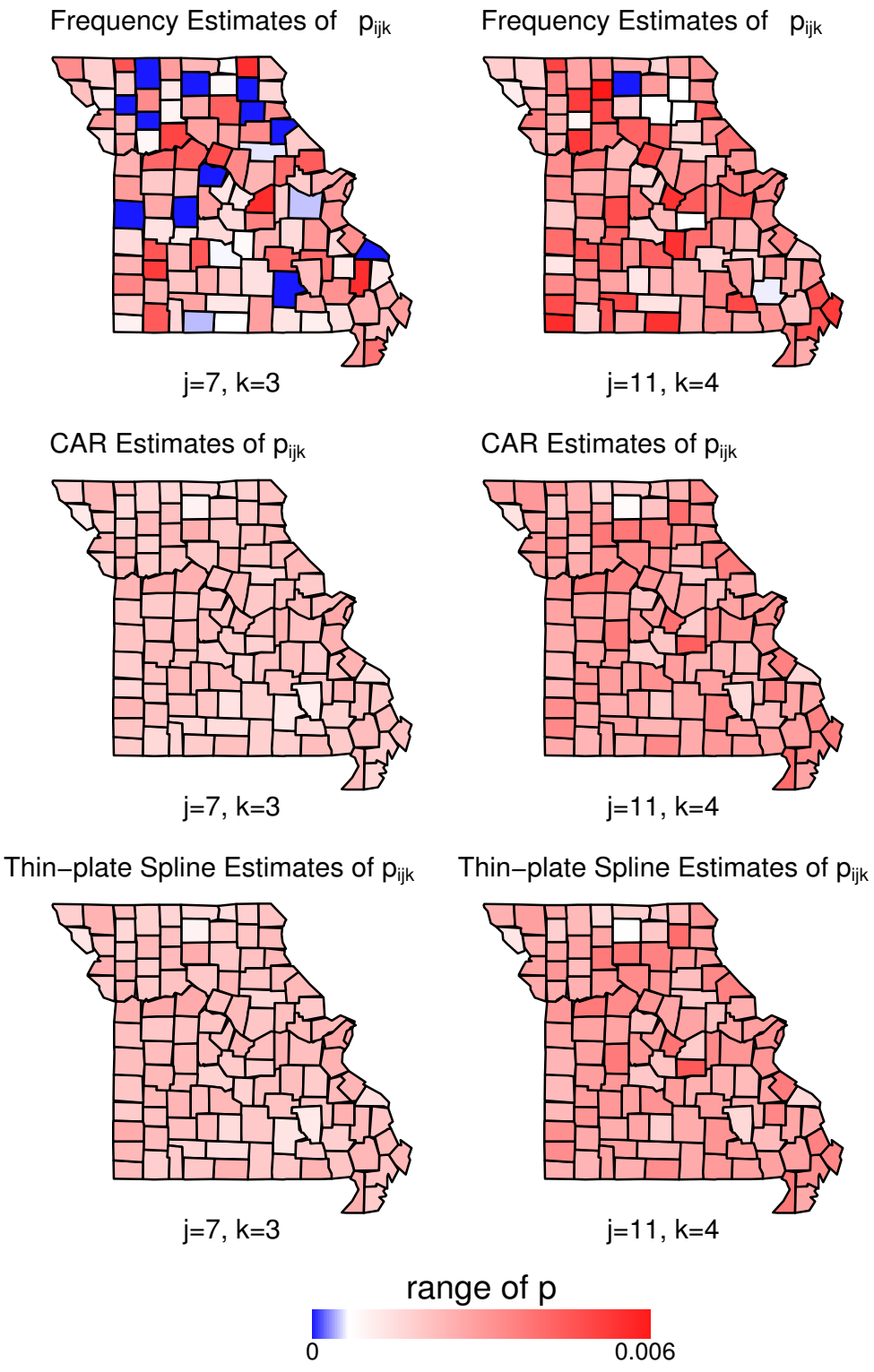

Figure 4.5: Maps of Frequency and Bayesian Estimates of Mortality Rates p_{ij} from the Joint Thin-plate Splines Model for Female Breast Cancer in Missouri from 1996-2000 for $(j,k) = (7,3)$ and $(j,k) = (11,4)$.

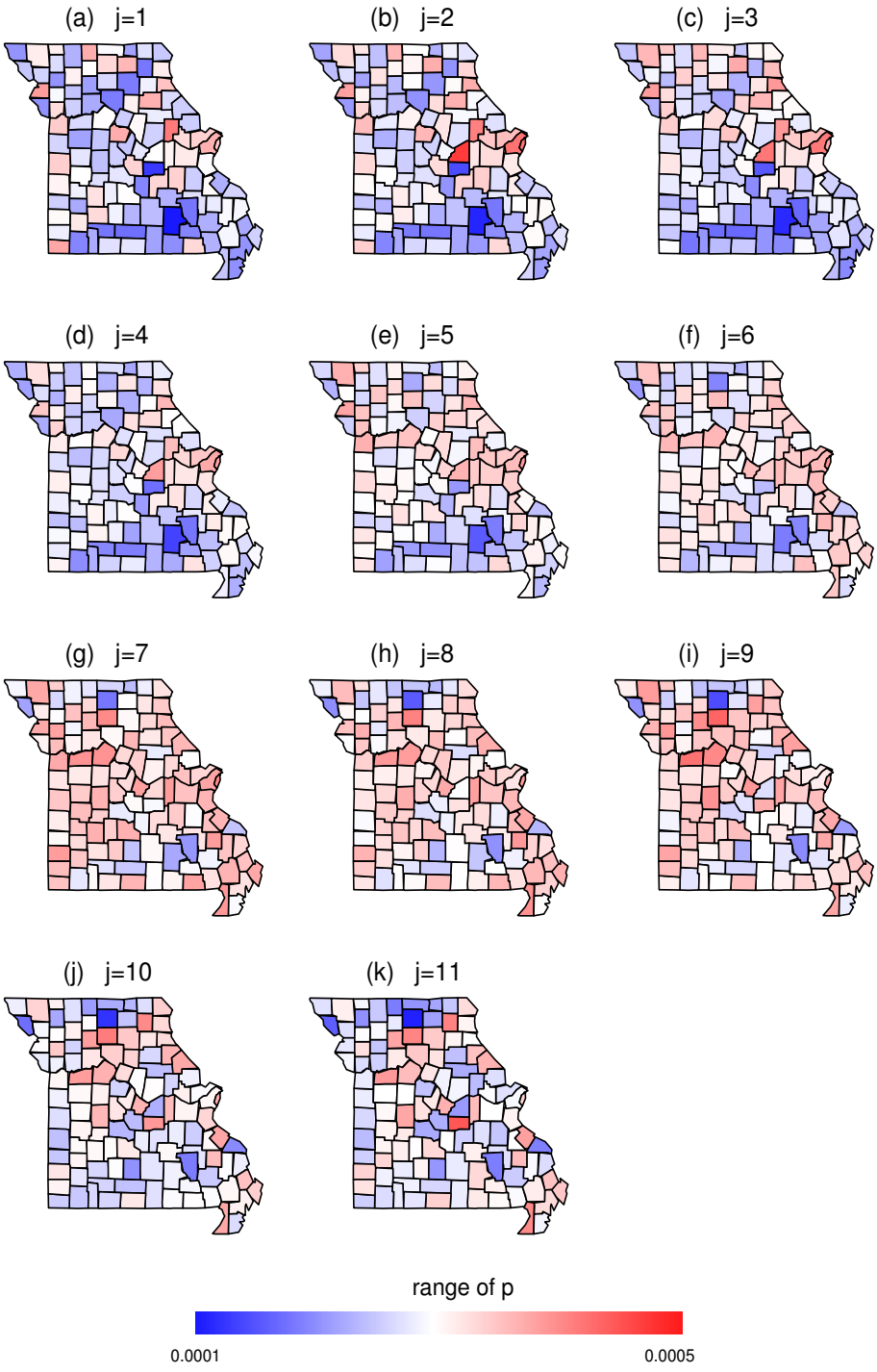

Figure 4.6: Maps of Bayesian Estimates of Mortality Rates p_{ij} from the Joint Thin-plate Splines Model for Female Breast Cancer in Missouri from 1996-2000 for $k = 1$ and $j = 1, \ldots, 11$.

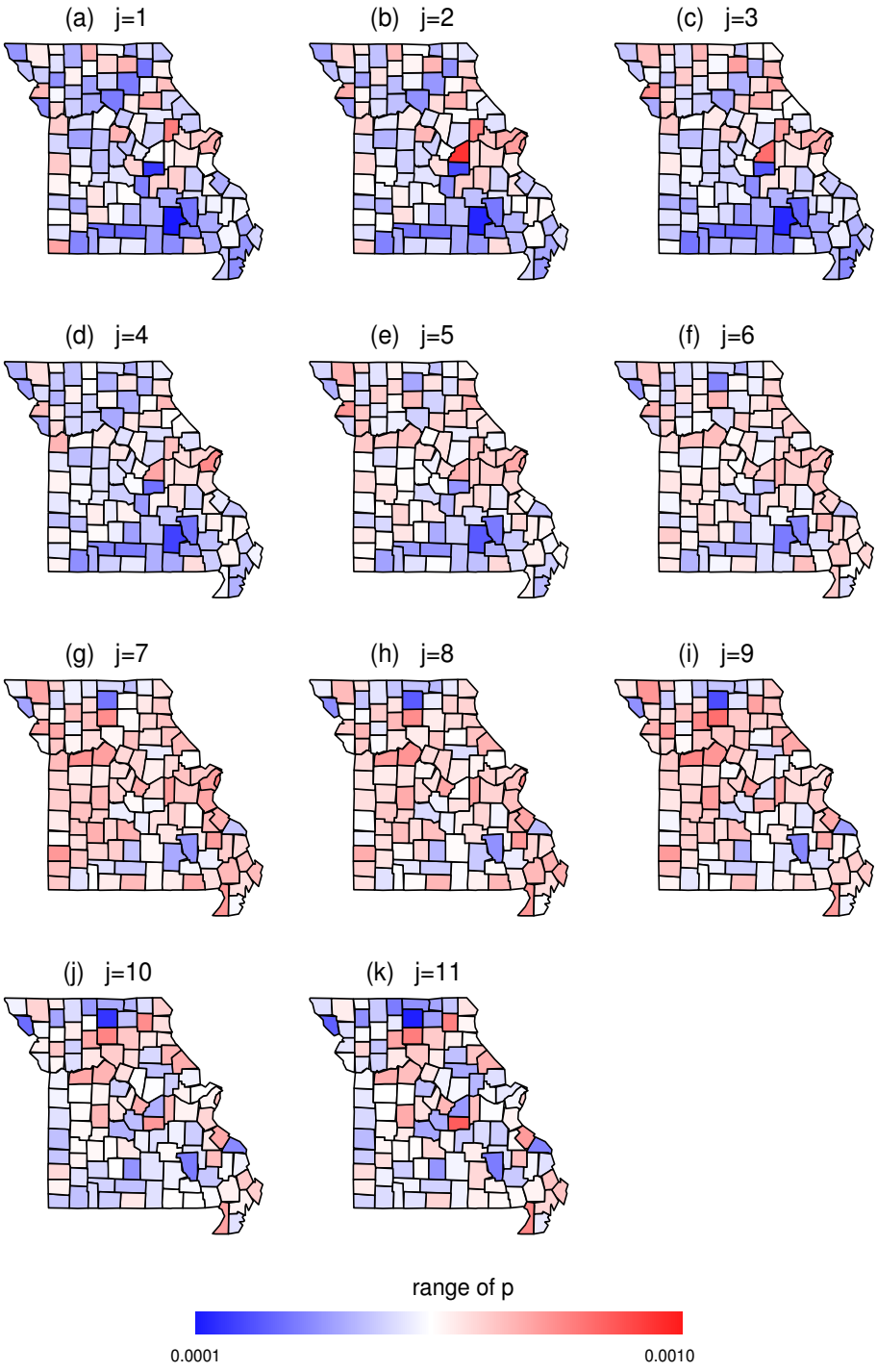

Figure 4.7: Maps of Bayesian Estimates of Mortality Rates p_{ij} from the Joint Thin-plate Splines Model for Female Breast Cancer in Missouri from 1996-2000 for $k = 2$ and $j = 1, \ldots, 11$.

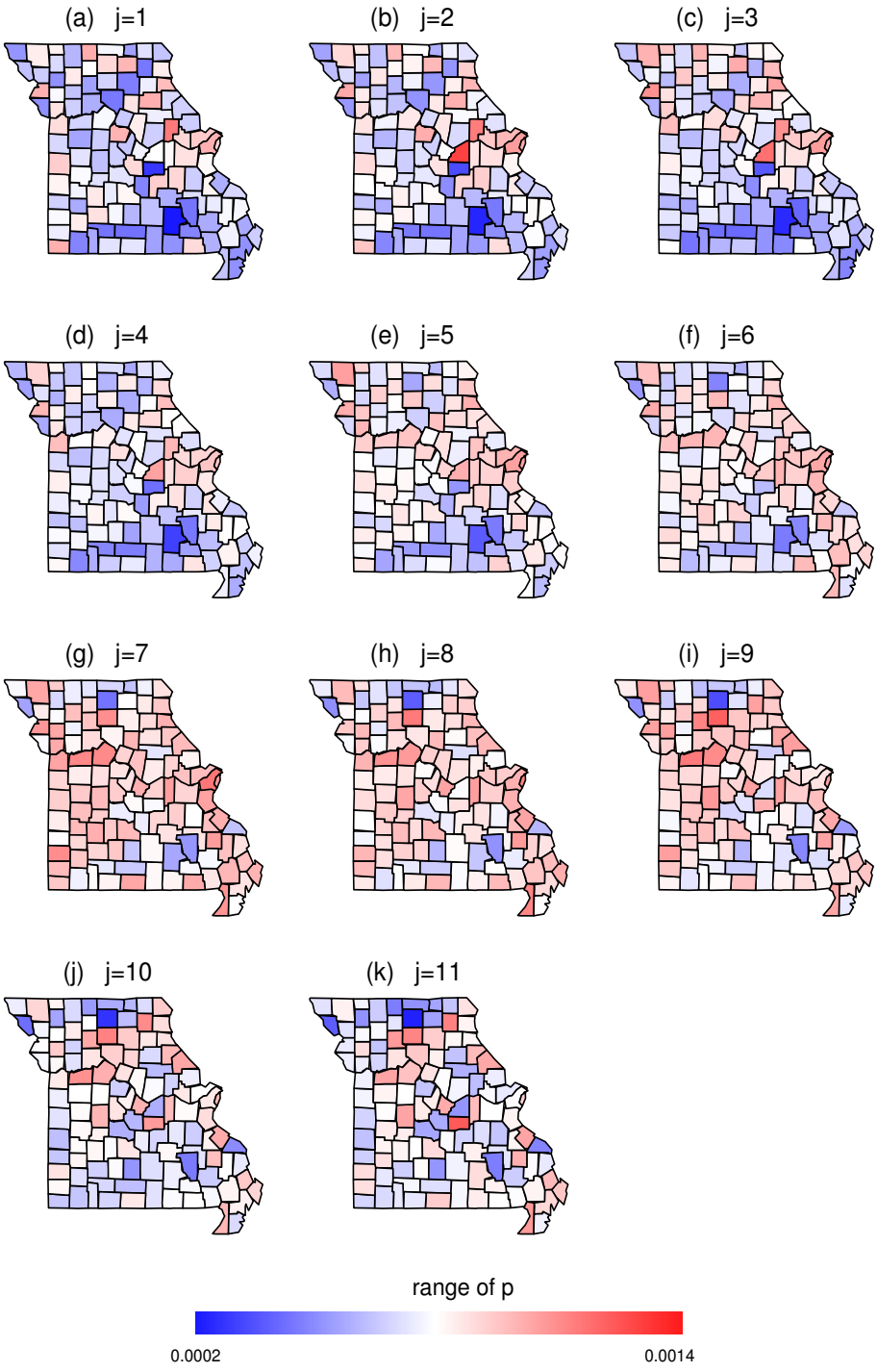

Figure 4.8: Maps of Bayesian Estimates of Mortality Rates p_{ij} from the Joint Thin-plate Splines Model for Female Breast Cancer in Missouri from 1996-2000 for $k = 3$ and $j = 1, \ldots, 11$.

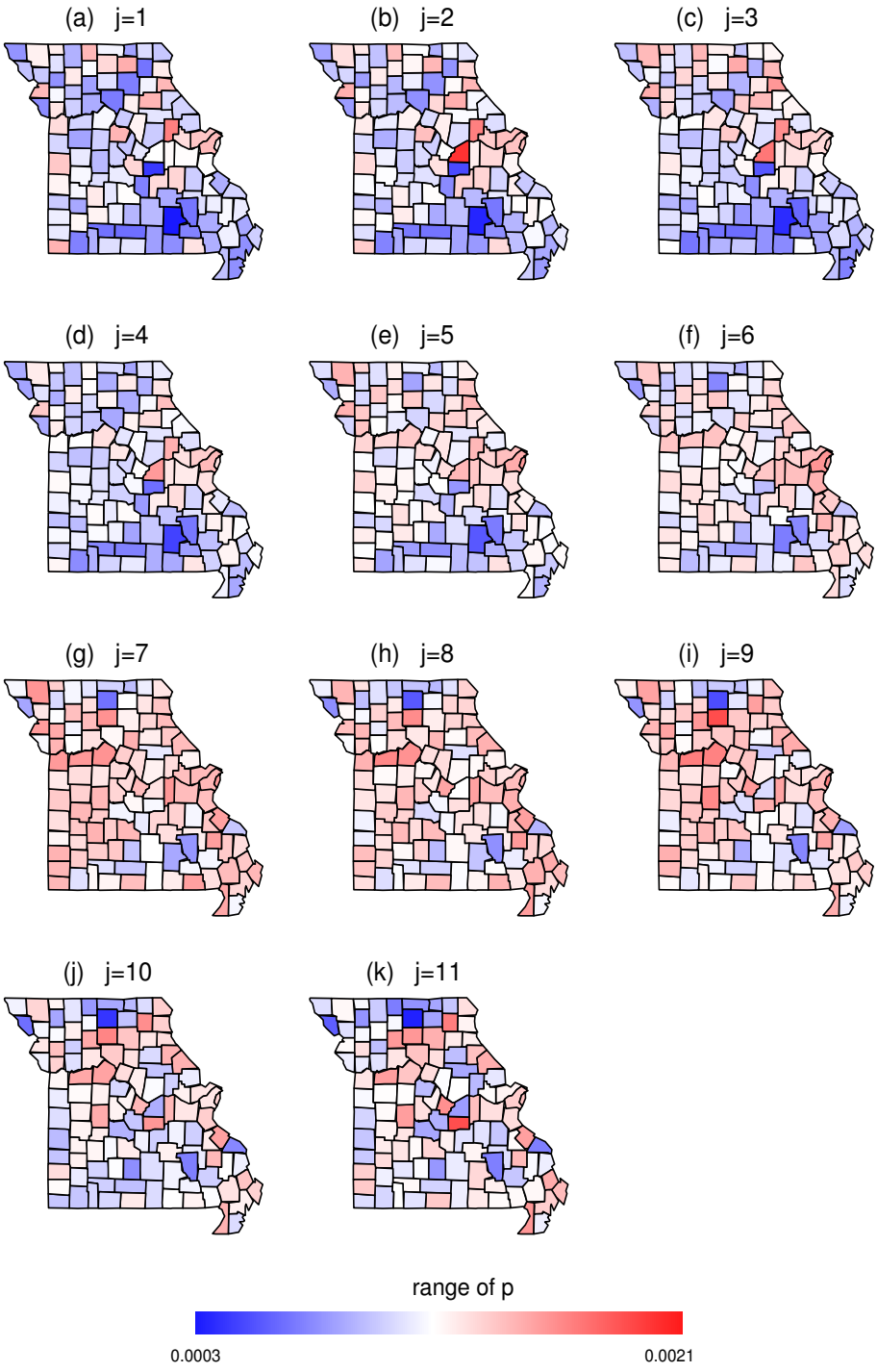

Figure 4.9: Maps of Bayesian Estimates of Mortality Rates p_{ij} from the Joint Thin-plate Splines Model for Female Breast Cancer in Missouri from 1996-2000 for $k = 4$ and $j = 1, \ldots, 11$.

Chapter 5

Using a g-prior for Model Comparison and Refinement

5.1 The Problem

Comparison in Section 4.3.2 of the models using the thin-plate spline prior and the CAR prior for spatial effects was done using the DIC and $D(m)$ model selection criteria. These model selection criteria are limited in their effectiveness in part because they have no definitive means of determining if a difference between two values is significant. The natural choice to decide between two models is to use Bayes factors, but Bayes factors are not defined for improper priors such as the thin-plate spline model of Section 4.1.2. In order to allow for comparison between the model based on the CAR prior and the thin-plate spline prior, this chapter presents an alternative parameterization of the thin-plate splines based prior using proper priors for the unique parameters and demonstrates how the Bayes factors can be calculated using bridge sampling, thus allowing for the comparison of the two models.

