2015

Model Selection for Hierarchical Poisson Modeling in Disease Mapping

Rachel Moss Carroll
Medical University of South Carolina

Model Selection for Hierarchical Poisson Modeling in Disease Mapping

Rachel Moss Carroll

A dissertation submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy.

Department of Public Health Sciences

College of Graduate Studies

Medical University of South Carolina

December 2015

APPROVED BY:	
COMMITTEE CHAIR	Andrew Lawson
SIGNED	_____
MEMBER	Christel Faes
SIGNED	_____
MEMBER	Russell Kirby
SIGNED	_____
MEMBER	Elizabeth Garrett-Mayer
SIGNED	_____
MEMBER	Bethany Wolf
SIGNED	_____

Acknowledgements

This dissertation would not have been possible without the help and support from a number of teachers, mentors, family, and friends.

I would like to thank my advisor, Dr. Andrew Lawson, for his instruction, encouragement, and dedication to this project. I am also indebted to my committee, Drs. Christel Faes, Russell Kirby, Beth Wolf, and Elizabeth Garrett-Mayer, whose passions for biomedical research have inspired me to take my own passion seriously. Further, I extend thanks to the many teachers, professors, and collaborators from Florence School District One, Charleston Southern University, Medical University of South Carolina, and beyond who aided in my development as a student, researcher, and biostatistician. Additionally, June Watson has been a saving grace throughout my time at MUSC.

I dedicate this dissertation to my family. I am immensely grateful to my husband, Jon. Without his love and support, I most certainly would not have made it this far. My mom (Vicki Moss), dad (John Moss), and sisters (Chrissy Welch and Carly Moss) have been a constant support system since the day I came screaming into this world. Without their patience and guidance, I know that this achievement would not have been possible. Even my in-laws and extended family, who are too numerous to name, were incredibly supportive through this entire process.

I would also like to thank my friends, classmates, fellow cyclists, and the choir members from St. Philip's in Charleston as well as Christ and St. Luke's in Norfolk. Outings, messages, and phone calls do a great deal towards keeping a girl sane. My specific thanks to Joye, Lauren, Alex, Chawarat, Harriet, Marilyn, and Jess. And finally, though they cannot possibly know how much of a help they have been, I would like to thank Galileo and Kayla, the ultimate companion animals.

This research was supported in part by funding under grant NIH R01CA172805.

Contents

List of Tables

List of Figures

Related Manuscripts

Carroll R, Lawson AB, Faes C, Kirby RS, Aregay M, Watjou K. Spatio-temporal Bayesian model selection for disease mapping. Stat Methods Med Res. 2016; Submitted.

Carroll R, Lawson AB, Faes C, Kirby RS, Aregay M, Watjou K. Spatially-dependent Bayesian model selection for disease mapping. Stat Methods Med Res. 2016; Submitted.

Carroll R, Lawson AB, Faes C, Kirby RS, Aregay M, Watjou K. Bayesian model selection methods in modeling small area colon cancer incidence. Ann Epidemiol. 2015;In print. doi: 10.1016/j.annepidem.2015.10.011.

Carroll R, Lawson AB, Faes C, Kirby RS, Aregay M, Watjou K. Comparing INLA and OpenBUGS for hierarchical Poisson modeling in disease mapping. Spat Spatiotemporal Epidemiol. 2015;14-15:45-54. doi: 10.1016/j.sste.2015.08.001.

Abstract

In disease mapping where predictor effects are to be modeled, it is often the case that sets of predictors are fixed and the aim is to choose between fixed model sets. In this dissertation, I focus on this dimension reduction objective by applying the Poisson data model commonly used for disease mapping of small area health data. I begin with a software comparison of the recently developed `R` package `INLA` (Integrated Nested Laplace Approximation) to the MCMC approach by way of the `BRugs` package in `R`, which calls `OpenBUGS`. This software comparison leads to choosing the appropriate platform for carrying out the second portion of this work: a methodology comparison of my proposed non-spatial and spatial approaches of Bayesian model selection to Bayesian Model Averaging. Following that, for the third and final aim, I extend my Bayesian model selection methodologies to the spatio-temporal setting and evaluate the benefit and usefulness of four different modeling approaches. These explorations demonstrate the importance of altering the defaults in `INLA` and the flexibility of the BUGS software. Additionally, they offer a novel way of determining appropriate linear predictors in the context of non-spatial, spatial, and spatio-temporal small area health data in disease mapping.

1. Introduction

Bayesian methods in Biostatistics have become more conventional in recent years due to advances in computing technology and software availability; this, in conjunction with the development and adoption of the Markov chain Monte Carlo (MCMC) method, has truly broadened the uses of this paradigm. The importance of MCMC methods were first noted in review articles for medical applications as early as 1993.[1] The software package Bayesian inference Using Gibbs Sampling (BUGS), explicitly `OpenBUGS` and `WinBUGS`, implements these methods in a way that has led to considerable wide spread use.[2, 3]

The topic of disease mapping exists within the Bayesian paradigm. This subject capitalizes on the importance of geographical location in predicting disease outcomes. While this is not a new notion, technological improvements, such as geographical information system (GIS), coupled with the development of the conditional autoregressive (CAR)[4] model have greatly improved the capabilities of this area of research.

In the application of disease mapping, it is often of interest to determine the best set of predictors associated with a particular disease outcome. Here, I propose a method for doing so using model selection. These model selection procedures can be applied in non-spatial, spatial, or spatio-temporal framework to allow for maximum flexibility, and is characterized by the notion that there is a predefined set of potentially important linear predictors. From the model selection process, I calculate model weights or probabilities that aid in determining the most appropriate linear predictor for the given area whether it be a non-spatial, spatial, or spatio-temporal unit. The general aims are as follows:

1. To compare `INLA`[5] and `OpenBUGS` in the disease mapping framework.
2. To develop non-spatial and spatial model selection methods.
3. To extend in the model selection methods into the spatio-temporal setting.

Aim 1 considers two prominent software packages available for disease mapping in Bayesian inference: `INLA` and `OpenBUGS`. In this examination, a range typical of model scenarios are examined for their fit in

each software package. Initially the priors are based on the default settings available in `INLA`, then altered priors are fit to determine the scenario that achieves nearly identical results in the two packages. Ultimately, I determine that `INLA` can reproduce results found in `OpenBUGS` when the default settings are altered, but I elected to continue using the BUGS software in this project as it offers more flexible modeling options for exploring the model selection techniques.

Aim 2 focuses on developing both the non-spatial and spatial model selection methodologies. These methods provide a way to establish the best of a pre-defined set of linear predictors, and when looking at the spatial method, separate linear predictors could be chosen for different counties of interest. In addition to developing the model selection methods, Bayesian Model Averaging (BMA) is also performed on the linear predictors of interest. This allows for a comparison between my Bayesian model selection (BMS) and the BMA methods. Both BMS and BMA procedures are performed in `OpenBUGS` as a result of the findings in Aim 1. Additionally, a real data example using colon cancer incidence in the state of Georgia, USA is performed with these techniques. Based on this exploration of model selection techniques, I determine that BMS is the better option over BMA, but using them in conjunction can lead to even more information about the linear predictors of interest.

Aim 3 extends Aim 2 into space-time to allow for a spatio-temporal model selection approach to BMS within the disease mapping frame work. In this extension, demographic data from the state of Georgia, USA is used to create the simulated outcomes that vary by the predictors included, the fixed parameter estimates, and the random effects involved. Four different spatio-temporal fitted model scenarios are explored and examined for how they apply to a range of different simulated model scenarios. In the end, I believe that a special case of the model selection methods, a mixture model, is the paramount choice for the most simulated data scenarios.

2. Aim 1: Comparing INLA and OpenBUGS

2.1. Introduction

In Bayesian modeling there are many challenges in conventional use of posterior sampling via MCMC for inference.[6] One challenge is the need to evaluate convergence of posterior samples, which often requires extensive simulation and can be very time consuming. Software for implementing MCMC is now widely used and the packages WinBUGS, OpenBUGS, as well as certain SAS procedures and select R[7] packages such as MCMCpack,[8] can provide access to these methods.

An Integrated Nested Laplace Approximation[5, 9-13] has been implemented as an R package (INLA) targeted at performing Bayesian analyses without having to use posterior sampling methods. Unlike MCMC algorithms, which rely on Monte Carlo integration, the INLA package performs Bayesian analyses via numerical integration, and so does not require extensive iterative computation. In most cases, Bayesian estimation using the INLA methodology takes much less time as compared to estimation using MCMC. However, there have been few, if any, attempts at comparison of these packages' performance capabilities with respect to spatial models in a disease mapping context. In this study, I compare how INLA performs in different modeling situations to OpenBUGS[3] via the BRugs package in R.[13]

Here, I am particularly interested in comparing these packages for conventional Poisson data models when spatial structures are present in the covariates as well as through uncorrelated and correlated spatial random effects. I have designed models that express these attributes in different ways and apply them to mimic their use for disease mapping. These models are all commonly implemented for Bayesian analysis. I am also interested in exploring the different options available in INLA to optimize estimation as well as goodness of fit (GoF) for these particular models to determine if INLA can perform in an equivalent fashion to OpenBUGS. For the analyses, I used the following versions of software and packages: R version 3.0.3, OpenBUGS version 3.2.3 rev 1012 (using default sampler settings), INLA version 0.0.1403203700,

and BRugs version 0.8.3. I am aware that there are now more recent versions of these programs, but, as of August 2014, I was able to reproduce these results in the latest versions as well.

This chapter is developed as follows. First, I discuss the development of my simulated data set and the different models used for comparison. Next, I describe the methods used for comparing the performance of R-INLA and OpenBUGS. Following that, I display results, and finally, I discuss the benefits of one package versus the other under different scenarios.

2.2. Methods

2.2.1. Simulated Data and Models

In this study, focus is on the use of INLA versus MCMC in a disease mapping context. Performance of the two methods is compared for the commonly used convolution model with spatially-varying and non-spatially-varying predictor variables. To evaluate the performance of these alternative approaches, I simulate data to establish realistic ground truth for disease risk variation. A common model for small area counts of disease is the Poisson data model. Specifically, I define a count outcome as y_i in the i^{th} small area. I assume a map of m small areas. In addition, I assume an expected count (e_i) is available in each small area. Thus, my outcome has the distribution:

$$y_i \sim Pois(\mu_i)$$
$$\mu_i = e_i \theta_i$$

In my simulations, I fix the expected rate for the area, and hence focus on the estimation of relative risk. To complete the parameterization, I assume a relative risk (θ_i) which is parameterized with a range of different risk models.

To satisfy this I chose the county map of the state of Georgia USA, which has 159 areas (counties). Hence, $i = (1,...159)$ for this county set. I fix e_i at one for all models; while this is a simplifying assumption, it

allows me to reduce the amount of variability present in the simulations. Furthermore, the assumption represents data associated with a fairly sparse disease presence.

I examined six basic models for risk (S1 up to S6) which have different combinations of covariates and random effects as might be found in common applications. First, I generated four predictors with different spatial patterns. The four chosen were median age (x_1), median education (x_2), median income (x_3), and a binary predictor representing presence/absence of a major medical center in a county (x_4). These variables are county-level measures for the 159 counties in the state of Georgia. I chose these variables because it is important to represent a range of different spatial structures and types of predictors. Furthermore, observed predictors/covariates could have spatial correlation and so I included this in the definition of two of the predictors (median age and major medical center). Table 1 displays the predictors generated via simulation and their parameterization where the Gaussian parameters are the mean and variance. Also, note that the spatial predictors have a covariance structure applied, and this is explained in detail later.

Table 1: Description of predictor variables and their simulation.

Variable	Spatial	Distribution (marginal)
Median Age (years)	Yes	$x_1 \sim Norm(40,4)$
Median Education (years)	No	$x_2 \sim Norm(13,4)$
Median Income (thousands)	No	$x_3 \sim Norm(45,1)$
Major Medical Center (Yes/No)	Yes	$\pi \sim Norm(0,25)$ $\text{logit}(p) = \pi$ $x_4 \sim Bern(p)$

These distributions lead to measures that reflect typical values for these variables. For example, the median age for a county in the U.S. is roughly 40, and the values do not vary much from one county to the next. Similar explanations can be applied to both the median education and median income variables. For the major medical center variable, this indicated marginal distribution leads to roughly half of the counties answering ‘yes’ to having a major medical center. These variables have been selected as placeholders, and

should be thought of as representing any of the typical variables one might utilize in building disease mapping models.

The spatially-structured covariates were generated using the `RandomFields` package in `R`.[14] The simulation uses the county centroid as a location to create a Gaussian Random Field (GRF), which is defined via a covariance structure. I assume a Gaussian covariance structure, and this assumption leads to a stationary and isotropic process.[15] I specify this structure by using the `RMgauss()` command. I assume a power exponential covariance model of the form:

$$C(r) = e^{-\alpha r^2}$$

where r is the Euclidean distance between two centroids and the covariance parameter (α) is set as $\alpha = 1$. Following the selection of the covariance structure, I must also set the mean of the GRF to create the same marginal distributions as described in Table 1 using the `RMtrend()` command. Finally, to simulate the GRF, I use the `RFsimulate()` command to create a GRF and assign a value to the spatial covariate. There is only a slight extension that must be applied when the spatial covariate is binary such as x_4. To create this variable, I use a GRF to simulate π mentioned in Table 1, rather than the covariate itself. From there, I simulate from a Bernoulli distribution to give the binary indicators for each county using expit(π) for the probabilities.

The distribution of all covariates, $x_1, ..., x_4$, on the Georgia county base map is illustrated in Figure 1. Notice that the median age and major medical center covariates appropriately appear to have spatially dependent distributions. Though I have defined the mean and covariance structure of these GRFs, the distribution can still take on many forms, and these variables reflect only one realization of the distribution.

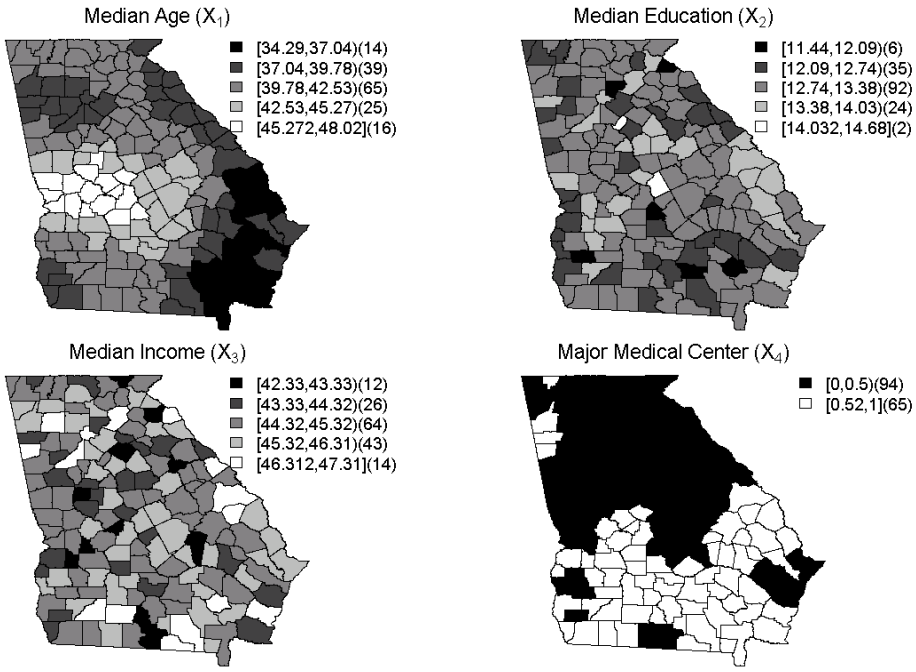


Figure 1: Display of the spatial distribution of simulated covariates per county.

Furthermore, picking these two variables (x_1 and x_4) to have the spatial structure also leads to at least one variable with spatial structure included in all of the first 5 models that I simulated which are shown in Table 2. Note that v_i follows an intrinsic CAR model with precision τ_v . M6 is simply a convolution model that allows me to explore how the spatial covariates may be affecting the spatial random effects, u and v . For these simulations, I fix the covariates as one realization from the distributions described in Table 1 above and generate the outcomes using a fixed set of parameters ($e_i = 1, \alpha_0 = 0.1,\ \alpha_1 = 0.1,\ \alpha_2 = -0.05,$ $\alpha_3 = -0.05,\ \alpha_4 = 0.5,\ \tau_u = 1$ and $\tau_v = 1$). While the magnitude of the α 's are quite small, this guarantees that the outcome variable has sparse disease patterning. I also use only one realization of the uncorrelated and correlated random effects described in Table 2 below. Then, $\log(\theta_i)$ is calculated based on the fixed parameters and realizations. Finally, I generate the outcome as a Poisson variate with mean θ_i since e_i is fixed at one. The simulated data sets consist of sets of counts: $\{y_{ij}^*\}, j = (1,\ldots,200)$ where j denotes the j^{th} simulated data set.

Table 2: Description of simulated model contents.

Model	Log relative risk
M1	$\log(\theta_i) = 1 + 0.1x_{1i} - 0.05x_{2i}$
M2	$\log(\theta_i) = 1 + 0.1x_{1i} - 0.05x_{2i} + u_i$ $u_i \sim Norm(0,1)$
M3	$\log(\theta_i) = 1 + 0.1x_{1i} - 0.05x_{2i} + u_i + v_i$ $u_i \sim Norm(0,1)$, $v_i \sim CAR(\tau_v), \tau_v = 1$
M4	$\log(\theta_i) = 1 + 0.1x_{1i} - 0.05x_{2i} - 0.05x_{3i} + 0.5x_{4i}$
M5	$\log(\theta_i) = 1 + 0.1x_{1i} - 0.05x_{2i} - 0.05x_{3i} + 0.5x_{4i} + u_i + v_i$ $u_i \sim Norm(0,1), v_i \sim CAR(\tau_v), \tau_v = 1$
M6	$\log(\theta_i) = 1 + u_i + v_i$ $u_i \sim Norm(0,1), v_i \sim CAR(\tau_v), \tau_v = 1$

For the uncorrelated and correlated spatial effects, u and v , mentioned in Table 2, I fix both precisions, τ_u and τ_v , to be one during the simulation process. Their equality guarantees that one of the spatial effects will not dominate the model and lead to identifiability issues.[16] The uncorrelated spatial effect is distributed $N(0,1)$, this is specified as such for simplicity as well as easy identification in the model fitting process. The correlated spatial random effect in these models is generated using the `R` package `BRugs`[13] such that they have an Improper CAR[4] structure as follows:

$$v_i \sim N\left(\frac{1}{n_i}\sum_{i\sim l} v_l, \frac{1}{n_i \tau_v}\right)$$

where $i \neq l$, n_i is the number of neighbors for county i , and $i \sim l$ indicates that the two counties i and l are neighbors.[4] This set of neighbors simply includes the immediate neighborhood. Including these types of effects in spatial disease mapping models is very common as there is typically an uncorrelated random noise that varies from county to county as well as a correlated random structure that induces correlation based on neighborhoods.

Additional Simulation Variants

In the analysis described above, I mean centered the predictors to help in model GoF. Another standardization technique for these types of analysis involves mean centering then dividing by the standard deviation per predictor. I created such a data set using the M5 model (full predictor set with convolution random effects) to assess the effect of this standardization.

In addition, I also examined the effect of varying the precision of random effects to assess performance of model fit. Now, the true correlated spatial effect has a precision of 0.5. The uncorrelated spatial effect still has a precision of one, but I simulate a new realization of the variable. Because the precisions are no longer equal, this could lead to the masking, or domination effect eluded to earlier.[16] Following the simulations of the new spatial effects, I created new Poisson outcomes with the same six models, aside from the spatial random effects, indicated in Table 2. These data sets are considered the variant data sets.

2.2.2. Fitted Models

The fitted models F1–F6 are described in Table 3. Note that v_i is an intrinsic CAR model with precision τ_v . These models are based on the default prior distributions for `INLA` and vary by the number of covariates considered as well as the spatial random effects included to give a wide range of models as reflected in the simulation data section above. As part of examining the ability of `INLA` and `OpenBUGS` to recover true risk, I considered a variety of prior specifications. I changed the Gamma prior distributions on the precisions, τ_u and τ_v , to the following: $Gam(2,1)$, $Gam(1,1)$, and $Gam(1,0.5)$. Of these options, $Gam(1,0.5)$ offers the best alternative to the default prior distribution, $Gam(1,5e-05)$, as it is the most non-informative of the prior distributions explored. Since $Gam(1,5e-05)$ is the default non-informative prior distribution for precisions in `INLA`, I assumed this as default for my comparable `OpenBUGS` models. I would as like to note that when the outcome is Gaussian distributed, the default setting for `INLA` are a reasonable choice, and this is likely why it is set as such.

Table 3: Fitted model description.

Model	Description
F1	$\log(\theta_i) = a_0 + a_1 x_{1i} + a_2 x_{2i}$ $a_j \sim Norm(0, \tau_\alpha)$ where τ_α is fixed
F2	$\log(\theta_i) = a_0 + a_1 x_{1i} + a_2 x_{2i} + u_i$ $a_j \sim Norm(0, \tau_\alpha)$, where τ_α is fixed $u_i \sim Norm(0, \tau_u), \tau_u \sim Gam(1, 5e-05)$
F3	$\log(\theta_i) = a_0 + a_1 x_{1i} + a_2 x_{2i} + u_i + v_i$ $a_j \sim Norm(0, \tau_\alpha)$, where τ_α is fixed $u_i \sim Norm(0, \tau_u), \tau_u \sim Gam(1, 5e-05)$ $v_i \sim CAR(\tau_v), \tau_v \sim Gam(1, 5e-05)$
F4	$\log(\theta_i) = a_0 + a_1 x_{1i} + a_2 x_{2i} + a_3 x_{3i} + a_4 x_{41}$ $a_j \sim Norm(0, \tau_\alpha)$, where τ_α is fixed
F5	$\log(\theta_i) = a_0 + a_1 x_{1i} + a_2 x_{2i} + a_3 x_{3i} + a_4 x_{41} + u_i + v_i$ $a_j \sim Norm(0, \tau_\alpha)$, where τ_α is fixed $u_i \sim Norm(0, \tau_u), \tau_u \sim Gam(1, 5e-05)$ $v_i \sim CAR(\tau_v), \tau_v \sim Gam(1, 5e-05)$
F6	$\log(\theta_i) = a_0 + u_i + v_i$ $a_j \sim Norm(0, \tau_\alpha)$, where τ_α is fixed $u_i \sim Norm(0, \tau_u), \tau_u \sim Gam(1, 5e-05)$ $v_i \sim CAR(\tau_v), \tau_v \sim Gam(1, 5e-05)$

To attempt recovering the ground truth indicated in the simulation data section, I build these models in both `INLA` and `WinBUGS` to assess the two packages' recovery abilities. These abilities are assessed as described in the comparison techniques section. Also, note that I mean center the fitted continuous predictors to aid in model fit.

In this study, I do not consider model misspecification; I simply use the appropriate fitted model applied to the corresponding simulated data. From here on, I will refer to these results with respect to the fitted model and simulated data that are being used. For example, M1 simulated data with F1 fitted model will be referred to as M1F1.

Fitting Models to Simulation Variants

I also fit the models to these variants using the $Gam(1, 0.5)$ prior distribution in a similar fashion to that described previously in this section to determine how this data set characteristic affect the model fits. These models will be referred to as, for example, M5F5S and M5F5V respectfully. S denotes standardization and V denotes spatial precision variant.

2.2.3. Comparison Techniques

For this study, I build equivalent models in `INLA` and `OpenBUGS` based on the prior distributions indicated in Table 2 and fit them to the 200 data sets all simulated in the same manner. For the `OpenBUGS` model, this is accomplished using the `BRugs` package in `R`.

Comparisons of the models on the two different packages is accomplished by calculating the mean squared error (MSE) of the parameter estimates, fitted y values, and the relative risk (θ) as well as mean squared predictive error (MSPE), the number of effective parameters (pD), mean deviance ($\overline{D(\theta)}$) and deviance information criterion (DIC) for each model.[17]

I compute *MSE* for the parameter estimates, outcome measures, and relative risk as well as *MSPE* as follows:

$$MSE(\alpha) = \sum_{j=1}^{J} \frac{\left(\alpha_j - \hat{\alpha}_j\right)^2}{J},$$

$$MSE(M) = \sum_{i=1}^{n} \frac{\left(y_i^* - \hat{y}_i\right)^2}{n},$$

$$MSE(\theta) = \sum_{i=1}^{n} \frac{\left(\theta_i^* - \theta_i\right)^2}{n}, \text{and}$$

$$MSPE = \sum_{i=1}^{n} \frac{\left(y_i^* - y_{i,pred}\right)^2}{n}$$

where α_j, θ_i^*, and y_i^* are the true fixed and simulated values respectively such that the * indicates simulated values. Furthermore, J is the number of covariates considered in each specific model, and n is the number of counties. $MSE(\alpha)$ uses the posterior values of the parameter estimates since sampling is not available in INLA. These measurements are all averaged over the 200 simulated data sets. To access $\hat{y}$ in OpenBUGS, I must collect the posterior values of μ_i from the samples; I use this same value for θ_i since they are equivalent because the expected rate, e_i, is set at one. To access $y_{i,pred}$ I simply set up code such that it is distributed Poisson with mean μ_i. Note that y_{pred} is initialized rather than data-read to generate predictive values. An example of my OpenBUGS model code is located in the Appendix section A.1.1. To calculate the $MSPE$ in INLA, I must create copies of all of the model components, append these copies to the original vectors, add in an additional "iid" random effect, and create an empty vector, rather than a copy, to append to the outcome variable. Now, the vector that typically only contains $\hat{y}_i$ will have length $2n$ where the second half is the predicted y values to use as $y_{i,pred}$ in the $MSPE$ calculation. An example using M3F3 can be found in Appendix section A.1.1.

Computations for the $D(\theta), pD$, and DIC are built into and easily attainable in both packages. They are calculated as follows:

$$
\begin{aligned}
D(\theta) &= -2\sum_i \log\left(p(y_i|\theta)\right), \\
pD_{OpenBUGS} &= \overline{D(\theta)} - D(\theta), \\
pD_{INLA} &= n - tr\left\{Q(\theta)Q^*(\theta)^{-1}\right\}, \text{and} \\
DIC &= \overline{D(\theta)} + pD
\end{aligned}
$$

respectfully. The first two formulas show the definitions of deviance in OpenBUGS and INLA respectively. To calculate $\overline{D(\theta)}$ from $D(\theta)$ produced in OpenBUGS, I simply average this value over a sample from the converged Markov chain.[18] Using $D(\theta)$ from INLA to calculate $\overline{D(\theta)}$ requires a two-step process that initially computes the conditional mean using univariate numerical integration for each $i = (1,...159)$.[5]

Next, θ is integrated out of the expression with respect to $p(\theta|y)$. Furthermore, the deviance for INLA is calculated at the posterior mean (or mode in the case of hyperparameters) of the latent field rather than the posterior mean of all parameters as seen in OpenBUGS. $pD_{OpenBUGS}$ is the classical definition of pD, and $D(\theta)$ is calculated as the deviance computed at the posterior mean estimates. For pD_{INLA}, n is the number of observations, $Q(\theta)$ is the prior precision matrix, and $Q^*(\theta)$ is posterior covariance matrix of the Gaussian approximation.[5] I average these estimates over the 200 simulated data sets to gain an overall assessment of performance.

For all of these measures of GoF, lower values indicate a superior model. However, valid comparisons can only be made within models as the likelihoods change when the outcome being modeled changes. Furthermore, a lower value in one software package does not necessarily mean that it fits better than the other package. Another measure of GoF that I could consider is WAIC which includes a smaller penalty for number of parameters, but I decided to limit my options to those described above.[19]

2.3. Results

In order to have comparable results, I must run the OpenBUGS models to convergence. While I am not able to confirm convergence for all 200 data sets easily, I do check a representative percentage of the data sets by way of the Brooks-Gelman-Rubin (BGR) diagnostic plots available in OpenBUGS. All plots indicate that I achieve convergence for these 6 models, and I always extend the model runs for 2500 iterations per chain beyond the convergence point in these test data sets.

Table 4: Parameter estimates associated with both statistical software packages, based on the default precision prior distribution.

Model	α_0	α_1	α_2	α_3	α_4	τ_u	τ_v
M1F1 `OpenBUGS`	0.081 (0.08)	0.097 (0.03)*	-0.055 (0.15)	---	---	---	---
M1F1 INLA(SL)	0.085 (0.08)	0.097 (0.03)*	-0.056 (0.16)	---	---	---	---
M1F1 INLA(FL)	0.082 (0.08)	0.097 (0.03)*	-0.055 (0.16)	---	---	---	---
M2F2 `OpenBUGS`	0.128 (0.12)	0.163 (0.03)*	-0.052 (0.23)	---	---	0.955 (0.20)	---
M2F2 INLA(SL)	0.152 (0.12)	0.164 (0.04)*	-0.054 (0.22)	---	---	18654 (18573)	---
M2F2 INLA(FL)	0.148 (0.12)	0.164 (0.04)*	-0.054 (0.22)	---	---	18654 (18573)	---
M3F3 `OpenBUGS`	0.122 (0.12)	0.166 (0.04)*	-0.018 (0.23)	---	---	0.938 (0.19)	209.47 (177)
M3F3 INLA(SL)	0.145 (0.12)	0.165 (0.04)*	-0.018 (0.23)	---	---	12490 (12472)	18768 (18541)
M3F3 INLA(FL)	0.141 (0.12)	0.166 (0.04)*	-0.018 (0.23)	---	---	12490 (12472)	18768 (18541)
M4F4 `OpenBUGS`	-0.123 (0.10)	0.101 (0.02)*	-0.040 (0.15)	-0.054 (0.07)	0.481 (0.14)*	---	---
M4F4 INLA(SL)	-0.122 (0.11)	0.101 (0.02)*	-0.041 (0.16)	-0.054 (0.07)	0.486 (0.14)*	---	---
M4F4 INLA(FL)	-0.127 (0.11)	0.101 (0.02)*	-0.040 (0.16)	-0.054 (0.07)	0.487 (0.14)*	---	---
M5F5 `OpenBUGS`	0.016 (0.14)	0.160 (0.04)*	-0.05 (0.21)	-0.100 (0.10)	0.237 (0.21)	0.940 (0.20)	189.57 (165)
M5F5 INLA(SL)	0.020 (0.15)	0.161 (0.04)*	-0.061 (0.23)	-0.105 (0.11)	0.279 (0.21)	18345 (18245)	18446 (18385)
M5F5 INLA(FL)	0.014 (0.12)	0.161 (0.04)*	-0.061 (0.23)	-0.105 (0.11)	0.279 (0.21)	18345 (18245)	18446 (18385)
M6F6 `OpenBUGS`	0.149 (0.12)	---	---	---	---	0.955 (0.20)	202.77 (172)
M6F6 INLA(SL)	0.172 (0.12)	---	---	---	---	17989 (17800)	18712 (18558)
M6F6 INLA(FL)	0.169 (0.12)	---	---	---	---	17989 (17800)	18712 (18558)
Truth	0.1	0.1	-0.05	-0.05	0.5	1	1

The notation is as follows: mean of means, (mean of standard deviations), and * indicates a well estimated variable.