This methodology is illustrated using the data from Missouri Turkey Hunting Survey of 1996. This dataset consists of responses from a survey mailed at random to individuals who purchased a hunting permit during the two-week turkey hunting season of 1996. Respondents indicated in which county they hunted during each week of the season, and if they harvested a turkey during that week. The surveys were mailed at random and the responses were then post-stratified by county. As a result, several counties have missing or sparse data. This makes traditional frequentist methods of rate estimation for the individual counties erratic. It is desirable to use a spatial smoother to reduce the noise in the observed rate estimates and create a map of smoothed estimates that depict the spatial variation in hunter success rates. Previously we have used a hierarchical model with a thin-plate splines prior and compared the results of that model to the results from a similar model using a CAR prior. Comparing two models using Bayes Factors requires that the two models to be compared are nested and that the unique parameters have proper priors. In this case the two models are not nested so a third null model that is nested to both models is introduced for the purposes of computation. The two models considered differ slightly from their previous parameterizations but do provide substantially the same results and allow a comparison of the two models using both the DIC and Bayes Factors.

5.2 The Data and Likelihood

The data here consist of, y_{ij}, the number of turkeys harvested in county i during week j from the survey and n_{ij}, the corresponding number of hunters for county i during week j, where $i = 1, \ldots, I$ and $j = 1, \ldots, J$. In this example for the data from the 1996 MTHS $I = 114$ and $J = 2$.

Given n_{ij}, the y_{ij} are assumed to follow independent binomial distributions

$$(y_{ij} \mid n_{ij}, p_{ij}) \overset{indep.}{\sim} Binomial(n_{ij}, p_{ij}), \quad i = 1, \ldots, I, j = 1, J, \tag{5.1}$$

where p_{ij} is the probability of harvesting a turkey in county i during week j.

5.3 The Hierarchical Models

In order to model the rates p_{ij}, we make use of the logit transformation of p_{ij}, $\nu_{ij} = \log(p_{ij}/(1-p_{ij}))$, for $i = 1, \cdots, I$, $j = 1, \cdots, J$. Following He & Sun (2000) , the first stage prior for ν_{ij} has the form,

$$\nu_{ij} = \theta_j + z_i + \epsilon_{ij}, i = 1, \ldots, I, j = 1, \ldots, J, \tag{5.2}$$

where

$$\epsilon_{ij} \overset{iid}{\sim} N(0, \delta_0). \tag{5.3}$$

Here θ_j is the effect for week j and z_i is the county effect. We write $\boldsymbol{\nu} = (\nu_{11}, \ldots, \nu_{I1}, \ldots, \nu_{IJ})'$, $\boldsymbol{X}_0 = \boldsymbol{I}_J \otimes \boldsymbol{1}_I$, and $\boldsymbol{X}_1 = \boldsymbol{1}_J \otimes \boldsymbol{I}_I$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)'$, and $\boldsymbol{z} = (z_1, \ldots, z_I)'$. Then the first stage prior $\boldsymbol{\nu}$ given $(\boldsymbol{\theta}, \boldsymbol{z}, \delta_0)$ is equivalent to

$$(\boldsymbol{\nu} \mid \boldsymbol{\theta}, \boldsymbol{z}, \delta_0) \sim N_{IJ}(\boldsymbol{X}_0\boldsymbol{\theta} + \boldsymbol{X}_1\boldsymbol{z},\ \delta_0 \boldsymbol{I}_{IJ}), \tag{5.4}$$

For the second stage prior, we assume the following density for $(\boldsymbol{\theta}, \delta_0)$,

$$\pi(\boldsymbol{\theta}, \delta_0) \quad \propto \quad \frac{1}{\delta_0}. \tag{5.5}$$

In the following two subsections, we will present two different models for the spatial effects $\boldsymbol{z}$, namely the CAR prior and the thin-plate spline prior. He & Sun (2000) showed that the CAR prior is preferable to many other priors. The main purpose of this chapter is compare the CAR prior with the thin-plate spline prior.

5.3.1 Model M_1 for $\boldsymbol{z}$

The first model M_1 uses the CAR prior for spatial effects $\boldsymbol{z}$ written as

$$(\boldsymbol{z} \mid \rho, \delta_0, \eta) \;\sim\; N_I\Big(\mathbf{0}, \frac{\delta_0}{\eta}(\boldsymbol{I} - \rho\boldsymbol{C})^{-1}\Big). \tag{5.6}$$

This is the same prior given for $\boldsymbol{z}$ in (2.6), with $\delta_1 = \delta_0/\eta$. The prior for ρ is given in (2.12). The prior for η uses a Pareto distribution, whose density is given by

$$\pi(\eta \mid a) \;=\; \frac{a}{(a+\eta)^2},\ \eta > 0, \tag{5.7}$$

where $a > 0$ is a positive constant. Although this prior is proper, it has neither mean nor variance. Furthermore, it has the following hierarchical structure.

Lemma 4 *Assume that $\eta \mid \phi \sim Exp(\phi)$ and $\phi \sim Exp(a)$. The marginal prior for η has the density (5.7).*

Proof. Note that,

$$\pi(\eta \mid \phi) \;=\; \phi e^{-\phi\eta},\ \text{and}\ \pi(\phi \mid a) = a e^{-a\phi}. \tag{5.8}$$

The result holds immediately. □

Note that the hyper-parameter a is the median of the Pareto distribution. Possible choices for a will be discussed in a later section. We summarize the Model M_1 as follows:

M_1 : $\boldsymbol{z}$ has the hierarchical structure (5.6), (2.12) and (5.7).

5.3.2 Model M_2 for $\boldsymbol{z}$

The second model M_2 uses a prior for $\boldsymbol{z}$, the county effects that is in the spirit of a decomposition of the thin-plate spline prior (4.27). . Note that the matrix $\boldsymbol{T}$

as defined in (4.8) contains a column of ones corresponding to a constant term. In our model (5.4), the column space of $\boldsymbol{X}_0$ contains a constant vector $\mathbf{1}$. To avoid identifiability issues, we remove the first column of ones in $\boldsymbol{T}$ and define,

$$\boldsymbol{T}_{-} = (T_{ij}),\ i = 1, \ldots, I,\ j = 2, \ldots, t, \tag{5.9}$$

$$\boldsymbol{\beta}_{-} = (\beta_2, \ldots, \beta_t)'. \tag{5.10}$$

Now we could define a modified prior for $\boldsymbol{z}$ based on (4.27),

$$\boldsymbol{z} = \boldsymbol{T}_{-}\boldsymbol{\beta}_{-} + \boldsymbol{X}_2\boldsymbol{u}, \tag{5.11}$$

where $\boldsymbol{X}_2\boldsymbol{u}$ term represents the part of informative normal. To define the prior for u let $\boldsymbol{D}$ be the Cholesky decomposition of $\boldsymbol{F}_2'\boldsymbol{K}\boldsymbol{F}_2$, i.e $\boldsymbol{D}$ is a lower triangular matrix with positive diagonal elements such that

$$\boldsymbol{F}_2'\boldsymbol{K}\boldsymbol{F}_2 = \boldsymbol{D}'\boldsymbol{D}. \tag{5.12}$$

Define

$$\boldsymbol{X}_2 = \boldsymbol{K}\boldsymbol{F}_2\boldsymbol{D}^{-1}. \tag{5.13}$$

The prior for $(\boldsymbol{u} \mid \eta_2, \delta_0)$ is $N_{I-t}(\mathbf{0}, \delta_0/\eta_2\boldsymbol{I})$, i.e.

$$[\boldsymbol{u} \mid \eta_2, \delta_0] = \left(\frac{\eta_2}{2\pi\delta_0}\right)^{(I-t)/2} \exp\left(-\frac{\eta_2}{2\delta_0}\boldsymbol{u}'\boldsymbol{u}\right). \tag{5.14}$$

The prior for η_2 uses the Pareto distribution as in (5.7),

$$[\eta_2 \mid b] = \frac{b}{(b+\eta_2)^2}, \tag{5.15}$$

where $b > 0$ is a fixed constant.

In (4.27) we noted that the prior for $\boldsymbol{\beta}$ is flat, in this case we need a proper prior for $\boldsymbol{\beta}_{-}$ in order to calculate Bayes factors for model selection. Recently Liang et al.

(2005) studied the properties of Zellner's g-prior (Zellner 1986), specifically a Zellner-Siow prior (Zellner & Siow 1980), for testing if a group of regression coefficients of a normal linear model were zero. We adapt such a prior in our case defining $g = \eta_2$, and the prior for $\boldsymbol{\beta}_-$ as

$$[\boldsymbol{\beta}_- \mid \delta_0] = \frac{(I/2)^{\frac{1}{2}}}{(2\pi\delta_0)^{\frac{t-1}{2}}} \frac{\Gamma(\frac{t}{2})}{\Gamma(1/2)} |\boldsymbol{T}_-'\boldsymbol{T}_-|^{\frac{1}{2}} \left[\frac{I}{2} + \frac{1}{2\delta_0}\boldsymbol{\beta}_-'(\boldsymbol{T}_-'\boldsymbol{T}_-)^{-1}\boldsymbol{\beta}_-\right]^{-\frac{t}{2}}. \qquad (5.16)$$

Note that (5.16) is a multivariate t-distribution and has the following hierarchical structure.

Lemma 5 *Assume that* $(\boldsymbol{\beta}_- \mid \eta_1, \delta_0) \sim N(\mathbf{0}, \frac{\delta_0}{\eta_1}(\boldsymbol{T}_-'\boldsymbol{T}_-)^{-1})$ *and* $\eta_1 \sim Gamma(1/2, I/2)$. *The marginal prior* $\boldsymbol{\beta}_-$ *has the density (5.16).*

Proof. Note that,

$$\begin{aligned} \pi(\boldsymbol{\beta}_- \mid \eta_1, \delta_0) &= |\boldsymbol{T}_-'\boldsymbol{T}_-|^{1/2} \left(\frac{\eta_1}{2\pi\delta_0}\right)^{(t-1)/2} \exp\left(-\frac{\eta_1}{2\delta_0}\boldsymbol{\beta}_-'(\boldsymbol{T}_-'\boldsymbol{T}_-)^{-1}\boldsymbol{\beta}_-\right), \\ \pi(\eta_1) &= \frac{(I/2)^{1/2}}{\Gamma(1/2)}\eta_1^{1/2}\exp(-\eta_1(I/2)). \end{aligned}$$

The result holds immediately. □

Let These results define the prior for $\boldsymbol{z}$ in model M_2:

$M_2 : \boldsymbol{z}$ has the hierarchical structure of (5.11)-(5.16) and (5.14)-(5.15).

Note that the prior for $\boldsymbol{z}$ in model M_1 is an analog for Zellner's g-prior for part of the spatial effects in model M_2 (Liang et al. 2005).

5.4 Bayesian Computation

5.4.1 Bayes Factor Computation and the Null Model M_0

Since models M_1 and M_2 are not nested, we introduce a null model M_0 given as

$$M_0 : \ (\boldsymbol{\nu} \mid \boldsymbol{\theta}, \delta_0) \sim N_{IJ}(\boldsymbol{X}_0\boldsymbol{\theta}, \ \delta_0 \boldsymbol{I}_{IJ}), \tag{5.17}$$

where the priors for $(\boldsymbol{\theta}, \delta_0)$ is defined in (5.5). If we evaluate Bayes factor of model M_k versus M_0,

$$BF_{k0} \ = \ \frac{\pi(\boldsymbol{y} \mid M_k)}{\pi(\boldsymbol{y} \mid M_0)}, \ k = 1, 2,$$

the Bayes Factor B_{21} is

$$BF_{21} \ = \ \frac{BF_{20}}{BF_{10}}. \tag{5.18}$$

In order to evaluate BF_{k0}, we apply Meng & Wong (1996)'s bridge sampling algorithm. Let $\boldsymbol{\gamma} = (\boldsymbol{\nu}, \boldsymbol{\theta}, \delta_0)$ be the common parameters between the models M_k and M_0. Let $\gamma_{km} = (\boldsymbol{\nu}^{(m)}, \boldsymbol{\theta}^{(m)}, \delta_0^{(m)})$, $m = 1, \ldots, n_k$, be the output of the Gibbs sampling for model M_k. The algorithm is iterated until convergence, and we set $B_{k0}^{(0)} = 1$. The estimate of B_{k0}, the Bayes factor comparing models M_k and M_0 at the $(g+1)$ iteration is

$$\hat{B}_{k0}^{(g+1)} = \frac{\frac{1}{n_0}\sum_{m=1}^{n_0} \frac{l_{0m}}{d_k l_{0m} + d_0 \hat{B}_{k0}^{(g)}}}{\frac{1}{n_k}\sum_{m=1}^{n_k} \frac{1}{d_k l_{km} + d_0 \hat{B}_{k0}^{(g)}}}, \tag{5.19}$$

where $d_k = 1 - d_0 = n_k/(n_k + n_0)$, $l_{0m} = q_k(\gamma_{0m})/q_0(\gamma_{0m})$, $l_{km} = q_k(\gamma_{km})/q_0(\gamma_{km})$, and q_k is the product of the likelihood and the marginal prior density of $\boldsymbol{\gamma}$ under model M_k. We choose $n_k = n_0$, so that $d_k = d_0 = 1/2$. In the case of B_{10}, after some

analytic integration we have

$$\begin{aligned}
\frac{q_1(\boldsymbol{\gamma})}{q_0(\boldsymbol{\gamma})} &= \int\int\int \frac{L(\boldsymbol{y} \mid \boldsymbol{\nu})\pi(\boldsymbol{\nu} \mid \boldsymbol{\theta}, \boldsymbol{z}, \delta_0)\pi(\boldsymbol{\theta})\pi(\boldsymbol{z} \mid \eta, \rho, \delta_0)\pi(\delta_0)\pi(\eta)\pi(\rho)}{L(\boldsymbol{y} \mid \boldsymbol{\nu})\pi(\nu \mid \boldsymbol{\theta}, \delta_0)\pi(\boldsymbol{\theta})\pi(\delta_0)} d\boldsymbol{z}d\eta d\rho \\
&= \int\int\int \frac{\pi(\boldsymbol{\nu} \mid \boldsymbol{\theta}, \boldsymbol{z}, \delta_0)\pi(\boldsymbol{z} \mid \eta, \rho, \delta_0)\pi(\eta)\pi(\rho)}{\pi(\boldsymbol{\nu} \mid \boldsymbol{\theta}, \delta_0)} d\boldsymbol{z}d\eta d\rho \\
&= \int_0^{\infty}\int_{\lambda_1^{-1}}^{\lambda_I^{-1}} \sqrt{\frac{\mid \boldsymbol{I} - \rho\boldsymbol{C} \mid}{|\boldsymbol{G}_{\eta,\rho}|}} \frac{a\eta^{I/2}}{(\eta + a)^2} \exp\left\{ \frac{1}{2\delta_0} (\boldsymbol{\nu} - \boldsymbol{X}_0\boldsymbol{\theta})' \boldsymbol{X}_1 \boldsymbol{G}_{\eta,\rho}^{-1} \boldsymbol{X}_1' (\boldsymbol{\nu} - \boldsymbol{X}_0\boldsymbol{\theta}) \right\} d\rho d\eta,
\end{aligned}$$

where $\boldsymbol{G}_{\eta,\rho} = J\boldsymbol{I}_I + \eta(\boldsymbol{I} - \rho\boldsymbol{C})$. In the case of B_{20}, after some analytic integration

$$\begin{aligned}
&\frac{q_2(\boldsymbol{\gamma})}{q_0(\boldsymbol{\gamma})} \\
&= \frac{L(\boldsymbol{y} \mid \boldsymbol{\nu})\pi(\boldsymbol{\theta})\pi(\delta_0)}{L(\boldsymbol{y} \mid \boldsymbol{\nu})\pi(\boldsymbol{\theta})\pi(\delta_0)\pi(\nu \mid \boldsymbol{\theta}, \delta_0)} \\
&\times \int\int\int\int \pi(\boldsymbol{\nu} \mid \boldsymbol{\theta}, \boldsymbol{\beta}_-, \boldsymbol{u}, \delta_0)\pi(\boldsymbol{\beta}_- \mid \eta_1, \delta_0)\pi(\boldsymbol{u} \mid \eta_2, \delta_0)\pi(\eta_1)\pi(\eta_2) d\boldsymbol{\beta}_- d\boldsymbol{u} d\eta_1 d\eta_2 \\
&= \int_0^{\infty}\int_0^{\infty} \frac{\eta_1^{1/2}(I/2)^{1/2}}{(1 + \eta_1)\sqrt{|\boldsymbol{G}_{\eta_2}|}\Gamma(\frac{1}{2})} \frac{b\eta_2^{(I-3)/2}}{(\eta_2 + b)^2} \exp\left(-\frac{I\eta_1}{2} + \frac{1}{2\delta_0} \widetilde{\boldsymbol{\nu}}' \boldsymbol{X}_2 \boldsymbol{G}_{\eta_2}^{-1} \boldsymbol{X}_2' \widetilde{\boldsymbol{\nu}} \right) d\eta_1 d\eta_2,
\end{aligned}$$

where

$$\widetilde{\boldsymbol{\nu}} = \boldsymbol{\nu} - \boldsymbol{X}_0\boldsymbol{\theta} \text{ and } \boldsymbol{G}_{\eta_2} = \boldsymbol{I} + \boldsymbol{X}_2' \left(\boldsymbol{I} + \frac{1}{1 + \eta_2} \boldsymbol{T}_- (\boldsymbol{T}_-' \boldsymbol{T}_-)^{-1} \boldsymbol{T}_-' \right) \boldsymbol{X}_2.$$

The remaining integration can be carried out numerically and the resulting Bayes Factors calculated.

5.4.2 Full Conditional Distributions

Evaluation of the models M_0, M_1 and M_2 are carried out using Gibbs sampling. In order to do this the necessary conditional posterior distributions are needed. The following are given whose proofs are standard and omitted.

Lemma 6 *Under model M_0, the conditional distributions are given below.*

(a) For given $(\boldsymbol{\theta}, \delta_0; \boldsymbol{y})$, $(\nu_{11}, \ldots, \nu_{IJ})$ *are independent. The conditional density of* ν_{ij} *depends only on* y_{ij} *and is given by*

$$[\nu_{ij} \mid \theta_j, \delta_0, y_{ij}] \;\; \propto \;\; \exp(\nu_{ij} y_{ij} - n_{ij} \log(1 + e^{\nu_{ij}}) - (\nu_{ij} - \theta_j)^2 / 2\delta_0).$$

(b) The conditional posterior density of ν_{ij} *in (a) is log-concave.*

(c) $(\boldsymbol{\theta} \mid \boldsymbol{\nu}, \delta_0; \boldsymbol{y}) \sim N(\boldsymbol{X}_0' \boldsymbol{\nu} / I, \;\; \delta_0 \boldsymbol{I}_J / I)$.

(d) $(\delta_0 \mid \boldsymbol{\nu}, \boldsymbol{\theta}; \boldsymbol{y}) \sim IG(IJ/2, (\boldsymbol{\nu} - \boldsymbol{X}_0 \boldsymbol{\theta})'(\boldsymbol{\nu} - \boldsymbol{X}_0 \boldsymbol{\theta}) / 2\delta_0)$.

Lemma 7 *Under model* M_1*, we have the following conditional distributions.*

(a) For given $(\boldsymbol{\theta}, z_i, \delta_0; \boldsymbol{y})$, $(\nu_{11}, \ldots, \nu_{IJ})$ *are independent. The conditional density of* ν_{ij} *depends only on* y_{ij} *and is given by*

$$[\nu_{ij} \mid \theta_j, z_i, \delta_0, y_{ij}] \;\; \propto \;\; \exp\Big(\nu_{ij} y_{ij} - n_{ij} \log(1 + e^{\nu_{ij}}) - (\nu_{ij} - \theta_j - z_i)^2 / 2\delta_0\Big).$$

(b) The conditional posterior density of ν_{ij} *in (a) is log-concave.*

(c) $(\boldsymbol{\theta} \mid \boldsymbol{\nu}, \boldsymbol{z}, \delta_0; \boldsymbol{y}) \sim N(\boldsymbol{X}_0' (\boldsymbol{\nu} - \boldsymbol{X}_1 \boldsymbol{z}) / I, \;\; \delta_0 \boldsymbol{I}_J / I)$.