Table 4 shows the parameter estimates and their significances associated with each model as well as the truth. Compared to the truth, most models do well with recovery for the fixed effects, with default precision prior distribution. The estimates associated with the spatial variables, x_1 and x_4, are well estimated much

of the time while this is not true for the non-spatial ones. This could be due to the fact that the true values of these parameters have larger magnitudes. These issues are common across both `OpenBUGS` and `R-INLA`. As far as τ_u and τ_v are concerned, the estimates show that `INLA` is far overestimating the true values of the precision parameters; `OpenBUGS` is also overestimating τ_v but not nearly as much as `INLA`. The `INLA` Laplace models (INLA(FL)) indicate that I specified for `INLA` to use the full Laplace strategy rather than the default simplified Laplace strategy (INLA(SL)) [20]. With this specification, I see that there are not many differences with respect to the fixed effect estimates. It is interesting that the standard deviations associated with INLA(SL) and INLA(FL) are the same in nearly all situations when rounded to the hundredth decimal place. I also note that the estimates for α_1 change by a fair margin when comparing M1F1 to M3F3 or M4F4 to M5F5. Similarly, the estimates for α_4 change when comparing M4F4 to M5F5. As both α_1 and α_4 are spatial covariates, this may indicate influence from the spatial random effect estimates.

Supplemental Table 1 in Appendix Section A.1.2. shows the GoF measures associated with each model and software package. Some of these estimates appear to be exactly the same after rounding, but in actuality they are different numbers. Here, I see some differences among models both within and across platforms. For the models without random effects (M1F1 and M4F4), the pD values are very close, and INLA (FL) produces significantly lower values than INLA (SL). The $MSPE$ and $MSE(M)$ estimates produced by both packages are nearly identical, but when the Laplace strategy is applied in `INLA`, higher estimates for the spatial models (M2F2, M3F3, M5F5, and M6F6) are recorded. This table also includes results for the models using the alternative precision prior distributions. When I adjust the precision prior distributions for `OpenBUGS`, I begin to see significant differences in the pD and thus the DIC produced for the convolution models (M3F3, M5F5, and M6F6). Note that when running `OpenBUGS` for the altered prior distributions, there are a few (no more than 3 per model) data sets that lead to negative pD values in M3F3, M5F5, and M6F6. I remove these results from the mean calculations.

Table 5 illustrates how the precision parameters change after alterations to the default settings. In addition to changing the default prior distributions, the `inla.hyperpar()` command is also explored to try and attain better estimates of the hyperparameters involved with estimating the random components of the models. While this function only affects the hyperparameter estimates, all others remain the same. It is obvious that these estimates are different from the original estimates located in the first row, but they still do not reflect the true values. In fact, they are further away from the truth by making the effects more precise.[21]

Table 5: Altered `INLA` precision estimates compared to the original estimates as well as the truth.

Model	Parameter	M2F2	M3F3	M5F5	M6F6	True value
`INLA` Gam(1,5e-05)	τ_u	18656 (18577)	12491 (12475)	18485 (18381)	17990 (17804)	1
	τ_v	---	18770 (18546)	18570 (18440)	18714 (18563)	1
`INLA` Ga(1,5e-05) `inla.hyperpar()`	τ_u	22376 (15350)	15037 (10514)	21962 (15924)	21886 (15046)	1
	τ_v	---	21938 (16641)	21920 (15989)	22525 (15636)	1
`INLA` Gam(1,0.5)	τ_u	0.98 (0.20)	1.03 (0.23)	1.07 (0.25)	1.06 (0.24)	1
	τ_v	---	3.92 (2.44)	3.42 (2.22)	3.94 (2.43)	1
`OpenBUGS` Gam(1,0.5)	τ_u	0.94 (0.19)	1.00 (0.23)	1.03 (0.25)	3.34 (2.08)	1
	τ_v	---	1.04 (0.26)	3.79 (2.24)	3.76 (2.25)	1
`OpenBUGS` Gam(1,5e-05)	τ_u	0.96 (0.20)	0.94 (0.19)	0.94 (0.20)	0.96 (0.20)	1
	τ_v	---	209.47 (177)	189.57 (165)	202.77 (172)	1

The next set of precision estimates relate to the models where I specified $Gam(1,0.5)$ as the prior distribution. Results for the $Gam(2,1)$ and $Gam(1,1)$ priors are located in Supplemental Table 2 Appendix section A.1.2. Changing these prior distributions greatly alters the results with respect to the precision estimates while the fixed and GoF estimates for these models remain nearly identical. In fact, the standard deviations of the estimates remain very close to the originals. It is obvious that `R-INLA` improves

drastically with respect to the precision estimates related to both effects, correlated and uncorrelated. While `OpenBUGS` produces more precise estimates under the default settings, changing the prior distributions continues to improve the accuracy of those estimates in this package as well. I see now that the estimates produced in `INLA` reflect those produced in `OpenBUGS` very closely. This is true for all of the alternative prior options explored. Again, the $Gam(1,1)$ prior seems to perform the best as it produces estimates that are the closest to the truth in all cases and for both packages. Unlike the GoF measures, though, the differences seen here are not significant within the alternative prior distributions.

The last alteration I attempt involves scaling the models based on the Scaling IMGRF models tutorial on the `INLA` website, and these results can also be found in Supplemental Table 2 Appendix section A.1.2.[22] I make this alteration using the $Gam(1,5e-05)$ and $Gam(1,0.5)$ prior distributions only. This tutorial suggests implementing the following global command: `inla.setOption(scale.model.default = TRUE)` to scale the global variance of the model such that $\sigma^2_{GV}(x)=1$, then the reference variance, $\sigma^2_{ref}(x)$, is used to scale the hyperprior as follows: $\frac{\tau}{\sigma^2_{ref}(x)}$. Based on the results below, I can see that this modification does not always ensure that the precisions are closer to the truth.[23]

Figure 2 shows how the models in `OpenBUGS` and `INLA` recover the uncorrelated and correlated spatial effects respectively using M5F5. The other model estimate maps can be found in Supplemental Figures 2-5 Appendix A.1.2. The values displayed are averaged over the 200 data sets; furthermore, the 'truth' is only one realization of the random effect, but this is what I used to simulate the outcomes. Note that all of these maps have been scaled in the same way as the plot displaying the true effect to gain a better comparison and understanding of the relationship being displayed, and this is why the `INLA` plots appear to have no variation. `OpenBUGS` seems to be recovering both effects better than `INLA` under the default settings, but neither set of models are preforming as well as I would like with respect to the correlated random effect,

v . The INLA estimates seem to be much closer to zero than they should be, and this is supported by the larger precision estimates shown in Tables 3 and 5 above. Also, the discrepancies among the correlated effects may be reflecting the spatial nature of some of the covariates. Furthermore, others have noted correlated spatial effects behaving too smoothly in certain cases.

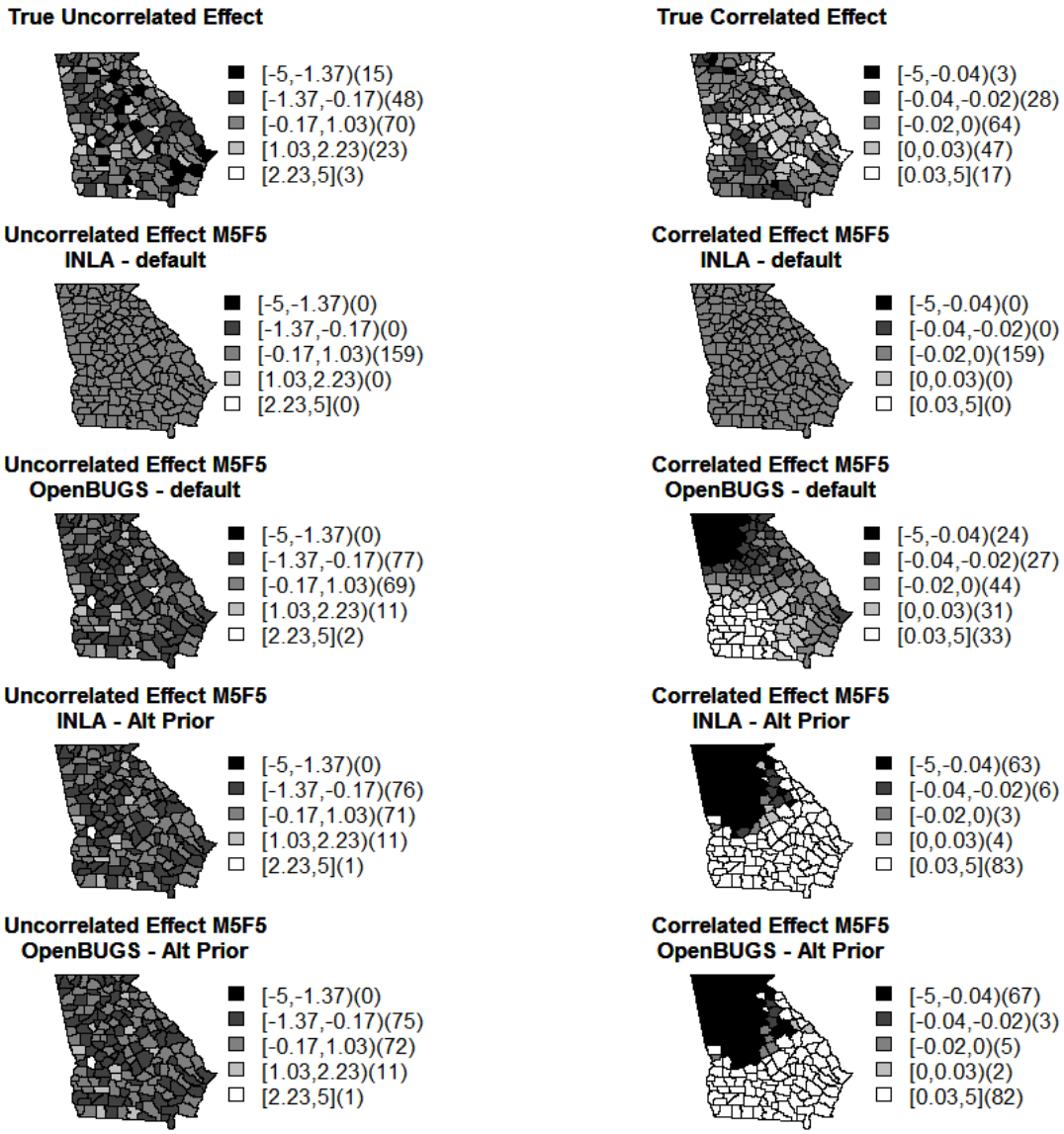


Figure 2: The true and mean estimated spatial effects as calculated in INLA and OpenBUGS under default and alternative priors using M5F5.

The maps notated with 'Alt Prior' display the uncorrelated and correlated effects produced in INLA and OpenBUGS when the prior distributions for τ_u and τ_v are changed from $Gam(1,5e-05)$, the default, to $Gam(1,0.5)$. This prior alteration results in a much better reproduction of the spatial random effects, as reflected in the precision estimates displayed in Table 5. The uncorrelated effects are very similar for both INLA and OpenBUGS, and they also reflect the truth much better than before with the default prior distributions. As far as the correlated effects are concerned, I see that the estimations for these effects are almost identical when looking at the results produced in OpenBUGS and INLA, but they still do not reflect the truth as well as I may like. As mentioned before, though, these effects do reflect the smoothness expected among correlated effects as well as the spatial structures present in some of the covariates. These figures also reflect the results produced for the other alternative precision prior distributions, as they are alike. Finally, Supplemental Figure 6 Appendix A.1.2. is the sum of the correlated and uncorrelated random effects seen in Supplemental Figures 4 and 5. This figure looks much alike Supplemental Figure 4 as the magnitude of the uncorrelated effect is larger than that of the correlated one. This may be one of the factors inhibiting my ability to appropriately recover the correlated random effect.

Simulation Variants

The results in Supplemental Table 3 Appendix A.1.2 show the fit of M5F5S when the covariates are standardized rather than only mean centered. These results show that standardization leads to a well estimated α_4, but also some larger standard deviations with respect to OpenBUGS for α_2 and α_3. Note that α_1 for INLA(FL) is very close to being well estimated. The precision estimates for OpenBUGS continue to be slightly closer to the truth though this is not statistically significant.

Supplemental Table 4 Appendix A.1.2. shows the GoF measures for the standardized models, and these results show similar patterns. I cannot truly compare the deviance and DIC measures presented here to the previous results since the outcomes are different, but I can look at the patterns among the other measures. I

continue to see a separation with respect to the *MSPE* while the different *MSE* estimates are nearly identical. Supplemental Figure 7 Appendix A.1.2 illustrates the maps of the uncorrelated and correlated effects produced for the standardized model, M5F5S. These maps look comparable to what I saw previously for M5F5, and I still see much likeness when comparing the `INLA` results to the `OpenBUGS` results.

The results in Supplemental Table 5 Appendix A.1.2. show the re-fit of the models with a $Gam(1,0.5)$ prior on the precisions of the spatial random effects with the validation data set, which offers a different distribution for the correlated random effect, and a different realization of the uncorrelated random effect. For the `INLA` models, I continued to use the Laplace strategy for a better fit. Based on these results, I am still seeing changes in the estimates for α_1 when comparing models M1F1V and M3F3V; this time I actually see this estimate fail to be well estimated in M3F3V while it is well estimated in M1F1V. I also see this occur for α_1 when comparing M4F4V to M5F5V. For α_4 in M4F4V and M5F5V, the estimates are different, but they are not well estimated in either model. Also, the standard deviations are larger. With respect to the precision estimates, I still see an overestimation of the correlated effect, τ_v, while the uncorrelated effect, τ_u, is recovered fairly well. Even though the truth for τ_v is smaller in this validation study compared to the first simulated data set, some of the estimates seen in the table below are actually higher in this situation.

Supplemental Table 6 Appendix A.1.2 shows the GoF estimates for the models using the validation data set. I see `OpenBUGS` producing at least slightly lower $MSE(M)$ and *MSPE* values than INLA(FL). For $MSE(\theta)$, the relationship is alike the models with the original data in that `OpenBUGS` consistently produces slightly lower estimates when spatial random effect are considered. The mean deviance is very

similar for all models and packages. Note that there were also some negative pD values for M5F5V and M6F6V in `OpenBUGS`; these were removed in the same way as before.

Supplemental Figures 8 and 9 in Appendix A.1.2. demonstrate how well the models recover the true uncorrelated and correlated spatial random effects when applied to the validation data set. Here I see an analogous situation to the first set of simulated data in that the uncorrelated effect is recovered very well while the correlated is not recovered quite as well as I would like it to be. I also notice, again, that the correlated effect is smoother than the true correlated effect. Furthermore, the INLA(FL) results are very similar to the `OpenBUGS` results just as I noted. One difference is that there is not as much of a change in the estimation when it comes to comparing M6F6V with the others that include covariates. This may be because the true correlated spatial effect is distributed more similarly to the spatial covariates. Following that, Supplemental Figure 10 Appendix A.1.2. shows the sum of the correlated and uncorrelated random effects, and as seen before, this looks well estimated because the correlated effect is of a much smaller magnitude compared to the uncorrelated effect. Because of this, it also looks analogous to Supplemental Figure 8.

2.4. Discussion

The results above show some substantial differences in the performances of `INLA` and `OpenBUGS`. For the default settings, I see that many of the fixed effect parameter estimates are alike for the two software packages, but the differences become more evident when looking at the GoF measures and spatial random effect estimates. `OpenBUGS` seems to outperform `INLA` when spatial random effects are included in the model. This is shown in the $MSE(M)$ and $MSPE$ when spatial random effects are added to the models as well as the individual plots of the produced estimated spatial effects which is to be expected based on the precision estimates. Furthermore, I note improvements to the INLA(FL) models when I specify the full Laplace strategy and even more still altering the default precision prior settings.

INLA does have an advantage over OpenBUGS in computation time. The computing time for these 200 simulations is no more than a couple of hours for INLA while it takes OpenBUGS several days. Obviously, these times fluctuate depending on the computer specifications, what other processes are actively running on the system in use, and, in the case of OpenBUGS, the number of parameters that are being collected. If I did not need to recover u, v, y_{pred} or μ, the computation time could be much shorter, but even taking that into account, OpenBUGS will never be as fast as INLA. Note that I run all simulations on the same server, so these arguments are appropriate. This indicates that, especially for simpler models, if computation time is an issue, then INLA may be the better option. Additionally, R-INLA is noted for being able to apply non-discrete spatial effects.[24]

A shortcoming for INLA involves the ability to use hyperparameters as flexibly as in OpenBUGS. For example, I was unable to implement prior distributions for the standard deviations in INLA, while this can be done easily in OpenBUGS. Placing prior distributions on the standard deviations rather than fixing them or placing them on the precisions can lead to better model fits in some situations. Furthermore, there is not an easy way to place hyperprior distributions on the precisions of the fixed effects.

There are many options in INLA for improving the models. Initially, I explore specifying the use of a full Laplace approximation strategy in INLA, but this does not lead to different parameter estimates and takes a longer time to complete for the 200 data sets and 6 models. Specifying the full Laplace strategy did, however, lead to different GoF measures that were closer to those produced with OpenBUGS. Furthermore, the simplified Laplace strategy is not sufficient for computing predictive measures.[21] There is one other strategy that can be implemented, the Gaussian, and it is the most efficient but least accurate.[20] Following that, I explore using the `inla.hyperpar()` option to no avail. This option is supposed to improve the hyperparameter estimates, but, as displayed in Table 5, these results actually brought the estimates further from the truth. This function is defined as a way to retrieve more precise estimates of the hyperparameters,[25] and this is how it behaves in this simulation study. I also consider changing the default

prior distributions with respect to the precision parameters for both the INLA(FL) and OpenBUGS models, and the results are much improved. Both the uncorrelated and correlated random effect estimates are much closer to the truth. There also appears to be less separation between the estimates produced in `OpenBUGS` verses those produced in INLA(FL). Of the options attempted for these altered prior distributions, the $Gam(1,1)$ prior preformed the best, but it is also the most informative. The $Gam(1,0.5)$ option is a good choice as it is still fairly non-informative. Furthermore, it still shows substantial improvements over the default settings. Finally, implementing the model scaling feature did not aid in the models gaining estimates closer to the truth, which is what I am aiming for. It seems as though this feature is mostly for assisting in model interpretability and comparability.

When I apply the models to the validation data sets I see very similar patterns to those present in the first simulation data sets, but there were also some differences. Fewer of the covariates were well estimated in the validation data set. Also, the correlated random effect estimates did not vary as much from model to model as in the first data sets, but they were still not as close to the truth as I would like for them to be. This could be due to the fact that the true correlated spatial effect is slightly closer in its distribution to the spatially structured covariates. Furthermore, the uncorrelated effect estimates do not seem to be recovered quite as well. These differences may be due to the masking effects from either the spatially structured covariates or the unequal true precision values. I see similarities with the patterns present in the GoF measures though these values cannot be compared directly because the model ingredients are not the same.

During this study, I uncovered many interesting properties associated with both packages for these types of spatial disease mapping models. One relates to having a mixture of spatially structured covariates as well as spatial random effects. There appears to be a type of masking that occurs when both are present in the model. This is seen in two ways. First, by looking at the parameter estimates presented in Table 3; here, I notice that the estimates for α_1 change from M1F1 to M3F3. A comparable change occurs for α_4 when considering M4F4 verses M5F5. In the second case, α_4 changes from being well estimated to no longer

being well estimated. Second, the masking effect is seen by comparing the correlated spatial effect estimates associated with M6F6 to all other models. Since M6F6 does not have any covariates, spatial or otherwise, I can see that the covariates may be playing a role in changing the estimates produced. Note that this was not as evident in the validation study, especially with respect to the plots of the correlated spatial random effect estimates.

Another interesting property of these models is that, in general, the correlated spatial effect tends to be overestimated and well as overly smooth which is evident in the figures produced above. I also do not see the non-spatial covariates being well estimated in any of the models. This could be due to the fact that the magnitudes of the true estimates are smaller than the others. Furthermore, when considering the validation data set, I see that the models have a more difficult time recovering the true effects, and this may be due to the occurrence of a masking effect from the unequal true precisions.[16]

2.5. Conclusion

Ultimately, `INLA`, in its default state, does not perform as well as `OpenBUGS` with respect to the precision parameter estimates for the spatial random effects, but it is much more computationally efficient. Through this simulation study, I learned that by specifying the full Laplace strategy, I result in better fitting models that are equivalent to `OpenBUGS`. Furthermore, altering the precision prior distributions for correlated and uncorrelated random effects brings these estimates much closer to the truth as well as to the values produced by `OpenBUGS` when the same prior distributions are used. Thus, for Poisson modeling in disease mapping, it is of utmost importance to adjust the default settings when using `INLA` as an alternative for Bayesian Analysis, especially when spatial random effects are included in the models.

3. Aim 2: Spatial Model Selection

3.1. Introduction

There are many instances in the disease mapping frame work where one may wish to select between two or more linear predictors, or models, of interest. In certain situations, the model selection process may be more applicable than simply using variable selection; for example, if there are prior beliefs that these particular linear predictors could be informative.

In this chapter, I discuss a way to implement BMS in comparison to BMA[18, 26, 27] using the BUGS software byway of `BRugs` and `R2WinBUGS`.[3, 13, 28] Both of these methods are considered to be model selection techniques and can be used in lieu of variable selection as they alleviate some of the issues related to variable selection (e.g. co-linearity).[29-34] Here, if two or more predictors are known to be collinear, I suggest that they be included in the different linear predictors that the method is selecting between. The structure of my proposed BMS procedure is similar to BMA, and thus, this commonly used method makes for a respectable comparison.

There are several useful variable selection procedures available in the Bayesian paradigm,[35, 36] but I believe that model selection is the better alternative, especially when there are particular linear predictors of interest. Similar comparisons between BMA and variable selection has been performed in the past. Viallefont's simulation study first shows that variable selection methods often produce too many variables selected as 'significant.' Then, they determined that BMA produces easy to interpret, precise results displaying the posterior probability that a variable is a risk factor. They also caution users about interpreting the averaged posterior parameters as each of the alternative linear predictors has been adjusted for different confounders.[37]

Spatial model selection, which allows different linear predictors to be selected for each spatial unit, is the main focus of this chapter. The data is partitioned into different spatial areas to determine how well the

model selection procedure recovers the truth when spatial structure is present in the data. I make a comparison of BMS to BMA and additionally, calculate GoF measures for each of the partitioned areas.

This chapter is developed as follows. First, I describe the methods associated with the BMS and BMA processes. Next, I discuss the development of my simulated data set and the different models used for exploring this methodology. Finally, I discuss the benefits of using the model selection method under different scenarios in the disease mapping context.

3.2. Methods

This chapter focusses on the context of disease mapping in *m* predefined small areas. I make the conventional assumption that an aggregate count of disease (y_i) is observed in the i^{th} small area and that these outcome counts are conditionally independent Poisson distributed outcomes, i.e. $y_i \mid \mu_i \sim Pois(\mu_i)$. This is a commonly assumed model for small area counts in disease mapping.[17] In what follows I examine two types of model formulation: complete and partial models. For the complete models, it is assumed that the linear predictor applies to the whole study region, in the sense that the underlying model is the same for all areas. In the partial models, it is assumed that the underlying linear predictor is different for different partitions of the set of spatial units.

3.2.1 Bayesian Model Selection

To evaluate a number of alternative linear predictor models, I can adopt a method which fits a variety of models, and the selection of weights allows each model to be evaluated for its appropriateness. In general, for $d = 1,....,D$ complete models, the following structure applies:

$$y_i \mid \mu_i \sim Poi(\mu_i)$$
$$\mu_i = e_i \theta_i$$
$$\log(\theta_i) = \sum_d w_d \varphi_{id}$$
$$\text{logit}(p_d) \sim Norm(0, \tau_d)$$

$$w_d \sim Bern(p_d)$$

where φ_{id} is my d^{th} model's suggested linear predictor complimented with a possible random effects. In general, I write φ_{id} as $x_i^T \beta\psi + u_i\psi_{d,J+1} + v_i\psi_{d,J+2}$ with x_i^T, the vector of J possible covariates, $j = 1,\ldots, J+2$, u_i the uncorrelated heterogeneous (UH) term, v_i the correlated heterogeneous (CH) term, and ψ_{dj} an indicator for if the j^{th} predictor or random effect is to be included in the linear predictor of the d^{th} model. Model priors here as well as in the following models are such that: $\beta_{dj} \sim Norm(0,1)$, $u_i \sim Norm(0,\tau_u)$, $\tau_u \sim Gam(1,0.5)$, and $\tau_d^{-1/2} \sim Unif(0,5)$. Hence, for a variable not included in the d^{th} model, ψ_{dj} would be zero, otherwise it would be one. Further, w_d is a model selection indicator, equal to one if the d^{th} model is selected and zero otherwise. The model selection probability is given by the probabilities p_d. For the partial models, the following structure applies:

$$y_i \sim Poi(\mu_i)$$

$$\mu_i = e_i\theta_i$$

$$\log(\theta_i) = \sum_d w_{id}\varphi_{id}$$

$$\text{logit}(p_{id}) \sim Norm\left(\frac{1}{n_i}\sum_{i\sim l}\text{logit}(p_{idl}), \frac{1}{n_i\tau_{id}}\right)$$

$$w_{id} \sim Bern(p_{id})$$

in which the model selection indicator is spatially structured. In the equations, $i \neq l$, n_i is the number of neighbors for county i, and $i \sim l$ indicates that the two counties i and l are neighbors. This is an intrinsic CAR model, which adds the desired spatial structure to the model selection process and a new level of complexity in comparison to the complete models.

3.2.2 Bayesian model averaging

Bayesian model averaging is similar to the BMS technique described in the previous section.[38] This method averages over the D possible models, $M_1,\ldots,M_D$ to find the posterior distribution of θ as follows:

$$P(\theta|y_1,...,y_m)=\sum_{d=1}^{D}P(\theta|y_1,...,y_m,M_d)P(M_d|y_1,...,y_m)$$

where $P(M_d|y_1,...,y_m)$ is the model probability for model *d*, and $P(\theta|y_1,...,y_m,M_d)$ is determined by marginalizing the posterior of the model parameters. The posterior probability for selection model M_d is given by:[27, 37, 38]

$$P(M_d\,|\,y_1,...,y_m)=P(y_1,...,y_m\,|\,M_d)P(M_d)/C$$

$$\text{where } C=\sum_{l=1}^{D}P(y_1,...,y_m\,|\,M_l)P(M_l)$$

$$\text{and } P(y_1,...,y_m\,|\,M_d)=\int P(y_1,...,y_m\,|\,\theta_d,M_d)P(\theta_d\,|\,M_d)d\theta_d$$

Alternatively, I can estimate the model probabilities $P(M_d|y_1,\ldots,y_m)$ using the deviance information criteria (DIC). The model probabilities can be approximated by:[39]

$$p_{DIC1}(M_d\,|\,y_1,...,y_m)\approx\frac{e^{-DIC_d}}{\sum_k e^{-DIC_k}}$$

$$p_{DIC2}(M_d\,|\,y_1,...,y_m)\approx\frac{\sum_i e^{-DIC_{id}}}{\sum_k\sum_i e^{-DIC_{ik}}}$$

In these expressions, DIC_d is the DIC associated with the d^{th} alternative linear predictor. Note that in the first expression the *DICs* are calculated overall (over all areas), whereas in the second expression, the *DICs* are calculated locally (per region). I calculate this measure by collecting the deviance for each of the alternative linear predictors using the Poisson likelihood as follows:

$$D_i(\theta_i) = -2\left(l_i(\theta_i)\right) = -2\left(y_i \log(e_i\theta_i) - e_i\theta_i - \log(y_i!)\right)$$

Next, $\overline{D_i}$ is computed from the sampler and supplied to the following function to calculate the local DIC for county i: $DIC_i = \overline{D_i} + pD_i$ where $pD_i = \overline{D_i} - D_i(\bar{\theta})$ and $\bar{\theta}_i$ is the posterior mean of θ_i. Finally, the model DIC is simply the sum of the local $DICs$.

To use the BMA framework with the partial models, I apply the same type of spatial structure seen with BMS on the model probabilities by way of the CAR model. Then, I simply use the local DIC to calculate the alternative model probabilities measure for each of the m areas.

3.3. Simulated Data and Fitted Models

In this and the following sections, models will be referred to by the contents of both the simulated data and the fitted model applied. Table 6 is a list of notations for describing models, and Supplemental Table 1 in Appendix A.2. is a list of all simulated data models explicitly defined.

Table 6: Notation for describing model contents.

Notation	Definition
E1	Model with $e_{ij} = 1$
E2	Model with $e_{ij} \sim Gam(1,1)$
C	Complete model
P	Partial model
SX	Simulated data model 'X'
FX	Fitted model 'X'
RE	Model with an uncorrelated random effect included
CV	Model with a convolution component included

3.3.1. Simulated Data

To evaluate the performance of these alternative approaches, I simulate data to establish realistic ground truth for disease risk variation. To match the methods described above, I define a count outcome as y_i in the i^{th} small area. I assume a map of m small areas. In addition, I assume that the expected count (e_i) is

available in each small area. Thus, my outcome follows a Poisson distribution with expectation μ_i in county i.

In the simulations, I fix the expected rate for the areas in order to focus on the estimation of relative risk θ_i. To complete the parameterization, I assume a relative risk which is parameterized with a range of different risk models. I examined nine basic models for risk (S1 up to S9) which have different combinations of covariates and random effects as might be found in common applications. First, I generated four predictors with different spatial patterns. The four chosen variables were median age (x_1), median education (x_2), median income (x_3), and a binary predictor representing presence/absence of a major medical center in a county (x_4). These variables are the same county-level measures for the 159 counties in the state of Georgia used in Chapter 2.

For the simulations, I fix the covariates as one realization from the distributions described in Table 6 and generate the outcomes using a fixed set of parameters unique to each of the 9 models. The β's, seen in Table 7 are quite small, particularly for the first six models, and this guarantees that the outcome variable maintains a fairly small value to continue representing a sparse disease. Note that S7, S8, and S9 are alike S2, S4, and S6 respectively with the exception of higher magnitudes associated with the fixed parameter estimates. Further, note that models S4 and S8 do not pick any of the variable that have a spatial structure. In the next step, $\log(\theta_i)$ is calculated based on the fixed parameters and realizations. Finally, I generate the outcome as a Poisson distributed variable with mean $e_i\theta_i$. The simulated data sets consist of sets of counts: $\{y_{ij}^*\}$, $\mathrm{k} = (1,\ldots,300)$ where k denotes the k^{th} simulated data set. Note that I simulated two batches of the covariates with 150 data sets per batch and still allow for Poisson variation between the 300 simulate data sets. This reduces the amount of variation in the simulation study which allows the main focus to be on the different model selection techniques.

Table 7: Model coefficients for the simulated data.

Model	β_{x1}	β_{x2}	β_{x3}	β_{x4}
S1	0.1	0.1	0.1	0.1
S2	0.1	-0.1	0.1	-0.1
S3	0.1	0.0	0.2	0.1
S4	0.0	-0.1	0.1	0.0
S5	-0.1	0.2	0.0	0.0
S6	0.0	0.0	0.0	0.1
S7	0.2	-0.2	0.2	-0.3
S8	0.0	-0.3	0.3	0.0
S9	0.0	0.0	0.0	0.3

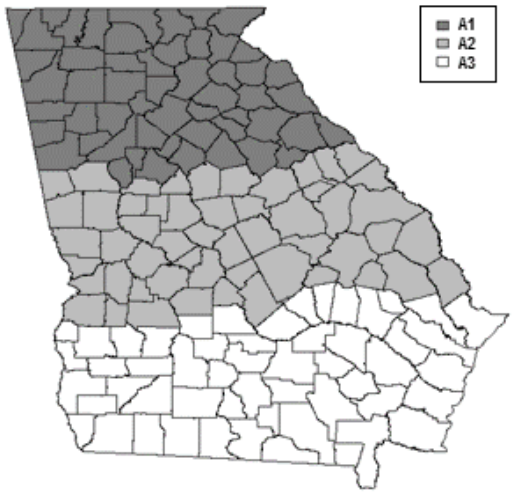


Figure 3: Display of the partitioned regions for the partial models.

In addition to the 9 model variants described in Table 7, I have also considered a variant that allows me to limit the regions of the state that certain predictors have an effect. This will allow me to study spatially-varying model selection or spatially-varying model averaging. The state is partitioned into three regions, A1, A2, and A3, each containing 53 counties as illustrated in Figure 3. Next, I will consider partial models, notated as 'P,' for each of the 9 models described in Table 7. These scenarios are described in Table 8.

Finally, I also explored the case that there is some unmeasured confounding in the data. This was accomplished by including a UH term and a CH term defined as $u_i \sim Norm(0, 0.01)$ and

$v_i \sim N\left(\frac{1}{n_i}\sum_{i\sim l} v_l, \frac{1}{n_i \tau_v}\right)$ respectively. The fixed precision values for these components are the same $\left(1/\tau_v = 0.01\right)$ so that one term does not dominate the other and lead to identifiability issues.[40] Two additional simulated model scenarios result from inclusion of these random effects; the first assumes that there is simply random error present in the data, thus only the UH term is involved. This model is denoted as RE. The second model is a convolution model, denoted CV, and includes both the UH and CH terms; this model assumes that there is random error as well as error with a spatial structure. For simplicity, I only looked at this application for the first simulated model scenario (S1).

Table 8: Scenarios for partial models.

Notation	Meaning
PS1	M1 with x1 and x2 missing from A1
PS2	M2 with x3 and x4 missing from A2
PS3	M3 with x1 and x4 missing from A2 and A3
PS4	M4 with x2 missing from A1
PS5	M5 with x2 missing from A1
PS6	M6 with x4 missing from A2
PS7	M7 with x1 and x2 missing from A1
PS8	M8 with x2 missing from A1
PS9	M9 with x4 missing from A2

3.3.2. Fitted Models

In Table 9, I illustrate the 3 linear predictors fitted using both the model selection and BMA techniques described above. These options are appropriate for the models listed in Table 7 such that S1 up to S9 is associated with F1 up to F9. For the complete models, the suffix 'Alt1' represents the true model for all counties. For the partial models, the suffixes 'Alt1' or 'Alt2' represent true models for some counties depending on the descriptions in Table 8. 'Alt3' is never a true model. The different alternatives offer some models that include random noise to determine if that has an effect on which model is selected. When the 'RE' and 'CV' simulated models are being fitted, all linear predictor alternatives include the appropriate random effects.

Table 9: Model Alternatives.

Model	β_{x1}	β_{x2}	β_{x3}	β_{x4}	Rand. Noise
F1Alt1	Yes	Yes	Yes	Yes	No
F1Alt2	No	No	Yes	Yes	No
F1Alt3	Yes	No	No	No	No
F2Alt1	Yes	Yes	Yes	Yes	No
F2Alt2	Yes	Yes	No	No	No
F2Alt3	No	No	No	Yes	Yes
F3Alt1	Yes	No	Yes	Yes	No
F3Alt2	No	No	Yes	No	No
F3Alt3	No	No	Yes	No	Yes
F4Alt1	No	Yes	Yes	No	No
F4Alt2	No	No	Yes	No	No
F4Alt3	No	No	Yes	Yes	Yes
F5Alt1	Yes	Yes	No	No	No
F5Alt2	Yes	No	No	No	No
F5Alt3	No	Yes	Yes	No	No
F6Alt1	No	No	No	Yes	No
F6Alt2	No	No	No	No	No
F6Alt3	No	Yes	Yes	No	No
F7Alt1	Yes	Yes	Yes	Yes	No
F7Alt2	No	No	Yes	Yes	No
F7Alt3	Yes	No	No	No	No
F8Alt1	No	Yes	Yes	No	No
F8Alt2	No	No	Yes	No	No
F8Alt3	No	No	Yes	Yes	Yes
F9Alt1	No	No	No	Yes	No
F9Alt2	No	No	No	No	No
F9Alt3	No	Yes	Yes	No	No

3.4. Results

The results below are in relation to the implementation of the methods described above. First I present the BMS results for the complete and the partial models followed by the BMA results for the complete and partial models. Finally, I present a real data example of implementing these methods.

Note that I also collected the parameter estimates, produced for each of the alternative models, during each implementation. None of these estimates are well estimated as seen in Supplemental Table 2 in Appendix A.2., but obtaining these estimates is not the goal of the model selection process. I suggest that one should refit with the selected model to obtain appropriate parameter estimates.

3.4.1. Complete Models

All results in this section are associated with the complete models described in Table 7 above. The models do not allow for variation from one county to the next, and the first linear predictor alternative is the true linear predictor, thus, model weight p_1 should be the highest.

Figure 4 presents some the results for model scenarios S1-S6 fitted with corresponding model F1-F6 under the scenario of a constant expected rate (E1) and varying expected rates (E2). The figure illustrates the distribution of the model weights for the BMS procedure. These figures suggest that most models correctly select the linear predictor associated with p_1. When the evidentiary support for a particular model increases, as is the case in scenarios S7F7-S9F9, this becomes even more evident. The model selection probabilities decrease from the first to the second and third model, showing a better distinction between the models. Only in the scenario S6F6 and S4F4, an incorrect model is selected. Note, however, that the evidence in these scenarios for the underlying model is small.

The model probabilities corresponding to scenarios E1 and E2 are similar. In most settings, it is observed that the model probability of the correct model decreases when comparing scenario E1 to E2. This indicates that model selection is more difficult when the variation in the data increases. But, differences between the model probabilities are only small, and conclusions about the selected model hardly change. The main difference I observed when comparing models in this way is that the E2 models produce lower DICs. This fact does not necessarily indicate that the E2 models fit better than the E1 models as the models have different likelihoods. These GoF measures are displayed in Supplemental Table 4.

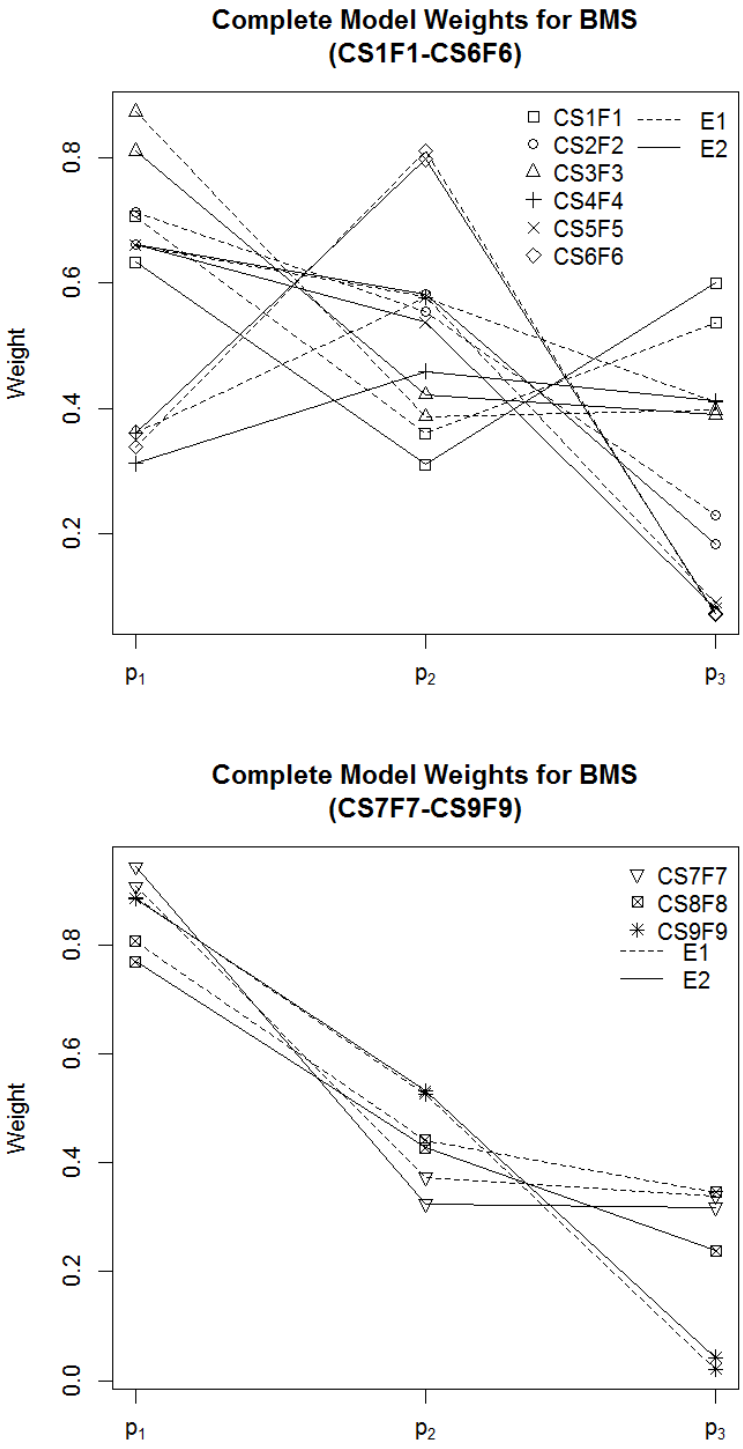


Figure 4: Models weights associated with the complete models using BMS.

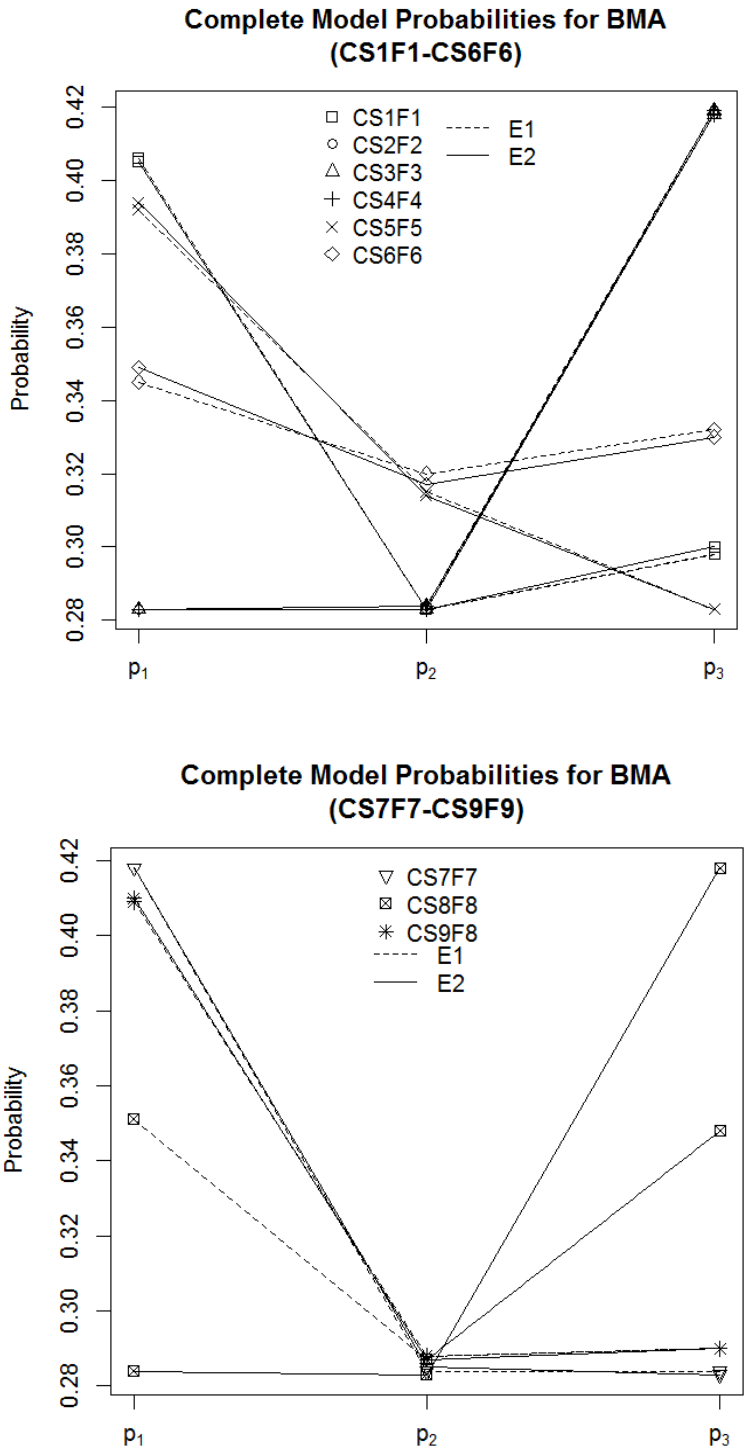


Figure 5: Model probabilities associated with the complete models using BMA.

Figure 5 displays the model probabilities associated with the complete models using the BMA technique. A completely different picture as compared to BMS is observed. These plots suggest that there is not a consistent pattern in the way that the different models perform as far as selecting linear predictors. Typically, p_1 still obtains the highest probability value, but there are several mismatches (7 of the 18 models, 38.9%). As with the BMS models, the E1 and E2 estimates continue to be nearly identical for the model probabilities while they differ when considering GoF measures. These GoF measures are displayed in Supplemental Table 6.

3.4.2. Partial Models

All results in this section are associated with the partial models described in Table 8. These models allow for variation in the linear predictors applied in different areas on the county map. Furthermore, for all models the first or second alternatives are true for certain areas of the county map, thus model weights p_1 or p_2 should be the highest in their appropriate areas. Here, I display only a sample of the models fitted. The resulting county maps for the full range of models fit with the BMS and BMA methodology can be seen in Supplemental Figures 1 and 2 in Appendix A.2. respectively.

The first set of maps shown in Figure 6 are associated with the E1PS4F4 and E1PS8F8 scenarios using BMS. These models in particular illustrate the relationship I hope to see in the partial models. S4F4 does well to estimate p_1, but is not as accurate with p_2. In comparison, S8F8 improves the estimations of both p_1 and p_2, in that it attains higher values in the appropriate regions of the county map. In both cases, though, I continue to see a residual present in p_3. This indicates that the model can only select the correct underlying model if the parameter effect is strong enough.

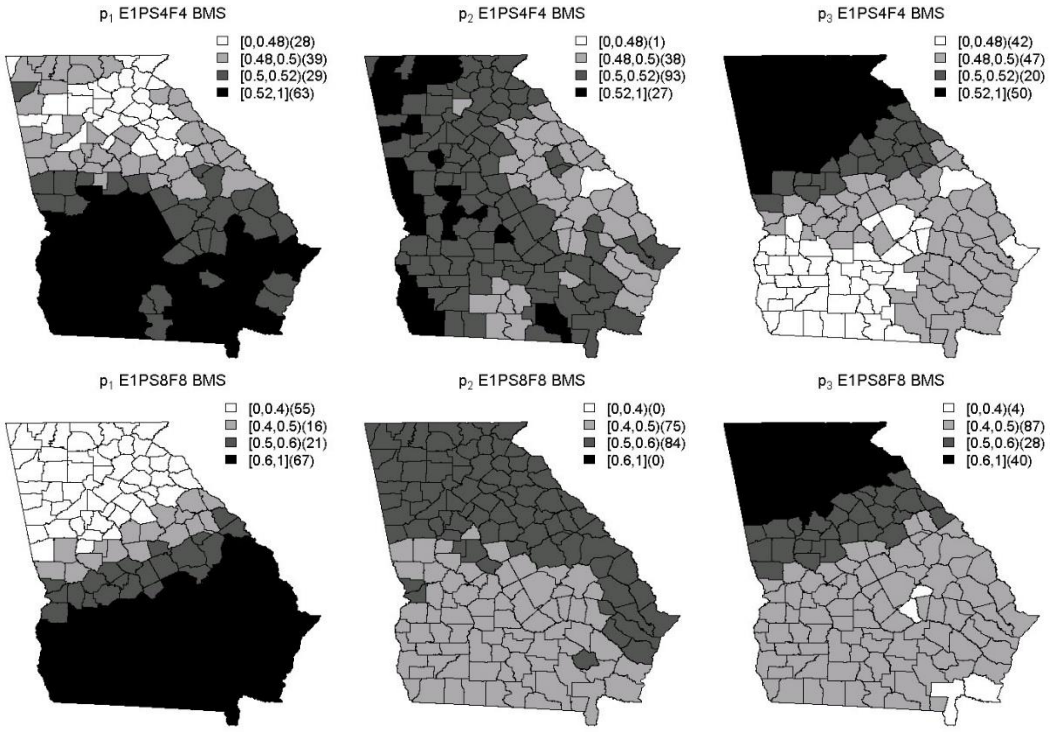


Figure 6: Model weights associated with E1PS4F4/S8F8 for BMS.

The second set of maps involving S6F6 and S9F9 displayed in Figure 7 are not as convincing, and we believe this is largely due to the fact that the true regions associated with p_1 are strictly separated to the extreme North and South of the map. There is still evidence of an improvement when moving from S6F6 to S9F9 since the regions are becoming slightly more defined as far as the maps are concerned. It may be the case that we need even more evidentiary support when the areas are separated in this manner.

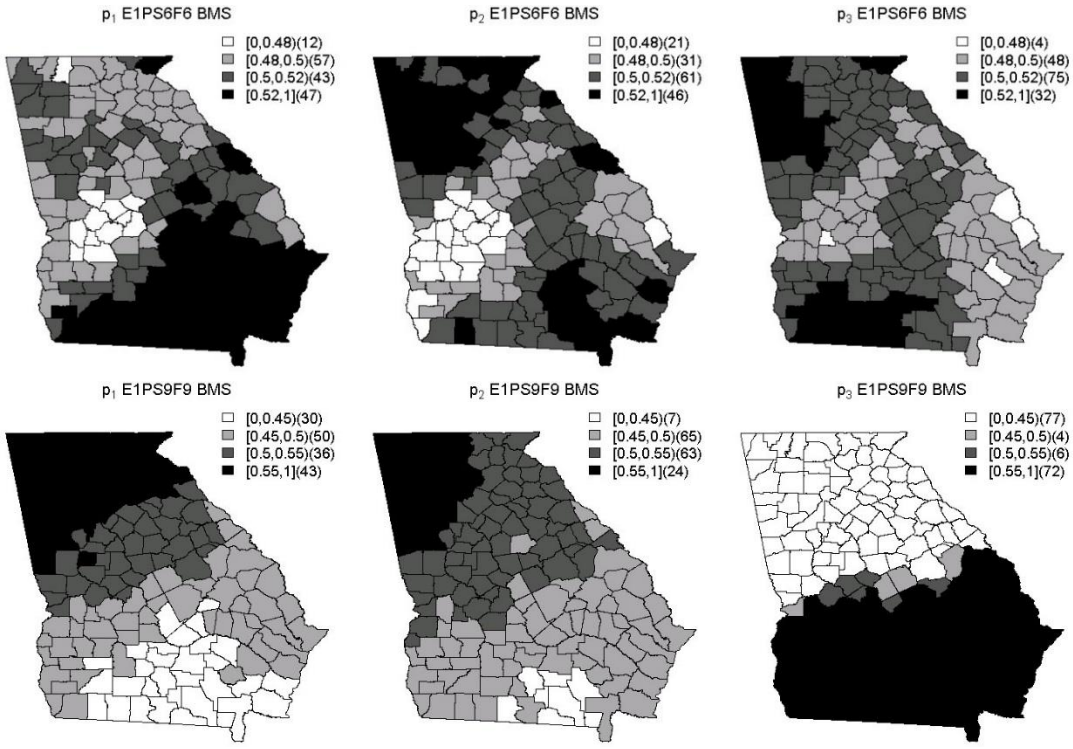


Figure 7: Model weights associated with E1PS6F6/S9F9 for BMS.

Table 10 displays the GoF measures associated with the BMS technique applied to the partial models. I continue to see smaller DIC estimates associated with the E2 models while the MSE and MSPE values are often larger. Larger MSE and MSPE estimates for E2 models are appropriate in the sense that there is more variance present in these models, thus estimation should be more difficult. Furthermore, MSE and MSPE values are nearly identical for the majority of models. In some cases, though, these estimates are quite extreme, such as for scenarios E2PS8F8 and E2PS9F9. Scenario E2PS4F4 demonstrates that outlier values in MSE are not necessarily reflected in the MSPE. Additionally, the estimates associated with A1 are typically smaller than A2 or A3. I believe this is due to the fact that this region has smaller, closer together counties. Finally, the results for the E1PS1F1RE and E1PS1F1CV model indicate that it performs very well; in fact, the most of the MSE and MSPE values are smaller than those produced with E1PS1F1.

Table 10: BMS GoF measures for partial models.

Model	$\bar{D}$	DIC	$\frac{Var(\bar{D})}{2}$	MSE_{A1}	MSE_{A2}	MSE_{A3}	$MSPE_{A1}$	$MSPE_{A2}$	$MSPE_{A3}$
E1PS1F1	574.08	652.70	78.62	1.33	1.76	1.45	1.38	1.78	1.49
E2PS1F1	489.81	560.66	70.85	1.38	1.51	1.45	1.44	1.57	1.51
E1PS2F2	562.72	644.58	81.86	1.10	1.37	1.19	1.18	1.44	1.26
E2PS2F2	481.59	551.13	69.54	1.36	1.46	1.44	1.42	1.53	1.51
E1PS3F3	561.03	644.91	83.87	1.17	1.23	1.17	1.25	1.30	1.24
E2PS3F3	478.93	549.57	70.64	1.44	1.45	1.44	1.50	1.51	1.50
E1PS4F4	561.33	641.73	80.39	1.25	1.32	1.26	1.32	1.40	1.32
E2PS4F4	480.43	548.68	68.25	6.93	5.76	7.08	1.30	1.26	1.41
E1PS5F5	564.12	639.13	75.01	1.35	1.52	1.43	1.39	1.59	1.48
E2PS5F5	481.46	540.73	59.27	1.55	1.65	2.12	1.60	1.69	2.16
E1PS6F6	567.53	646.31	78.78	1.22	1.30	1.21	1.27	1.37	1.25
E2PS6F6	487.98	550.77	62.79	1.26	1.24	1.48	1.33	1.29	1.55
E1PS7F7	546.88	623.93	77.04	1.04	1.65	0.92	1.12	1.73	0.99
E2PS7F7	467.59	526.52	58.92	1.03	1.80	1.15	1.08	1.85	1.20
E1PS8F8	561.90	642.71	80.81	1.25	1.20	1.19	1.33	1.27	1.26
E2PS8F8	476.94	544.74	67.80	1.13	5.18	3.56	1.17	5.35	3.56
E1PS9F9	591.67	676.12	84.45	1.23	1.36	1.53	1.31	1.43	1.61
E2PS9F9	498.65	568.11	69.45	1.73	5.46	3.93	1.73	5.46	3.93
E1PS1F1RE	579.21	678.05	98.84	1.04	1.29	1.16	1.10	1.35	1.23
E1PS1F1CV	573.74	662.31	88.57	1.22	2.01	1.31	1.29	2.14	1.37

The sets of maps included in Figure 8 are associated with E1PS4F4 and E1PS8F8 when fitted using the BMA technique. These maps are for comparison to the BMS fitted maps shown in Figure 6. In comparison, I see that both fits of these models fail to be as accurate as those attained using the BMS technique. There are improvements to be noted when moving from S4 to S8 as it seems that the model probabilities are becoming slightly closer to the truth, but there is still some misrepresentation present in the estimation of p_3 since I do not expect Alt3 to be selected for any counties. Furthermore, the maps produced using the model probabilities calculated via local DIC do not produce reasonable results.

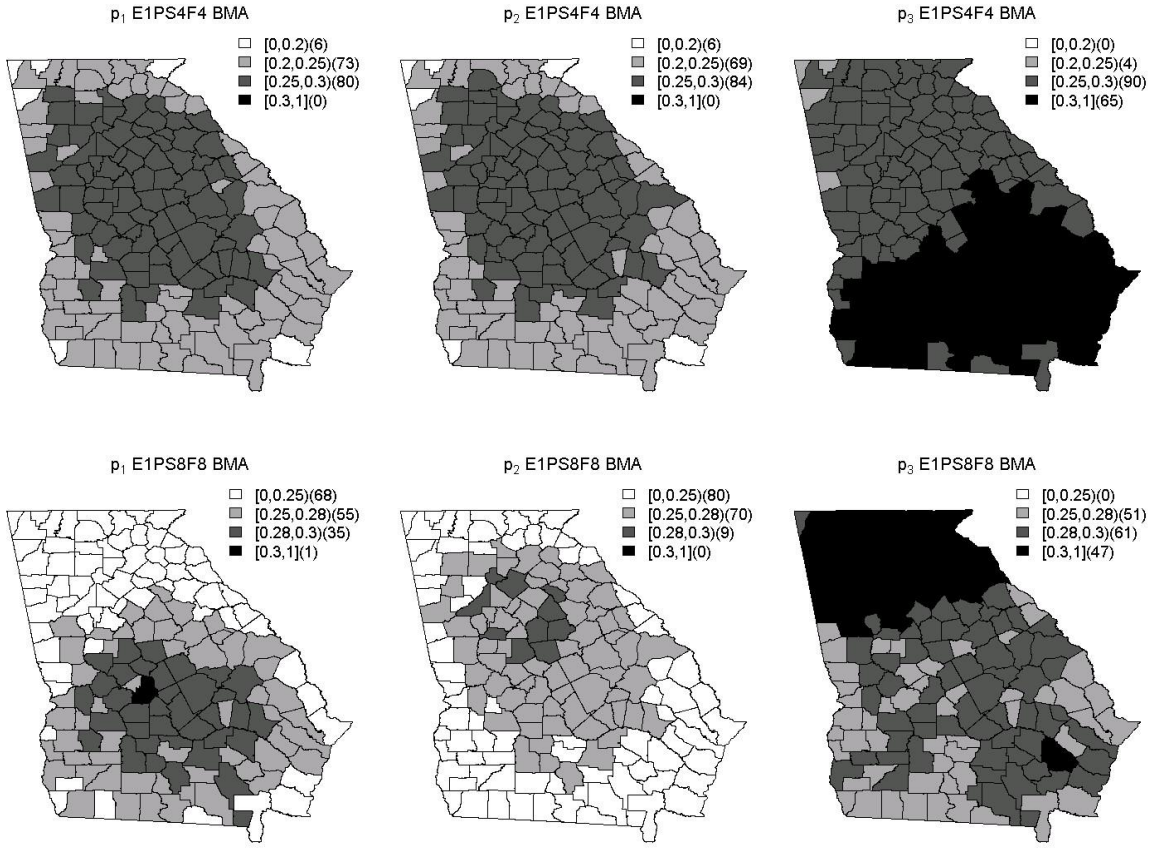


Figure 8: Model probabilities associated with E1PS4F4/S8F8 for BMA.

As before, the next set of maps displayed in Figure 9 illustrate the fits associated with E1S6F6 and E1S9F9 when using the BMA technique. Here, I see no improvement when comparing S6 to S9, and none of the maps tend to suggest a certain region relating to one linear predictor versus another. The BMS technique continues to handle this situation better than BMA. Additionally, the DIC probability maps still fail to supply reasonable results.