(d) $(\boldsymbol{z} \mid \boldsymbol{\nu}, \boldsymbol{\theta}, \eta, \delta_0, \rho; \boldsymbol{y}) \sim N(\boldsymbol{a}, \delta_0 \boldsymbol{B})$*, where* $\boldsymbol{B} = (\eta(\boldsymbol{I}_I - \rho \boldsymbol{C}) + J \boldsymbol{I}_I)^{-1}$ *and* $\boldsymbol{a} = \boldsymbol{B} \boldsymbol{X}_1' (\boldsymbol{\nu} - \boldsymbol{X}_0 \boldsymbol{\theta})$.

(e) $(\delta_0 \mid \boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{z}, \eta; \boldsymbol{y}) \sim IG(I(J+1)/2, c/2)$*, where*

$$c = (\boldsymbol{\nu} - \boldsymbol{X}_0 \boldsymbol{\theta} - \boldsymbol{X}_1 \boldsymbol{z})'(\boldsymbol{\nu} - \boldsymbol{X}_0 \boldsymbol{\theta} - \boldsymbol{X}_1 \boldsymbol{z}) + \eta \boldsymbol{z}' (\boldsymbol{I}_I - \rho \boldsymbol{C}) \boldsymbol{z}.$$

(f) $(\eta \mid \boldsymbol{z}, \phi, \rho; \boldsymbol{y}) \sim Gamma(I/2, \phi + \boldsymbol{z}' (\boldsymbol{I}_I - \rho \boldsymbol{C}) \boldsymbol{z} / 2\delta_0)$.

(g) $(\phi \mid \eta; \boldsymbol{y}) \sim Gamma(2, \eta + a)$.

(h) $[\rho \mid \boldsymbol{z}, \eta, \delta_0; \boldsymbol{y}] \propto |\boldsymbol{I} - \rho \boldsymbol{C}|^{1/2} \exp(\eta \rho \boldsymbol{z}' \boldsymbol{C} \boldsymbol{z} / 2\delta_0)$.

(i) The conditional density of ρ *in part (h) is log-concave.*

Lemma 8 *Under model* M_2*, we have the following conditional distributions.*

(a) For given $(\boldsymbol{\theta}, \boldsymbol{\beta}_-, \boldsymbol{u}, \delta_0; \boldsymbol{y})$*,* $(\nu_{11}, \ldots, \nu_{IJ})$ *are independent. The conditional density of* ν_{ij} *depends only on* y_{ij} *and is given by*

$$[\nu_{ij} \mid \theta_j, \boldsymbol{\beta}_-, \boldsymbol{u}, \delta_0, y_{ij}] \propto$$
$$\exp\Big(\nu_{ij} y_{ij} - n_{ij} \log(1 + e^{\nu_{ij}}) - (\nu_{ij} - \theta_j - (\boldsymbol{T}_- \boldsymbol{\beta}_- + \boldsymbol{X}_2 \boldsymbol{u})_i)^2 / 2\delta_0\Big).$$

(b) The conditional posterior density of ν_{ij} *in (a) is log-concave.*

(c) $(\boldsymbol{\theta} \mid \boldsymbol{\nu}, \boldsymbol{\beta}_-, \boldsymbol{u}, \delta_0; \boldsymbol{y}) \sim N(\boldsymbol{X}_0'(\boldsymbol{\nu} - \boldsymbol{X}_1(\boldsymbol{T}_- \boldsymbol{\beta}_- + \boldsymbol{X}_2 \boldsymbol{u}))/I, \delta_0 \boldsymbol{I}_J / I)$.

(d) $(\boldsymbol{\beta}_- \mid \boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{u}, \eta_1, \delta_0; \boldsymbol{y}) \sim N(g(\boldsymbol{T}_-' \boldsymbol{T}_-)^{-1} \boldsymbol{T}_-'(\boldsymbol{\nu} - \boldsymbol{X}_0 \boldsymbol{\theta} - \boldsymbol{X}_1 \boldsymbol{X}_2 \boldsymbol{u}), \delta_0 g (\boldsymbol{T}_-' \boldsymbol{T}_-)^{-1})$,
where $g = 1/(1 + \eta_1)$.

(e) $(\boldsymbol{u} \mid \boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\beta}_-, \eta_2, \delta_0) \sim N(\boldsymbol{G} \boldsymbol{X}_2' \boldsymbol{X}_1'(\boldsymbol{\nu} - \boldsymbol{X}_0 \boldsymbol{\theta} - \boldsymbol{X}_1 \boldsymbol{T}_- \boldsymbol{\beta}_-), \delta_0 \boldsymbol{G})$,
where $\boldsymbol{G} = (\eta_2 \boldsymbol{I}_{I-3} + J \boldsymbol{X}_2' \boldsymbol{X}_2)^{-1}$.

(f) $(\delta_0 \mid \boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\beta}_-, \boldsymbol{u}, \eta_1, \eta_2; \boldsymbol{y}) \sim IG((IJ + I - 3)/2 + 1,\ d/2)$,
where $d = \tilde{\boldsymbol{\nu}}' \tilde{\boldsymbol{\nu}} + \eta_1 \boldsymbol{\beta}_-' (\boldsymbol{T}_-' \boldsymbol{T}_-) \boldsymbol{\beta}_- + \eta_2 \boldsymbol{u}' \boldsymbol{u}$ *and* $\tilde{\boldsymbol{\nu}} = (\boldsymbol{\nu} - \boldsymbol{X}_0 \boldsymbol{\theta} - \boldsymbol{X}_1 (\boldsymbol{T}_- \boldsymbol{\beta}_- + \boldsymbol{X}_2 \boldsymbol{u}))$.

(g) $(\eta_1 \mid \boldsymbol{\beta}_-, \delta_0; \boldsymbol{y}) \sim Gamma(3/2, I + \boldsymbol{\beta}_-' \boldsymbol{T}_-' \boldsymbol{T}_- \boldsymbol{\beta}_- / 2\delta_0)$.

(h) $(\eta_2 \mid \boldsymbol{u}, \phi; \boldsymbol{y}) \sim Gamma((I - 3)/2 + 1, \phi + \boldsymbol{u}' \boldsymbol{u} / 2\delta_0)$.

(i) $(\phi \mid \eta_2; \boldsymbol{y}) \sim Gamma(2, \eta_2 + a)$.

The resulting full conditionals can all be evaluated using Gibbs sampling. For the log-concave densities, the ARS algorithm in Gilks & Wild (1992) is used.

5.4.3 Choices for the Prior Median for η and η_2

The parameters a and b in (5.7) and (5.15) are the medians of the Pareto distribution. While this distribution is vague in that it has no mean or variance, the choice of median is still informative. This leads to the question as to what is a reasonable choice for a or b? Interpreting the prior median a is the more complex case so as a simpler illustration we will examine the case of choosing b, the prior median for η_2. From a modeling perspective, we can consider our choice as being one where we want to choose apriori, how complex our model is. Model complexity can be expressed as the effective number of parameters. In the case of thin-plate spline models this is the trace of the smoother matrix, as shown in Hastie & Tibrishani (1990). In this example the smoother matrix is written in terms of the posterior covariance matrix for $\boldsymbol{u}$, and is given as,

$$\boldsymbol{X}_2(\eta_2\boldsymbol{I}_{I-3} + J\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'. \tag{5.20}$$

As $\eta_2 \to \infty$, the trace of (5.20) approaches 0, and as $\eta_2 \to 0$ it approaches $I-3$, the rank of $\boldsymbol{X}_2$. The larger η_2, the smoother the model fit, the smaller η_2, the rougher the model fit. The value of the trace may not have any direct relation to the number of parameters in the model, but it does give a relative scale with which to choose a value for the prior median b. So we want to pick a value for the prior model complexity, between those limits, and then choose the prior median corresponding to that value. Obviously the upper limit assumes that the model has a large degree of roughness, and is equivalent to a saturated model. Conversely if $\eta_2 = 0$, we assume that there are no non-linear effects and that the spatial pattern can be described in linear terms. In the first case, assuming the saturated model is true defeats the purpose of modeling at all. If we assume that the model is purely linear, then the addition of a non-linear

term is unnecessary and shouldn't even be in the model. Instead we should pick what we think is a reasonable number of non-linear terms to describe the data, given the number of observations. Given that there is a maximum of $I - 3 = 111$ non-linear terms in the model, it is reasonable to assume apriori that 20 or so would suffice to adequately describe the data. If we choose $\eta_2 = 1$ the trace of (5.20) is 18.23, which for our purposes is close enough to 20, thus we will choose $a = b = 1$.

5.4.4 Decorrelation to Improve Mixing of the Posterior Sampling Chain

Preliminary evidence indicates that the MCMC chain for $\boldsymbol{\theta}$ and $\boldsymbol{z}$ mix poorly in the CAR model, and that the chains for $\boldsymbol{\theta}$ and $\boldsymbol{u}$ mix poorly in the g-prior model. In order to improve the mixing of these chains we apply decorrelation steps as described in Graves et al. (n.d.). These decorrelation steps are extra steps in the MCMC chain that help reduce or eliminate autocorrelation in the chain. These steps work by adding moves to the parameters of interest in directions that are likelihood invariant after each MCMC cycle is complete. The basis for this procedure is from the work of Liu & Sabatti (2000). A thorough explanation of the theoretical reasons for this procedure and why it works is beyond the scope of this paper. Instead, we will present the algorithms for the application of the decorrelation steps in each case.

For the CAR model, $\nu_{ij} = \theta_j + z_i$, and the likelihood for $\boldsymbol{\theta}$ and $\boldsymbol{z}$ is invariant under the transformation $\boldsymbol{\theta} \to \boldsymbol{\theta} + \mathbf{1}c,\ \boldsymbol{z} \to \boldsymbol{z} - \mathbf{1}c$ for any scalar c according to Liu & Sabatti (2000), such a move is valid if c is chosen with distribution

$$(c \mid \boldsymbol{z}, \rho, \eta_1, \delta_0) \sim N(s\mathbf{1}'(\boldsymbol{I}_I - \rho\boldsymbol{C})\boldsymbol{z}, s\delta_0/(\eta_1)). \tag{5.21}$$

where $s = 1/(IJ - \rho\mathbf{1}'\boldsymbol{C}\mathbf{1})$. The following augmented MCMC algorithm displayed

much better mixing and converged significantly faster than the traditional Gibbs algorithm.

Algorithm with Decorrelation Step for Model M_1

- Generate one full cycle, obtaining updated $\boldsymbol{\theta}$ and $\boldsymbol{z}$.
- Update the value for c from (5.21).
- Replace $\boldsymbol{\theta}$ with $\boldsymbol{\theta} + \mathbf{1}_J c$.
- Replace $\boldsymbol{z}$ with $\boldsymbol{z} + \mathbf{1}_I c$.

For the g-prior model the likelihood of $\boldsymbol{\theta}$ and $\boldsymbol{u}$ given $\boldsymbol{\nu}$ is nearly invariant under the transformation $\boldsymbol{\theta} \to \boldsymbol{\theta} + \mathbf{1}c$ and $\boldsymbol{u} \to \boldsymbol{u} - \boldsymbol{w}c$, where $\boldsymbol{w} = (\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\mathbf{1}_{I-3}$. According to Liu & Sabatti (2000) this transformation is valid if c is chosen from the distribution

$$(c \mid \boldsymbol{u}, \eta_2, \delta_0) \sim N(sa, \delta_0 s). \tag{5.22}$$

where $a = [(\boldsymbol{X}_0\mathbf{1}_J - \boldsymbol{X}_1\boldsymbol{X}_2\boldsymbol{w})'(\boldsymbol{\nu} - \boldsymbol{X}_0\boldsymbol{\theta} - \boldsymbol{X}_1(\boldsymbol{T}_-\boldsymbol{\beta}_- + \boldsymbol{X}_2\boldsymbol{u})) + \eta_2\boldsymbol{w}'\boldsymbol{u}]$. and $s = [(\boldsymbol{X}_0\mathbf{1}_J - \boldsymbol{X}_1\boldsymbol{X}_2\boldsymbol{w})'(\boldsymbol{X}_0\mathbf{1}_J - \boldsymbol{X}_1\boldsymbol{X}_2\boldsymbol{w}) + \eta_2\boldsymbol{w}'\boldsymbol{w}]^{-1}$.

Algorithm with Decorrelation Step for Model M_2

- Generate one full cycle, obtaining updated $\boldsymbol{\theta}$ and $\boldsymbol{u}$.
- Update the value for c from (5.22).
- Replace $\boldsymbol{\theta}$ with $\boldsymbol{\theta} + \mathbf{1}_J c$.
- Replace $\boldsymbol{u}$ with $\boldsymbol{u} + \boldsymbol{w}c$.

These decorrelation steps are very easy to implement in existing code and are very effective at reducing the correlation within chains and improving mixing.

5.4.5 Bayes Factor Comparison to the DIC

The DIC or deviance information criterion for each of the models is calculated as outlined in Spiegelhalter et al. (2002). The results are shown in Table 5.1.

Table 5.1: DIC Values for M_0 M_1 and M_2

Model Comparison Using DIC			
Model	P_D	$\bar{D}$	DIC
M_0	100.89	13065.43	13166.32
M_1	68.80	13029.69	13098.49
M_2	38.85	13047.60	13086.45

According to these results, the models are in rank of ascending DIC, M_2, M_1 and M_0. According to the guidelines given by Spiegelhalter et al. (2002) the null and CAR models should not be considered as candidate models for the data over the g-prior model.

The Bayes Factors[1] provide a clearer picture than the DIC

$$\begin{aligned} BF_{10} &= 1627.087, \\ BF_{20} &= 2.73 \times 10^{14}, \\ BF_{21} &= 1.59 \times 10^{11}. \end{aligned}$$

These results indicate that the CAR model is significantly better than the null model and that the g-prior model is significantly better than the CAR model. In this case, the Bayes Factor results and the DIC results are comparable and reinforce the interpretation that the g-prior model is the better model to select in this case.

[1]Preliminary analysis, may not reflect exact final results.

5.5 Results

The maps provides an illustration of the differences in the raw rates, the rates smoothed by the CAR model and the rates smoothed by the g-prior model. From the maps of the raw rate estimates for weeks one and two, it is apparent that the success rates for week one are higher than those for week two, and that the rates of success are higher in the northern portion of the state during both weeks. There are a few counties in the southeastern portion of the state that have no data for either week. The spatial pattern shown by the raw rate maps does appear noisy, and while there does appear to be a pattern, it is partially obscured.

The results for the smoothed rate estimates from both the smoothed models show considerable improvement over the raw rate estimates. The pattern of success rates across the state appears more homogeneous in general and it is easier to see the pattern overall. The difference between the CAR model and the g-prior model is more subtle and reveals the difference in how these two models model spatial structure. The CAR model results appear at first glance to be rougher than the g-prior results. The g-prior model has preserved the large scale difference between regions across the state while smoothing the localized differences between counties more than the CAR model. It is also interesting to note that the counties in the southeastern corner of the state where there are no data have been smoothed toward the overall mean in the case of the CAR model, where in the g-prior model they have been smoothed toward the regional mean. This is especially apparent for week one. The results show that the g-prior model appears to preserve the regional large scale trends better than the CAR model. The trade off to this may be that the g-prior model does not do as well when it comes to identifying individual regions that are truly anomalous; this capability is

sometimes referred to as edge detection. In this instance, it may be that the CAR model is the superior model for detecting edges, or areas of transition. This is a question for further exploration and discussion.

5.6 Discussion and Further Work

The results of these models show that in this case the g-prior model is in fact the better smoother. This is demonstrated using both DIC and Bayes Factors. The results show that the g-prior model preserves overall spatial patterns better than the CAR model, while the CAR model may do better at individual anomalies. This underlying difference is apparent in the structure of the two priors. Recall that in the case of the g-prior, the matrix $\boldsymbol{K}$ contains the spatial structure of the data, and that the entries of $\boldsymbol{K}$ are basis functions for every pair coordinate points. In the case of the CAR model, information about the spatial structure is contained in the matrix $\boldsymbol{C}$, which only contains information about the first order adjacency of a given region. This intuitively explains why the g-prior model smooths more across regions of the state than the CAR model, and likewise why the CAR model preserves individual anomalies better than the g-prior model.

The usefulness of either of the models for the Missouri Department of Conservation is evident, in light of the departments goals and desires in wildlife management. The results presented here suggest that the g-prior model is the better model to use. The g-prior model is easy to implement and the results are straightforward to interpret. The issue of prior selection for the median of η_2 is one to be discussed further. Over time, it is reasonable that the professionals implementing this model for similar data each year may develop a reasonable informative choice based on their experience. In

addition, it is possible that an appropriate objective choice of prior median may also be found. In either case, the lack of a non-informative prior should not be considered an impediment to implementation and use of the g-prior for spatial models.

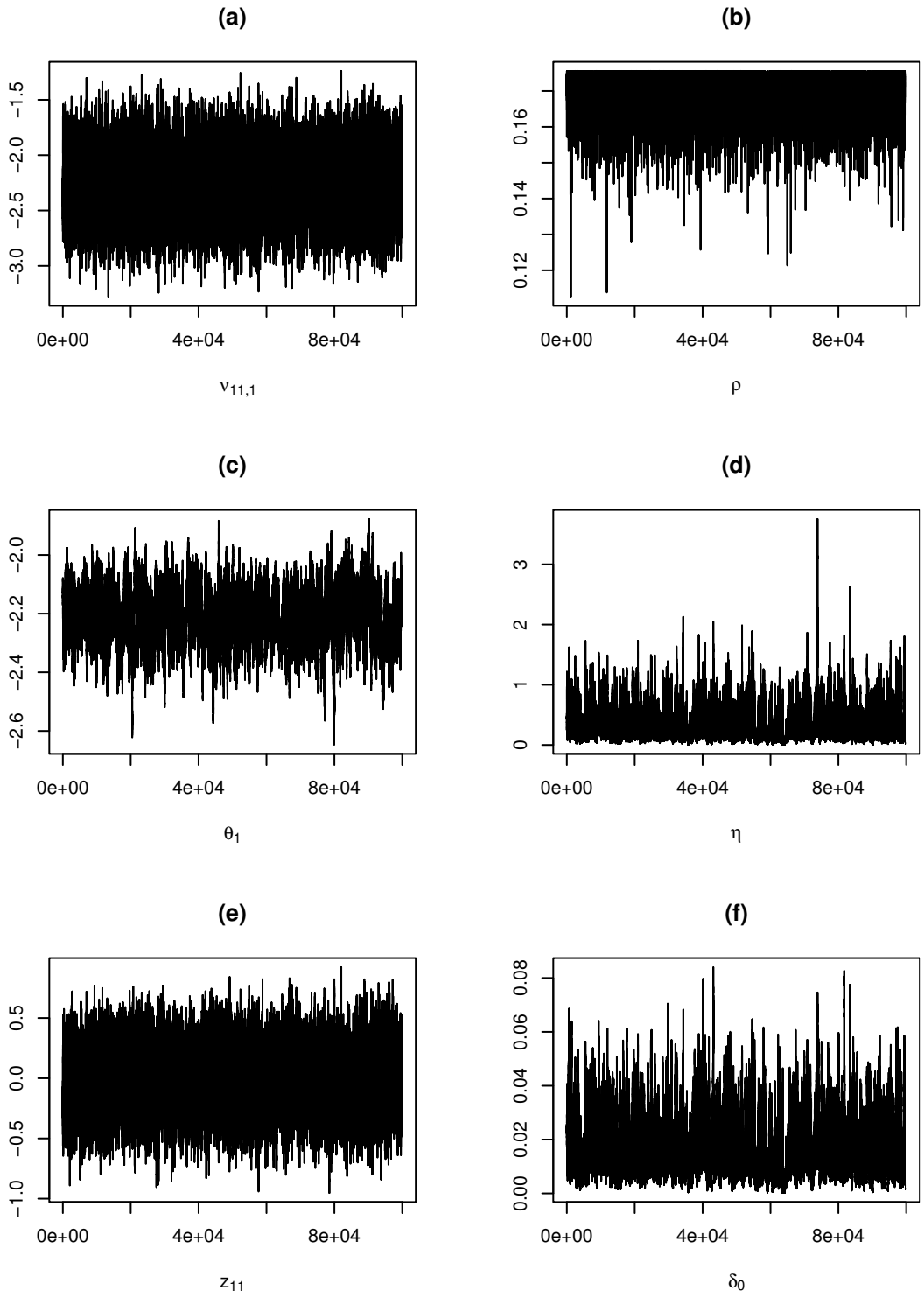

Figure 5.1: Trace Plots of (a) $\nu_{11,1}$, (b) ρ, (c) θ_1, (d) η, (e) z_{11} and (f) δ_0 from CAR Model for 1996 MTHS Without Decorrelation Step.

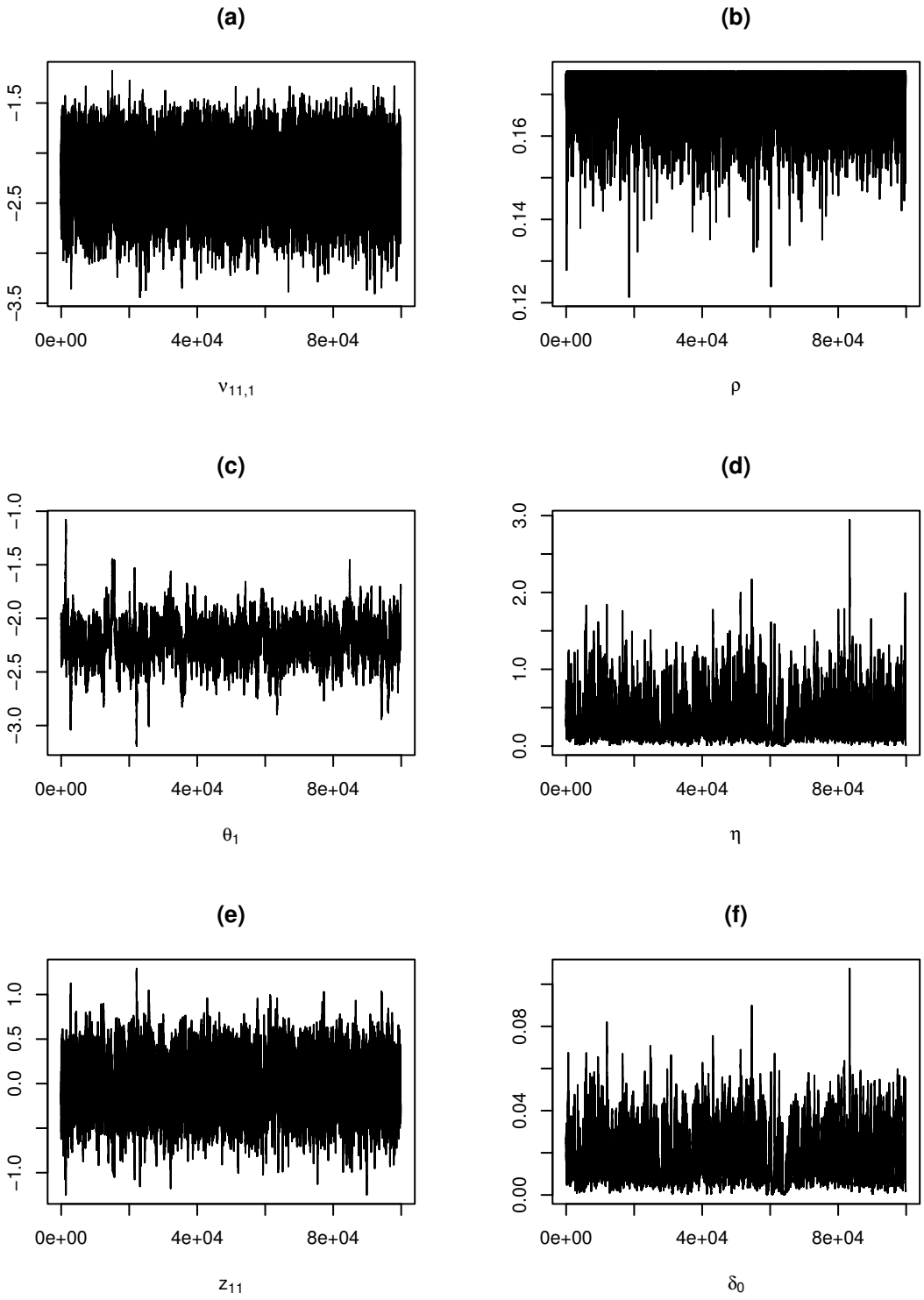

Figure 5.2: Trace Plots of (a) $\nu_{11,1}$, (b) ρ, (c) θ_1, (d) η, (e) z_{11} and (f) δ_0 from CAR Model for 1996 MTHS Without Decorrelation Step.

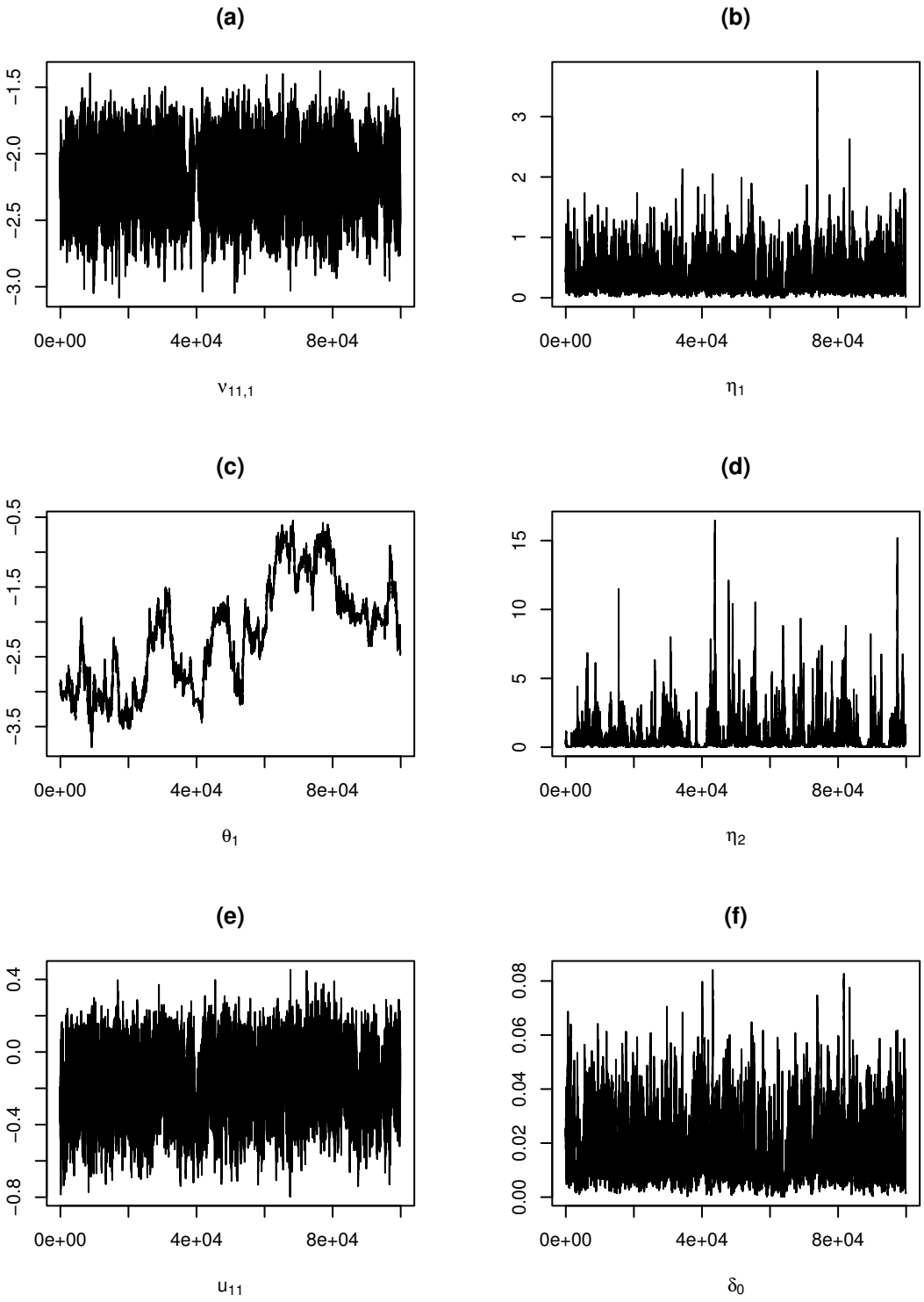

Figure 5.3: Trace Plots of (a) $\nu_{11,1}$, (b) η_1, (c) θ_1, (d) η_2, (e) u_{11} and (f) δ_0 from g-Prior Model for 1996 MTHS Without Decorrelation Step.

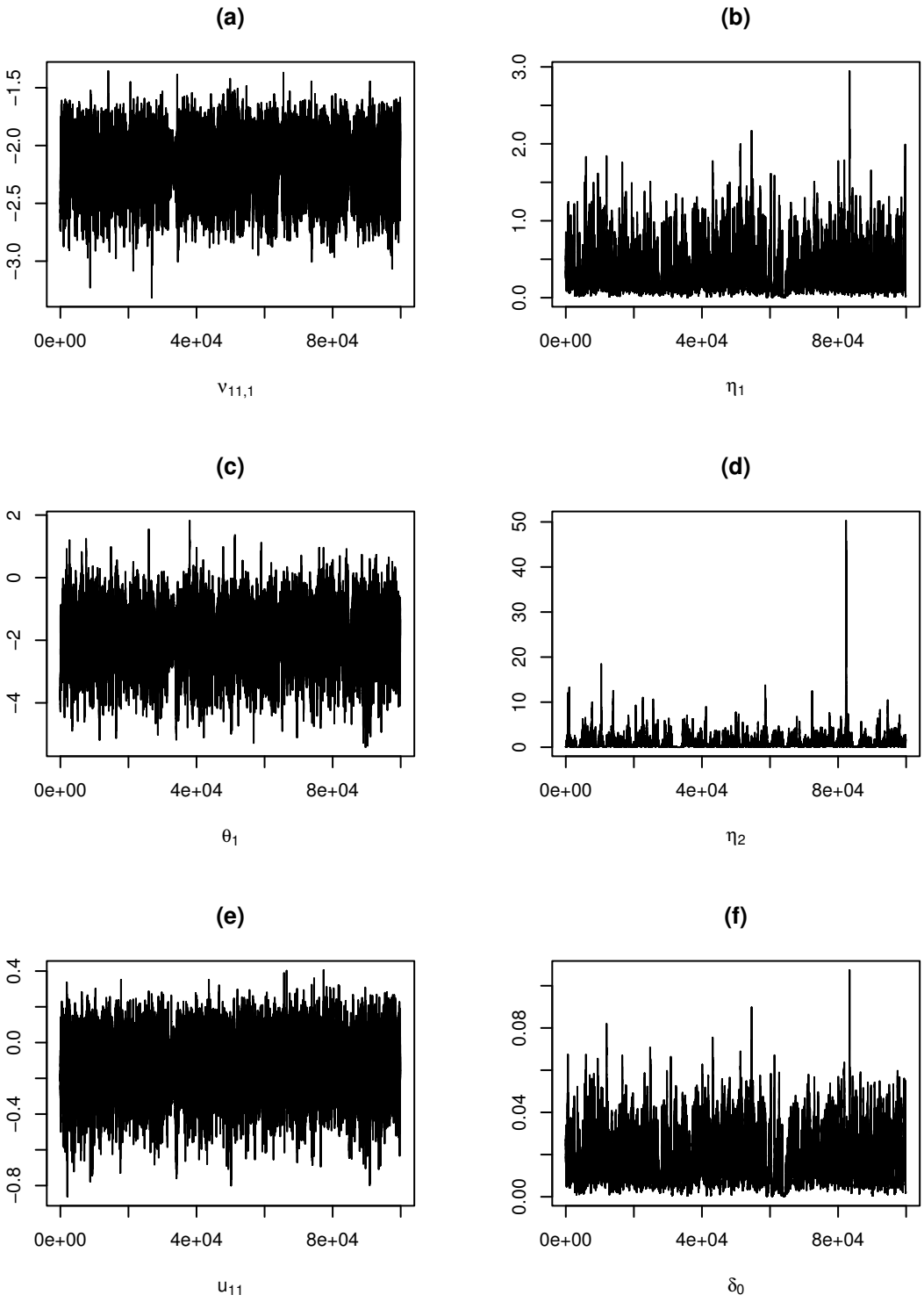

Figure 5.4: Trace Plots of (a) $\nu_{11,1}$, (b) η_1, (c) θ_1, (d) η_2, (e) u_{11} and (f) δ_0 from g-Prior Model for 1996 MTHS With Decorrelation Step.

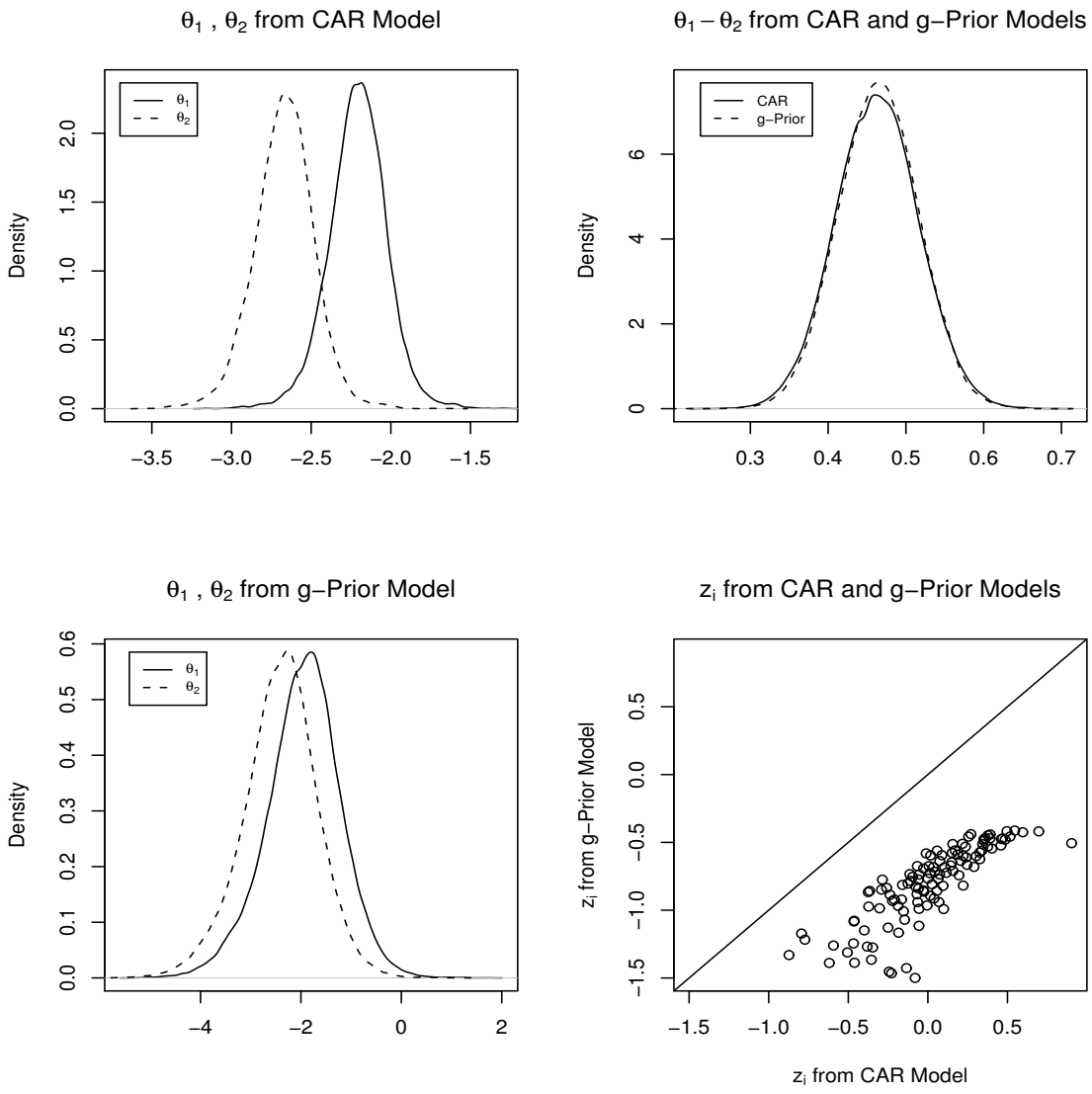

Figure 5.5: Comparison of Results from CAR and g-Prior Models for 1996 MTHS for θ_1, θ_2 and z_i.

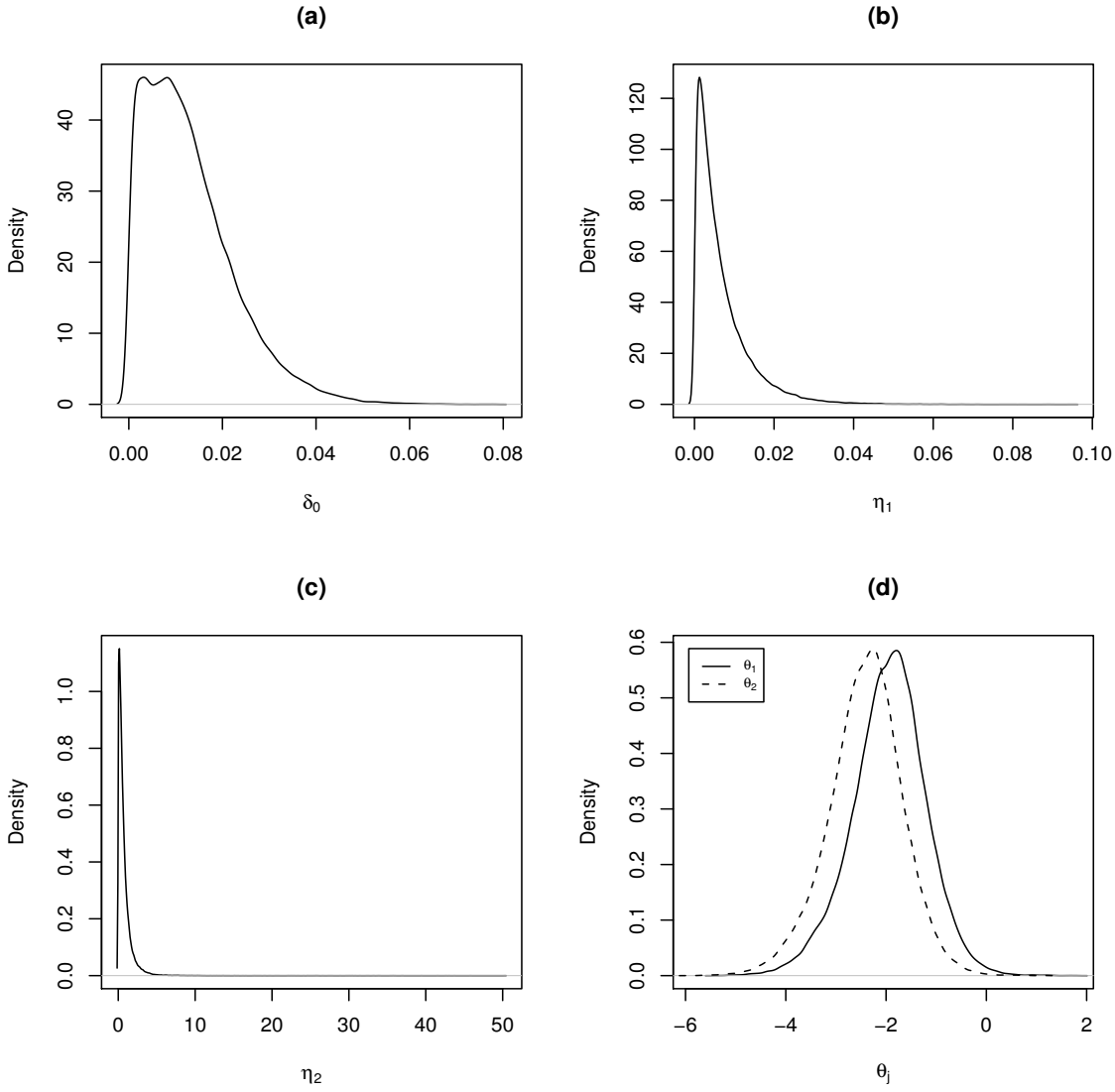

Figure 5.6: Posterior Densities for δ_0, δ_1, λ, θ_1 and θ_2 from g-Prior Model for 1996 MTHS.