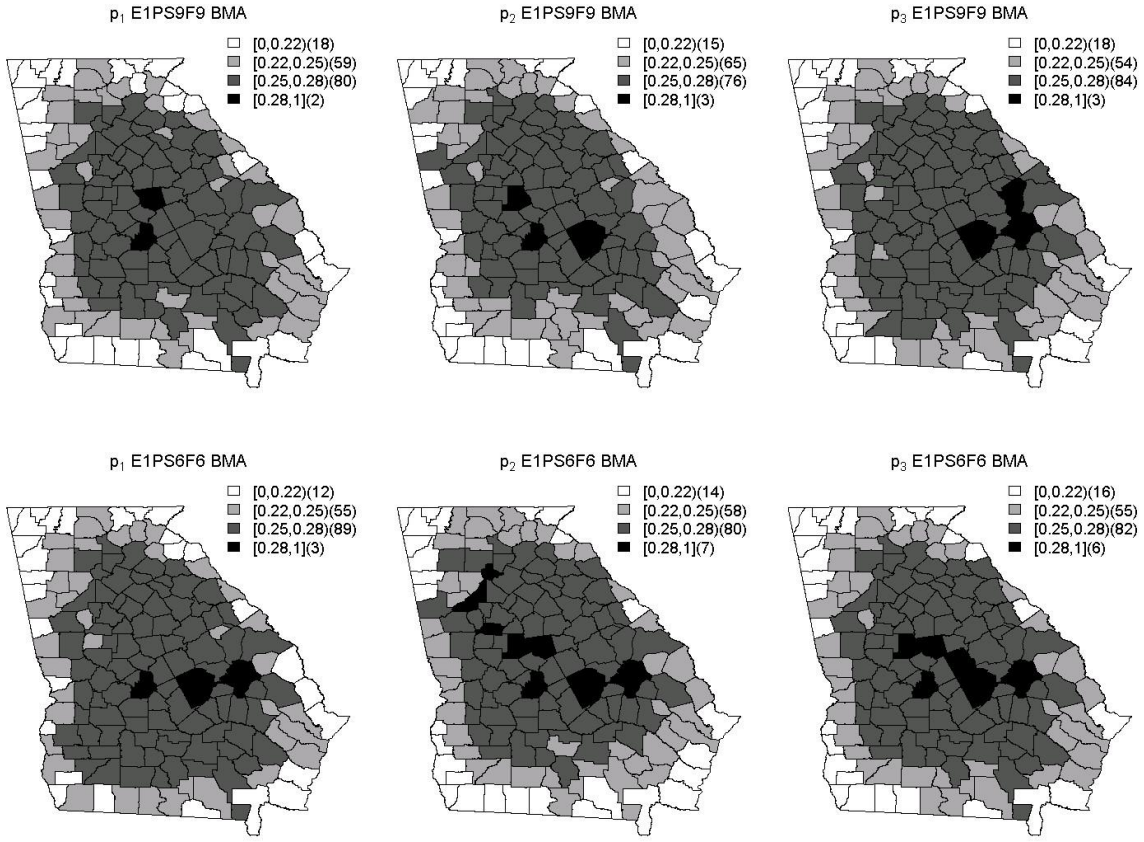


Figure 9: Model probabilities associated with E1PS6F6/S9F9 for BMA.

Figure 10 illustrates a comparison of both model selection methods' abilities to accurately recover the truth. This comparison is accomplished by associating the probability from a specific county that should be selected to the mean probability of all counties that should not be selected for that particular model; this is described with the following formulation:

$$\pi_M = \frac{\sum_{i=1}^{n_A} I_{p_{ki} \in A}\left(p_{ki} > \overline{p_{kA^C}}\right)}{n_A}$$

where π_M are the probabilities displayed in Figure 9, I is an indicator function that is 1 when the enclosed logical function is true and zero otherwise, $k = 1,2$ for the model probabilities p_1 and p_2, A is the area for which p_k should be highest, A^C is area A compliment, $\overline{p_{kA^C}}$ is the mean of the model

probabilities in area A^C, and n_A is the number of counties in area A thus $i = 1,\ldots,n_A$. So, for example, with p_1 from PS1F1, a county in A2 or A3 should have a higher model probability estimate than the mean of model probabilities calculated for the counties in A1. In these plots, I see BMS preforming the same or better than BMA for 21 out of the possible 36 model combinations.

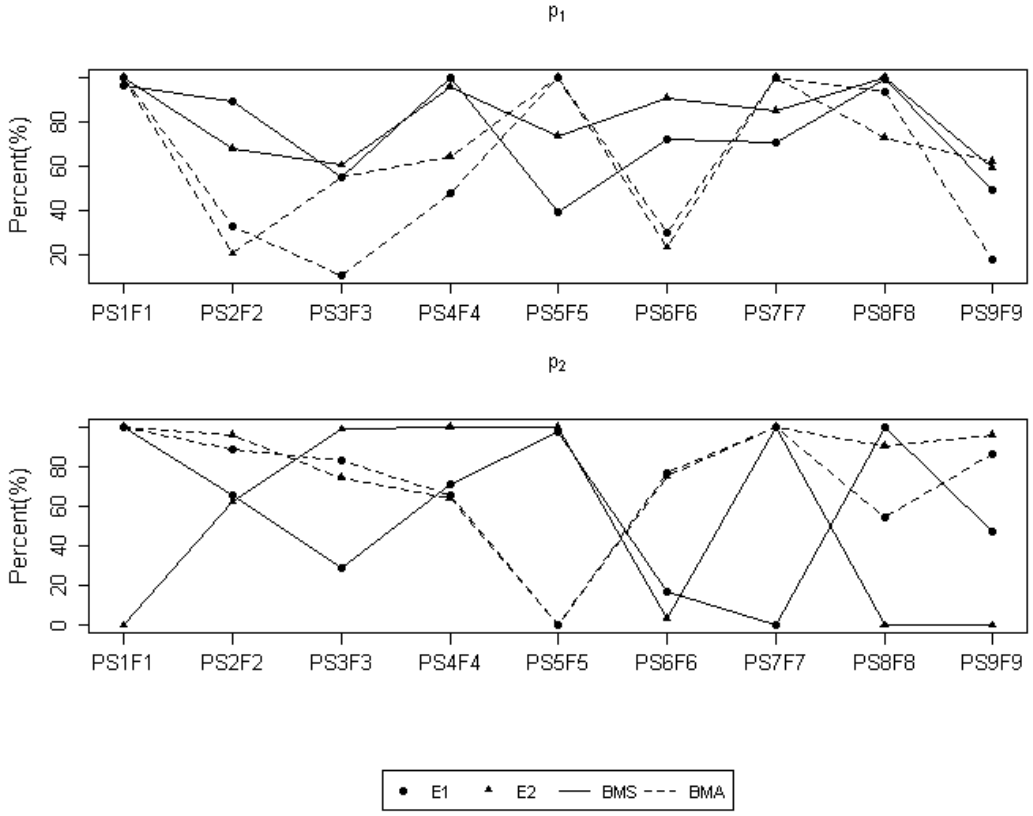


Figure 10: Representation of the models' ability to recover the truth.

Table 11: GoF measures for BMA model fits.

	$\bar{D}$	DIC	$\frac{Var(\bar{D})}{2}$	DIC_{loc}	DIC_{A1}	DIC_{A2}	DIC_{A3}	MSE_{A1}	MSE_{A2}	MSE_{A3}
E1PS1F1	2455.38	2476.58	21.20	3071.80	1003.99	1045.74	1022.07	0.97	1.50	1.18
E2PS1F1	2097.75	2117.27	19.52	2134.62	659.53	735.26	739.83	4.84	5.01	4.46
E1PS2F2	2298.06	2487.52	189.45	2768.65	909.81	938.10	920.75	1.13	1.52	1.27
E2PS2F2	1963.62	2117.66	154.04	1882.42	587.75	630.22	664.45	5.04	4.66	3.98
E1PS3F3	2287.83	2460.36	172.52	2766.46	924.12	923.80	918.54	1.24	1.26	1.17
E2PS3F3	1954.24	2088.57	134.32	1859.04	592.31	614.52	652.21	5.62	5.59	4.24
E1PS4F3	2282.17	2451.06	168.88	2762.14	919.01	919.46	923.68	1.06	1.10	1.11
E2PS4F3	1957.81	2086.93	129.13	1871.96	592.61	622.52	656.83	4.70	3.55	4.16
E1PS5F5	2406.87	2424.65	17.78	3010.12	1009.08	995.17	1005.88	1.06	1.18	1.14
E2PS5F5	2051.09	2067.67	16.59	2045.27	660.59	676.37	708.31	4.96	4.01	5.49
E1PS6F6	2394.56	2402.73	8.17	2976.58	990.09	985.44	1001.04	0.96	0.93	0.98
E2PS6F6	2049.45	2057.78	8.33	2048.56	646.80	681.73	720.04	4.84	3.53	4.39
E1PS7F7	2445.02	2481.95	35.93	3135.98	1000.75	1157.03	978.19	1.05	2.59	1.23
E2PS7F7	2093.38	2125.59	32.21	2250.22	656.83	816.61	776.79	3.68	7.89	3.98
E1PS8F8	2291.10	2465.30	174.20	2773.70	924.77	919.42	929.51	1.46	1.39	1.43
E2PS8F8	1949.31	2079.34	133.03	1859.86	601.57	597.40	660.88	4.32	4.61	5.25
E1PS9F9	2513.67	2525.15	11.48	3137.71	1043.27	1005.20	1089.23	1.08	2.88	1.46
E2PS9F9	2126.08	2137.42	11.34	2150.69	692.09	689.33	769.27	4.11	4.58	6.30
E1PS1F1RE	2240.84	2602.07	361.23	2544.09	426.93	438.86	434.01	1.10	1.71	1.34
E1PS1F1CV	2240.70	2585.74	345.03	2506.28	411.33	395.44	418.21	1.84	3.06	1.85

Table 11 displays the GoF measures for the partial models using the BMA technique. The MSPE measures can be viewed in Supplemental Table 7. One of these properties is that, in general, the E2 models produce smaller DIC values but higher MSE and MSPE values for the majority of models. The difference between the MSE and MSPE values when comparing E1 and E2 for the BMA case is much larger than what I saw when using BMS. This again indicated that model selection is more difficult when there is more variation in the data. Another of these properties involves lower MSE and MSPE values for area A1; additionally, this is typically reflected in the local DIC measure for A1 as well. One new comparison that I can make with the BMA results is the local DIC measure to the BUGS calculated DIC measure, and these are different for every model. Additionally, the local DIC measure is always higher than the other DIC measure for the E1 models while this is not typically the case for the E2 models. I believe that all of these properties combined suggest that they perform better when there is no variation in the expected rates. Finally, when comparing E1PS1F1RE and E1PS1F1CV to E1PS1F1 for the BMA models, I see that the results are still good but not better as I saw with the BMs technique, except for the local DIC measurements.

3.4.3. Misspecified Models

In my research, I also misspecified models such that I fit the complete simulated data sets with the partial method and vice versa. In particular, I did this with model E1CS1 and E1PS1 because they performed well initially in both the BMS and BMA methodology. These appropriately specified results can be seen in Supplemental Tables 2 and 4 for the complete models as well as Supplemental Figures 1 and 2 for partial models.

The results from fitting partial simulated data (E1PS1) using the appropriate complete model (CF1) are in Table 12. They show the BMS method choosing the linear predictor associated with p_1 while BMA selected the linear predictor associated with p_3. The true linear predictors here could be p_1 or p_2 depending on the county. I suspect that the BMA method incorrectly selects p_3 because the first two linear predictors are alternating as the true model for the different counties. For example, if $p_3 = 0.4$ for all

counties while p_1 and p_2 alternate between the values 0.1 and 0.5 depending on the county of interest, I could get similar results. The GoF measures are quite different for these two methods, but this is similar to the previous results in Section 3.4.1.

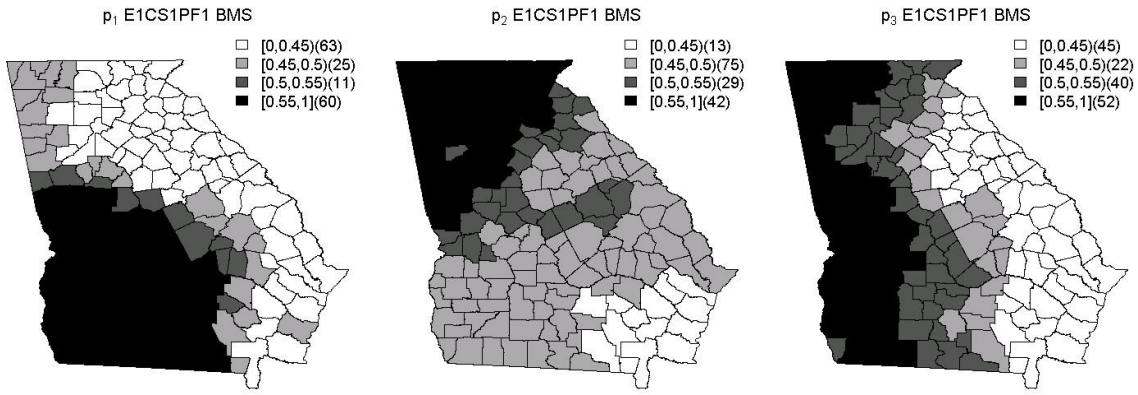
Table 12: Model weights, probabilities, and GoF measures for the misspecified models.

	p_1	p_2	p_3	$\bar{D}$	DIC	$\frac{Var(\bar{D})}{2}$
E1PS1CF1 BMS	0.551	0.333	0.470	609.15	614.43	5.27
E1PS1CF1 BMA	0.283	0.283	0.419	2315.38	2483.11	167.73
E1CS1PF1 BMS	0.498*	0.500*	0.497*	575.54	674.23	98.69
E1CS1PF1 BMA	0.256*	0.248*	0.253*	2437.01	2459.70	22.70

*These values are means of the weights and probabilities that are actually varying across counties.

When fitting the complete simulated data (E1CS1) to the appropriate partial fitted model (PF1), I expect to see the linear predictor associated with p_1 selected for all counties, and the county maps for this misspecification are displayed in Figure 11. Both methods' results still show some variability among the different counties, but the BMA method clearly selects the linear predictor associated with p_2 for the northern counties. The GoF measures as well as mean probabilities and weights are shown in Table 12.

.



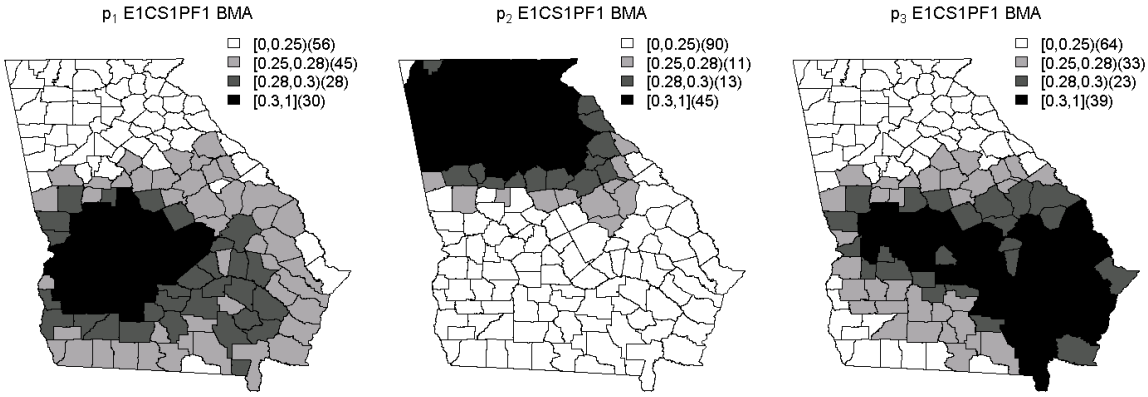


Figure 11: Model weights and probabilities associated with the misspecified models fit with E1CMSPF1.

3.5. Colon Cancer Data Example

3.5.1. Introduction

Colon cancer (ICD-9-CM code: 153), accompanied by rectum cancer (ICD-9-CM code: 154.1), is ranked as the third most common tumor type in the United States, with colon cancer being the more frequent of the two. Routine screening for this cancer, particularly after the age of 50, is encouraged since a good prognosis typically accompanies an early diagnosis. Important risk factors of colon cancer include: nutritional inclinations, age, smoking status, inflammatory bowel diseases, previous incidence of malignant disease, and some genetic traits.[41-43] Research examining the geography of some of these risk factors suggests that there may be an underlying spatial structure to the incidence of colon cancer.[44, 45]

The data of interest in this study is the 2003 colon cancer incidence for the 159 counties in the state of Georgia, USA. The Area Health Resource Files (AHRF)[46] data set provides ecological predictors useful for explaining the variation in this outcome. The chosen predictors are as follows: median household income (in thousands of dollars), percent persons below poverty level (PPBPL), unemployment rate of those aged 16 or greater (UER), and percent African American (AA) population. Other studies indicate that poverty and race are associated with colon cancer incidence.[41, 47] Of the chosen variables, there is evidence to

suggest that median income and PPBPL may be correlated (section 3.5.2). This same evidence could also be an indicator of the underlying spatial effect that I believe may play a role in colon cancer incidence. The age cut off associated with the unemployment variable may be criticized since much of the younger population in this age range may not hold steady jobs as they are full time students. In the individual level data used to create this county level variable, 'student' is an option as an employment status.

3.5.2. Data and Linear Predictor Alternatives

The colon cancer data comes from the online analytical statistical information system (Oasis) of the Georgia Department of Public Health. For the 1332 diagnosed colon cancers across the state in the year 2003, there was approximately a mean incidence of 8.38 cases per county where the minimum county level value was zero and the maximum value was 102. In these data, there are no missing values at the county level.

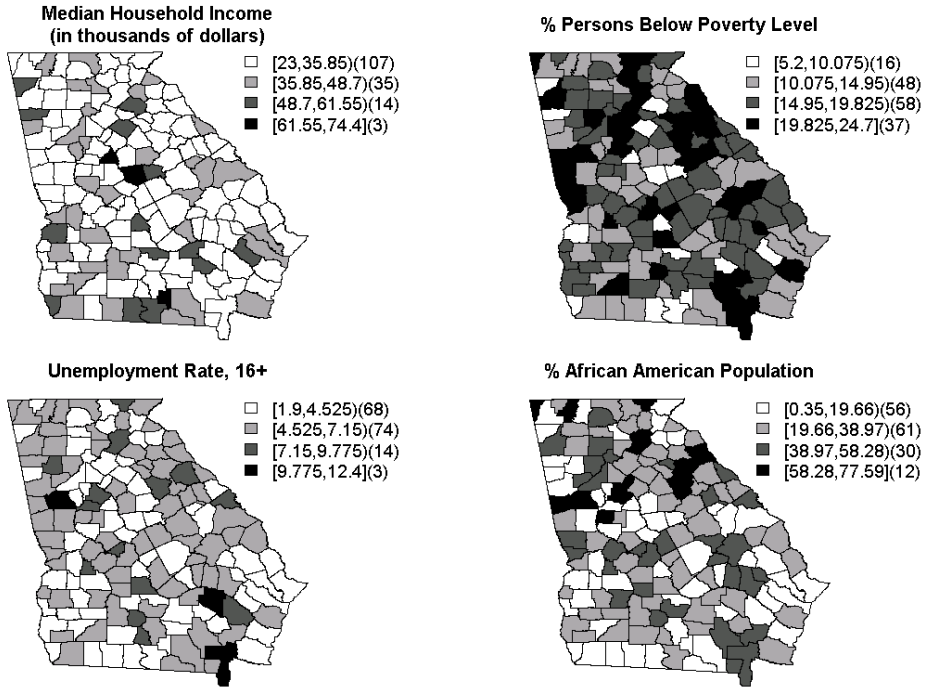


Figure 12: Geographical distribution of predictors from the AHRF data set.

The geographical distributions of the chosen predictors are displayed in Figure 12 and suggests some spatial clustering. An additional indicator of the underlying spatial structure is made evident by the pattern of standardized incidence ratios (SIR) displayed in Figure 13. The SIR is calculated as the ratio of the observed colon cancer incidences to the expected rates for each of the 159 counties and can be useful as a first step in data analysis.[48] Qualitatively, for these data, there does appear some spatial structure.

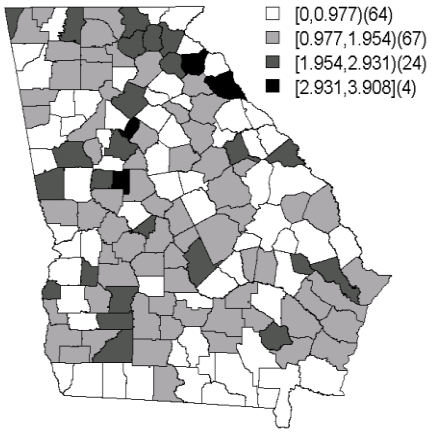


Figure 13: Map of the Standardized Incidence Ratio for the 2003 colon cancer data.

Based on the chosen predictors (median income in thousands - x_1, PPBPL - x_2, UER - x_3, and percent AA population - x_4), I have employed three possible linear predictors for use with both the BMS and BMA methods. Table 13 displays these alternative predictor options. The first linear predictor (Alt1) includes all of the covariates. The second (Alt2) includes only income and percent AA population. The third and final linear predictor (Alt3) includes PPBPL and percent AA population. Note that all of my possible linear predictors contain an uncorrelated random effect to aid in accounting for any uncontrolled for parameters or extra noise present in the data, and they differ by the predictors included. Additionally, for all of these linear predictor alternatives, the prior distributions are such that:

$$u_{id} \sim Norm(0, \tau_u),\ \tau_u \sim Gam(1, 0.5), \text{ and } \alpha_{jd} \sim Norm(0,1)$$

where $i = 1,\ldots,159$, $d = 1,\ldots,D$ such that D is the number of linear predictors to be selected between, and $j = 0,\ldots,J$ such that J is the number of predictors for the d^{th} model.

Table 13: Alternative linear predictor contents.

Model	Contents
Alt1	$\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + u_i$
Alt2	$\alpha_0 + \alpha_1 x_1 + \alpha_4 x_4 + u_i$
Alt3	$\alpha_0 + \alpha_2 x_2 + \alpha_4 x_4 + u_i$

I alternate income and PPBPL in the second two linear predictors because there is evidence to suggest that they may be correlated. This is not an uncommon assumption as, typically, when income is higher, poverty is lower, as shown in Table 14. This table illustrates, through individual Poisson model fits, that median income and PPBPL are collinear with respect to the incidence of colon cancer outcome because PPBPL becomes well estimated when median income is removed from the model. I also note some changes in percent AA population when PPBPL is used in place of median income. These individual model fits were performed using Bayesian approximation techniques by way of the `R` package `INLA`.[5, 24]

Table 14: Individual fits of possible linear predictors. Posterior mean and standard deviations are displayed.

	Alt1 Mean (SD)	Alt2 Mean (SD)	Alt 3 Mean (SD)
Intercept	1.49 (0.08)*	1.49 (0.08)*	1.48 (0.09)*
Median income (x_1) (in thousands)	0.69 (0.19)*	0.65 (0.09)*	---
PPBPL (x_2)	0.10 (0.23)	---	-0.69 (0.11)*
UER (x_3)	-0.14 (0.10)	---	---
% AA population (x_4)	0.28 (0.13)*	0.26 (0.09)*	0.41 (0.12)*

*Indicates that a predictor is well estimated or 'significant' and SD stands for standard deviation

In addition to collinearity, the changes seen in the parameter estimates could also indicate that some of these predictors may be more important in certain regions of the county map. This indication will be further explored with the application of the BMS and BMA techniques. The covariates were standardized prior to fitting the models.

3.5.3. Results

The results below illustrate the application of BMS and BMA as described above. Following the application of the model selection techniques, I re-fit the selected linear predictors to the appropriate counties.

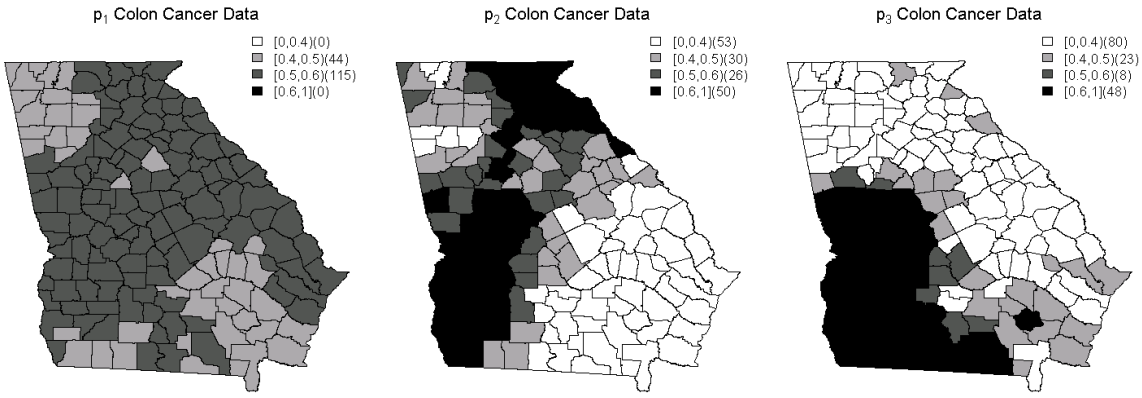


Figure 14: County-specific model selection probabilities corresponding to Alt1, Alt2 and Alt3, based on the BMS procedure.

The results from fitting these models with the real data using the BMS method are displayed in Figure 14 and suggest that it may be beneficial to use the second alternative linear predictor option in the Northern and Western counties of the state as the weights produced for those counties are fairly large. By the same guidelines, the third alternative linear predictor option may be optimal for the Southern and Western counties. Given the fact that there is some overlap in these results, this indicates that either predictor could be appropriate for these counties. Thus, there is not a clear best alternative linear predictor for those counties. From these results, I also see that it is beneficial to place correlated covariates in separate linear predictors and allow the BMS process to determine which is most appropriate across the county map. Additionally, I note that the distribution of the county weights across the county map for p_1 and p_2 are very similar to each other as were the parameter estimates, α_{j1} and α_{j2}, associated with the initial individual model fits shown in Table 12, Section 3.2.1. This indicates that median income may have a more

palpable relationship with colon cancer incidence compared to PPBPL. Furthermore, these results do suggest that there is a spatial relationship between these predictors and colon cancer incidence.

This application of the BMA method produced different results from those obtained using the BMS method. I use the 0.3 cut off to determine model significance in this instance because in a situation where there is no most appropriate model, the model probabilities would equal 0.3. This value is determined by summing across the three county weights produced by simulation, dividing by 3, and taking the mean of the summed values across all counties. Thus, the BMA model probability results shown in Figure 15 suggest that the second alternative linear predictor option should be used for the Northern counties while the third linear predictor option seems appropriate for the majority of the county map, particularly the Southern and Western counties. Additionally, the first alternative linear predictor appears important in some of the central counties. Again, I see that there is some overlap in the appropriate linear predictors, but in the case of BMA, the differences are trivial for the majority of counties. This suggests that the BMA method designates all of the linear predictors as interchangeable across the county map. I also continue to see some similarities between the distributions produced for p_1 and p_2. These similarities are not as distinct as they were in the BMS method results, however. Additionally, the results below do still suggest that there is a spatial relationship present for colon cancer incidence.

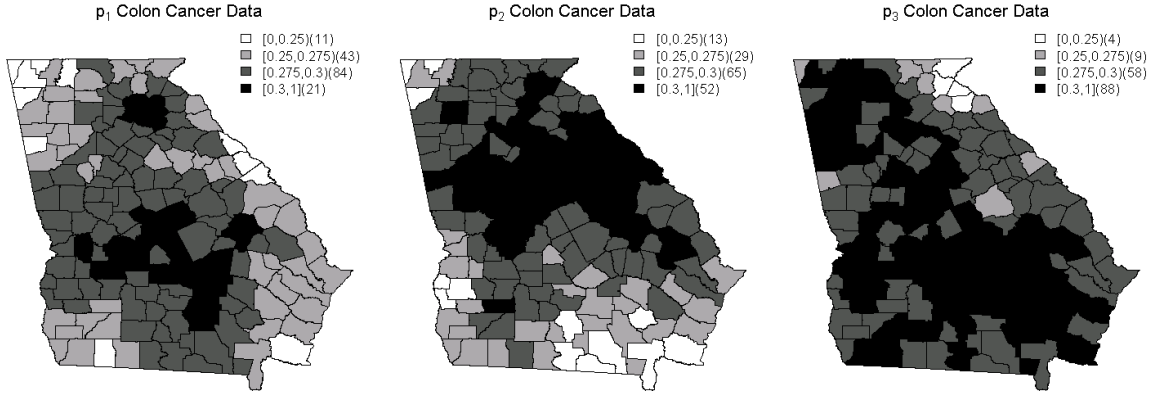


Figure 15: County probabilities based on the BMA procedure.

Table 15 illustrates individual model re-fits for the appropriate counties with the selected linear predictors based on both the BMS and BMA results. From the results, first I chose to fit the 53 most Northern counties with the second linear predictor and the remaining Southern counties with the third alternative linear predictor. In comparison to the initial individual model fits in Table 12 in Section 3.2.1, in Table 13 I observe increases in the magnitudes of the parameter estimates for Alt2 while Alt3 stays roughly the same.

Next, I consider applying the linear predictors based exactly on the BMS results. Here, if BMS produced a weight greater than 0.5, that county was included in the model re-fit for the associated alternative linear predictor. This is not a strict cutoff for the BMS method, it is simply what seemed appropriate for these data as this is the value that all weight would acquire if there were no true model. This value is determined in the same way as the BMA 0.3 cutoff value described above. The appropriate counties selected for Alt2 are considered to be in area A2 while the area for the counties selected with Alt3 is named A3. Based on these definitions and in comparison to the initial model fits (Table 14), the estimates associated with Alt2 and Alt3 decrease in magnitude as well as value. Here, for both linear predictors, the parameter estimates associated with percent AA population are no longer well estimated.

Table 15: Selected linear predictor re-fits.

Counties included (number)	North (53) Mean (SD)	South (106) Mean (SD)	A2 (50) Mean (SD)	A3 (48) Mean (SD)
Linear Predictor	Alt2	Alt3	Alt2	Alt3
Intercept	1.50 (0.15)*	1.49 (0.10)*	1.62 (0.13)*	1.55 (0.15)*
Median Income (in thousands)	0.87 (0.17)*	---	0.66 (0.17)*	---
PPBPL	---	-0.63 (0.13)*	---	-0.37 (0.21)*
% AA population	0.28 (0.15)*	0.36 (0.14)*	0.02 (0.14)	-0.06 (0.21)

*Indicates that a predictor is well estimated or 'significant,' SD stands for standard deviation, and the definitions of A2 and A3 are as follows: The appropriate counties selected for Alt2 are considered to be in area A2 while the area for the counties selected with Alt3 is named A3.

3.5.4. Discussion

The results above present evidence suggesting that there is a spatial structure in the distribution of colon cancer incidence. These results also show that the model selection techniques are useful in determination of the appropriate linear predictors for different areas of the county map. Additionally, there are many reasons why some of the selected linear predictors may not perform as well as expected. These include: 1) the larger size and more separation among the Southern counties; 2) the limited number of counties selected when restricting the included counties to those with a weight of 0.5 or more; and 3) the strength of association in the data: in general, these predictors may not have a very strong relationship with colon cancer incidence. Thus, they are difficult to first select and then fit to produce well estimated parameter estimates.

Based on the results above, both techniques suggest that median income and percent AA population are useful in predicting colon cancer incidence in the Northern counties of Georgia. Alternatively, these results also suggest that PPBPL and percent AA population are useful in predicting incidence of colon cancer in the Southern counties of the state of Georgia. After applying the appropriate transformations defined in Section 2.2.1, the explicit interpretations of the parameter estimates from the individual model re-fits are as follows. For the Northern counties, every $1000 increase in median income indicates that 1.09 times as many incidences of colon cancer occur, and each 1% increase in AA population indicates that 1.02 times as many incidences of colon cancer occur. For the Southern counties, every one unit increase in PPBPL indicates 0.87 times as many incidences of colon cancer occur, and every one percent increase in percent AA population indicates that 1.02 times as many incidences of colon cancer occur. These estimates are not very large in magnitude because they are incremental, continuous increases per unit of the parameter of interest; they are also displayed in Table 16.

Table 16: Transformed mean parameter estimates for linear predictor re-fits.

Counties included (number)	North (53)	South (106)	Predictor Standard Deviation
Linear Predictor	Alt2	Alt3	---
Median Income (in thousands)	$\exp(0.87/9.75)=1.09$	---	9.75
PPBPL	---	$\exp(-0.63/4.58)=0.87$	4.58
% AA population	$\exp(0.28/17.47)=1.02$	$\exp(0.36/17.47)=1.02$	17.47

For BMS, I receive very clear indications that the linear predictors which include either PPBPL or median income are preferred over the linear predictor that includes both. I also see that these two linear predictors are clustered in specific areas, with some overlap, of the county map. Once I look into re-fitting the selected models, I see that the second linear predictor seems to have a slightly stronger relationship with colon cancer incidence than the third in their appropriately selected counties. Furthermore, A1 and A2 produce even less substantial results. These losses in substantiality may be due to the limited number of counties selected or the smaller size and closer proximity to each other involved with the counties in the Northern part of the state, or both.

Some of my simulation studies shown in Section 3.3.4 have suggested that the BMA technique does not perform as well in detecting smaller levels of association in the data. The results here also suggest that this may be an issue for these data. My predictors are not among the most important risk factors mentioned previously, thus they are not considered of high association with incidence of colon cancer. BMA does not appear to choose one linear predictor clearly over the other, though it does seem to have a slight preference of the second two alternatives in comparison to the first. In general, BMA will not produce probabilities of the same magnitude as the weights produced by BMS because they are scaled such that they add to 1. This leads to slightly more interpretable results produced with BMA. Regarding measures of GoF measures, because in BMA each of the linear predictors are fit individually then averaged to create the averaged posterior, this method offers the ability to examine how well each of the alternative linear predictors perform, while BMS does not.

Both BMS and BMA techniques illustrate the importance of keeping collinear variables in separate alternative linear predictors. In both sets of results, I see that the first alternative linear predictor, which contains two collinear predictors, is somewhat under-stimulated in comparison to the other two alternatives. This is most distinguished in the BMS results. Additionally, the results given here indicate the importance of using these spatial models for incidence of colon cancer. I received meaningful, important results when performing the initial individual, non-spatial model fits, but I gain even more information by allowing the model selection techniques to determine the appropriate linear predictor for each individual county. Furthermore, the model re-fits illustrate how these techniques perform better when the regions of interest are smaller and closer together. This issue has been noted in my simulation studies as well. Both model selection techniques suggested that Alt2 was a good choice for the Northern counties, and when I re-fit that linear predictor for only those Northern counties, I saw the parameter estimates become even more substantial. For the A1 and A2 counties, however, this was not as clear even though I still gain some meaningful results for these counties.

3.5.5. Conclusion

Based on this exploration of the spatial structure of colon cancer incidence, my findings that there is much information to gain by employing spatial model selection techniques to determine the appropriate linear predictor that best explains the variation in the data. Through the application of these techniques, I determined the important predictors for the different areas of the county map, and these indicate that median income and percent AA population are important predictors of colon cancer incidence in the Northern counties of the state while PPBPL and percent AA population are important for the Southern counties of the state. By employing these two methods in combination, I was able to detect some interesting and important aspects of these data.

3.6. Discussion

Based on the simulation results and from a qualitative assessment, I believe that the BMS technique outperforms BMA in terms of selecting the appropriate linear predictors. The results for the BMS method

are also more consistent when comparing different scenarios such as E1 versus E2, and S1F1-S9F9 to their appropriate counterparts. Furthermore, I discovered that the BMS technique is more robust to misspecified models. For both techniques, the complete models tend to recover the truth more efficiently and accurately than the partial models, but this is to be expected as these models are not as complex.

I also see significant improvements in recovering the appropriate estimates when comparing the models whose data sets have true parameter estimates with larger magnitudes, meaning there is more evidentiary support in the data. This is evident in complete, partial, E1, and E2 models as well as both BMS and BMA techniques. As far as GoF measures are concerned, DICs cannot be compared across models due to different outcome variables, and thus different likelihoods. The MSE and MSPE estimates suggest that, for the most part, the E1 models fit better than the E2 models. I also note that the A1 region often offers the lowest MSE and MSPE measures. I believe that this occurs because the counties in the A1 region are smaller and closer together than the others. Additionally, more often than not, S7F7-S9F9 models produce lower MSE and MSPE values. This suggests that the techniques perform better when there is more support in the data.

There are several obstacles that can be encountered when preforming both model selection methods. The first of these involves the strength of association present in the data. If there is not enough evidence in the data, I have seen that both BMS and BMA fail to perform well.

Another issue involves extra noise in the data. As in many statistical applications, extra variation in the data can lead to difficulties in estimation; BMS and BMA are not immune to this issue. By the same token, including an uncorrelated random effect in one of the alternative linear predictors when it is not truly needed can also result in an improper selection of a linear predictor. In many cases, though, there is random noise present in data, and including that random effect can be helpful. This is why random effects were included in all linear predictor alternatives for my real data example. Furthermore, I did include two scenarios in my simulation study (E1PS1F1RE and E1PS1F1CV) that introduced this extra noise and fit the

models such that it was reflected in the true and alternative linear predictors. In this situation, the results suggest that both methods seem to perform comparably when extra variation was imposed upon the data. Here, the BMS method actually produced slightly smaller MSE and MSPE values than those produced with E1PS1F1.

One shortcoming of both of these methods is that they must be performed in sequence with an additional model fit to determine the parameter estimates associated with the selected linear predictor or predictors in the case that one predictor is more appropriate in a certain region of the state. This is a shortcoming that adds to the complexity of the model fitting process, but it is worth pursuing to obtain the most appropriate results.

3.7. Conclusion

From this comparison between my proposed BMS and BMA, I conclude that the BMS application qualitatively produces more accurate as well as more precise results than those produced by BMA in terms of selecting the appropriate linear predictors across maps. There still may be some instances, though, where BMA is the preferred method because one is able to calculate the local DICs in that situation.

4. Aim 3: Spatio-temporal Model Selection

4.1. Introduction

It has long been established that model selection methods are useful in determining the most appropriate linear predictor for a set of data. Model selection techniques are often used in lieu of variable selection as they alleviate some of the issues related to variable selection (e.g. co-linearity, an excess of parameters).[29, 32, 34, 49-51] The results presented in Chapter 3 have shown that model selection methods can be helpful for spatial data. However, researchers often have access to data across time as well as space. These model selection methods allow for different models to be employed at each temporal unit, spatial unit, or both, depending on how the model is specified.[52, 53]

One challenge indicated by previous studies involving spatio-temporal variable selection, includes having more parameters than the possible number of MCMC iterations.[54] My model selection methods minimize this issue as it is not necessary to apply a parameter to each individual predictor when determining its inclusion probability. Rather, the inclusion probability is applied to the linear predictors to determine their appropriateness as a whole.

A common problem with large scale MCMC simulation is computation time. Computationally fast spatio-temporal analyses can be performed using approximate Bayesian inference via the `R` package `INLA`,[5, 7, 10, 20, 21, 24, 55], however this package does not offer the degree of flexibility required for my model selection methods, as discussed in Chapter 2. Alternatively, to improve the computation time in this instance, the `R` package `snowfall`, which allows for parallelizing code, has been implemented.[56]

The aim of this chapter revolves around a spatio-temporal extension to my previously developed model selection methods from Chapter 3 via the BUGS software `BRugs` which calls `OpenBUGS` from the statistical software program, `R`.[13, 28] Here, I attempt fitting my model selection methods to ten different simulated model scenarios and assess them for GoF. I also apply one additional fitted non-model selection

method to determine how well I am able to recover the simulated random effects. I also include a special case of the model selection methods described as a mixture model. My objective is to determine which fitted model selection option performs best depending on the type of data applied to the model using a variety of simulated data sets.

This chapter is developed as follows. I first describe the different simulated data models. Next, I define the specific types of space-time fitted models I wish to explore as well as my methods for comparing those models. Finally, I present, discuss, and draw conclusions relating to the results.

4.2. Methods

This chapter focuses on the context of disease mapping in *m* predefined small areas. For all models defined below, both simulated and fitted, I conventionally assume that the outcome, an aggregated count of disease (y_{ij}), is observed in the i^{th} small area at time j and that these counts are conditionally independent Poisson distributed outcomes. In addition, I assume that the expected count (e_{ij}) is available in each small area at each time. This is a commonly assumed model for small area counts in disease mapping,[17] and is defined as follows:

$$y_{ij} \mid \mu_{ij} \sim Pois(\mu_{ij})$$

$$\mu_{ij} = e_{ij}\theta_{ij}$$

$$\log\left(\theta_{ij}\right) = X_{ij}^{T}\alpha + R_{ij}$$

where the mean of the Poisson model, μ_{ij} , is defined as the expected rate of disease times the risk. Further, X_{ij}^{T} is the ij^{th} element of the design matrix, α is the vector of parameter estimates, and R_{ij} is a linear combination of different types of random effects which are described in more detail in Section 4.2.1 and Table 14. This is a basic definition of the model of interest and will be explained in more detail as well as with more options in the following sections. In what follows I describe my simulated data scenarios and

fitted model alternatives. While these models represent a subset of the spatio-temporal model possibilities for these data, I believe that they are an appropriate representative subset.

4.2.1. Simulated Data

To simulate the data, I fix the expected rate $\left(e_{ij}\right)$ for the areas across the simulated data sets as a single realization of a $Gam(1,1)$ distribution. This allows the primary focus to be on the estimation of risk θ_{ij}. To complete the parameterization, a relative risk which is parameterized with a range of different risk models is assumed. Simulated data scenarios are notated with an 'S' preceding the model number.

The predictors used to simulate the outcome data come from the AHRF[46] data set for the state of Georgia, USA which was presented in Chapter 3. Here, this analysis covers the ten year time frame 1998 to 2007, and this leads to a ten year time frame for my data. The chosen ecological predictors are also the same as used previously: median household income (in thousands of dollars), PPBPL, UER, and percent AA population. These predictors are considered x_1, x_2, x_3, and x_4 respectively; additionally, they are standardized for simulation purposes. Depending on the model of interest, some of these predictors will only vary spatially while others will vary temporally as well as spatially. For the strictly spatially varying covariates, I use a central year of the data, year 2003 or $j = 6$ and assume that this covariate remains the same across the study time frame.

The ten basic models for risk (S1 up to S10) are presented in Table 17. The parameters that make up these models have the same prior distributions for each scenario, but over the 200 simulated data sets, I allow the temporal trend (γ_j), the spatio-temporal interaction term (ψ_{ij}), and the UH term (u_i) to vary from one simulated data set to the next. The respective hierarchical prior distributions of these random effects are as follows: $N(\gamma_{j-1}, \tau_\gamma^{-1})$, $N(0, \tau_\psi^{-1})$, and $N(0, \tau_u^{-1})$. The CH term, $v_i \sim N\left(\frac{1}{n_i}\sum_{i\sim l} v_l, \frac{1}{n_i \tau_v}\right)$, is a conditional autoregressive (CAR) distribution,[4] and unlike the other random effects, it remains constant as one

realization of this distribution so that the spatial correlation is known and the same for all models and simulated data sets. This definition of the correlated effect is such that $i \neq l$, n_i is the number of neighbors for county i, and $i \sim l$ indicates that the two counties i and l are neighbors. The standard deviations associated with these prior distributions, $\tau^{-1/2}$, are all uniformly distributed over the interval [0, 4]. In some ways, this uniform distribution on the standard deviation could be considered a strong assumption, but since I am dealing with a logit model, the interval still leads to a fairly non-informative prior distribution for the standard deviation. Further, other studies have shown that relative risks are not overly influenced by prior distributions placed on the variance parameters.[57] Ultimately, the first two simulated models (S1 and S2) are simply to establish my ability to recover the random effect parameters, as they do not contain any predictors.

Table 17: Simulated model scenarios.

Model	Contents
S1	$\log(\theta_{ij}) = \alpha_0 + u_i + v_i + \gamma_j + \psi_{ij}$
S2	$\log(\theta_{ij}) = \alpha_0 + u_i + v_i + \psi_{ij}$
S3	$\log(\theta_{ij}) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i + \gamma_j$
S4	$\log(\theta_{ij}) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i + v_i + \gamma_j$
S5	$\log(\theta_{ij}) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i}$
S6	$\log(\theta_{ij}) = \alpha_0 + \alpha_3 x_{3ij} + \alpha_4 x_{4ij}$
S7	$\log(\theta_{ij}) = \alpha_0 + \alpha_{3j} x_{3ij} + \alpha_{4j} x_{4ij}$
S8	$\log(\theta_{ij}) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_{3j} x_{3ij} + \alpha_{4j} x_{4ij}$
S9	$\log(\theta_{ij}) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_{3j} x_{3ij} + \alpha_{4j} x_{4ij} + u_i + v_i + \psi_{ij}$
S10	$\log(\theta_{ij}) = \alpha_0 + \alpha_{1j} x_{1i} + \alpha_{2j} x_{2i} + u_i + v_i + \gamma_j$

As far as the parameter estimates are concerned, the α's are fixed across the 200 simulated data sets. For the non-temporal varying model scenarios (S3 to S6), they are set to be $\{0.2,\ -0.3,\ 0.2,\ -0.4\}$ so that they are forced to be fairly large and thus have a detectable relationship with the outcome, yet not so large resulting in simulated outcome counts that tend towards infinity. For the scenarios where the α's vary over

time (S7-S10), a single random walk trend, $\alpha_{k,j} \sim N\left(\alpha_{k,j-1}, 0.5\right)$, is employed for the temporally varying parameters such that the following matrix represents the parameter estimates that are fixed across the simulated data sets:

$$\begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ -0.3 & -0.3 & -0.3 & -0.3 & -0.3 & -0.3 & -0.3 & -0.3 & -0.3 & -0.3 \\ 0.2 & 0.598 & 1.082 & 1.352 & 1.996 & 1.357 & 1.542 & 1.606 & 1.800 & 1.514 \\ -0.4 & -0.616 & 0.421 & -0.383 & -0.738 & -1.376 & -1.413 & -0.889 & -0.763 & -0.010 \end{bmatrix}.$$

This matrix is the fixed set of values for the scenarios that allow for some temporally varying parameter estimates (S7 to S10). Fixing the α's is for simplicity, consistency, and because they are not the main focus of the model selection application to begin with. The results in Chapter 3 suggests that model selection should be performed ultimately to select the best linear predictors, and re-fits of the chosen models should be executed to gain appropriate parameter estimates. Additionally, S10 differs from the other models that only include the spatially varying predictors in that the parameter estimates are allowed to vary over time; in this instance, I simply use the bottom two rows of fixed parameter estimates from the matrix above for simulation purposes. For all simulated model scenarios, the intercept, α_0, is set to be 5. These ten different scenarios furnish a wide range of models, and thus, a good basis to judge which fitted models perform best under the different conditions. Some models contain all of the predictors while others contain only a subset or none at all; I expect certain fitted models to perform better for specific scenarios based on this fact alone.

4.2.2. Fitted Models

The fitted models described below are applied to different combinations of the ten simulated models mentioned above based on their appropriateness. A model is deemed appropriate based on if it can accommodate spatio-temporal covariates. Fitted models are distinguished from the simulated model scenarios by an 'F' preceding the model number rather than an 'S.'

4.2.2.1. The Knorr-Held Model

The first fitted model, F1, is not a model selection method, rather its purpose is to determine how well the spatio-temporal trends and random effects are being recovered. It is known as the Knorr-Held model[58] and is as follows:

$$\log\left(\theta_{ij}\right) = \alpha_0 + u_i + v_i + \gamma_j + \psi_{ij}$$

where the hierarchical prior distributions are such that $\alpha_0 \sim N\left(0, \tau_\alpha^{-1}\right)$, $\tau_\alpha^{-1/2} \sim Unif\left(0,4\right)$, $u_i \sim N\left(0, \tau_u^{-1}\right)$, $\tau_u^{-1/2} \sim Unif\left(0,4\right)$, $v_i \sim N(\bar{u}_{\delta_i}, \tau_v^{-1} / n_{\delta_i})$, $\tau_v^{-1/2} \sim Unif\left(0,4\right)$, $\gamma_j \sim N\left(\gamma_{j-1}, \tau_\gamma^{-1}\right)$, $\tau_\gamma^{-1/2} \sim Unif\left(0,4\right)$, $\psi_{ij} \sim N\left(0, \tau_\psi^{-1}\right)$, and $\tau_\psi^{-1/2} \sim Unif\left(0,4\right)$. These priors will continue to apply for the rest of the fitted models when appropriate. This fitted model will be fitted with simulated data model scenarios S1 to S4, S9, and S10. Additionally, for the models, $i = 1,\ldots,159$ represents the spatial unit, counties, and $j = 1,\ldots,10$ represents the temporal unit, years.

4.2.2.2. Model Selection Methods

The second fitted model that I propose, F2, is the first of the model selection techniques and takes the stance that there are only spatially varying predictors with outcomes that are measured over time. It is described as follows:

$$\log\left(\theta_{ij}\right) = \alpha_0 + \sum_{l=1}^{L} w_{lij}\omega_{lij} + \psi_{ij}$$

where the indicators, w_{lij}, vary over space and time, thus different linear predictors could be chosen for each of the counties as well as years. The Bernoulli probabilities associated with these indicators, p_{lij}, are defined such that $w_{lij} \sim Bern\left(p_{lij}\right)$, $p_{lij} = z_{lij} / \sum_{l=1}^{L} z_{lij}$, $\text{logit}\left(z_{lij}\right) \sim CAR\left(\tau_l^{-1}\right)$, and $\tau_l^{-1/2} \sim Unif\left(0,4\right)$. The model probabilities for the following model selection methods are defined similarly, though some methods do not allow for temporal or spatial variation. Furthermore, the set of L linear predictors are temporally dependent such that the parameter estimates are allowed to vary over time. Therefore, while the covariates

remain the same, their relationship with the outcome has the ability to change over time. Explicitly, the set of alternative linear predictors is as follows:

$$\omega_{ij} = \begin{cases} \alpha_1 x_{1i} + \alpha_2 x_{2i} \\ \alpha_{1j} x_{1i} + \alpha_{2j} x_{2i} \\ \alpha_{1j} x_{1i} + \alpha_{2j} x_{2i} + \gamma_j \\ \alpha_{1j} x_{1i} + \alpha_{2j} x_{2i} + \gamma_j + u_i + v_i \\ \alpha_1 x_{1i} + \alpha_2 x_{2i} + \gamma_j + u_i + v_i \end{cases} .$$

These linear predictor alternatives are specified such that they are true for a selection of the simulated data scenarios defined in Table 7. F2 will be applied to simulated model scenarios S3- S5, and S10.

F3 simply allows separate linear predictors for the spatial and temporal components of the model as well as a random residual term to represent the spatio-temporal interaction. This model is described with the following notation:

$$\log\left(\theta_{ij}\right) = \alpha_{0j} + \sum_{l=1}^{L} w_l^S \omega_{li}^S + \sum_{k=1}^{K} w_k^T \omega_{kij}^T + \psi_{ij}$$

where L and K indicate the number of linear predictors associated with the spatial and temporal model sets, ω_{li}^S and ω_{kij}^T respectively; these are the models to be selected from based on the probabilities associated with indicators w_l^S and w_k^T. Since these probabilities do not vary over space, they have the same distribution and it is described as follows with w_l^S being the example: $w_l^S \sim Bern\left(p_l\right)$, $p_l = z_l / \sum_{l=1}^{L} z_l$, $\log it\left(z_l\right) \sim N\left(0, \tau_l^{-1}\right)$, and $\tau_l^{-1/2} \sim Unif\left(0,4\right)$; this will be the model probability distribution associated with the remainder of the fitted models. This model will be fitted to simulated data models S3 through S10. Lastly, the sets of linear predictors are as follows:

$$\omega_i^S = \begin{cases} \alpha_1 x_{1i} + \alpha_2 x_{2i} \\ \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i \\ \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i + v_i \end{cases}, \quad \omega_{ij}^T = \begin{cases} \alpha_1 x_{3ij} + \alpha_2 x_{4ij} \\ \alpha_{1j} x_{3ij} + \alpha_{2j} x_{4ij} \\ \alpha_{1j} x_{3ij} + \alpha_{2j} x_{4ij} + \gamma_j \end{cases}$$

F4 is an extension of F2 whereby, for example, $\alpha_1 = \left(\alpha_1 + \alpha_{1i} + \alpha_{1j}\right)$ and will be fitted with the same simulated data scenarios as F2. In this expression, there exists a spatial and temporal component to the linear predictors, α_{li} and α_{lj} respectively, that allows the parameter estimates to vary over space and time rather than being fixed or temporally varying as they were in F2. Because of this increase in complexity among the alternative linear predictors, I have limited the indicators, w, such that they no longer have spatio-temporal variation. Thus, the model is now described as follows:

$$\log\left(\theta_{ij}\right) = \alpha_0 + \sum_{l=1}^{L} w_l \omega_{lij} + \psi_{ij}$$

Fitting the predictors in this fashion allows for more flexibility within the alternative linear predictors, and thus, there are fewer linear predictors to be selected from. Explicitly, the linear predictors are defined such that:

$$\omega_{ij} = \begin{cases} x_{1i}'(\alpha_1 + \alpha_{1i} + \alpha_{1j}) + x_{2i}'(\alpha_2 + \alpha_{2i} + \alpha_{2j}) \\ x_{1i}'(\alpha_1 + \alpha_{1i} + \alpha_{1j}) + x_{2i}'(\alpha_2 + \alpha_{2i} + \alpha_{2j}) + \gamma_j \\ x_{1i}'(\alpha_1 + \alpha_{1i} + \alpha_{1j}) + x_{2i}'(\alpha_2 + \alpha_{2i} + \alpha_{2j}) + \gamma_j + u_i + v_i \end{cases}$$

4.2.2.3. Spatio-temporal Mixture Model

For F5, I consider a situation that applies a type of mixture model which allows the data to indicate the appropriate relationship, a spatially or spatio-temporally dependent model. This is considered to be a special case of the model selection methods and is similar in structure to the BaySTDetect model proposed by Li et al.[52] It is described such that:

$$\log\left(\theta_{ij}\right) = \alpha_0 + p[\alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i + v_i] + (1-p)[\alpha_{3j} x_{3ij} + \alpha_{4j} x_{4ij} + \gamma_j] + \psi_{ij}$$

where some of the predictors vary across space and time, such as x_{3ij}, while others only vary across space, x_{1i}. For this, the model probability p is large if a constant parameter spatial model is the preferred fit and small if a temporal dependence spatio-temporal model is favored. F5 will be fitted to simulated data models S3 through S10.

Table 18: Summary of the fitted models.

Model	Associated Simulated Models	Description	Linear Predictor Alternatives
F1	S1-S4, S9, S10	$\log(\theta_{ij}) = \alpha_0 + u_i + v_i + \gamma_j + \psi_{ij}$	
F2	S3-S5, S10	$\log(\theta_{ij}) = \alpha_0 + \sum_{l=1}^{L} w_{li}\omega_{lj} + \psi_{ij}$	$\omega_{ij} = \begin{cases} \alpha_1 x_{1i} + \alpha_2 x_{2i} \\ \alpha_{1j} x_{1i} + \alpha_{2j} x_{2i} \\ \alpha_{1j} x_{1i} + \alpha_{2j} x_{2i} + \gamma_j \\ \alpha_{1j} x_{1i} + \alpha_{2j} x_{2i} + \gamma_j + u_i + v_i \\ \alpha_1 x_{1i} + \alpha_2 x_{2i} + \gamma_j + u_i + v_i \end{cases}$
F3	S3-S10	$\log(\theta_{ij}) = \alpha_{0j} + \sum_{l=1}^{L} w_{li}^{S}\omega_{lj}^{S} + \sum_{k=1}^{K} w_{kj}^{T}\omega_{ki}^{T} + \psi_{ij}$	$\omega_i^S = \begin{cases} \alpha_1 x_{1i} + \alpha_2 x_{2i} \\ \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i \\ \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i + v_i \end{cases} \quad \omega_{ij}^T = \begin{cases} \alpha_1 x_{3ij} + \alpha_2 x_{4ij} \\ \alpha_{1j} x_{3ij} + \alpha_{2j} x_{4ij} \\ \alpha_{1j} x_{3ij} + \alpha_{2j} x_{4ij} + \gamma_j \end{cases}$
F4	S3-S5, S10	$\log(\theta_{ij}) = \alpha_0 + \sum_{l=1}^{L} w_{li}\omega_{lj} + \psi_{ij}$	$\omega_{ij} = \begin{cases} x_{1i}'(\alpha_1 + \alpha_{1i} + \alpha_{1j}) + x_{2i}'(\alpha_2 + \alpha_{2i} + \alpha_{2j}) \\ x_{1i}'(\alpha_1 + \alpha_{1i} + \alpha_{1j}) + x_{2i}'(\alpha_2 + \alpha_{2i} + \alpha_{2j}) + \gamma_j \\ x_{1i}'(\alpha_1 + \alpha_{1i} + \alpha_{1j}) + x_{2i}'(\alpha_2 + \alpha_{2i} + \alpha_{2j}) + \gamma_j + u_i + v_i \end{cases}$
F5	S3-10	$\log(\theta_{ij}) = \alpha_0 + p[\alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i + v_i] + (1-p)[\alpha_{1j} x_{3ij} + \alpha_{2j} x_{4ij} + \gamma_j] + \psi_{ij}$	

4.2.2.4. Summary of Fitted Models

Table 18 summarizes the fitted models described in the previous sections. This table shows the simulated models associated with the different fitted models, the model description of the fitted models, and the linear predictor alternatives. Through these different combinations of simulated and fitted models, I believe that a wide range of scenarios are being tested and compared for inference.

To compare these modeling options among the different scenarios, I will use a combination of qualitative and quantitative examinations. Maps and figures of inclusion probabilities will be presented for comparison in a qualitative way. For a quantitative approach, I will calculate *DIC* and its components along with *MSE*, variance, and mean bias squared for each of the simulated and fitted model combinations. These measures will aid in determining the best way of modeling these types of data scenarios. First, the DIC estimates are defined as follows:

$$D(\theta) = -2\sum_i \log\left(p(y_i|\theta)\right)$$
$$pD = Var(D(\theta))/2$$
$$DIC = \overline{D(\theta)} + pD$$

The effective number of parameters, pD, in this definition involves the more conservative calculation using the variance of the deviance[59] rather than the mean deviance minus the deviation of the means.[60] This is considered a penalty term for the models as models tend to fit better when more parameters are effectively used. Second, MSE and bias squared are defined such that:

$$MSE\left(M_j\right) = \sum_{i=1}^{m} \frac{\left(y_{ij}^{*} - \hat{y}_{ij}\right)^2}{m} \qquad Bias_j^{\,2} = \sum_{i=1}^{m}\sum_{k=1}^{200} \frac{\left(R_{ijk} - \hat{R}_{ijk}\right)^2}{200m}$$

where R_{ijk} is the true random effect value and $\hat{R}_{ijk}$ is the estimated random effect value for each county (i) at each time (j) in each simulated data set (k). Thus, the mean bias squared subtracts the estimated parameter value from the true parameter value and then averages that difference over the 200 simulated data sets and m spatial units. When the random effect is a temporal measure, such as γ_j, the bias squared

calculation is only averaged over the simulated data sets, and when the random effect is a spatial only measure, such as u_i and v_i, the bias squared calculation gives one value rather than J.

4.3. Results

The Knorr-Held model (F1) fits suggest that the random effects are recovered inconsistently based on the results displayed in Figures 16 and 17. Figure 16 displays a plot of the mean estimated variance against the mean bias squared for each parameter that is represented in the simulated model formulation. For example, simulated data model S2 does not include a random walk term, thus it is not displayed. I also display temporally varying estimates when applicable. This display indicates that the spatial and spatio-temporal random effects are behaving well while the temporal random walk effect is not. The larger values associated with both variance and mean bias squared for the temporal random walk effects are consistently for the later study years, thus the estimates appear to become more variable and extreme with time. It is also obvious from this display that the correlated random effect, v_i, has very small estimates associated with the bias measure as well as the variance meaning that the estimates are close to the truth and consistent across the simulated data sets. This pattern is due to the fact that this random effect is constant across the simulated data sets. Additionally, the estimates associated with the spatio-temporal interaction term are very close together among the different model scenarios. This suggests that the interaction term fits consistently across time.

Figure 17 illustrates the true and estimated values of the uncorrelated and correlated random effects for a single data set in model scenario S1F1. For the CH effect, v_i, the estimates appear to give more extreme values than the truth but the spatial structure of the truth is present in the estimates with higher values in the eastern area of the map. From the maps of the UH effect, the estimates appear to produce more positive values. This indicated poor estimation of the intercept for this model, which there is. The true intercept value is 5 while the estimated value averaged over the 200 simulated data sets is 2.57.

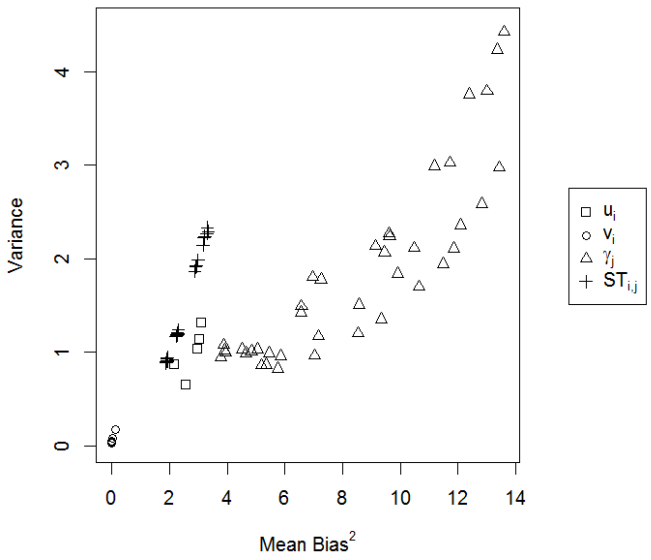


Figure 16: Variance versus mean bias squared for each random effect.