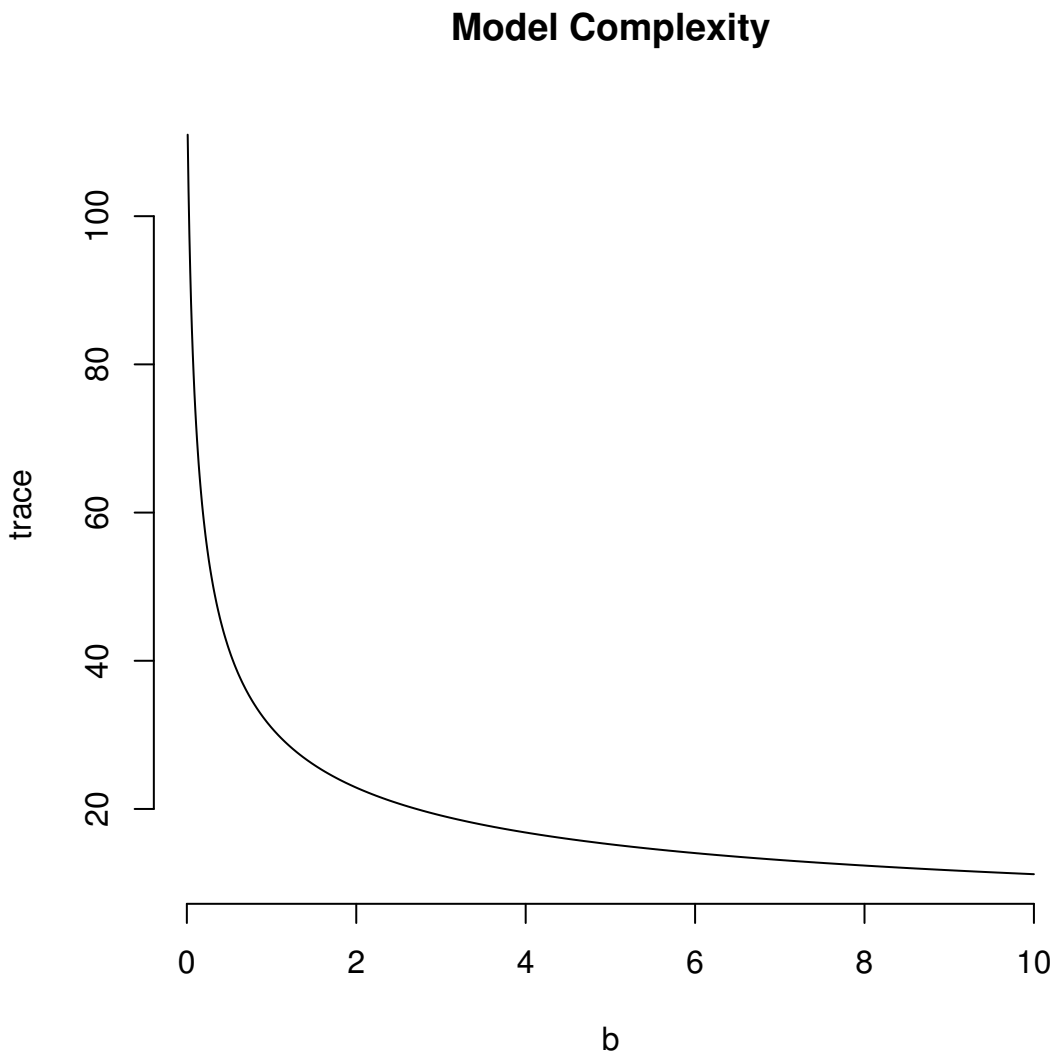

Figure 5.7: Prior Model Complexity as a Function of b for g-Prior Model of 1996 MTHS.

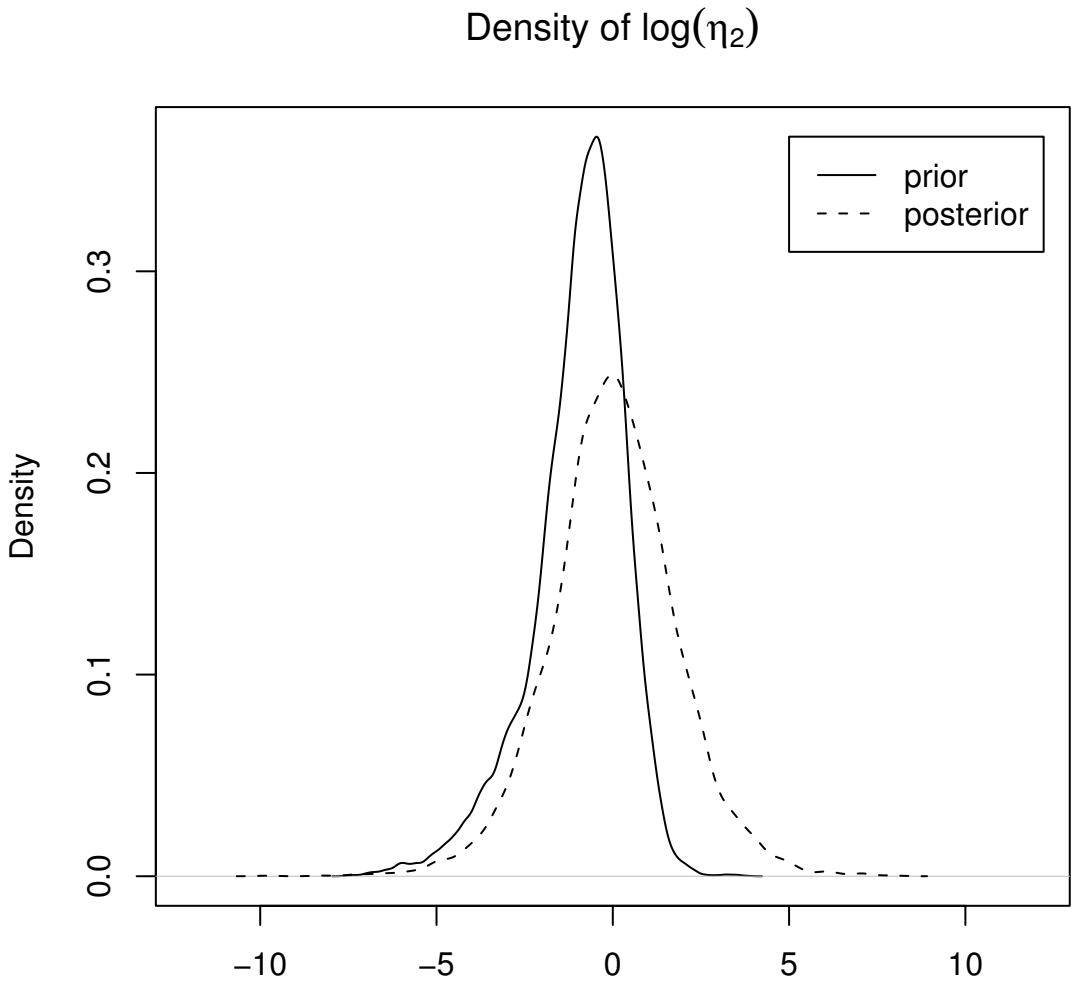

Figure 5.8: Prior and Posterior Densities of η_2 for $b = 1$ from g-Prior Model for 1996 MTHS.

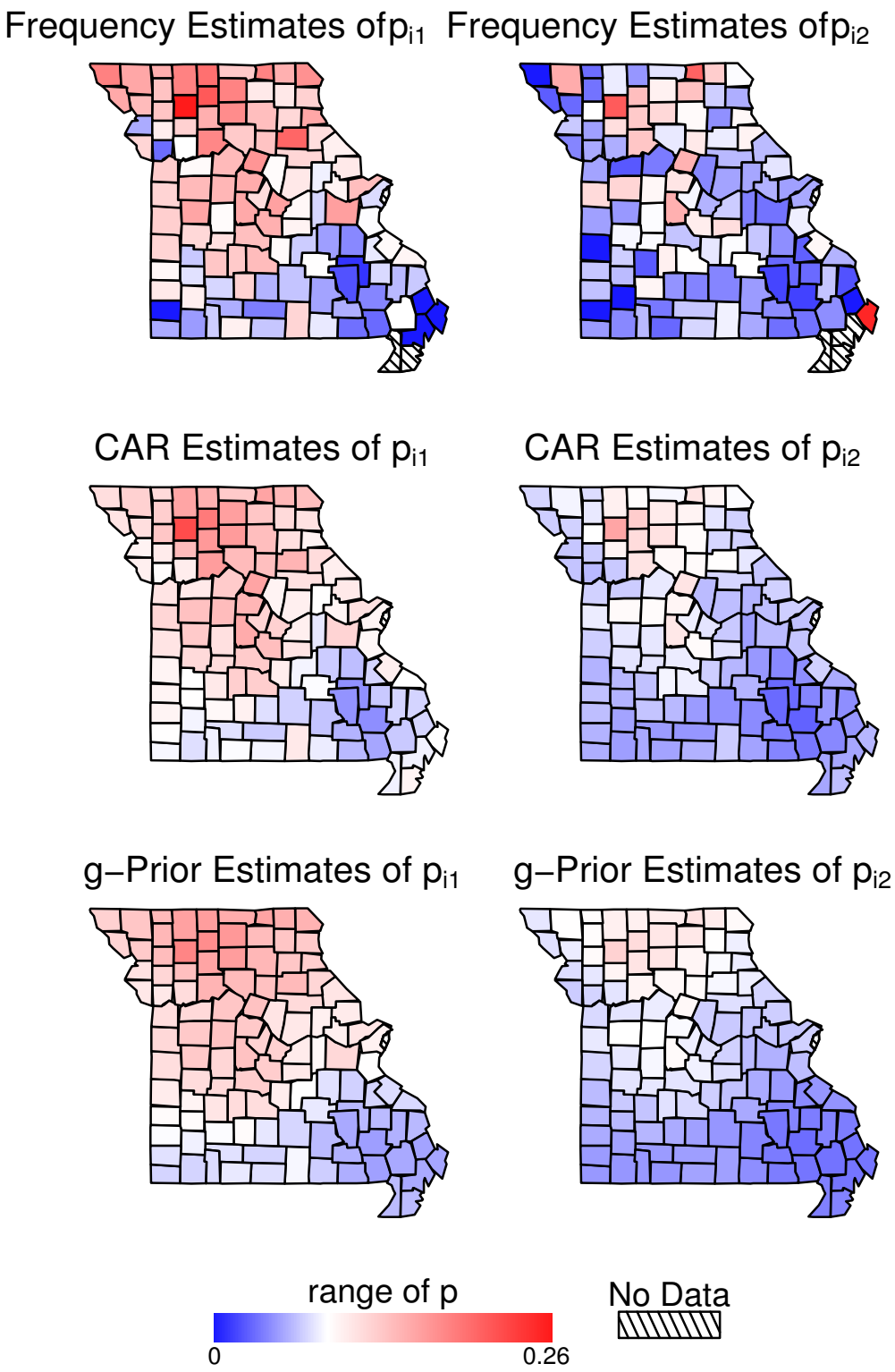

Figure 5.9: Maps of Frequency and Bayesian Estimates of Hunter Success Rates p_{ij} from the 1996 MTHS.

Chapter 6

Dimensional Reduction for a Smoothed National Cancer Map

Large scale problems in data analysis can often be impractical, as they require excessive computational time. Methods that reduce the dimension of the problem are desirable and of much interest. In the context of applying a thin-plate spline based prior for spatial effects updating at each step of the Gibbs sampler requires inverting a large matrix. While there are diagonalization procedures available, they rely on eigenvalue decomposition. These can still be computationally intense for large scale models and unreliable. Numerical limitations to eigenvalue decomposition intensify as dimension increases. In order to address this issue, we examine a method of basis reduction proposed in Nychka et al. (1996). While there are other methods of dimension reduction using principle components analysis based approaches as in Van Der Linde (2003), they suffer from the drawbacks inherent in the eigenvalue decomposition of a large matrix. We combine the reduced set of basis functions with the eigenvalue diagonalization first outlined in Section 3.5.1, and compare the results

for several different reduced bases and the full basis.

6.1 Introduction

As an example to implement this method, we use a data set of observed deaths due to colorectal cancer among men during the period 1999 through 2003 for each of the 3082 counties in the continental U.S. This may not be considered a large problem in some contexts, but the matrix that requires inversion at each iteration is 3079×3079, and is still computationally intensive. This example should be useful for exploring the possibilities of using this reduced basis technique for data reduction and its applicability to larger scale problems.

6.2 Data and Likelihood

The structure of the data and the model used here are familiar. Similar ones have been used in previous chapters. They provide a useful foundation for exploring how this methodology can be implemented and to evaluate its efficacy.

6.2.1 Data

The data set used here consists of the observed deaths due to colorectal cancer among men in each of the 3082 counties in the continental U.S. during the period 1999-2003 and each county's corresponding population size. There is some indication that there may be spatial trends in this data due to demographic and socio-economic differences across the country. For that reason, a spatial smoother on this scale is helpful to form a more coherent picture of mortality due to cancer.

6.2.2 Likelihood

The data as collected consists of y_i, the observed number of deaths in county $i = 1, \ldots, I = 3082$, and n_i, the corresponding population size. The observed number of deaths is assumed to follows a Poisson distribution

$$(y_i \mid p_i) \sim Poisson(n_i p_i), \tag{6.1}$$

where p_i is the mortality rate for county i.

6.2.3 Model

As previously we use the log of link function

$$\log(p_i) = \nu_i, \tag{6.2}$$

and consider the model

$$\boldsymbol{\nu} = \mu \mathbf{1}_I + \boldsymbol{X}_1 \boldsymbol{\beta} + \boldsymbol{X}_2 \boldsymbol{u} + \boldsymbol{\epsilon}, \tag{6.3}$$

where $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_I)'$ and $\boldsymbol{\epsilon} \sim N_I(\mathbf{0}, \delta_0 \boldsymbol{I}_I)$. Thus the first level prior for $\boldsymbol{\nu}$ is

$$(\boldsymbol{\nu} \mid \mu, \boldsymbol{\beta}, \boldsymbol{u}, \delta_0; \boldsymbol{y}) \sim N(\mu \mathbf{1}_I + \boldsymbol{X}_1 \boldsymbol{\beta} + \boldsymbol{X}_2 \boldsymbol{u}, \delta_0 \boldsymbol{I}_I). \tag{6.4}$$

The term μ represents an overall mean, and the matrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are constructed similarly to those in (4.8) and (5.12). The priors for $\boldsymbol{\beta}$ and $\boldsymbol{u}$ are given by (5.16), (5.14) and (5.15). The prior for (μ, δ_0) is

$$\pi(\mu, \delta_0) \propto \frac{1}{\delta_0}. \tag{6.5}$$

6.3 Computation

The model as given can be evaluated using a Gibbs sampler (Gelfand & Smith 1990) in conjunction with Gilks adaptive rejection sampler (Gilks & Wild 1992). The only necessary components are the full-conditional densities.

Lemma 9 *Assume that* $I > 5$,

(a) For given $(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{u}, \delta_0; \boldsymbol{y})$, $(\nu_1, \ldots, \nu_I)$ *are independent. The conditional density of* ν_i *depends only on* y_i *and is given by*

$$[\nu_i \mid \mu, \boldsymbol{\beta}, \boldsymbol{u}, \delta_0, y_i] \;\propto\; \exp\Big(\nu_i y_i - n_i e^{\nu_i} - \frac{1}{2\delta_0}(\nu_i - \mu - \boldsymbol{x}_{1i}\boldsymbol{\beta} - \boldsymbol{x}_{2i}\boldsymbol{u})^2\Big),$$

where $\boldsymbol{x}_{1i}$ *is the* i^{th} *row of* $\boldsymbol{X}_1$ *and* $\boldsymbol{x}_{2i}$ *is the* i^{th} *row of* $\boldsymbol{X}_2$.

(b) The conditional posterior density of ν_{ij} *in (a) is log-concave.*

(c) $(\boldsymbol{\mu} \mid \boldsymbol{\nu}, \boldsymbol{\beta}, \boldsymbol{u}, \delta_0; \boldsymbol{y}) \sim N(\mathbf{1}'(\boldsymbol{\nu} - \boldsymbol{X}_1\boldsymbol{\beta}_1 - \boldsymbol{X}_2\boldsymbol{u})/I, \delta_0/I)$.

(d) $(\boldsymbol{\beta} \mid \boldsymbol{\nu}, \boldsymbol{\mu}, \boldsymbol{u}, \eta_1, \delta_0; \boldsymbol{y}) \sim N(g(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'(\boldsymbol{\nu} - \mathbf{1}\mu - \boldsymbol{X}_2\boldsymbol{u}), \delta_0 g(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1})$, *where* $g = 1/(1+\eta_1)$.

(e) $(\boldsymbol{u} \mid \boldsymbol{\nu}, \mu, \boldsymbol{\beta}, \eta_2, \delta_0) \sim N(\boldsymbol{G}\boldsymbol{X}_2'(\boldsymbol{\nu} - \mathbf{1}\mu - \boldsymbol{X}_2\boldsymbol{\beta}), \delta_0\boldsymbol{G})$, *where* $\boldsymbol{G} = (\eta_2\boldsymbol{I} + \boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}$.

(f) $(\delta_0 \mid \boldsymbol{\nu}, \mu, \boldsymbol{\beta}, \boldsymbol{u}, \eta_1, \eta_2; \boldsymbol{y}) \sim IG(\frac{IJ+I-3}{2}, \frac{d}{2})$, *where* $d = \tilde{\boldsymbol{\nu}}'\tilde{\boldsymbol{\nu}} + \eta_1\boldsymbol{\beta}'(\boldsymbol{X}_1'\boldsymbol{X}_1)\boldsymbol{\beta} + \eta_2\boldsymbol{u}'(\boldsymbol{X}_2'\boldsymbol{X}_2)\boldsymbol{u}$ *and* $\tilde{\boldsymbol{\nu}} = \boldsymbol{\nu} - \mathbf{1}\mu - \boldsymbol{X}_1\boldsymbol{\beta} - \boldsymbol{X}_2\boldsymbol{u}$.

(g) $(\eta_1 \mid \boldsymbol{\beta}, \delta_0; \boldsymbol{y}) \sim Gamma(\frac{3}{2}, I + \frac{\boldsymbol{\beta}'\boldsymbol{X}_1'\boldsymbol{X}_1\boldsymbol{\beta}}{2\delta_0})$.

(h) $(\eta_2 \mid \boldsymbol{u}, \phi; \boldsymbol{y}) \sim Gamma(\frac{I-5}{2}, \frac{\boldsymbol{u}'\boldsymbol{u}}{2\delta_0} + \phi)$.

(i) $(\phi \mid \eta_2; \boldsymbol{y}) \sim Gamma(2, \eta_2 + a)$.

The full conditionals can be evaluated using Gibbs sampling (Gelfand & Smith 1990), with the exception of the conditional posterior of ν_i, which will be evaluated using the ARS algorithm of Gilks & Wild (1992).

6.3.1 Dimension Reduction

The most obvious impediment to computation in this instance is the posterior covariance matrix of $\boldsymbol{u}$, which requires inversion at each iteration of the Gibbs sampler. This inversion and updating can be facilitated by use of the algorithm presented in Section 3.5.1. This still requires the prior eigenvalue decomposition of the $(I-3) \times (I-3)$ matrix $\boldsymbol{X}_2'\boldsymbol{X}_2$. The accuracy of which is computationally limited. The obvious solution is to reduce the dimension of the matrix $\boldsymbol{X}_2'\boldsymbol{X}_2$. Nychka et al. (1996) suggests reducing the number of basis functions in (4.4). Recall that the matrix $\boldsymbol{K}$ is defined in (4.9) as $\boldsymbol{K} = (\psi_i(\boldsymbol{x}_j))$, where $\psi_i(\boldsymbol{x})$ is defined in (4.5). In all our previous examples we have assumed that in our definition of $\boldsymbol{K}$, $i = 1, \ldots, n$ and $j = 1, \ldots, n$. If we were to allow $k = 1, \ldots, q$ for some $q < n$, that is to use a smaller set of $\boldsymbol{\gamma}_k$s representative of the $\boldsymbol{x}_i$s to construct the matrices $\boldsymbol{K}$ and $\boldsymbol{T}$. We would ultimately have a smaller matrix $\boldsymbol{X}_2'\boldsymbol{X}_2$. Note that the matrix $\boldsymbol{X}_1$ is unchanged. For further explanation of this method of data reduction, see Nychka et al. (1996).