Table 19 displays the GoF measures associated with all of the model combinations of the fitted model selection methods averaged over the 200 simulated data sets. I can only truly compare these DIC measures across the fitted models as the data varies across the simulated models. Fitted model F2 does not appear to perform the best for any of the simulated data scenarios. F3 appears to be a good option for some of the simulated data scenarios; it has the flexibility to fit models with spatio-temporal parameters, and produces the lowest DIC values for scenarios S7 and S10. F3 also provides the lowest DIC estimate for S8, but it is not significantly lower than the estimate produced by F5. F4 and F5 attain much smaller pD values for the predominantly spatial models, with F4 performing the best overall for S3 and S4 while F5 performs the best for S5. Additionally, F5 produces the lowest DIC values for S6 and S9. From these estimates, I note that the deviances are quite close in value, and the large differences lie in the effective number of parameters, the penalty component to the DIC calculation. Altogether, these measurements accompanied by ability to fit a wide range of models indicates that the mixture model (F5) produces better fitting results overall since the DICs produced for this fitted model are consistently either the best or second best for all models except S10.

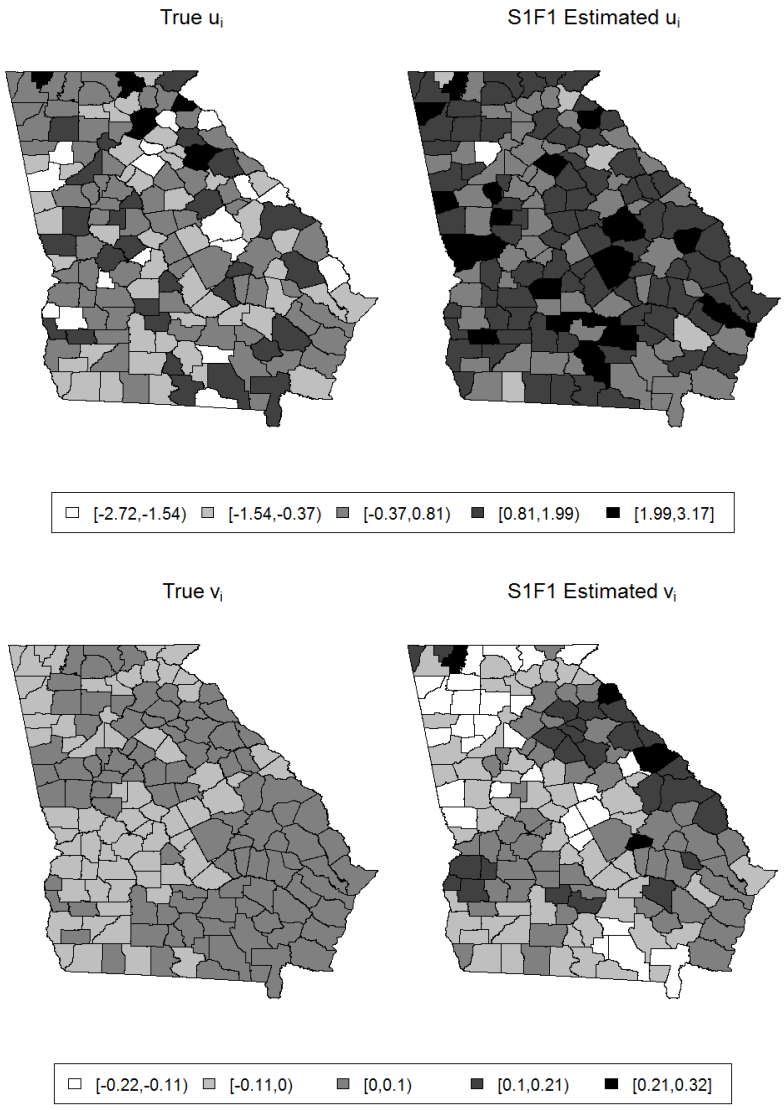


Figure 17: True and estimated random effect values for a single simulated data set with model S1F1.

Table 19: DIC, effective number of parameters, and mean deviance measures.

Model	F2			F3			F4			F5		
	DIC	pD	$\overline{D(\theta)}$	DIC	pD	$\overline{D(\theta)}$	DIC	pD	$\overline{D(\theta)}$	DIC	pD	$\overline{D(\theta)}$
S3	12777.94	1581.51	12777.94	12123.06	855.73	11267.32	11652.70	319.72	11332.97	11765.04	449.72	11315.32
S4	13067.43	1583.05	11484.38	12167.12	859.34	11307.78	10732.28	306.37	10425.91	11916.57	491.10	11425.47
S5	12404.88	978.86	11426.02	12629.71	1017.19	11612.52	12026.54	438.54	11587.99	11752.24	175.59	11576.65
S6	---	---	---	12674.43	1040.80	11633.64	---	---	---	11784.63	207.30	11577.33
S7	---	---	---	13262.81	1638.36	11624.45	---	---	---	13281.36	1659.66	11621.70
S8	---	---	---	13227.34	1607.54	11619.79	---	---	---	13230.00	1611.58	11618.42
S9	---	---	---	15035.67	3395.29	11640.38	---	---	---	14648.07	3012.52	11635.54
S10	13727.27	1877.75	11849.52	10095.13	1323.07	8772.06	12858.08	810.23	12047.85	13322.44	1522.15	11800.29

Figure 18 displays the MSE measures for a selection of the models. The MSE values are calculated for each year and averaged over the 200 simulated data sets, thus this display shows the distribution of those 10 years for a subset of the simulated and fitted model combinations. This indicates that the models whose boxplots are very wide have some years for which the models fit well while others still fit poorly. There are several model combinations that are excluded from this display because they produce MSE values that would dominate the display. The excluded values involve F2 fitted model scenarios, S10 simulated model scenarios, and S9F3. In sum, these plots suggest that fitted model F3 performs best across most model combinations, and that F5 performs slightly better or the same for some cases. Unlike for the DIC measures, F4 does not perform the best with respect to MSE for any model scenarios. The fits of S10 are displayed in the appendix and suggest that fitted model F3 is the best fitting option for S10, and this agrees with the DIC measures displayed in Table 19. For most models, particularly those that were excluded, the MSE values tend to increase over time.

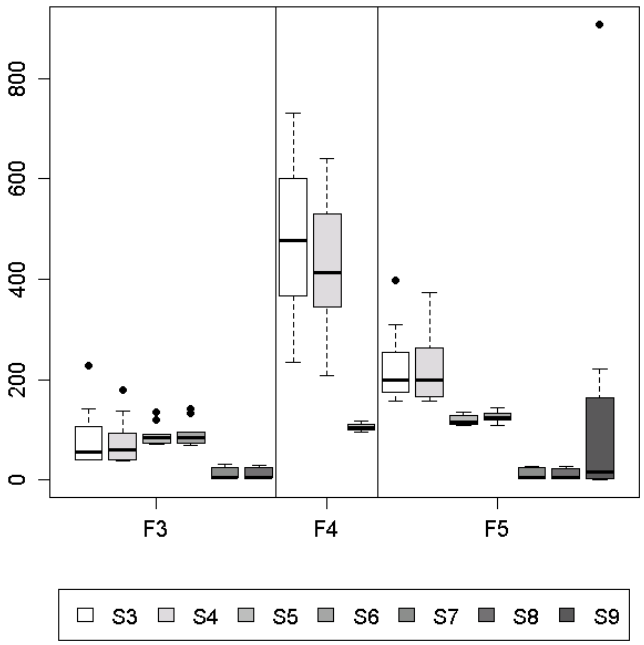


Figure 18: Averaged yearly MSE values.

The results for F2 are difficult to assess because the model probabilities vary over both space and time meaning that there are 1590 model probability estimates. I have included county maps displaying the model probabilities across time in the appendix. The maps show that there is not much variability across the counties, or across space, as the model probabilities are all very close in value.

Figure 19 plots the model probabilities associated with all simulated data set fits with fitted model F3. Most obviously, these show that there is more variance associated with the temporal linear predictors than with the spatial linear predictors. Additionally, the misspecified models (S3F3, S4F3, S9F3, and S10F3) appear to exhibit more variance overall than the appropriately specified models (S5F3 up to S8F3). Of the models that are truly specified, they seem to be somewhat well identified. Models S3F3 and S4F3 are only slightly misspecified in that they match spatial linear predictors, ω_2^S and ω_3^S respectively, but additionally have a temporal component, γ_j. This temporal component is included in ω_3^T, and that linear predictor does appear to be selected for these models. I also note that a large amount of models stay around the 0.333 range in probability estimates meaning that no linear predictor is favored. S9F3 is behaving differently from expected in that, based on the simulated model definition in Table 17, it seems as though ω_3^S and ω_2^T should be selected, and then the spatio-temporal interaction random effect would pick up the rest of the noise present in the data due to misspecification. Rather, it appears that the interaction term is attempting to explain all of the temporal noise since the temporal linear predictor that includes the appropriate temporal components, ω_3^T, is not being selected. Alternatively, S10F3 does appear to be selecting ω_3^T to account for the temporal variation.

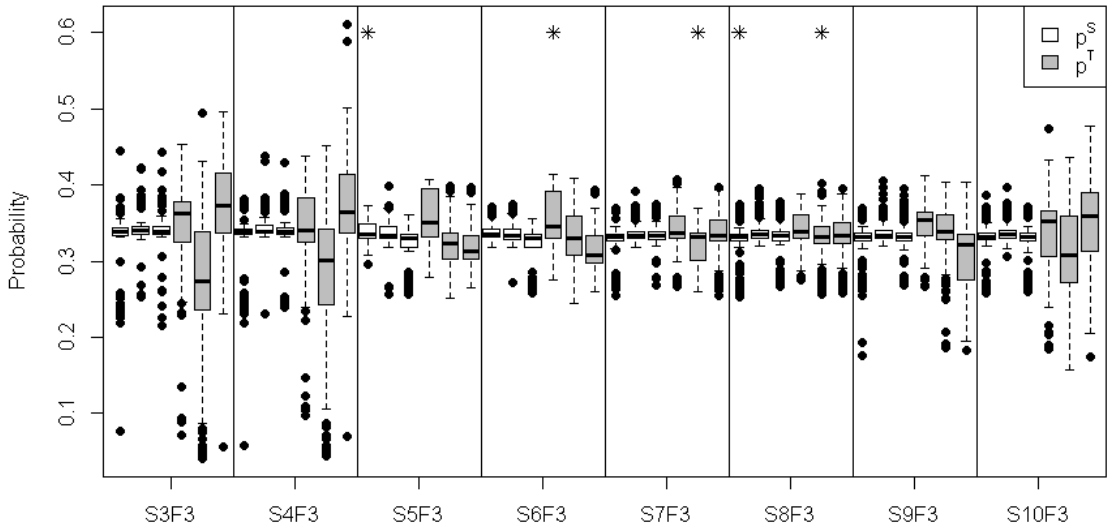


Figure 19: Model probabilities associated with the different fits of fitted model F3. Each boxplot refers to one of the six (3 spatial only, 3 spatio-temporal) linear predictor alternatives.

*indicates the true model for that particular model scenario

Figure 20 displays the model probabilities associated with fitted model F4. The distribution of the model probabilities are similar for all model scenarios. For M4F4 and M10F4, the median for p_3 is appropriately, though only slightly, higher than p_2. Additionally, the estimates produced for scenario S5F4 appear to be much more stable than the others. Both of these issues are likely due to the fact that the first linear predictor alternative is less complex and does not involve random effects which are present in the second and third linear predictor alternatives. Thus, the second two linear predictors dominate the selection process and there is more variation present when random effects are a part of the simulated data scenario.

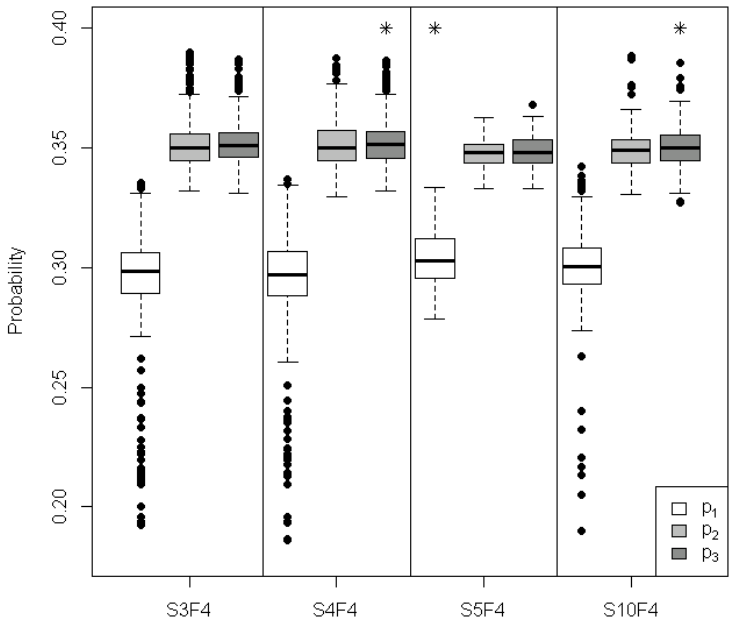


Figure 20: Model probabilities for F4 model fits.

* indicates the probability parameter associated with the true linear predictor.

Figure 21 displays violin plots of the model probabilities associated with the F5 model fits. These are produced using the `vioplot` package in R[61] and allow for viewing the distributions associated with the model probabilities for each scenario. Each volin plot image is a boxplot with a kernal density plot on each side such that the while circle is the median and the black rectangle represents the quartiles. This plot also displays a horizontal line at $probability = 0.5$ to aid in identifying the mixing effect of this fitted model. S3F5 and S4F5 are appropriately represented as spatial models with a bit of temporal mixture; their distributions are quite similar and this is to be expected because they only differ by one random effect, v_i; its inclusion in S4F5 causes that model to tend slightly more towards a predominantly spatial model. Alternatively, S5F5 is a simple, strictly spatial model that seems to be having a difficult time identifying that. Here, S6F5 and S7F5 are appropriately identified as spatio-temporal models based on their associated model probabilities; of these two models, S7F5 has a much smaller spread to its distribution because it is more truly specified within this model than S6F5 since the parameter estimates also vary temporally.

Interestingly, S8F5 appears to have a nearly identical distribution to S7F5, and while they do have common parameters, S8F5 additionally has spatial only covariates. Thus, it is expected to be represented as more of a mixed model. S9F5 and S10F5 also behave as expected by displaying mixture effects between the spatial and spatio-temporal components.

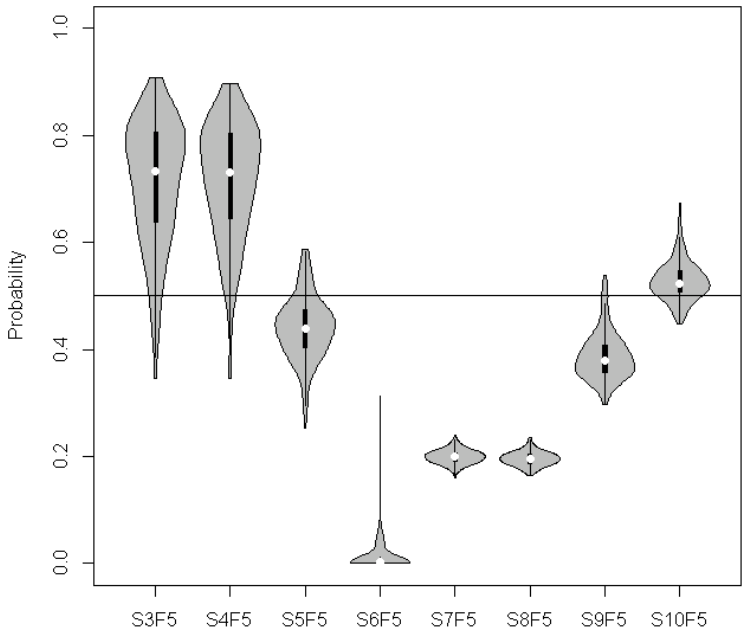


Figure 21: Violin plots of the model probabilities for F5.

4.4. Colon Cancer Data Example

For my real data example, I use incidence of colon cancer obtained from the Oasis of the Georgia Department of Public Health as an outcome with the same predictors from the AHRF data set.[46] These data overlap for a study time line of 2001 to 2007; as before, the spatially only varying covariates come from a central year of the study time line, 2004. Past studies suggest that colon cancer has a spatial structure and is related to these predictors, though they are not the main risk factors associated with colon cancer,[41, 44, 45, 47] and I are looking to explore the temporal structure as well. I apply these data using fitted models F3-F5 as these models appear to perform the best.

The GoF results from these different model selection alternatives are displayed in **Table 20** and Figure 22. Table **20** shows the DIC measures while the MSE measures are shown in Figure 22. In terms of DIC, F5 appears to be the best fitting model while F3 is the best with respect to MSE and mean deviance. However, there is not much of a difference in the MSE values from one fitted model to the next. Additionally, the MSE measures demonstrate that there is a similar relationship in model fit from one year to the next across the fitted models. This relationship appears to be influenced by the distribution of the outcome variable counts. The mean of those counts is lowest for year 5 ($\bar{y}_5 = 8.2$) and highest for year 7 ($\bar{y}_7 = 8.5$) while the median for all years is 4.0. This indicates 1) that the model fits poorly when there are outliers in the outcome variable and 2) that F3 is least vulnerable to the influence of outliers.

Table 20: DIC measures for the real data example.

Model	DIC	pD	$\bar{D}$
F3	5066.34	339.16	4727.18
F4	4911.83	176.64	4735.19
F5	4903.09	160.72	4742.36

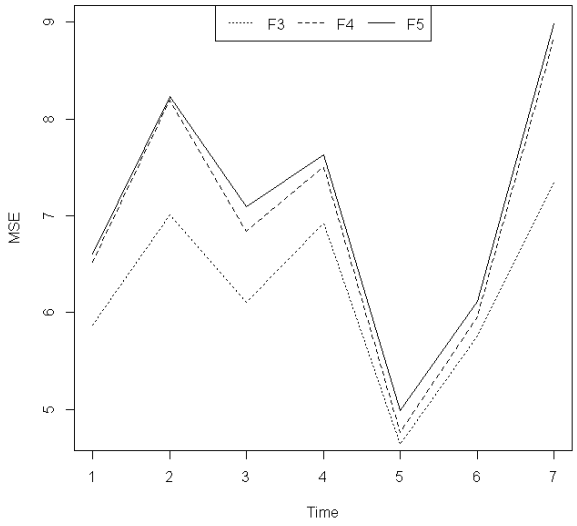


Figure 22: MSE measures for the colon cancer data example.

Table 21 displays the model probabilities produced as well as the resulting selected linear predictor. The F3 results indicate that there is more spatio-temporal structure in the data since that linear predictor is clearly selected while the spatial only linear predictors are not. F4 results look as though it is attempting to explain the temporal structure in the data using the γ_j term. Additionally, F5 continues to suggest that this is largely a spatio-temporal model with some spatial influence. The spatio-temporal suggestion here could be stemming from the covariates more so than the spatio-temporal random effect, this is suggested by F3's selection of the first spatio-temporal linear predictor alternative, and the real data example in Chapter 3 indicated that percent AA population is an important predictor for colon cancer.

Table 21: Model probabilities and selected linear predictors for the real data example.

Model		Model Probabilities			Selected Linear Predictor
		P1	P2	P3	
F3	Spatial	0.199 (0.00,0.50)	0.391 (0.10,0.75)	0.390 (0.10,0.73)	$\alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i$ or $\alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i + v_i$
	Spatio-Temporal	0.459 (0.13,0.88)	0.266 (0.00,0.59)	0.262 (0.01,0.60)	$\alpha_1 x_{3ij} + \alpha_2 x_{4ij}$
F4		0.244 (0.00,0.46)	0.366 (0.12,0.76)	0.359 (0.12,0.73)	$x_{1i}'(\alpha_1 + \alpha_{1i} + \alpha_{1j}) + x_{2i}'(\alpha_2 + \alpha_{2i} + \alpha_{2j}) + \gamma_j$ or $x_{1i}'(\alpha_1 + \alpha_{1i} + \alpha_{1j}) + x_{2i}'(\alpha_2 + \alpha_{2i} + \alpha_{2j}) + \gamma_j + u_i + v_i$
F5		0.237 (0.077,0.674)			$\log(\theta_{ij}) = \alpha_0 + p[\alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i + v_i] + (1-p)[\alpha_{3j} x_{3ij} + \alpha_{4j} x_{4ij} + \gamma_j] + \psi_{ij}$

Additional tables and figures included in the appendix display the estimates produced with the selected linear predictors. Supplemental Table 1 displays the temporal random walk term, γ_j. Overall, the estimates are very small; further, they are not well estimated. This agrees with the maps of the risk, θ_{ij}, which are displayed in Supplemental Figures 2, 4, and 6. These maps show some, though not much variation from one year to the next. Supplemental Figures 3, 5, and 7 are also maps that display the UH and CH effects for the selected linear predictors. All of the maps have a similar distribution displayed from one fitted model to the next, with the CH effect for F3 and F5 being the most similar. Additionally, while the magnitude of the estimates for θ_{ij} is also similar across fitted models, F5 produces random effect estimates with more

variation than the other models. For all fitted models, the UH effect appears to dominate the models, and this, accompanied with the consistent CH effect estimates, suggests that the models are performing well.

For the selected model re-fit, I have only chosen to use F3. F4 was eliminated because the inclusion of a spatio-temporal term appears to be important, and F4 cannot accommodate that appropriately. F5 is not being re-fit as there does appear to be mixing in the results, thus the model displayed above is the final model fit for F5. The re-fits of fitted model F3 involve both spatial linear predictor alternatives whose model probabilities were similar. The GoF measures produced in the re-fits are alike to each other as well as the values presented in Table 20 and Figure 22 in that F5 continues to be the best in terms of DIC while the F3 model with the third alternative spatio-temporal linear predictor is the best fitting in terms of MSE. The new displays of these GoF measures can be found in Supplementary Table 2 and Supplemental Figure 8. The θ_{ij} and random effect estimates produced for these model refits are nearly identical to those displayed for the model selection fits of these models. Additionally, the parameter estimates produced of these models are contained in Supplemental Table 3. These parameter estimates show that the spatio-temporal predictors are important since α_4 is well estimated for both F3 alternatives; this parameter estimate does not appear to be significant for F5, and that is likely due to the parameter being forced to vary over time. I also note that the intercepts are well estimated for all models, but they have opposite signs. This is due to the different structures present in F3 and F5. Additionally, both models suggest that colon cancer has a negative relationship with x_1, median household income, and F5 shows that there is also a positive relationship with PPBPL. Both of these interpretations agree with the previous colon cancer data example in Chapter 3.

4.5. Discussion

Based on the results in the previous section, the mixture model, F5, is the best option for the majority of model scenarios, but it is not without fault. Fitted models F3 and F4 also perform well for certain models, but each have their own weaknesses. F3 offers the ability to fit more combiations of linear predictor

alternatives by allowing for separate selection between the spatial and spatio-temporal components. F4, on the otherhand, allows for more complex linear predictor alternative within the spatial setting. The more complex linear predictor alternatives offered by both of these fitted model scenarios lead to slow fitting models and more difficulty with convergence. Fitted model F2 does not perform as well as these other models, and this is ultimately because of the complexity within the model probability estimates.

Concerning measures of GoF, most models fit best with F5 according to DIC. However, a few of the spatial and spatio-temporal models still perform better with F4 and F3 respectively. When taking MSE into consideration, models F3 and F5 perform best everytime with F3 performing slightly better overall. This suggests that they fit more consistently across the temporal span of the study region. These results are somewhat conflicting, so moving on to the models' ability to appropriately identify the true alternative linear predictors, it appears that while F3 performs well when the models are appropriately specified, but F5 seems to do best at identifying spatial only, spatio-temporal only, or both.

As mentioned previously, I suggest that these model selection procedures be performed in sequence with individual model re-fits. This continues to be true for F2-F4, but F5 could be the final model fit in certain cases. For example, a scenario such as S3F5 has both spatial and spatio-temporal components, and this mixture model may be an appropriate final fit in this case. On the other hand, scenarios that give results such as S7F5 may be better to re-fit as fully spatio-temporal models since there appears to be no spatial only component. However, the interpretation of the mixture models would be quite complex. The real data example illustrates that model F3-F5 produce nearly identical risk estimates, and models F3 and F5 produce random effect values that are alike.

The conventional model selection models are all quite sensitive to model misspecification, and this is to be expected since the models are very intricate. Fitted model F5 also superceeds the others in terms of misspecification because almost all models are truly specified when supplied to F5; there may be additional, unnecessary terms included with certain model fits, but this is true for the other fitted models as

well. Through the mixture component, F5 is able to accomodate most of the linear predictor alternatives the other fitted models offer. Fitted model F3 gives the worst, most variable results for misspecifed models, particularly when temporal misspecification is present.

Another issue among this study involves the random effects. With the amount of random effects included in these models, particularly for F2, F3, and F4, there can be issues with identifiability. Futhermore, all of the fitted models include a spatio-temporal interaction term as an uncorrelated random effect, and this may also be producing some identification issues with the model selection process. An example of this was distinguished involving model M9F3 selecting the inappropriate alternative linear predictor and letting the spatio-temporal intereaction term explain the variation instead. Even though the Knorr-Held model suggested that I am able to recover the random effects present in that model reasonably well, there is only one of each type of random effect in F1, thus no major identification concern. Additionally, this is another advantage of F5 because there are no multiple random effects present in that model.

One negative feature of fitted model F5 involves the results associated with model scenarios S5F5 and S8F5. S5 is fully spatial, but the median for that model probability distribution for S5F5 is below 0.5, however not significantly so. Correspondingly, S8 is spatio-temporal with a strong spatial component, but S8F5 does not appear to detect those spatial trends as the distribution of its model probabilities are quite stable and well below 0.5. This may be a fault in the data as the spatial only varying true parameter estimates do have smaller magnitudes than the temporally varying ones due to the random walk that was placed on these parameters. This fact may also be attributing to the extra variance noted among the temporal linear predictor alternatives in comparison to the spatial only alternatives for fitted model F3. Further, these simulated data model scenarios seem to have some temporal misidentification with F3 as well.

4.6. Conclusion

Following these study results, I believe that fitted model F5, the mixture model, is the best option of the model selection alternatives. This model is fast fitting, the least vulnerable to misspecification, the least vulnerable to identification issues, accommodates a wide range of linear predictors, and the one that appears to fit the best overall. Following F5, F3 is another worthwhile option as this model also accommodates an even wider range of linear predictor alternatives and offers the best fit in certain situations.

With respect to implementation in public health, my results suggest that both F5 and F3 are adequate starting points for spatio-temporal disease mapping situations in which there is a limited set of linear predictors that can be determined a priori. These alternative linear predictors 1) may be supposed optimal combinations of important predictors, 2) could be important in terms of the study design, and 3) can offer a comparison of alternating collinear predictors.

5. Summary

5.1. Conclusions

The motivation for this project stems from the need for a statistical method for dimension reduction in the spatial and spatio-temporal disease mapping setting. Dimension reduction is an important part of statistical inference because it aids in producing the ideal linear predictor for interpretation. Model selection furnishes a methodology that eliminates many of the issues associated with other alternatives for dimension reduction. In this project, I begin with a software comparison to determine the best platform for continuing with model selection methodologies that then are explored in the non-spatial, spatial, and spatio-temporal settings.

Aim 1, presented in Chapter 2, concluded that the default settings in INLA are not appropriate in the disease mapping context, but that it performs nearly identically to OpenBUGS when the defaults are altered as suggested in Chapter 2. These suggestions include: specifying the use of the full Laplace approximation technique and specifying the default hyperpriors associated with the random effects more appropriately, e.g. Gamma(1,0.5). I conclude that OpenBUGS offers more flexibility while INLA offers quick computation. Thus, I made the decision to continue with the OpenBUGS platform.

Aim 2, offered in Chapter 3, suggests that, based on the simulation results, BMS outperforms BMA in that it tends to recover the ground truth more appropriately, is more robust to misspecification, and produces more consistent results with respect to models with constant verses varying expected rates. For all models and in both methodologies, performance is improved for complete models over partial models, models with larger true parameter estimates, and areas where counties are smaller and closer together. From an application stand point, I believe that both BMS and BMA can be beneficial. Using the techniques in conjunction can lead to more information about the data of interest gained. Further, if BMS does not appear to clearly select one linear predictor over the others, BMA could produce a useful final model as it would be effectively averaging over the best linear predictor alternatives.

Aim 3, discussed in Chapter 4, shows that the mixture model, fitted model F5, offers the best results since it is fast fitting, least vulnerable to misspecification, and least vulnerable to identification issues. Further, F5 accommodates a wide range of models and offers the best overall fits. Following F5, F3 also offers the best fit for certain scenarios and accommodates an even wider range of linear predictors. From an application stand point, F5 could be the first and final model fit for a set of data if the model probability does not clearly suggest that the data structure is strictly spatial or spatio-temporal. For fitted models F2 through F4, one or more re-fits of the selected linear predictors for the appropriate counties are required to gain appropriate values for the parameter estimates.

5.2. Limitations

The explorations discussed in this dissertation contain several limitations. One that applies to all aims involves the limited number of fitted and simulated model scenarios explored. The number of possible scenarios is infinite, but I believe that this project explored a representative number of them. Two more issues that relate to all aims involve the varying county sizes and strength of association in the data. These two issues can lead to difficult model fits and inappropriate results in some areas of the spatial region.

For the model selection methods, more limitations are introduced. One of these limitations presented in the model selection methods includes identification issues related to including multiple random effects across the linear predictor alternatives. This, along with gaining the appropriate fixed parameter estimates, means that most of the model selection methods also require model refits of the appropriate linear predictors for the selected counties. An additional limitation that Aim 3 presents involves the more complex parameters associated with spatio-temporal models. When these parameters are allowed to vary over time, the models become more computationally demanding as well as more difficult to interpret.