The procedure is outlined as follows. Let $\boldsymbol{\gamma}_k$, $k = 1, \ldots, q$, $t < q < n$ be a

representative set for the total observed $\boldsymbol{x}$'s. Define the matrices

$$\begin{aligned}
\boldsymbol{K}_{-} &= (K_{ik})_{n\times q},\ K_{-ik} = \psi_i(\boldsymbol{\gamma}_k),\ k = 1,\ldots,q\ i = 1,\ldots,n, && (6.6)\\
\boldsymbol{K}^{*} &= (K^{*}_{il})_{q\times q},\ K^{*}_{il} = \psi_l(\boldsymbol{\gamma}_k).\ k = 1,\ldots,q,\ l = 1,\ldots,q, && (6.7)\\
\boldsymbol{T}^{*} &= (T^{*}_{ks})_{q\times t}\ T^{*}_{ks} = \phi_s(\boldsymbol{\gamma}_k),\ k = 1,\ldots,q,\ s = 1,\ldots,t, && (6.8)\\
\boldsymbol{T}^{*}\boldsymbol{T}^{*'} &= \boldsymbol{F}^{*}\boldsymbol{\Lambda}\boldsymbol{F}^{*'}, && (6.9)\\
\boldsymbol{T}_{-} &= (T_{ij})_{n\times(t-1)},\ T_{-ij} = \phi_j(\boldsymbol{x}_i)\ i = 1,\ldots,n,\ j = 2,\ldots,t. && (6.10)
\end{aligned}$$

Then $\boldsymbol{F}^{*}_1$ is the $q \times t$ matrix of vectors spanning the column space of $\boldsymbol{T}^{*}$ and $\boldsymbol{F}^{*}_2$ is the $q \times (q-t)$ matrix of columns vectors orthogonal to $\boldsymbol{T}^{*}$. Define $\boldsymbol{D}^{*}$ again using the Cholesky decomposition $\boldsymbol{F}^{*'}_2\boldsymbol{K}^{*}\boldsymbol{F}^{*}_2 = \boldsymbol{D}^{*'}\boldsymbol{D}^{*}$. Let

$$\boldsymbol{X}_{1_{n\times t}} = \boldsymbol{T}_{-}\ \textit{and}\ \boldsymbol{X}^{*}_{2_{n\times q-t}} = \boldsymbol{K}_{-}\boldsymbol{F}^{*}_2\boldsymbol{D}^{*-1}. \tag{6.11}$$

The vector $\boldsymbol{u}^{*} = (u^{*}_1,\ldots,u^{*}_{q-t})'$, and the matrix $\boldsymbol{X}^{*'}_2\boldsymbol{X}^{*}_2$ has dimension $(q-t)\times(q-t)$. The model then can be evaluated by substituting $\boldsymbol{X}^{*}_2$ for $\boldsymbol{X}_2$ and $\boldsymbol{u}^{*}$ for $\boldsymbol{u}$, with the same formal full conditional distributions and prior changes for the dimensional reduction. This reduces the computational burden of updates in each Gibbs cycle. Choices for values of q and the specific $\boldsymbol{\gamma}_k$'s will be discussed later.

6.4 Results

The full model as written is implemented along with three other reduced models. The results are compared and comments made on the specific choices for model reduction and how they influence the results.

6.4.1 Data Reduction

The dimensional reduction technique illustrated in Section 6.3.1 is implemented by creating three different grids of a reduced number of points that cover the area of the county centroids. These three different grids consist of 1600, 900 and 100 points uniformly distributed over the area of the range of the county centroid coordinates. In each case these grids, the reduced design matrix $\boldsymbol{X}_2^*$, and the eigenvalue decomposition of $\boldsymbol{X}_2^{*'}\boldsymbol{X}_2^*$ are calculated and the models implemented. As expected, the reduced dimension results in a significant reduction in computational time. The differences in computational time are roughly proportional to the reduction in the number of basis functions, since the computational burden is already eased by the use of diagonalization algorithm illustrated in Section 3.5.1. The differences in the qualitative and quantitative results are interesting. These are examined by looking at the resulting maps of smoothed values produced by each model as well as comparing their DIC and their ability to detect significant areas with high or low rates of mortality using the standardized mortality ratios (SMRs).

6.4.2 Calculating the Standardized Mortality Ratios

The standardized mortality ratio for county i is defined as the observed mortality divided by the expected mortality.

$$SMR_i = \frac{O_i}{E_i}, \tag{6.12}$$

where the expected mortality can be based on a reference population, in this case the national population as a whole

$$E_i = \frac{\sum_k^I y_k}{\sum_k^I n_k} n_i. \tag{6.13}$$

Clearly, (6.12) can be calculated at each iteration of the MCMC chain and sampled, thus providing samples from the densities of the SMRs, which can be used for detecting areas with significantly higher or lower mortality rates.

6.4.3 Data Mapping

Figures 6.3 - 6.7 show the frequency rate estimates calculated directly from the data and the results for the smoothed values obtained from each model. Examining the map for the full model first shows that the raw data rates seem to indicate a general trend of mortality rates increasing as we move from west to east across the continental U.S. What becomes clear in the map of smoothed values is that although this trend remains after smoothing. There are several areas of dark blue, indicating a low mortality rate centered at large urban areas such as Dallas, Austin, Houston, Atlanta, and Minneapolis, as well as other large metropolitan areas. This same phenomenon does not appear to manifest itself in some areas such as New York City or Los Angeles. The reason for this effect is one of interest and open for further discussion and exploration.

The remaining maps show similar spatial patterns and smoothing, though a careful examination of the maps does show that they do in fact become smoother as the number of basis functions are reduced. The lower rates around the cities mentioned above does remain a feature of these reduced model maps.

A more quantitative comparison between these models can be made by examining the DIC for each model. The results from Table 6.1 tend to agree with the maps. There appears to be little difference between the three reduced models' DIC. The DIC with 100 basis functions seems to indicate that we might be encountering the

Table 6.1: DIC values for $q = 3082, 1600, 900$ and 100

Model Comparison Using DIC			
q	P_D	$\bar{D}$	DIC
3082	1613.13	2658972.77	2660586.90
1600	1615.34	2658972.77	2660590.10
900	1615.01	2658974.40	2660589.41
100	1619.37	2658972.25	2660591.62

limit of dimension reduction. The P_D term for this model has increased noticeably, indicating that the variance in the estimates has increased. These results would seem to indicate that this strategy for dimensional reduction appears to be quite valid and robust, and that the potential for dimensional reduction is in fact quite significant. In order to further evaluate these results, we will use these four models to detect significant differences in mortality rates by using standardized mortality ratios.

6.4.4 Standardized Mortality Ratios

The results of the SMR estimates are shown in Figures 6.8-6.11. The areas in red are areas where the SMR is significantly greater than 1, the areas in blue are areas where the SMR is significantly less than 1. Significance is determined using the 95% posterior credible interval. What is shown agrees with the results shown in the rate maps; there are more areas of significantly higher mortality in the Eastern half of the continental U.S. than in the western half. Furthermore, the cities that appeared to show lower rates again appear to be areas of significantly lower mortality.

It can also be seen that there appears to be little difference in these four models in terms of there ability to detect significant SMRs. These results further reinforce the notion that the strategy for dimensional reduction implemented here provides an

excellent means of reducing computational burdens in implementing this thin-plate spline model for large scale problems.

6.5 Discussion

These results demonstrate that this method of data analysis through smoothing and the use of SMRs is a viable means of analysis for data on a large scale, such as national cancer data. The implementation of this method relies on the ease of use and stability of the models developed in order to gain acceptance in the scientific community at large. The results from this example indicate that this method of smoothing is robust to data reduction techniques and appears to be able to smooth data well while also retaining enough of the spatial heterogeneity to be able to detect significant differences between regions. These are properties, along with the relative ease of implementation, which make this method one that could achieve wider acceptance.

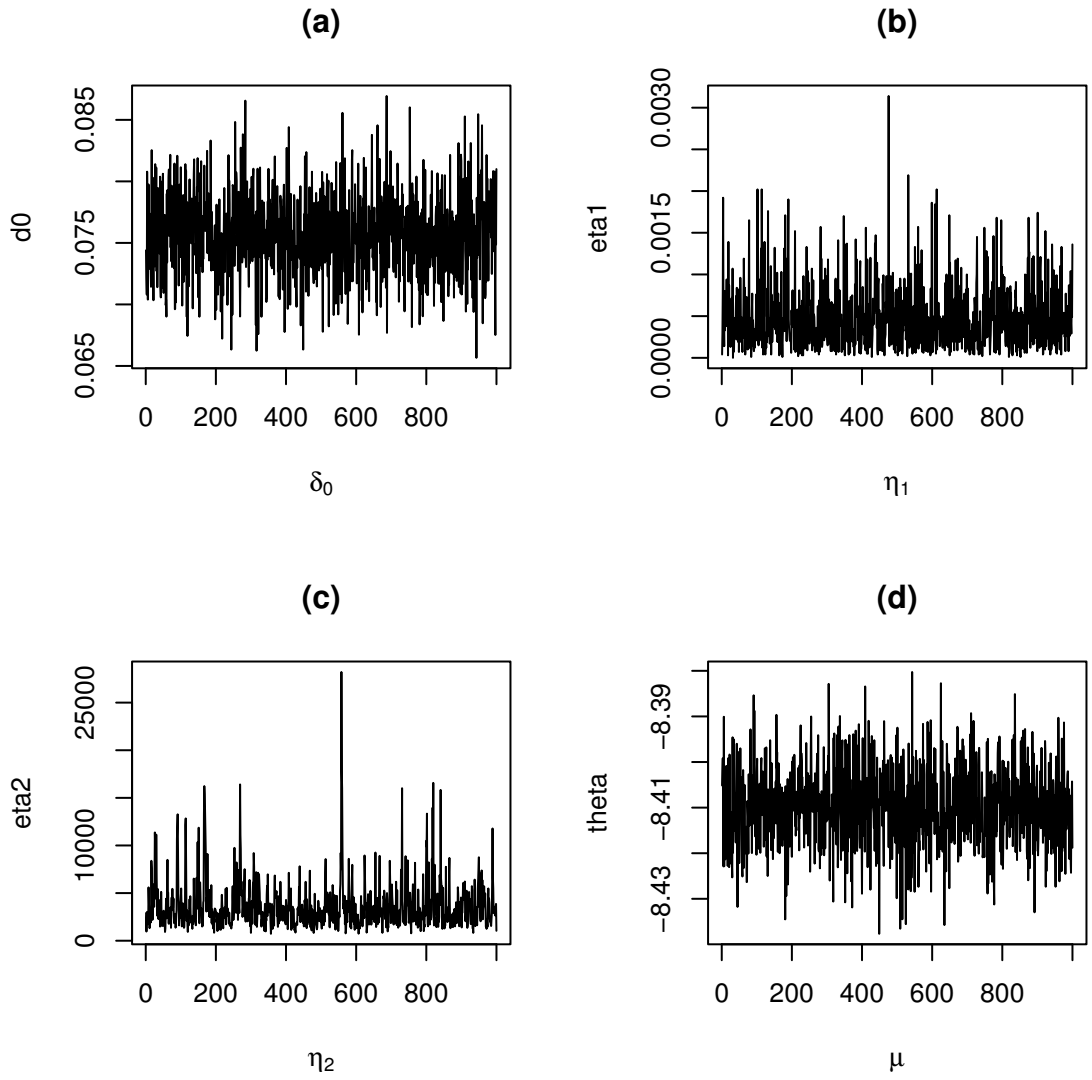

Figure 6.1: Trace Plots of (a) δ_0, (b) η_1, (c) η_2, and (d) μ from the g-Prior Model for p_i Male Mortality Rates due to Colorectal Cancer 1999-2003.

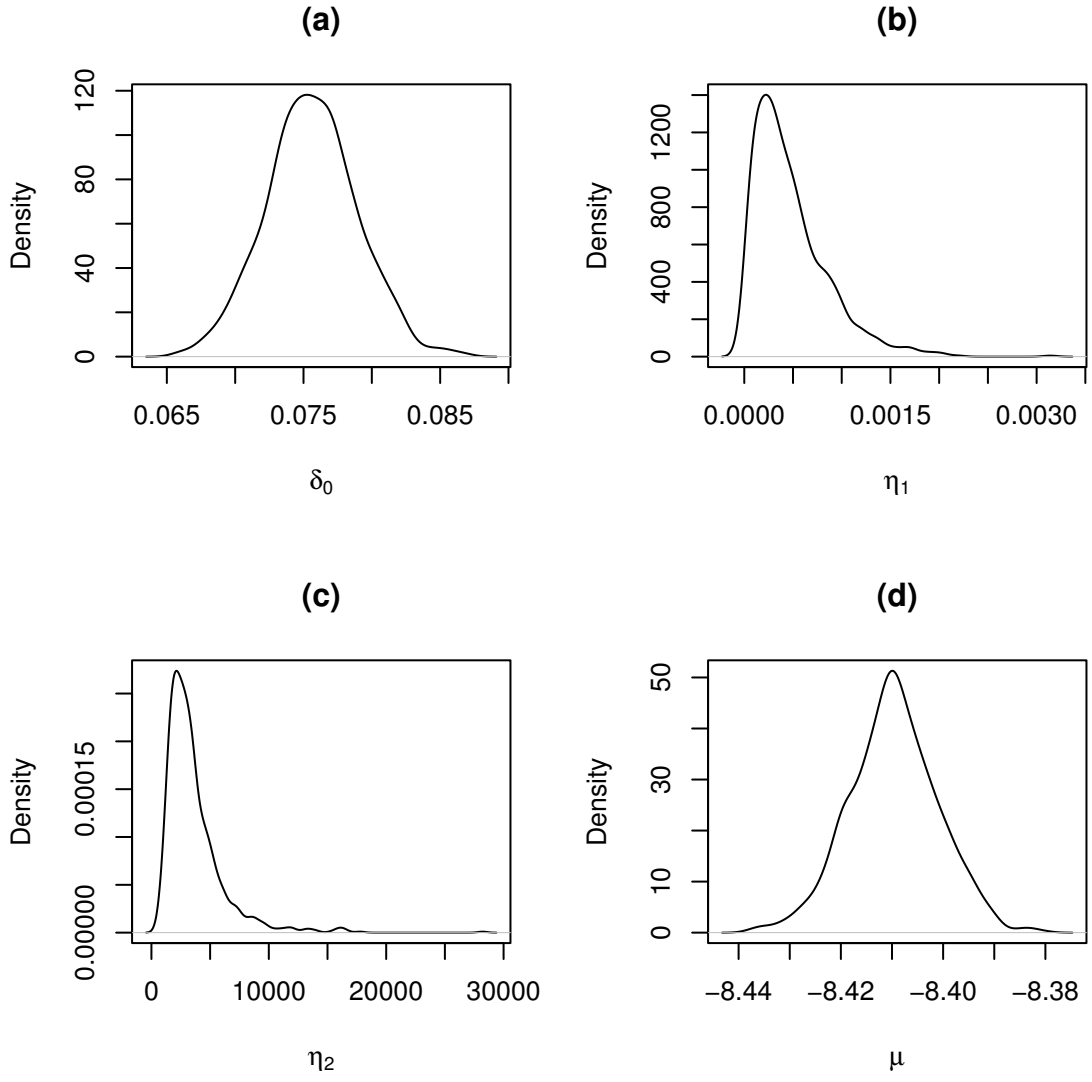

Figure 6.2: Posterior Densities for (a) δ_0, (b) η_1, (c) η_2 (d) and μ from the g-Prior Model for p_i Male Mortality Rates due to Colorectal Cancer 1999-2003.

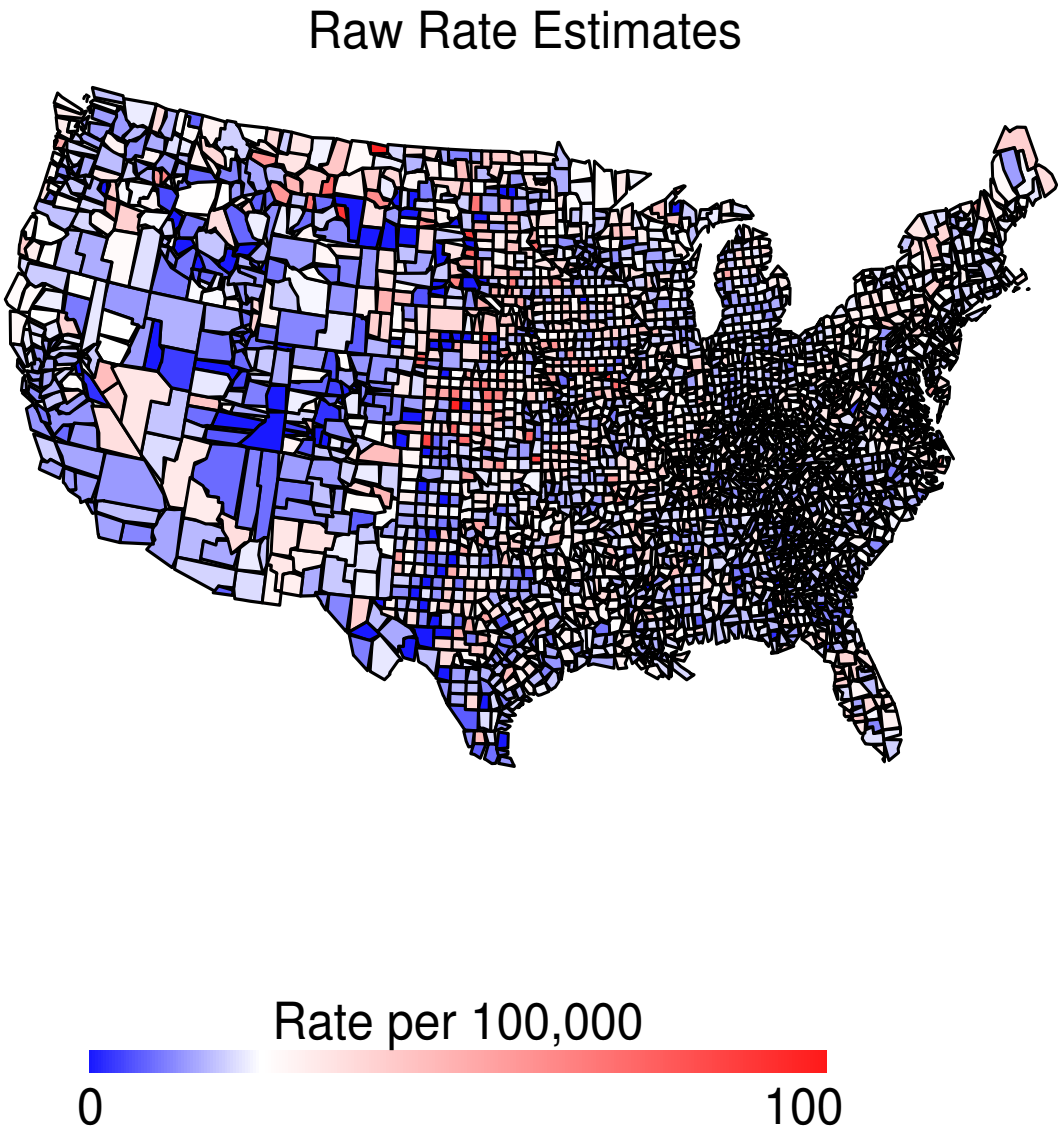

Figure 6.3: Map of Frequency Estimates for Male Mortality Rates due to Colorectal Cancer p_i per 100,000 from $1999 - 2003$.

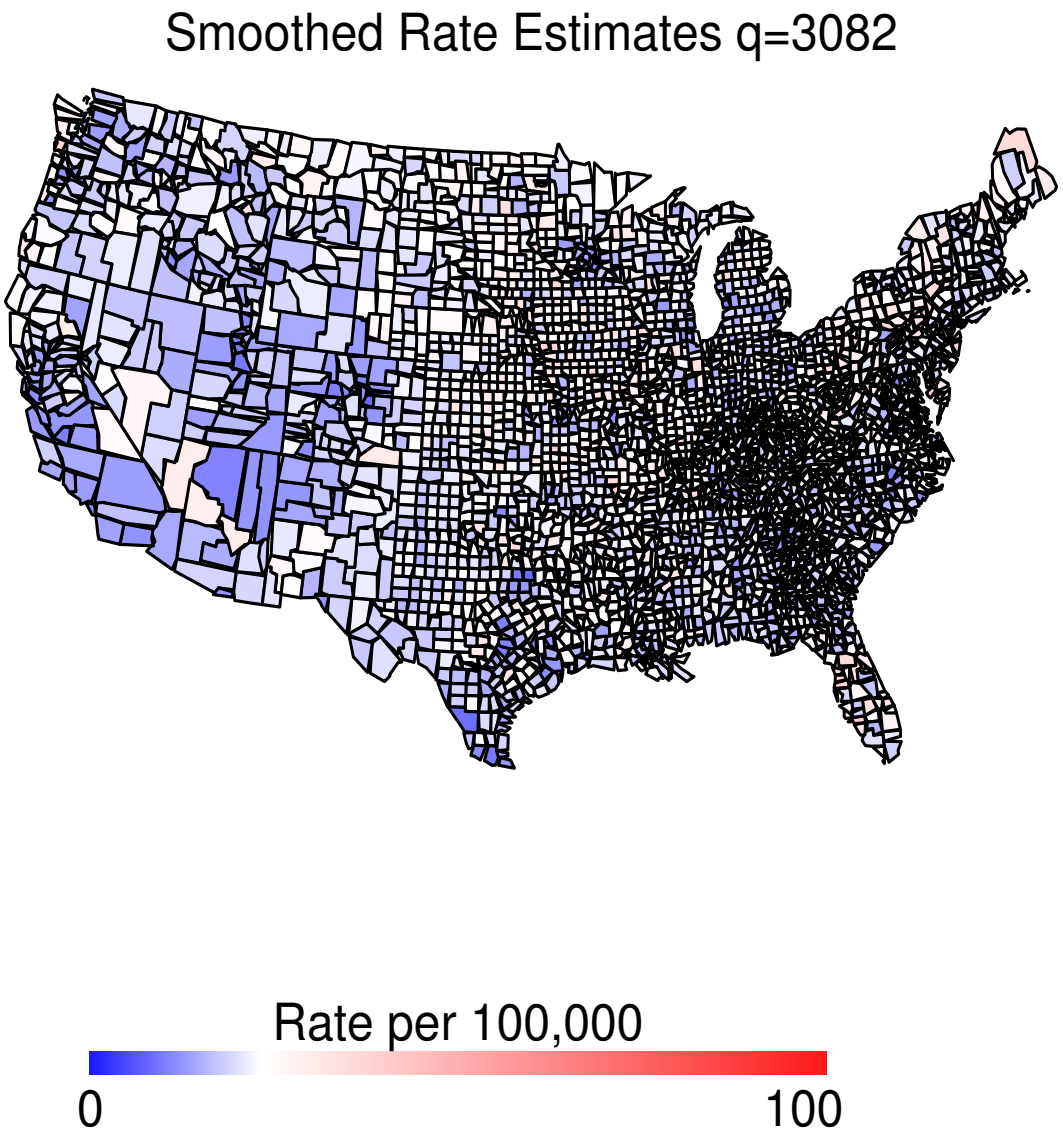

Figure 6.4: Map of Bayesian Estimates for Male Mortality Rates due to Colorectal Cancer p_i per 100,000 from $1999 - 2003$ from the Full Model.