5.3. Future Directions

Future work related to this project includes more a more in depth analysis of the mixture model, F5, from the spatio-temporal model selection methods in Aim 3. This model proved to be very suitable for my simulated and real data, and I am interested in looking at an extension that would allow the model probabilities to vary over space, time, or both. These extensions would offer an even more flexible structure

for the mixture model in that a different amount of mixing could be estimated for each county, year, or both.

In addition to the future work with fitted model F5 from Aim 3, there are many other spatio-temporal model selection methodological possibilities that could be explored for comparison to those that I have already analyzed. From an applied stand point, I also wish to explore how these methods can be used and optimized in other public health settings.

References

1. Gilks WR, Clayton DG, Spiegelhalter DJ, et al. Modeling Complexity: Applications of Gibbs sampling in medicine. *Journal of Royal Statistical Society, Series B*. 1993: 39-52.
2. Spiegelhalter D, Thomas A, Best N and Lunn D. WinBUGS manual. Cambridge, UK: MRC Biostatistics Unit, 2007.
3. Lunn D, Jackson C, Best N, Thomas A and Spiegelhalter D. *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Boca Raton, FL: CRC Press, 2013.
4. Besag J, York J and A. M. Bayesian image restoration with two applications in spatial statistics. *Ann Inst Stat Math*. 1991; 43: 59.
5. Rue H, Martino S and Chopin N. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *J R Stat Soc Series B*. 2009; 71: 319-92.
6. Gamerman D and Lopes H. *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. 2nd ed. New York: CRC Press, 2006.
7. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria. 2015.
8. Martin AD, Quinn KM and Park JH. MCMCpack: Markov Chain Monte Carlo in R. *J Stat Soft*. 2011; 42: 1-21.
9. Lindgren F, Rue H and Lindström J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J R Stat Soc Series B Stat Methodol*. 2011; 73: 423-98.
10. Martins TG, Simpson D, Lindgren F and Rue H. Bayesian computing with INLA: New features. *Comput Stat Data An*. 2013; 67: 68-83.
11. Simpson D, Lindgren F and Rue H. In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics*. 2012; 23: 65-74.
12. Simpson D, Lindgren F and Rue H. Think continuous: Markovian Gaussian models in spatial statistics. *Spat Stat*. 2012: 16-29.
13. Thomas A, O'hara B, Ligges U and Sturtz S. Making BUGS Open. *R News*. 2006; 6: 12-7.
14. Schlather M. *An introduction to positive definite functions and to unconditional simulation of random fields*. Lancaster University1999.
15. Diggle PJ and Ribeiro Jr. PJ. *Model-based Geostatistics*. New York: Springer, 2007.
16. Eberly LE and Carlin BP. Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Statistics in medicine*. 2000; 19: 2279-94.
17. Lawson AB. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. 2 ed. Boca Raton, FL: CRC Press, 2013.
18. Lesaffre E and Lawson AB. *Bayesian Biostatistics*. West Sussex, U.K.: Wiley, 2013.
19. Gelman A, Hwang J and Vehtari A. Understanding predictive information criteria for Bayesian models. *Statistics and Computing*. 2013: In Print.
20. Schrödle B and Held L. Spatio-temporal disease mapping using INLA. *Environmetrics*. 2011; 22: 725-34.
21. Schrödle B and Held L. A primer on disease mapping and ecological regression using INLA. *Computation Stat*. 2010; 26: 241-58.
22. Sørbye SH. Tutorial: Scaling IGMRF-models in R-INLA.: Department of Mathematics and Statistics, University of Tromso.
23. Sørbye SH and Rue H. Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spat Stat*. 2014; 8: 39-51.
24. Blangiardo M, Cameletti M, Baio G and Rue H. Spatial and spatio-temporal models with R-INLA. *Spat Spatiotemporal Epidemiol*. 2013; 4: 33-49.
25. Schrödle B and Held L. A primer on disease mapping and ecological regression using INLA. *Computational Statistics*. 2010; 26: 241-58.
26. Hoeting JA, Madigan D, Raftery AE and Volinsky CT. Bayesian Model Averaging: a Tutorial. *Statistical Science*. 1999; 14: 382-417.

27. Clyde M and Iversen ES. Bayesian model averaging in the M-open framework. In: Damien P, Dellaportas P, Polson NG and Stephens DA, (eds.). *Bayesian Theory and Applications*. Oxford: Oxford University Press, 2015.
28. Lunn D, Jackson C, Best N, Thomas A and Spiegelhalter D. *The BUGS Book: A Practical Introduction to Bayesian Analysis*. 1 ed. Boca Raton, FL: CRC Press, 2013, p.399.
29. Bondell HD, Krishna A and Ghosh SK. Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics*. 2010; 66: 1069-77.
30. Hoeting JA, Raftery AE and Madigan D. Simulataneous variable and transformation selection in linear regression. *Journal of Computational and Graphical Statistics*. 1995.
31. Scheel I, Ferkingstad E, Frigessi A, Haug O, Hinnerichsen M and Meze-Hausken E. A Bayesian hierarchical model with spatial variable selection: the effect of weather on insurance claims. *Appl Statist*. 2013; 62: 85-100.
32. Rockova V and George EI. Negotiating multicollinearity with spike-and-slab priors. *Metron*. 2014; 72: 217-29.
33. Garcia RI, Ibrahim JG and Zhu H. Variable selection for regression models with missing data. *Stat Sin*. 2010; 20: 149-65.
34. George EI and Clyde M. Model Uncertainty. *Stat Sci*. 2004; 19: 81-94.
35. Hans C. Model uncertainty and variable selection in Bayesian lasso regression. *Stat Comp*. 2009; 20: 221-9.
36. Li J, Das K, Fu G, Li R and Wu R. The Bayesian lasso for genome-wide association studies. *Bioinformatics*. 2011; 27: 516-23.
37. Viallefont V, Raftery AE and Richardson S. Variable selection and Bayesian model averaging in case-control studies. *Statistics in medicine*. 2001; 20: 3215-30.
38. Hoeting JA, Madigan D, Raftery AE and Volinsky CT. Bayesian model averaging: a tutorial. *Stat Sci*. 1999; 14: 382-417.
39. Wheeler DC, Hickson DA and Waller LA. Assessing local model adequacy in Bayesian hierarchical models using the partitioned deviance information criterion. *Computational statistics & data analysis*. 2010; 54: 1657-71.
40. Eberly LE and Carlin BP. Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Statistics in medicine*. 2000; 19: 2279-94.
41. American Cancer Society. Colorectal Cancer Facts & Figures. Atlanta, GA: American Cancer Society.
42. Ahmed RL, Schmitz KH, Anderson KE, Rosamond WD and Folsom AR. The metabolic syndrome and risk of incident colorectal cancer. *Cancer*. 2006; 107: 28-36.
43. Labianca R, Nordlinger B, Beretta GD, Brouquet A, Cervantes A and Group EGW. Primary colon cancer: ESMO Clinical Practice Guidelines for diagnosis, adjuvant treatment and follow-up. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2010; 21 Suppl 5: v70-7.
44. DeChello LM and Sheehan TJ. Spatial analysis of colorectal cancer incidence and proportion of late-stage in Massachusetts residents: 1995-1998. *International journal of health geographics*. 2007; 6: 20.
45. Elferink MA, Pukkala E, Klaase JM and Siesling S. Spatial variation in stage distribution in colorectal cancer in the Netherlands. *European journal of cancer*. 2012; 48: 1119-25.
46. Area Health Resource Files (AHRF). Rockville, MD: US Department of Health and Human Services, Health Resources and Services Administration, Bureau of Health Workforce, 2003.
47. Henry KA, Sherman RL, McDonald K, et al. Associations of census-tract poverty with subsite-specific colorectal cancer incidence rates and stage of disease at diagnosis in the United States. *Journal of cancer epidemiology*. 2014; 2014: 823484.
48. Breslow NE and Day NE. *The Design and Analysis of Cohort Studies*. New York: Oxford University Press, 1987.
49. Garcia RI, Ibrahim JG and Zhu H. Variable selection for regression models with missing data. *Statistica Sinica*. 2010; 20: 149-65.
50. Hoeting JA, Raftery AE and Madigan D. Simultaneous variable and transformation selection in linear regression. *J Comput Graph Stat*. 2002: 485-507.

51. Scheel I, Ferkingstad E, Frigessi A, Haug O, Hinnerichsen M and Meze-Hausken E. A Bayesian Hierarchical Model with Spatial Variable Selection: The Effect of Weather on Insurance Claims. *Journal of the Royal Statistical Society Series C, Applied statistics*. 2013; 62: 85-100.
52. Li G, Best N, Hansell AL, Ahmed I and Richardson S. BaySTDetect: detecting unusual temporal patterns in small area data via Bayesian model choice. *Biostatistics*. 2012; 13: 695-710.
53. Waldorp LJ, Huizenga HM, Nehorai A, Grasman RP and Molenaar PC. Model selection in spatio-temporal electromagnetic source analysis. *IEEE transactions on bio-medical engineering*. 2005; 52: 414-20.
54. Lee KJ, Jones GL, Caffo BS and Bassett SS. Spatial Bayesian variable selection models on functional magnetic resonance imaging time-series data. *Bayesian Anal*. 2014; 9: 699-732.
55. Ugarte MD, Adin A, Goicoa T and Militino AF. On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Stat Methods Med Res*. 2014; 23: 507-30.
56. Knaus J. snowfall: Easier cluster computing (based on snow). 2013, p. R package version 1.84-6.
57. Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M and Songini M. Bayesian analysis of space—time variation in disease risk. *Statistics in medicine*. 1995; 14: 2433-43.
58. Knorr-Held L. Bayesian modeling of inseperable space-time variation in disease risk. *Statistics in medicine*. 2000; 19: 2555-67.
59. Gelman A, Carlin JB, Stern HS and Rubin DB. *Bayesian Data Analysis*. 2nd ed.: CRC Press, 2004.
60. Spiegelhalter DJ, Best NG, Carlin BP and van der Linde A. Bayesian measures of model complexity and fit. *J Roy Statist Soc B*. 2002; 64: 583-639.
61. Adler D. vioplot: Violin plot. 2005.

Appendix

A.1. Aim 1
A.1.1. Code

`OpenBUGS` model file used for `BRugs`. Note that for different fits, I altered some of the parameters for τ_u and τ_v.

```
model{
for (i in 1:159){
x1c[i]<-x1[i]-mean(x1[])
x2c[i]<-x2[i]-mean(x2[])
x3c[i]<-x3[i]-mean(x3[])

        y1[i]~dpois(mu1[i])
        mu1[i]<-e[i]*theta1[i]
        log(theta1[i])<-a1[1]+a1[2]*x1c[i]+a1[3]*x2c[i]
        y1.pred[i]~dpois(mu1[i])

        y2[i]~dpois(mu2[i])
        mu2[i]<-e[i]*theta2[i]
        log(theta2[i])<-a2[1]+a2[2]*x1c[i]+a2[3]*x2c[i]+u2[i]
        u2[i]~dnorm(0,tauu2)
        y2.pred[i]~dpois(mu2[i])

        y3[i]~dpois(mu3[i])
        mu3[i]<-e[i]*theta3[i]
        log(theta3[i])<-a3[1]+a3[2]*x1c[i]+a3[3]*x2c[i]+u3[i]+v3[i]
        u3[i]~dnorm(0,tauu3)
        y3.pred[i]~dpois(mu3[i])

        y4[i]~dpois(mu4[i])
        mu4[i]<-e[i]*theta4[i]
        log(theta4[i])<-a4[1]+a4[2]*x1c[i]+a4[3]*x2c[i]+a4[4]*x3c[i]+a4[5]*x4[i]
        y4.pred[i]~dpois(mu4[i])

        y5[i]~dpois(mu5[i])
        mu5[i]<-e[i]*theta5[i]
        log(theta5[i])<-a5[1]+a5[2]*x1c[i]+a5[3]*x2c[i]+a5[4]*x3c[i]+a5[5]*x4[i]+u5[i]+v5[i]
        u5[i]~dnorm(0,tauu5)
        y5.pred[i]~dpois(mu5[i])

        y6[i]~dpois(mu6[i])
        mu6[i]<-e[i]*theta6[i]
        log(theta6[i])<-a6[1]+u6[i]+v6[i]
        u6[i]~dnorm(0,tauu6)
        y6.pred[i]~dpois(mu6[i])

        e[i]<-1
        }
for (j in 1:3){
        a1[j]~dnorm(0,taua)
```

```
        a2[j]~dnorm(0,taua)
        a3[j]~dnorm(0,taua)
        }
for (k in 1:5){
        a4[k]~dnorm(0,taua)
        a5[k]~dnorm(0,taua)
        }
a6[1]~dnorm(0,taua)

taua<-1

tauu2~dgamma(1,.5)
tauu3~dgamma(1,.5)
tauu5~dgamma(1,.5)
tauu6~dgamma(1,.5)
tau3  ~dgamma(1,.5)
tau5  ~dgamma(1,.5)
tau6  ~dgamma(1,.5)

v3[1:159]~car.normal(adj[],weights[],num[],tau3)
v5[1:159]~car.normal(adj[],weights[],num[],tau5)
v6[1:159]~car.normal(adj[],weights[],num[],tau6)
for(l in 1:sumNumNeigh){
        weights[l]<-1
        }
}
```

INLA code using M3F3 as an example. Note that for the different fits I altered the param= statement in the formula as well as the strategy= statement in the INLA call. Furthermore, I only show how to access y_{pred} and $\hat{y}$.

```
n=159
id=c(1:n,1:n)
id2=id
id3=id

data=list(y3=c(y[1:n,3],rep(NA,n)),
        x1=c(x[1:n,1],x[1:159,1])-mean(x[1:n,1]),
        x2=c(x[1:n,2],x[1:159,2])-mean(x[1:n,2]))

form3=y3~1+x1+x2+f(id,model="iid",param=c(1,.5))+
        f(id2,model="besag",graph="adjGA_INLA.txt",param=c(1,.5))+
        f(id3,model="iid")
resM3<-inla(form3,family="poisson",data=data3,
        control.compute   = list(dic=TRUE, mlik=TRUE,cpo=TRUE),
        control.fixed     = list(prec.intercept = 1, prec = 1),
        control.predictor = list(compute=TRUE),
        control.inla      = list(strategy = "laplace"))


yhat=resM3$summary.fitted.values[1:n,1]
ypred=resM3$summary.fitted.values[(n+1):(2*n),1]
```

A.1.2. Supplemental Tables and Figures

Table 1: GoF measures associated with each model and software package using the alternative prior distributions.

Model	Prior	$\overline{D}$	pD	DIC	MSE_{α}	MSE_{M}	MSE_{θ}	$MSPE$
S1F1 OpenBUGS	Default	433.45	2.92	436.38	0.01	1.13	0.02	1.13
S1F1 INLA(SL)	Default	433.47	2.98	436.45	0.01	1.13	0.02	1.13
S1F1 INLA(FL)	Default	433.47	2.97	436.44	0.01	1.13	0.02	1.13
S2F2 OpenBUGS	Default	450.50	85.50	536.00	0.01	0.53	1.90	0.53
S2F2 OpenBUGS	Gam(2,1)	450.61	85.51	536.13	0.01	0.54	1.90	0.53
S2F2 OpenBUGS	Gam(1,1)	449.46	86.37	535.83	0.01	0.51	1.90	0.51
S2F2 OpenBUGS	Gam(1,0.5)	450.18	85.91	536.09	0.01	0.53	1.90	0.53
S2F2 INLA(SL)	Default	455.12	87.90	543.02	0.01	0.51	1.91	0.58
S2F2 INLA(FL)	Default	452.49	83.10	535.58	0.01	0.60	1.99	0.79
S2F2 INLA(FL)	Gam(2,1)	451.43	83.58	535.02	0.01	0.57	1.99	0.76
S2F2 INLA(FL)	Gam(1,1)	450.46	84.74	535.20	0.01	0.54	1.98	0.73
S2F2 INLA(FL)	Gam(1,0.5)	451.04	84.36	535.40	0.01	0.53	1.98	0.74
S3F3 OpenBUGS	Default	449.60	84.65	534.25	0.01	0.52	1.91	0.52
S3F3 OpenBUGS	Gam(2,1)	449.43	56.54	505.96	0.01	0.51	1.92	0.51
S3F3 OpenBUGS	Gam(1,1)	448.05	50.43	498.48	0.01	0.48	1.92	0.48
S3F3 OpenBUGS	Gam(1,0.5)	448.78	60.84	509.63	0.01	0.50	1.92	0.50
S3F3 INLA(SL)	Default	454.29	88.35	542.64	0.01	0.50	1.92	0.57
S3F3 INLA(FL)	Default	451.51	83.43	534.94	0.01	0.59	1.99	0.78
S3F3 INLA(FL)	Gam(2,1)	449.79	85.12	534.91	0.01	0.54	2.00	0.73
S3F3 INLA(FL)	Gam(1,1)	448.59	86.53	535.12	0.01	0.51	1.99	0.69
S3F3 INLA(FL)	Gam(1,0.5)	449.61	85.70	535.31	0.01	0.56	1.99	0.72
S4F4 OpenBUGS	Default	433.11	4.75	437.87	0.02	1.15	0.05	1.15
S4F4 INLA(SL)	Default	433.22	4.94	438.16	0.21	1.15	0.05	1.15
S4F4 INLA(FL)	Default	433.22	4.92	438.15	0.21	1.15	0.05	1.15
S5F5 OpenBUGS	Default	448.76	83.26	532.02	0.03	0.52	1.84	0.52
S5F5 OpenBUGS	Gam(2,1)	449.08	42.11	491.19	0.05	0.51	1.84	0.51
S5F5 OpenBUGS	Gam(1,1)	447.61	35.30	482.91	0.06	0.48	1.85	0.48
S5F5 OpenBUGS	Gam(1,0.5)	448.49	46.54	495.04	0.08	0.50	1.84	0.50
S5F5 INLA(SL)	Default	453.67	88.90	542.57	0.14	0.49	1.85	0.56
S5F5 INLA(FL)	Default	450.73	83.85	534.58	0.14	0.58	1.91	0.77
S5F5 INLA(FL)	Gam(2,1)	449.21	85.02	534.23	0.05	0.54	1.91	0.72
S5F5 INLA(FL)	Gam(1,1)	448.05	86.36	534.41	0.06	0.50	1.91	0.69
S5F5 INLA(FL)	Gam(1,0.5)	448.98	85.64	534.62	0.05	0.52	1.91	0.71
S6F6 OpenBUGS	Default	449.84	84.64	534.48	0.01	0.54	1.60	0.54
S6F6 OpenBUGS	Gam(2,1)	449.63	52.46	502.09	0.01	0.53	1.61	0.53
S6F6 OpenBUGS	Gam(1,1)	448.01	46.30	494.30	0.01	0.50	1.61	0.50
S6F6 OpenBUGS	Gam(1,0.5)	449.17	56.55	505.72	0.01	0.53	1.61	0.52
S6F6 INLA(SL)	Default	454.61	88.35	542.96	0.01	0.51	1.62	0.60
S6F6 INLA(FL)	Default	451.90	83.45	535.35	0.01	0.61	1.69	0.81
S6F6 INLA(FL)	Gam(2,1)	450.07	85.12	535.19	0.01	0.56	1.69	0.76
S6F6 INLA(FL)	Gam(1,1)	448.79	86.56	535.35	0.01	0.53	1.69	0.72
S6F6 INLA(FL)	Gam(1,0.5)	449.84	85.73	535.58	0.01	0.55	1.69	0.74

Table 2: Altered INLA precision estimates compared the truth.

Model	Parameter	M2F2	M3F3	M5F5	M6F6	True value
INLA Ga(1,5e-05) Scaling option	τ_u	18654 (18572)	17429 (17309)	18808 (18650)	18657 (18531)	1
	τ_v	---	18670 (18533)	21005 (124899)	18704 (18551)	1
INLA Gam(1,0.5) Scaling option	τ_u	0.99 (0.20)	1.07 (0.25)	1.12 (0.27)	1.11 (0.26)	1
	τ_v	---	5.06 (2.71)	4.59 (2.47)	5.15 (2.69)	1
INLA Gam(2,1)	τ_u	1.01 (0.20)	1.06 (0.24)	1.10 (0.26)	1.10 (0.25)	1
	τ_v	---	3.08 (1.57)	2.77 (1.45)	3.08 (1.56)	1
INLA Gam(1,1)	τ_u	0.97 (0.19)	1.03 (0.23)	1.08 (0.25)	1.06 (0.24)	1
	τ_v	---	2.45 (1.33)	2.20 (1.22)	2.47 (1.32)	1
OpenBUGS Gam(2,1)	τ_u	0.96 (0.19)	1.03 (0.23)	1.07 (0.26)	1.06 (0.25)	1
	τ_v	---	3.00 (1.49)	2.73 (1.41)	3.00 (1.48)	1
OpenBUGS Gam(1,1)	τ_u	0.92 (0.18)	1.00 (0.24)	1.03 (0.25)	2.14 (1.15)	1
	τ_v	---	1.05 (0.27)	2.36 (1.25)	2.39 (1.25)	1

Figure 2: The true and average estimated uncorrelated spatial effects as calculated in `INLA` and `OpenBUGS`.

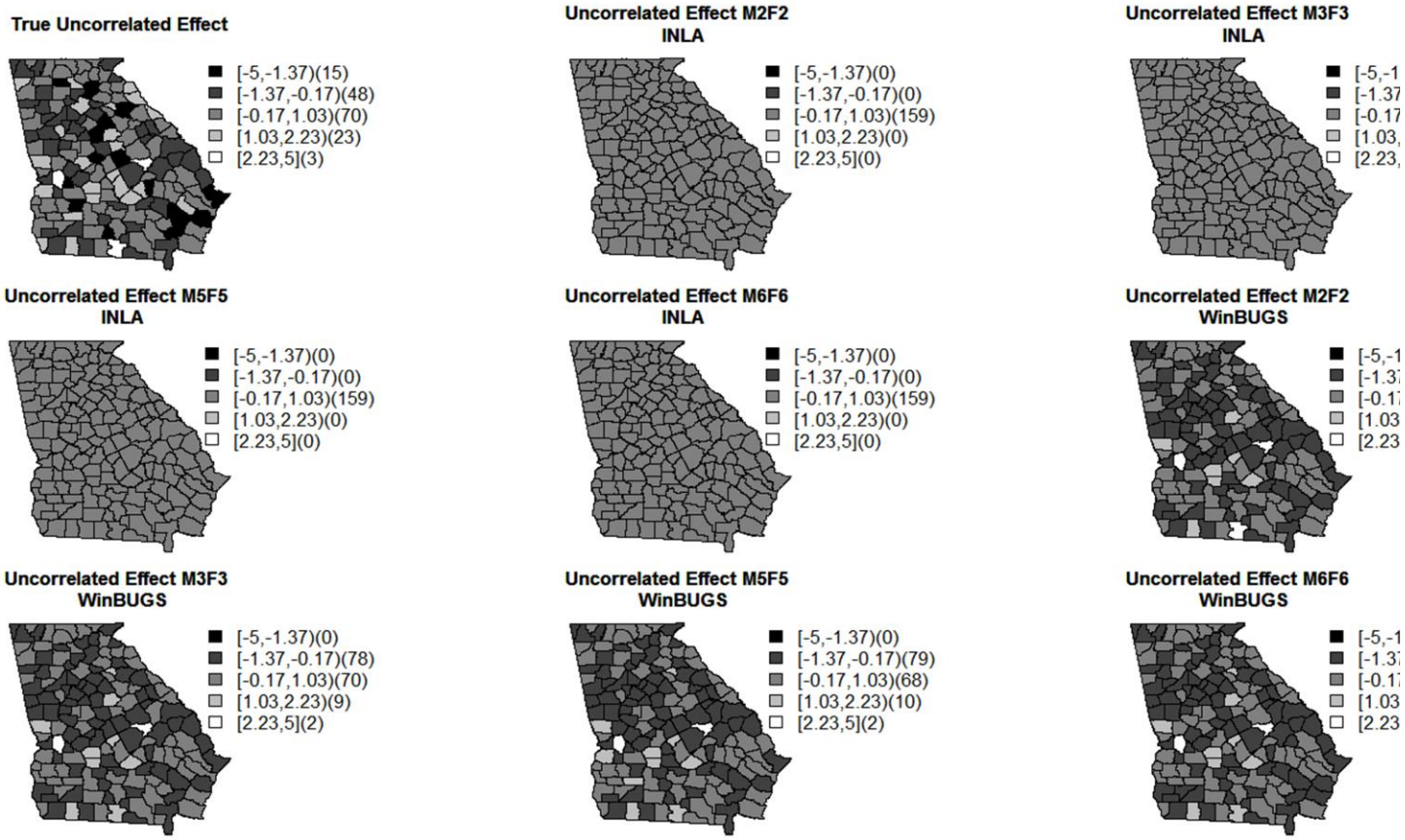


Figure 3: The true and average estimated correlated spatial effects as calculated in `INLA` and `OpenBUGS`.

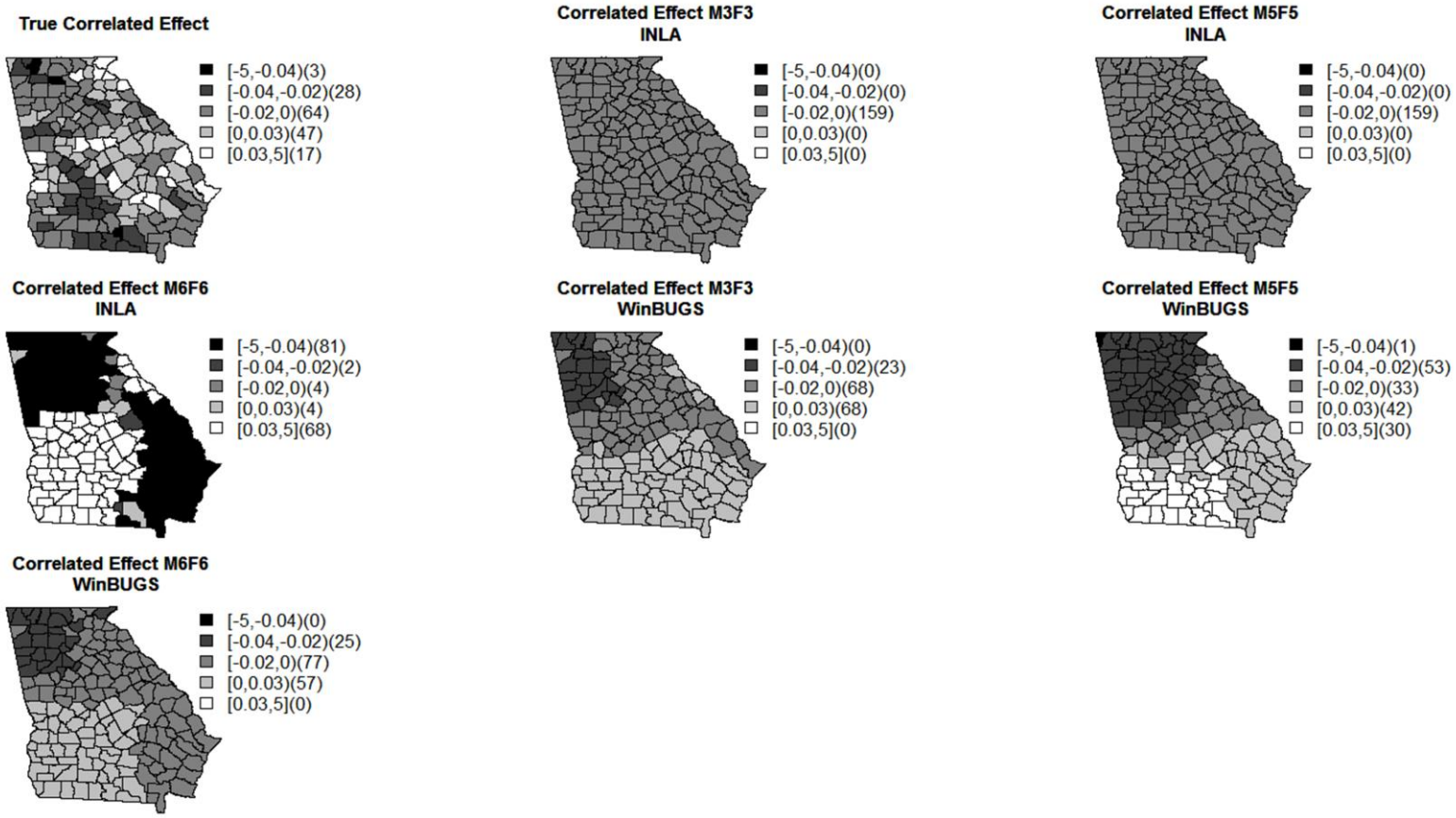


Figure 4: The true and average estimated uncorrelated spatial effects as calculated in `INLA` and `OpenBUGS` after changing the default prior distributions.