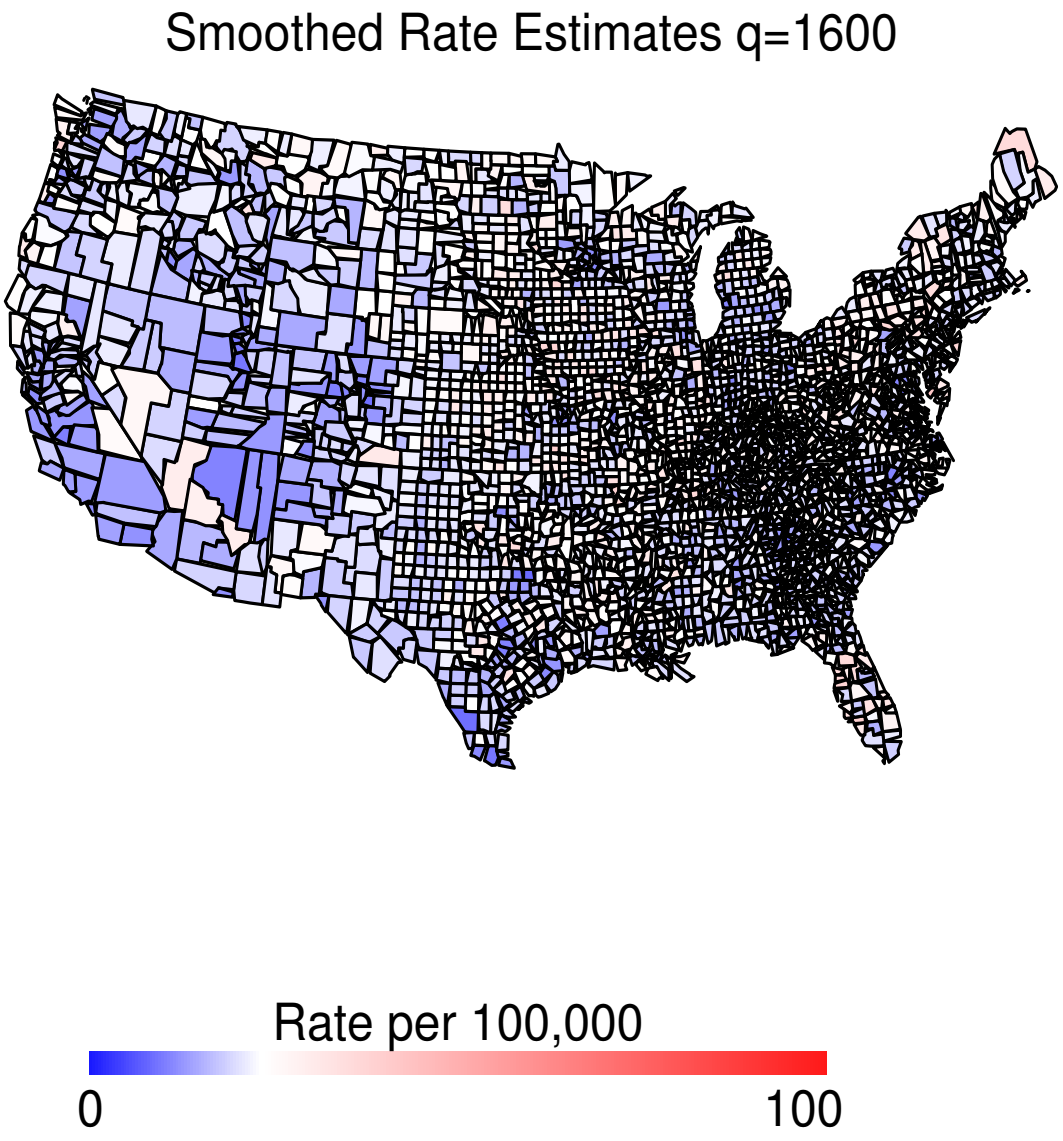

Figure 6.5: Map of Bayesian Estimates for Male Mortality Rates due to Colorectal Cancer p_i per 100,000 from $1999 - 2003$ from Reduced Model $q = 1600$.

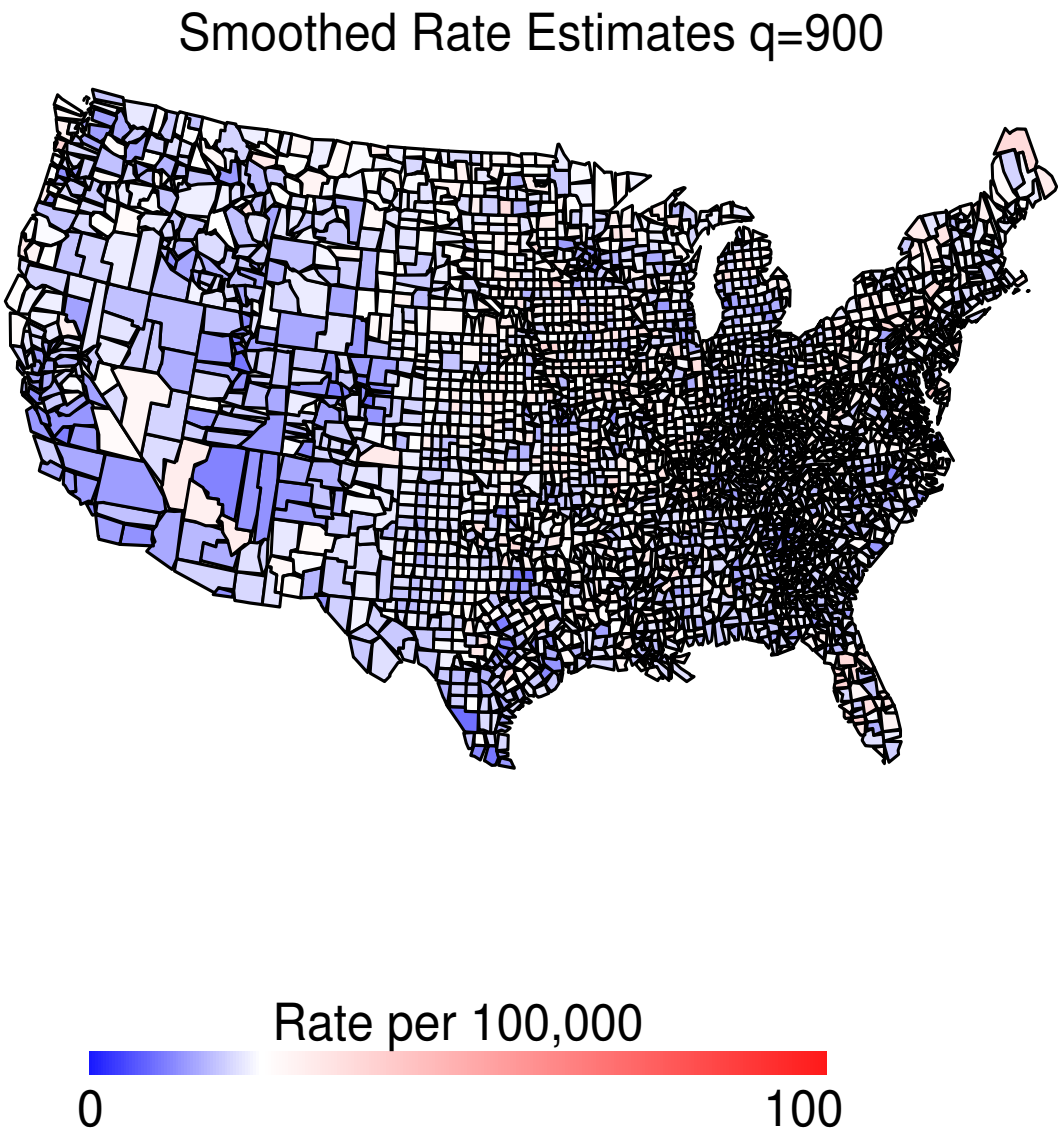

Figure 6.6: Map of Bayesian Estimates for Male Mortality Rates due to Colorectal Cancer p_i per 100,000 from $1999 - 2003$ from Reduced Model $q = 900$.

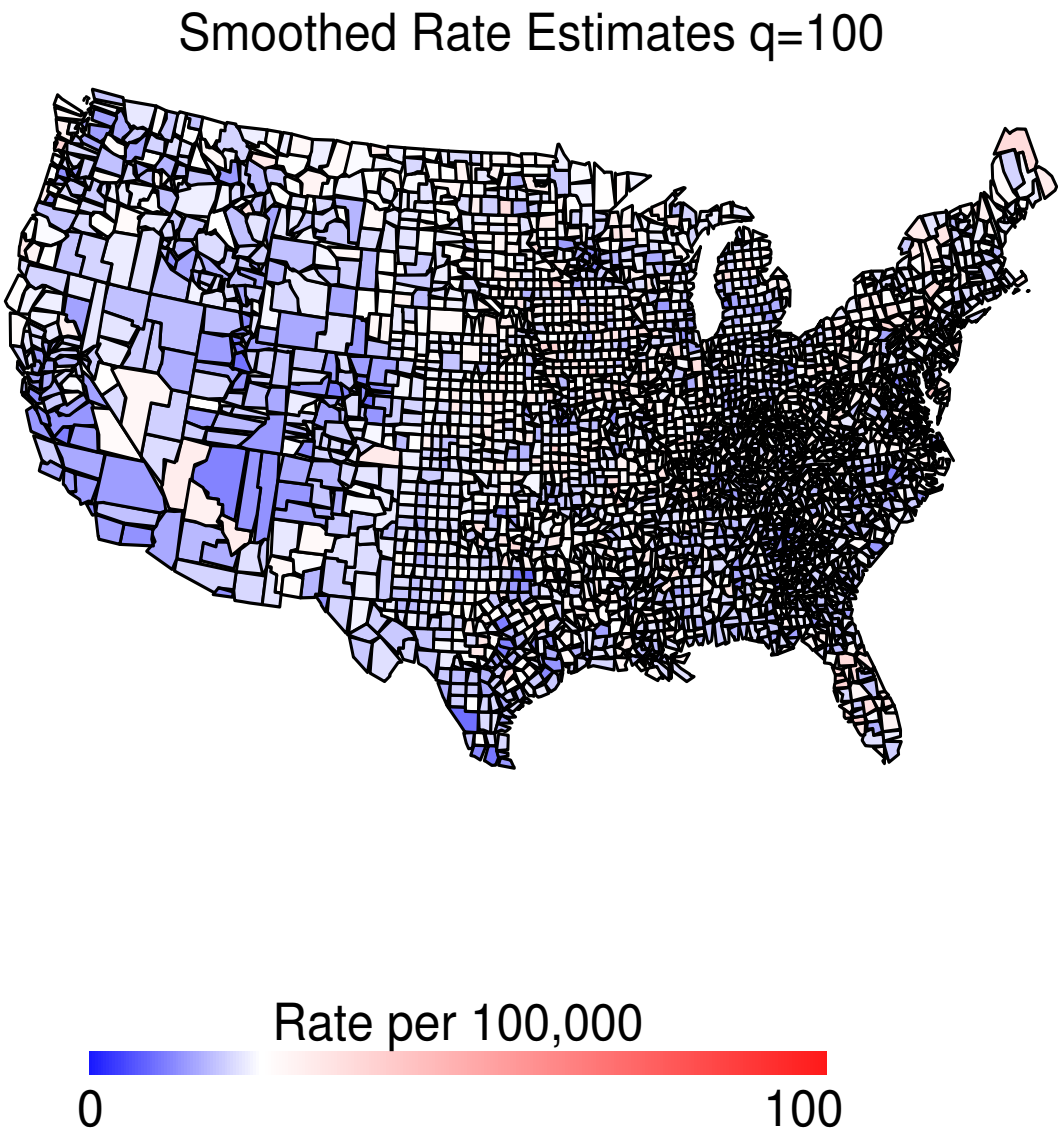

Figure 6.7: Map of Bayesian Estimates for Male Mortality Rates due to Colorectal Cancer p_i per 100,000 from $1999 - 2003$ from Reduced Model $q = 100$.

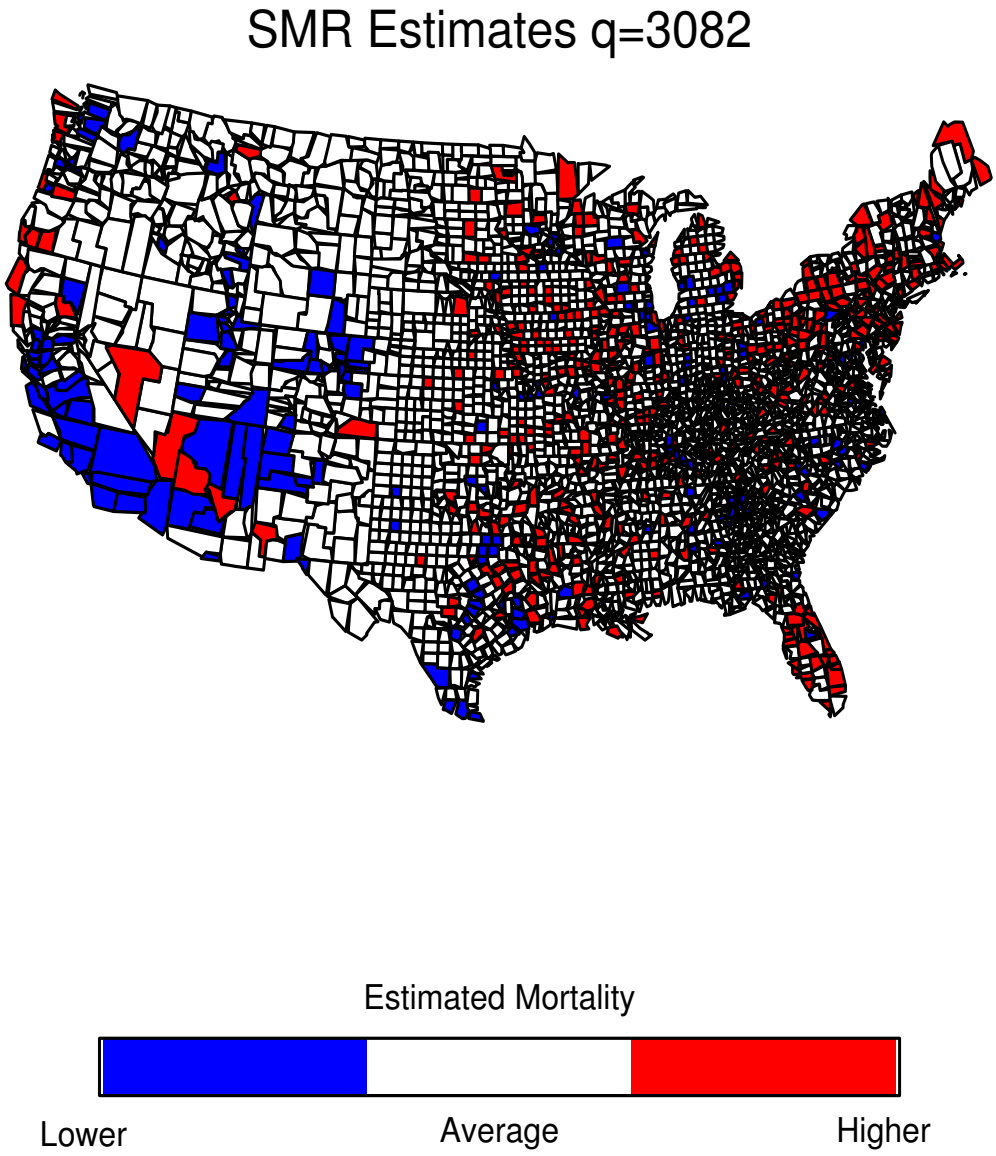

Figure 6.8: Significant Standardized Mortality Ratios for Male Mortality Due to Colorectal Cancer during $1999 - 2003$ from Full Model.

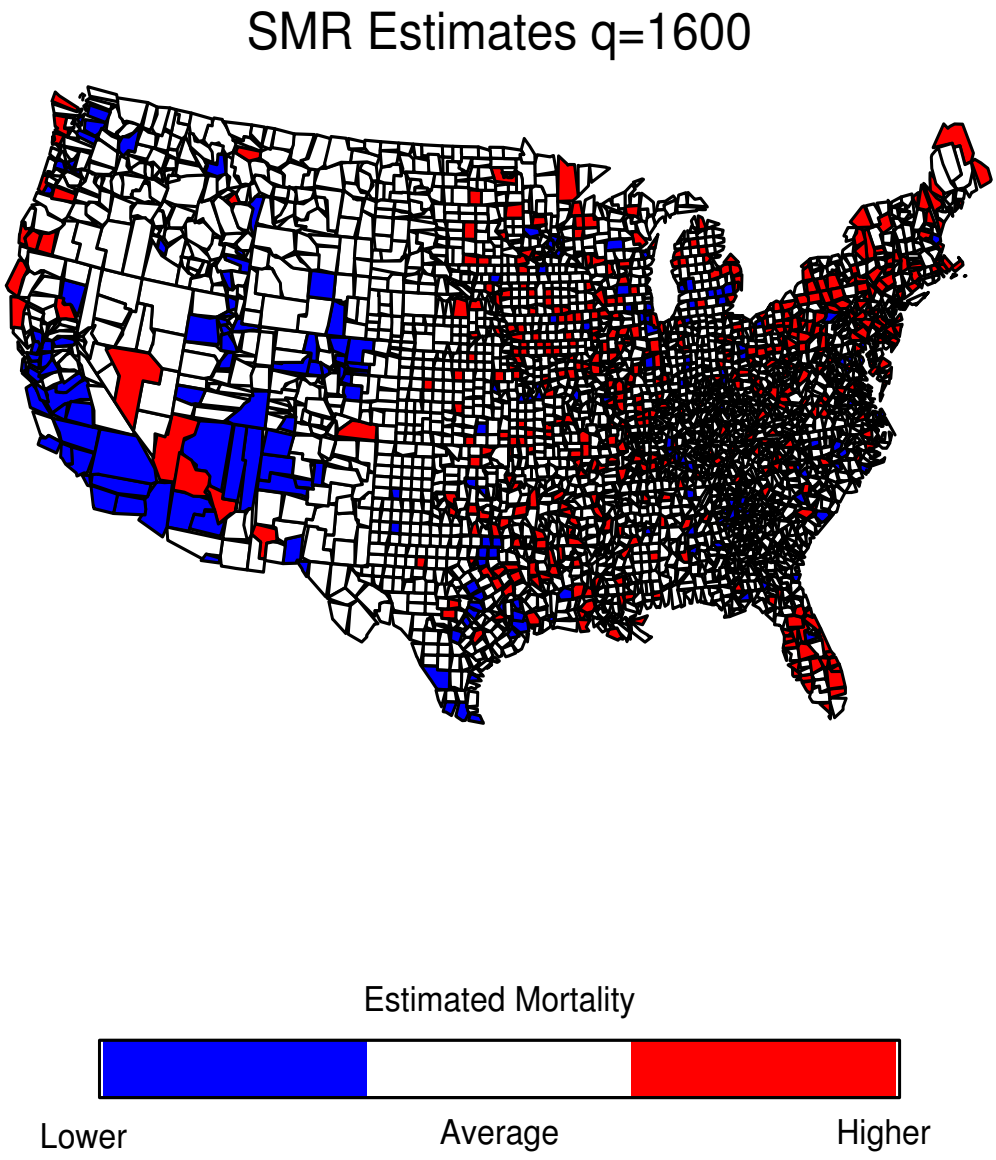

Figure 6.9: Significant Standardized Mortality Ratios for Male Mortality Due to Colorectal Cancer during $1999 - 2003$ from the Reduced Model, $q = 1600$.

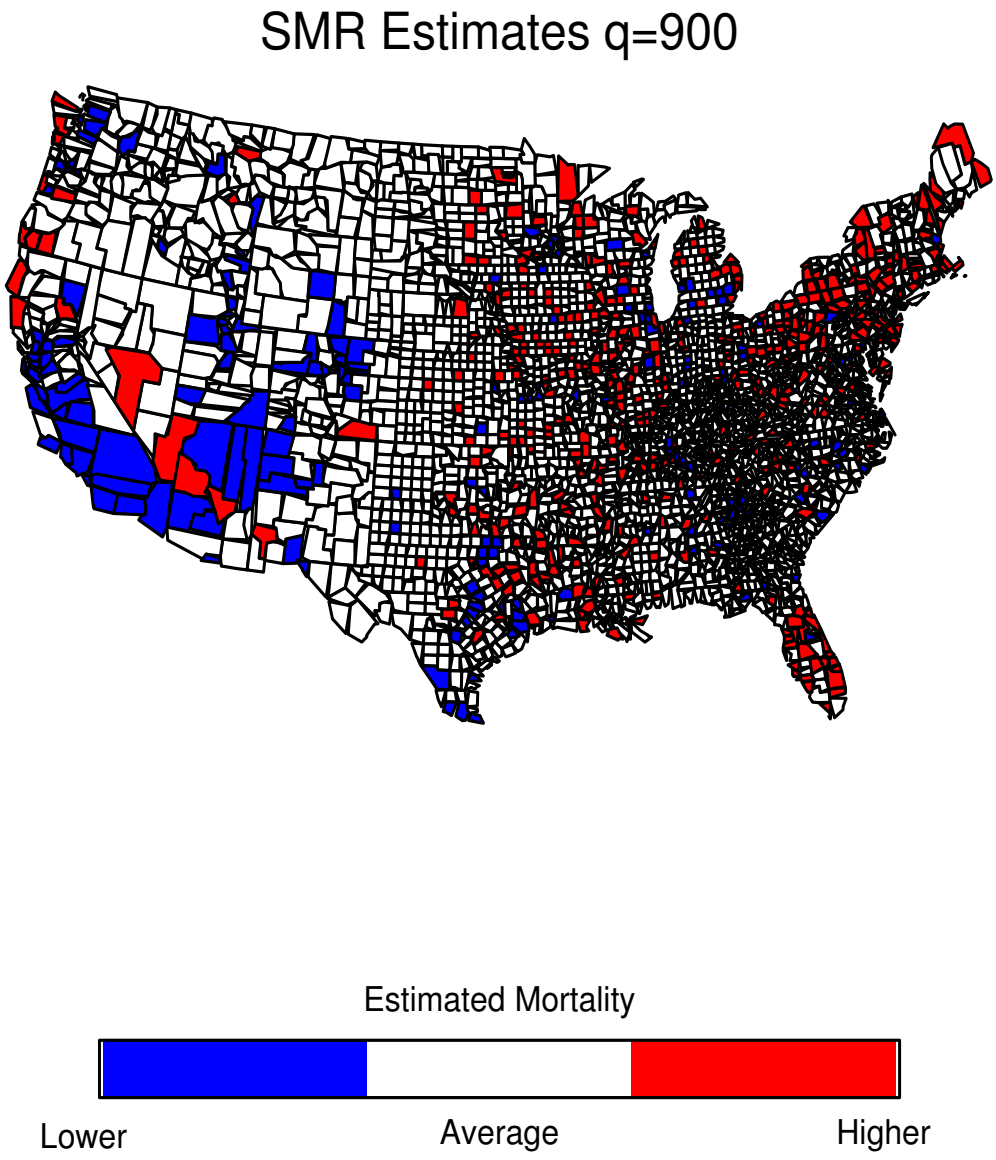

Figure 6.10: Significant Standardized Mortality Ratios for Male Mortality Due to Colorectal Cancer during $1999 - 2003$ from the Reduced Model, $q = 900$.

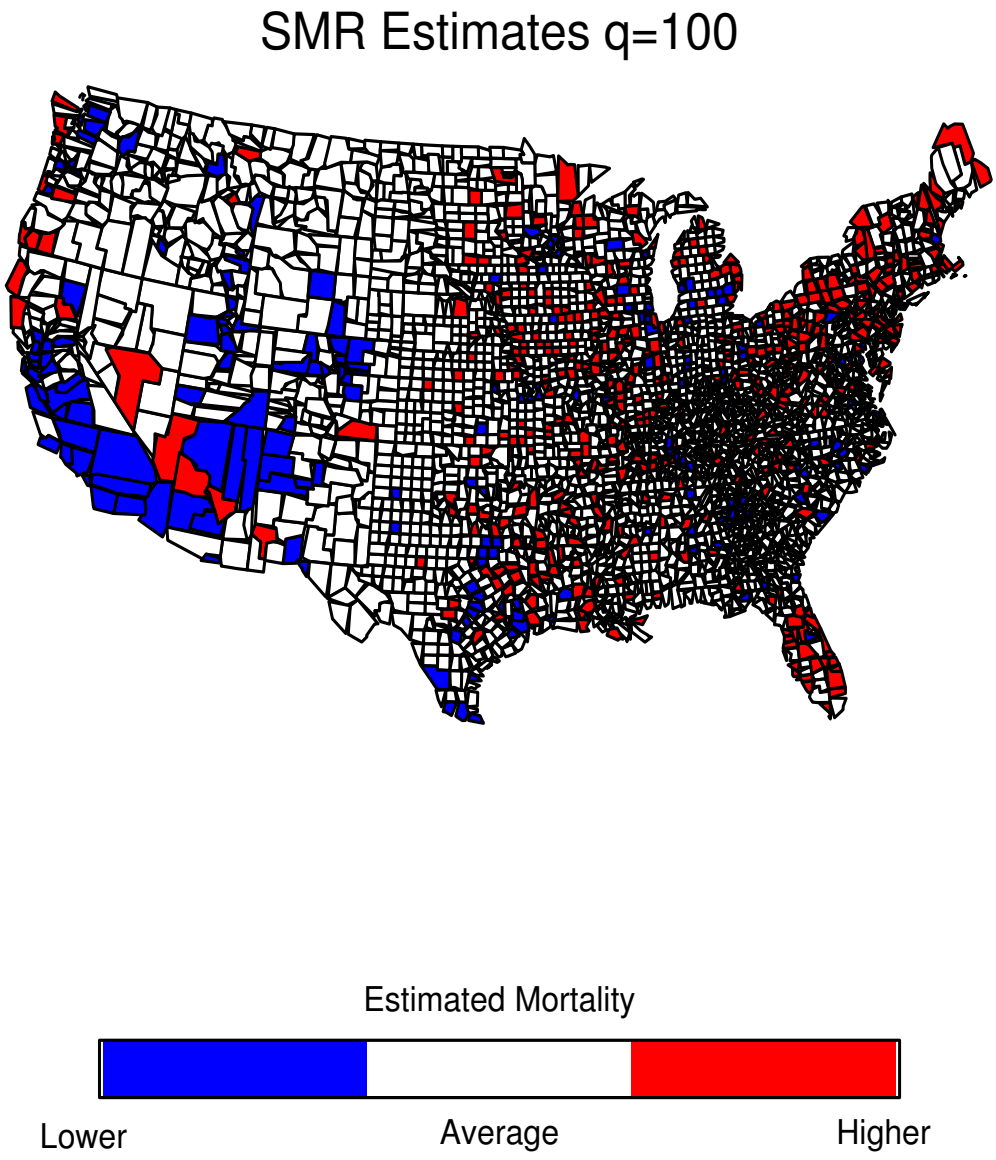

Figure 6.11: Significant Standardized Mortality Ratios for Male Mortality Due to Colorectal Cancer during $1999 - 2003$ from the Reduced Model, $q = 100$.

Bibliography

(2006*a*), website. http://seer.cacner.gov.

(2006*b*), website. http://www.cdc.gov/cancer/npcr.

(2006*c*), website. http://www.cdc.gov/cancer/npcr/uscs.

(2006*d*), website. http://www.naaccr.org.

(2006*e*), website.

(2006*f*), website.

(2006*g*), website.

(2006*h*), website.

(2006*i*), website.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M. & Songini, M. (1995), 'Bayesian Analysis of Space-Time Variation in Disease Risk', *Statistics in Medicine* **14**, 2433–2443.

Besag, J. (1974), 'Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussions)', *Journal of the Royal Statistical Society, Series B* **36**, 192–236.

Besag, J. & Kooperberg, C. (1995), 'On conditional and intrinsic autoregressions', *Biometrika* **82**, 733–746.

Besag, J., York, J. & Mollié, A. (1991), 'Bayesian image restoration with two applications in spatial statistcs', *Annals of the Institue of Statistical Mathematics* **43**, 1–59.

Clayton, D. & Bernardinelli, L. (1992), Bayesian Methods for Mapping Disease Risk, *in* P. Elliot, D. Cuzick, D. English & R. Stern, eds, 'Geographical and Environmental Epidemiology: Methods for Small-Area Studies', Oxford University Press, Oxford, U.K.

Clayton, D. & Kaldor, J. (1987), 'Empirical Bayes estimates of age-standardised relative risks for use in disease mapping', *Biometrics* **43**, 671–681.

Cressie, N. & Chan, N. H. (1989), 'Spatial modeling of regional variables', *Journal of the American Statistical Association* **84**, 393–401.

Devesa, S., Grauman, D., Blot, W. & Fraumeni, J. F. J. (1999), 'Cancer [s]urveillance [s]eries: [c]hanging [g]eographic [p]atterns of [l]ung [c]ancer [m]ortality in the [u]nited [s]tates, 1950 [t]hrough 1994', *Journal of the National Cancer Institute* **91**, 1040–1050.

Devine, O. J., Halloran, M. E. & Louis, T. A. (1994), 'Empirical Bayes Methods for Stabilising Incidence Rates Prior to Mapping', *Epidemiology* **5**, 622–630.

Devine, O. & Louis, T. A. (1994), 'A Constrained Empirical Bayes Esitmator for Incidence Rates in Areas with Small Populations', *Statistics in Medicine* **13**, 1119–1133.

Devine, O., Louis, T. A. & Halloran, M. E. (1994), 'Empirical Bayes Estimators for Spatially Correlated Incidence rates', *Environmetrics* **5**, 381–398.

Duchon, J. (1977), *Splines minimizing rotation-invariant semi-norms in Sobolev spaces*, Springer-Verlag, Berlin, pp. 85–100.

Ferrándiz, J., López, A., Llopis, A., Morales, M. & Tererizo, M. l. (1995), 'Spatial interaction between neighboring counties: Cancer mortality data in Valencia (Spain)', *Biometrics* **51**, 665–678.

Fharmeir, L. & Wagenpfeil, S. (1996), 'Smoothing hazard functions and time-varying effects in discrete duration and competing risk models', *Jouranl of the American Statistical Association* **91**, 1584=1594.

Gelfand, A. E. & Ghosh, S. (1998), 'Model choice: A minimum posterior predictive loss approach', *Biometrika* **85**, 1–11.

Gelfand, A. E. & Smith, A. F. M. (1990), 'Sampling-Based Approaches to Calculating Marginal Densities', *Journal of the American Statistical Association* **85**, 398–409.

Gilks, W. & Wild, P. (1992), 'Adaptive rejection sampling for Gibbs sampling', *Applied Statistics* **41**, 337–348.

Graves, T. L., Speckman, P. L. & Sun, D. (n.d.), Improving [m]ixing in [mcmc] [a]lgorithms for [l]inear [m]odels. submitted.

Green, P. & Silverman, B. W. (1994), *Nonparametric Regression and Generalized Linear Models*, Champman Hall, London, U.K.

Hastie, T. & Tibrishani, R. (1990), *Generalized [A]dditive [M]odels*, Chapman Hall, London, New York.

He, C. & Sun, D. (1998), 'Hierarchical [b]ayes estimation of hunting success rates', *Environmental and Ecological Statistics* **5**, 223=236.

He, Z. & Sun, D. (2000), 'Hierarchical Bayesian estimation of hunting success rates with spatial correlations', *Biometrics* **56**, 360–367.

Hutchinson, M. F. & Gessler, F. R. (1994), 'Splines - More Than Just a Smooth Interpolator', *Geoderma* **62**, 45–67.

Jackson-Thompson, J., Ahmed, F., German, R. R., Lai, S. & Friedman, C. (2006), 'Descriptive epidemiologu of colorectal cancer in the united states, 1998-2001', *Cancer* .

Kimmeldorf, G. & Wahba, G. (1970), 'A Correspondance Between Bayesian Estimation of Stochastic Processes and Smoothing by Splines', *Annals of Mathematical Statistics* **41**, 495–502.

Kimmeldorf, G. & Wahba, G. (1971), 'Some Results on Tchebychffan Spline Functions', *Journal of Mathematical Analysis Applications* **33**, 82–85.

Künsch, H. R. (1987), 'Intrinsic autoregressions and related models on the two-dimensional lattice', *Biometrika* **74**, 517–524.

Laslett, G. M. (1994), 'Kriging and Splines: An Empirical Comparison of Their Predictive Performance', *Journal of the American Statistical Association* **89**, 391–400.

Laslett, G. M. & McBratney, A. (1990), 'Further Comparison of Spatial Methods for Predicting Soil pH', *Journal of the Soil Science Society of America* **54**, 1553–1558.

Liang, F., Paulo, R., Molina, G., Clyde, M. & Berger, J. (2005), 'Mixtures of g-priors for Bayesian Variable Selection', *ISDS Technical Paper* **12**.

Liu, J. & Sabatti, C. (2000), 'Generalised [g]ibbs sampler and multigrid [m]onte [c]arlo for [b]ayesian computation', *Biometrika* **87,**, 353–369.

Manton, K. G., Woodbury, M. A., Stallard, E., Riggan, W. B., Creason, J. P. & Pellom, A. C. (1989), 'Empirical Bayes Procedures for Stabilising Maps of U.S. Cancer Mortality Rates', *Journal of the American Statistical Association* **84**, 637–650.

Manton, K., Woodbury, M., Tallard, E., Riggan, W. B., Creason, J. P. & Pellom, A. (1989), 'Empirical [b]ayes [p]rocedures for [s]tabilizing [m]aps of [u].[s]. [c]ancer [m]ortality [r]ates', *Journal of the American Statistical Association* **84**.

Marshall, R. J. (1991), 'Mapping disease and mortality rates using empirical Bayes estimators', *Applied Statistics* **40**, 283–294.

Matheron, G. (1973), 'The intrinsic random functions and their applications', *Advances in Applied Probability* **5**, 439–468.

Meinguet, J. (1979), 'Multivariate interpolation of arbitrary points made simple', *Journal of Applied Mathematical Physics* **30**, 292–304.

Meng, X.-L. & Wong, W. H. (1996), 'Simulating ratios of normalizing constants via a simple identity: A theoretical exploration', *Statistica Sinica* **6**, 831–860.

Mungiole, M. & Pickle, L. W.and Simonson, K. H. (1999), 'Application of a weighted head-banging algorithm to mortality data maps', *Statistics in Medicine* **18**, 3201–3209.

Nychka, D., Bailey, B., Ellner, S., Haaland, P. & O'Connel, M. (1996), 'Funfits [d]ata [a]nalysis and [s]tatistical [t]ools for [e]stimating [f]unctions', *North Carolina Institue of Statistics Mimeoseries* **2289**.

Nychka, D. W. (2000), *Smoothing and Regression: Approaches, Computation, and Application*, Chichester, New York, NY, chapter Spatial Process Estimators as Smoothers, pp. 393–424.

Picle, L., Feuer, E. & Edwards, B. (2003), U.[s]. [p]redicted [c]ancer [i]ncidence, 1999: [c]omplete [m]aps by [c]ounty and [s]tate [f]rom [s]patial [p]rojection [m]odels, monograph 5, National Cancer Institute, Bethesda, MD. NIH Publication No. 03-5435.

Sheriff, S., Sun, D., He, C. Z. & Vangilder, L. D. (n.d.), Estimating hunter success rates using small area estimation methods. Under revision for Journal of Wildlife Managment.

Shumway, R. H. & Stoffer, D. S. (2000), *Time Series Analysis and Its Applications*, Springer-Verlag, New York, NY.

Speckman, P. & Sun, D. (2003), 'Fully Bayesian spline smoothing and intrinsic autoregressive priors', *Biometrika* **90**, 289–302.

Spiegelhalter, D., Best, N., Carlin, B. & van der Linde, A. (2002), 'Bayesian measures of model complexity and fit', *Journal of the Royal Statistical Society, Series B* **64**, 583–616.

Sun, D., Tsutsakawa, R., Kim, H. & He, Z. (2000), 'Spatio-temporal interaction with disease mapping', *Statistics in Medicine* **19**, 2015–2035.

Van Der Linde, A. (2003), 'Pca-[b]ased [d]imension [r]eduction [f]or [s]plines', *Journal of Nonparmetric Statistics* **15**, 77–92.

van der Linde, A., Witzko, K. H. & Jöckel, K. H. (1995), 'Spatial-Temporal Analysis of Mortality Using Splines', *Biometrics* **51**, 1352–1360.

Wahba, G. (1978), 'Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression', *Journal of the Royal Statistical Society, Series B* **40**, 364–372.

Wahba, G. (1983), 'Bayesian "Confidence" Intervals for the Cross-Validated Smoothing Spline', *Journal of the Royal Statistical Society, Series B* **45**, 133–150.

Wahba, G. (1990), Spline Models for Observational Data, *in* 'Volume 59 of CBMC-NSF Regional Conference Series in Applied Mathematics', SIAM, Philadelphia.

Wahba, G., Wang, Y., Gu, C., Klein, R. & Klein, B. E. (1995), 'Smoothing Spline ANOVA fpr Exponential Families, with Applications to the Wisconsin Epidemiological Study for Diabetic Retinopathy', *Annals of Statistics* **23**, 1865–1895.

Waller, L. A., Carlin, B. P., Xia, H. & Gelfand, A. E. (1997), 'Hierarchical Spatio-Temporal Mapping of Disease Rates', *Journal of the American Statistical Association* **92**, 607–617.

White, G. & Sun, D. (2006), 'Simultaneous estimation of hunting pressure, harvest and hunter success rates using winbugs', *Far East Journal of Theoretical Statistics* **18**, 91–116.

Woodard, R., He, Z. & Sun, D. (2003), 'Bayesian Estimation of Hunting Success Rate

and Harvest for Spatially Correlated Post-stratified Data', *Biometrical Journal* **45**, 985–1005.

Woodard, R., Sun, D., He, C. & Sheriff, S. L. (1999), 'Estimating hunting success ratesvia [b]ayesian generalized linear models', *Journal of Agricultural, Biological and Environmental Statistics* **4**, 456–472.

Xia, H. & Carlin, B. P. (1998), 'Spatio-temporal models with errors in covariates: mapping Ohio lung cancer mortality', *Statistics in Medicine* **17**, 382–397.

Zellner, A. (1986), On assesing prior distributions and [b]ayesian regression analysis with g-prior distributions, *in* 'Bayesian [I]nference and [D]ecision [T[echniques: [E]ssays in [H]onor of [B]runo de [F]inetti', North-Holland Elsevier, pp. 223–243.

Zellner, A. & Siow, A. (1980), Posterior odds ratiosfor selected regression hypothesis, *in* '[B]ayesian [S]tatistics: [P]roceedimgs of the [F]irst [I]nternational [M]eeting held in [V]alencia ([S]pain)', pp. 585–603.

Zhang, S., Sun, D., He, Z. & Schootman, M. (2006), 'A Bayesian semiparametric model for colorectal cancer incidence', *Statistics in Medicine* **136**, 2873–2897.

VITA

Gentry White was born June 14, 1972 in Hinsdale, Illinois. He graduated with a A.B. in History from the University of Missouri-Columbia in 1994. He also received a B.S. in Mechanical and Aerospace Engineering and a M.A. in Statistics in 2001 and 2003 respectively, both from the University of Missouri-Columbia. Before he began his graduate studies he was employed as a Quality Engineer at Square D Company in Columbia, Missouri. He has accepted a Post-Doctoral VIGRE Position in the Department of Statistics at North Carolina State University.