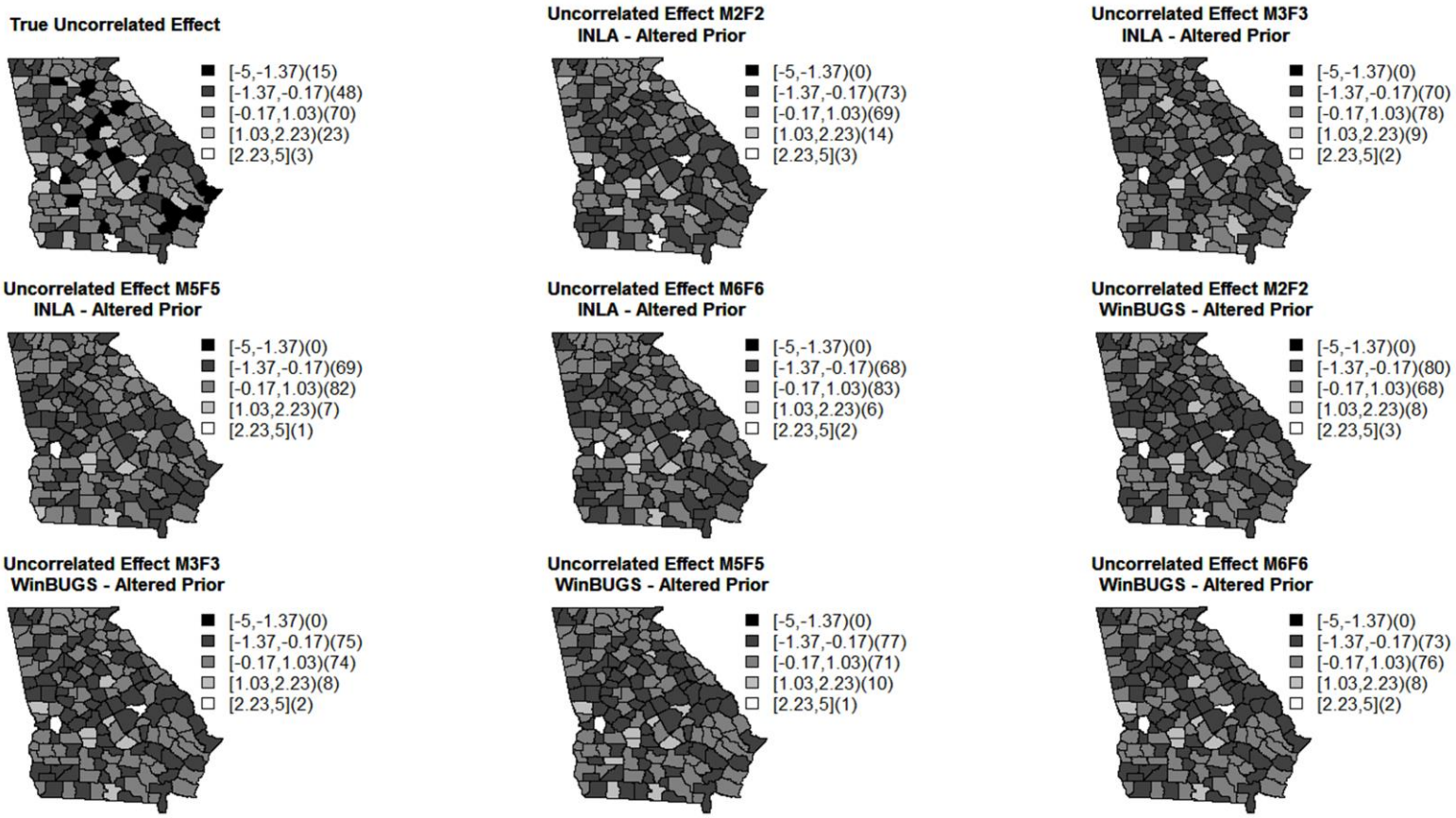


Figure 5: The true and average estimated correlated spatial effects as calculated in `INLA` and `OpenBUGS` after changing the default prior distributions.

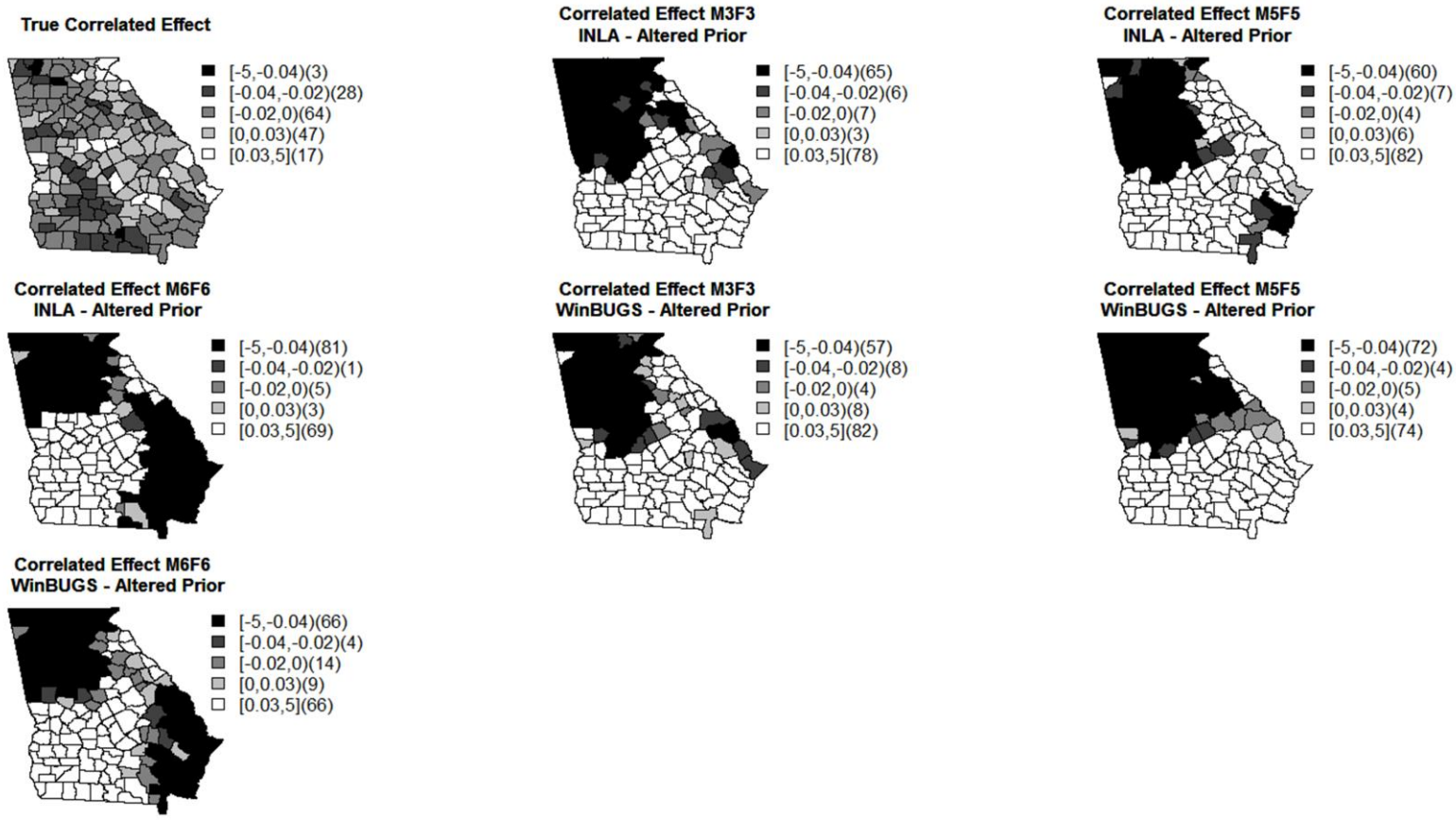


Figure 6: The true and average estimated sum of the random spatial effects as calculated in `INLA` and `OpenBUGS` after changing the default priors.

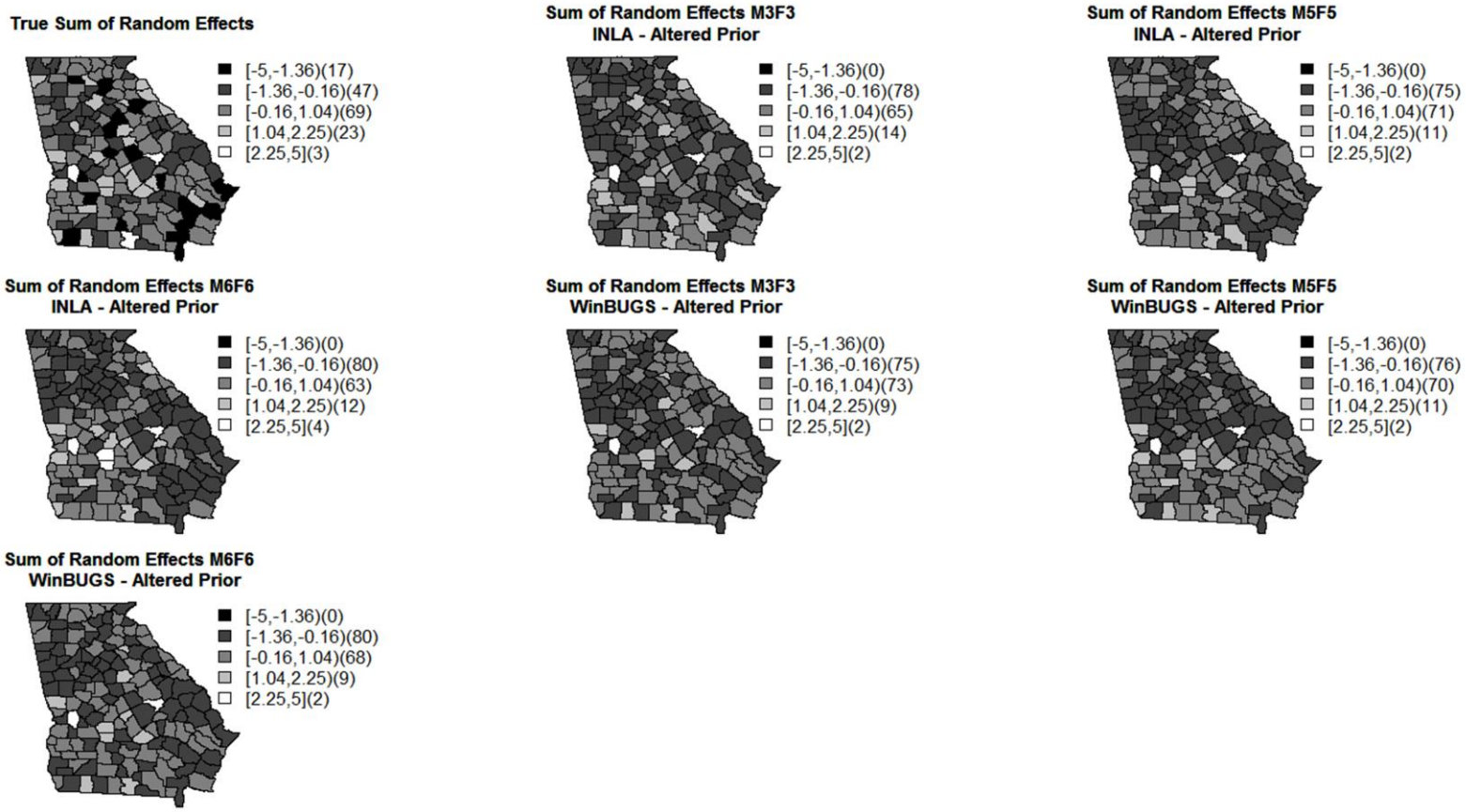


Table 3: Model re-fits with the $Gam(1,0.5)$ prior for the precisions of random effects and Laplace strategy for `INLA` with the standardized models.

	α_0	α_1	α_2	α_3	α_4	τ_u	τ_v
M5F5S `OpenBUGS`	-0.089 (0.16)	0.251 (0.13)*	-0.097 (0.51)	-0.282 (0.28)	0.540 (0.25)*	1.110 (0.28)	3.337 (2.00)
M5F5S `INLA`	-0.101 (0.17)	0.259 (0.13)	-0.023 (0.11)	-0.115 (0.11)	0.573 (0.26)*	1.134 (0.27)	3.502 (2.23)
Truth	0.1	0.1	-0.05	-0.05	0.5	1	1

Table 4: GoF estimates for model refits using the $Gam(1,0.5)$ prior and Laplace strategy with the standardized models.

Model	$\bar{D}$	pD	DIC	$MSE(\alpha)$	$MSE(M)$	$MSE(\theta)$	$MSPE$
M5F5S `OpenBUGS`	452.05	43.83	495.88	0.28	0.52	1.80	0.52
M5F5S `INLA`	452.25	85.44	537.69	0.26	0.54	1.88	0.73

Figure 7: M5F5S uncorrelated and correlated random effects as well as the truth.

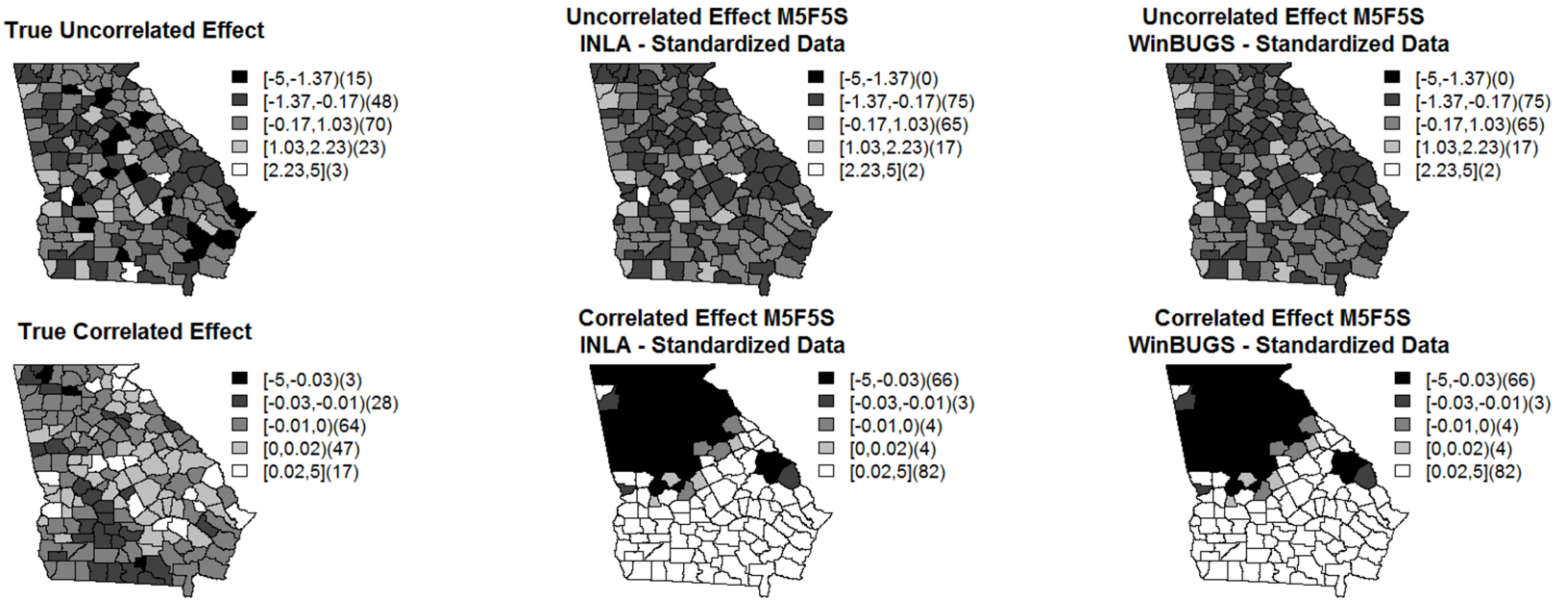


Table 5: Model re-fits with the $Gam(1, 0.5)$ prior for the precisions of random effects and Laplace strategy for `INLA` with the validation data set.

	α_0	α_1	α_2	α_3	α_4	τ_u	τ_v
S1F1V `OpenBUGS`	0.092 (0.08)	0.099 (0.03)*	-0.041 (0.15)	---	---	---	---
S1F1V `INLA`	0.092 (0.08)	0.099 (0.02)*	-0.043 (0.15)	---	---	---	---
S2F2V `OpenBUGS`	-0.020 (0.12)	0.071 (0.04)	0.103 (0.22)	---	---	0.993 (0.22)	---
S2F2V `INLA`	-0.012 (0.13)	0.071 (0.04)	0.105 (0.22)	---	---	1.050 (0.23)	---
S3F3V `OpenBUGS`	-0.009 (0.12)	0.085 (0.05)	0.137 (0.23)	---	---	1.182 (0.32)	1.133 (0.30)
S3F3V `INLA`	-0.002 (0.12)	0.085 (0.05)	0.141 (0.22)	---	---	1.229 (0.31)	3.740 (2.37)
S4F4V `OpenBUGS`	-0.129 (0.105)	0.098 (0.02)*	-0.044 (0.15)	-0.047 (0.14)	0.492 (0.14)*	---	---
S4F4V `INLA`	0.069 (0.08)	0.098 (0.02)*	-0.043 (0.16)	-0.048 (0.07)	0.498 (0.15)*	---	---
S5F5V `OpenBUGS`	-0.283 (0.17)	0.078 (0.04)	0.139 (0.22)	-0.081 (0.10)	0.658 (0.25)*	1.188 (0.32)	3.667 (2.20)
S5F5V `INLA`	-0.018 (0.12)	0.079 (0.05)	0.140 (0.23)	-0.082 (0.11)	0.708 (0.26)*	1.159 (0.28)	3.856 (2.43)
S6F6V `OpenBUGS`	-0.003 (0.12)	---	---	---	---	3.733 (2.26)	3.827 (2.26)
S6F6V `INLA`	0.003 (0.12)	---	---	---	---	1.234 (0.31)	3.867 (2.39)
Truth	0.1	0.1	-0.05	-0.05	0.5	1	0.5

Table 6: GoF estimates for models with the validation data set.

Model	$\overline{D}$	pD	DIC	$MSE(\alpha)$	$MSE(M)$	$MSE(\theta)$	$MSPE$
S1F1V OpenBUGS	437.08	2.93	440.01	0.01	1.17	0.02	1.17
S1F1V INLA	437.08	2.97	440.05	0.01	1.17	0.02	1.17
S2F2V OpenBUGS	422.98	80.05	477.14	0.02	0.51	1.24	0.51
S2F2V INLA	424.14	78.38	502.52	0.02	0.55	1.31	0.72
S3F3V OpenBUGS	426.22	50.91	477.14	0.02	0.52	1.25	0.51
S3F3V INLA	427.15	78.59	505.74	0.03	0.55	1.31	0.72
S4F4V OpenBUGS	431.38	4.70	436.0	0.22	1.12	0.04	1.12
S4F4V INLA	431.54	4.92	436.47	0.21	1.12	0.04	1.12
S5F5V OpenBUGS	431.38	4.70	436.07	0.32	0.50	1.31	0.51
S5F5V INLA	426.55	79.84	506.39	0.31	0.53	1.38	0.70
S6F6V OpenBUGS	425.27	50.37	425.27	0.02	0.53	1.24	0.53
S6F6V INLA	426.15	77.73	503.87	0.02	1.17	1.31	0.73

Figure 8: The true and average estimated uncorrelated spatial effects as calculated in `INLA` and `OpenBUGS` with the validation data set.

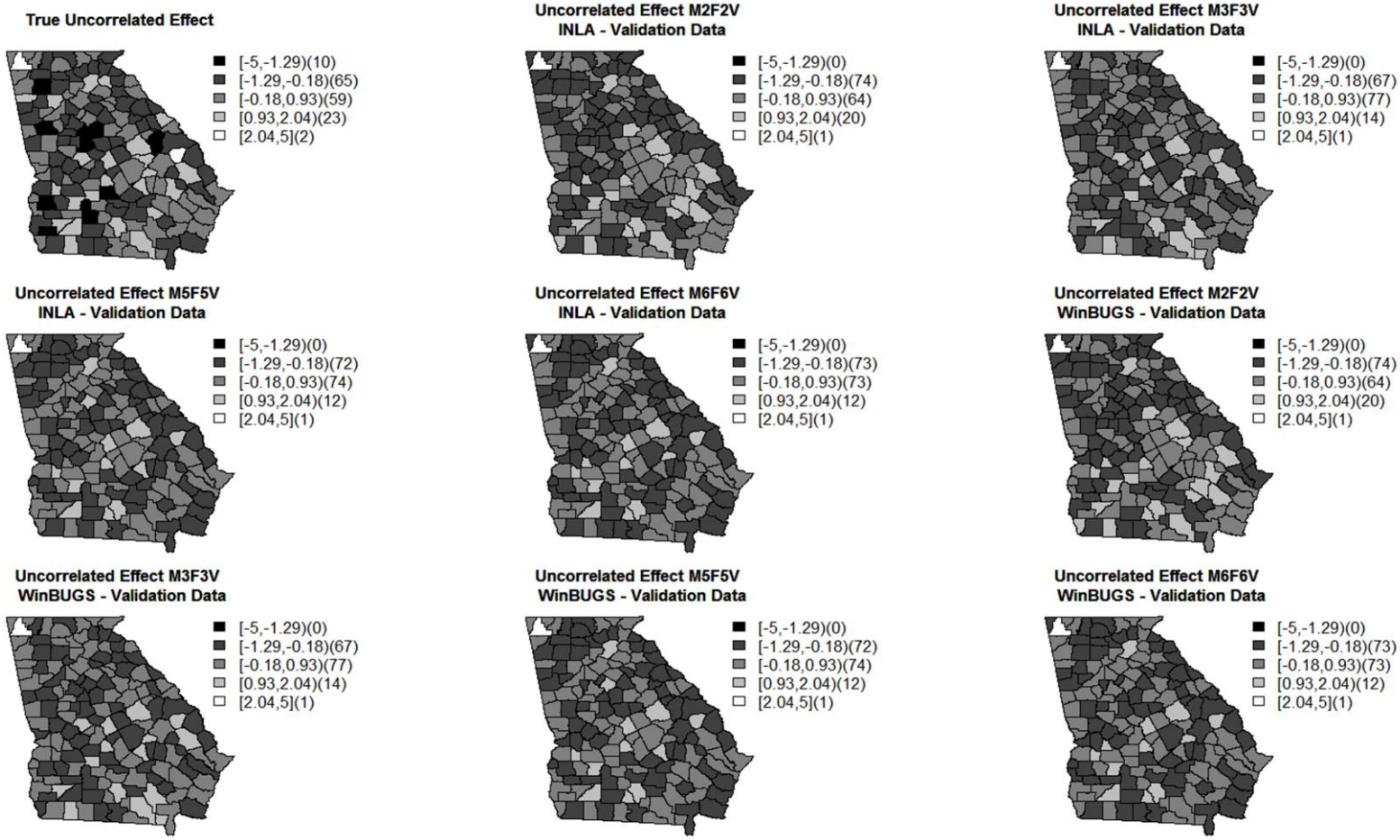


Figure 9: The true and average estimated correlated spatial effects as calculated in `INLA` and `OpenBUGS` with the validation data set.

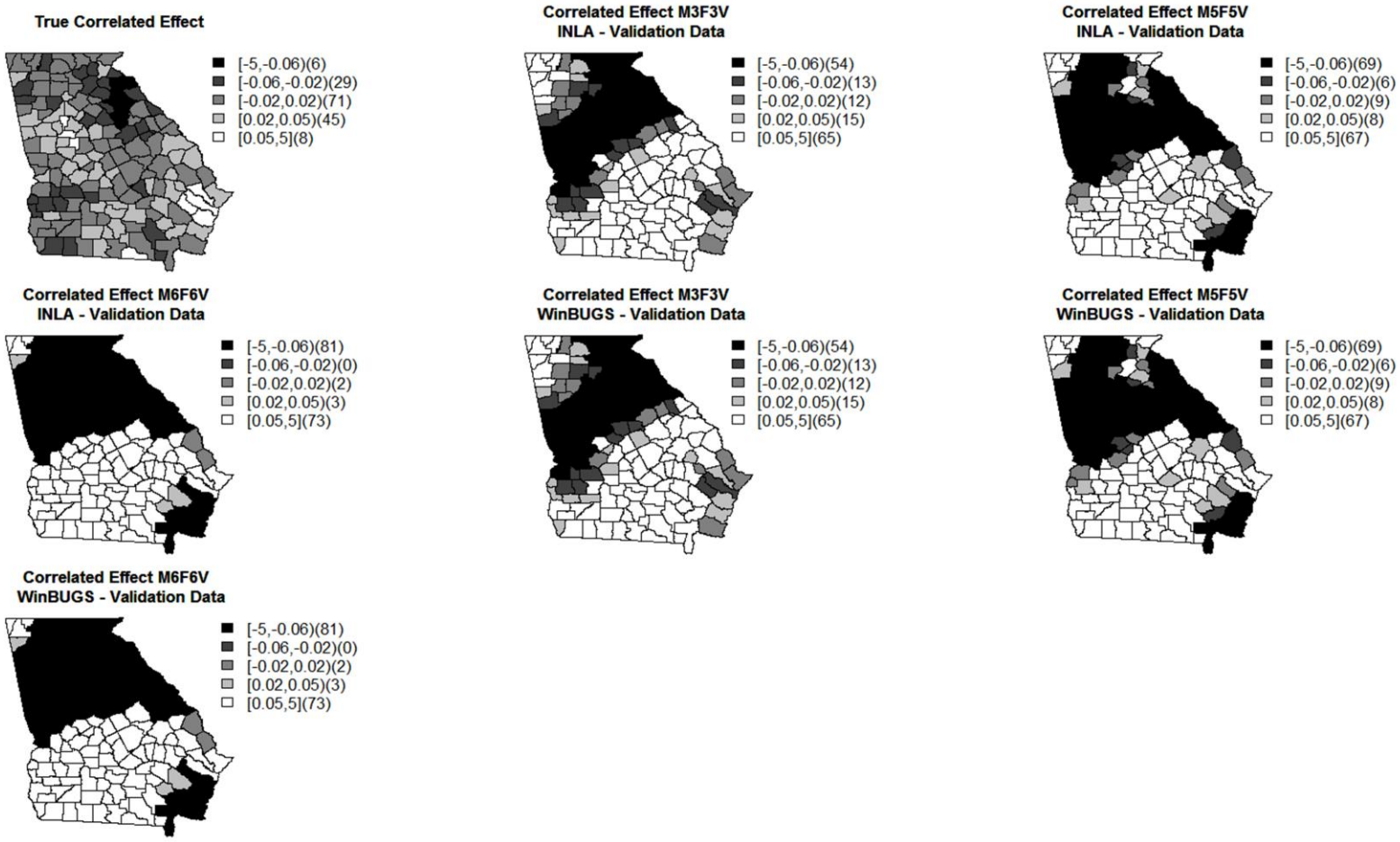
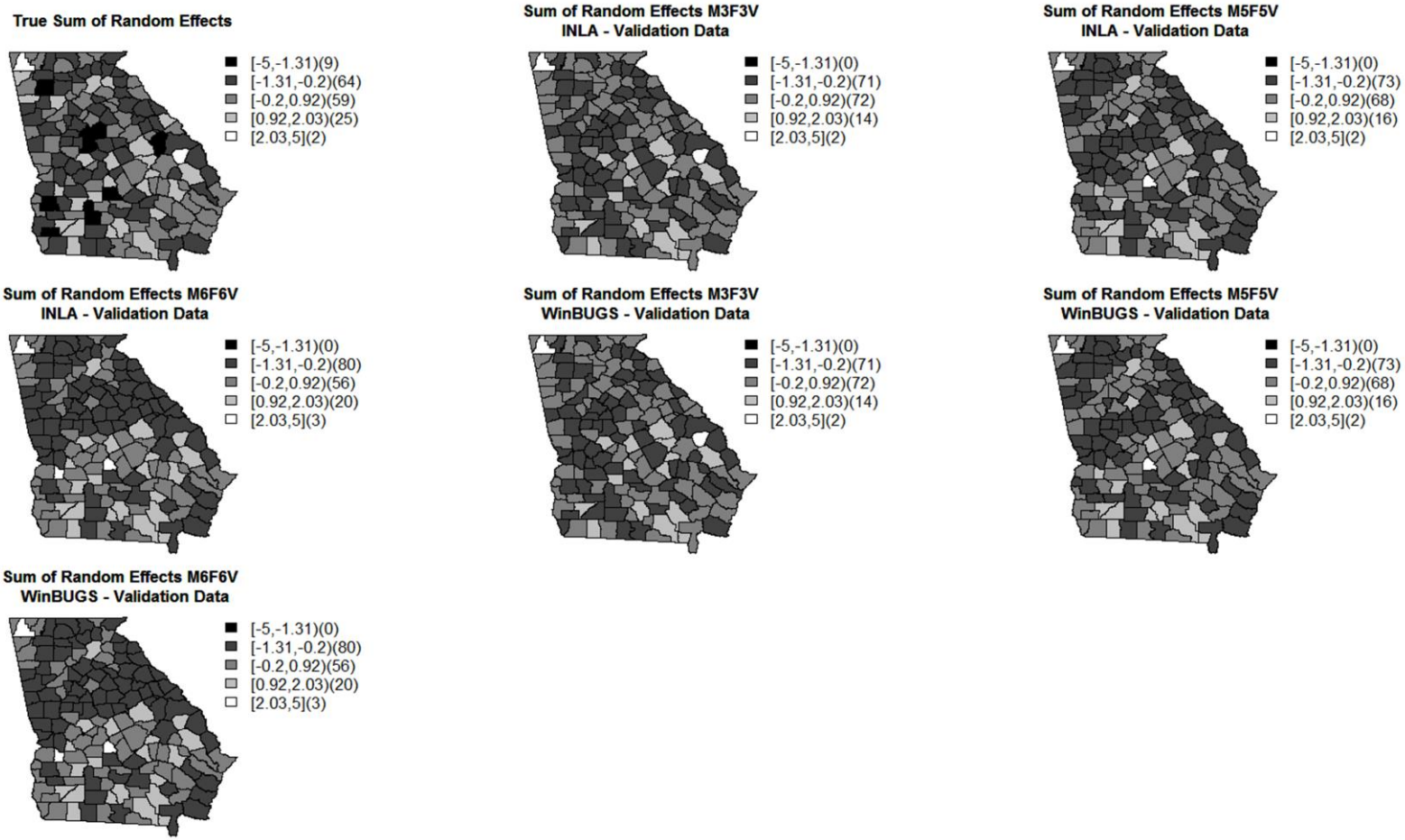


Figure 10: The true and average estimated sum of the random spatial effects as calculated in `INLA` and `OpenBUGS` with the validation data set.



A.2. Aim 2

Table 1: Descriptions of simulated data model contents

Model	Definition
E1CS1	Model S1 with $e_i = 1$ and β 's included for all regions
E2CS1	Model S1 with $e_i \sim Gam(1,1)$ and β 's included for all regions
E1CS2	Model S2 with $e_i = 1$ and β 's included for all regions
E2CS2	Model S2 with $e_i \sim Gam(1,1)$ and β 's included for all regions
E1CS3	Model S3 with $e_i = 1$ and β 's included for all regions
E2CS3	Model S3 with $e_i \sim Gam(1,1)$ and β 's included for all regions
E1CS4	Model S4 with $e_i = 1$ and β 's included for all regions
E2CS4	Model S4 with $e_i \sim Gam(1,1)$ and β 's included for all regions
E1CS5	Model S5 with $e_i = 1$ and β 's included for all regions
E2CS5	Model S5 with $e_i \sim Gam(1,1)$ and β 's included for all regions
E1CS6	Model S6 with $e_i = 1$ and β 's included for all regions
E2CS6	Model S6 with $e_i \sim Gam(1,1)$ and β 's included for all regions
E1CS7	Model S7 with $e_i = 1$ and β 's included for all regions
E2CS7	Model S7 with $e_i \sim Gam(1,1)$ and β 's included for all regions
E1CS8	Model S8 with $e_i = 1$ and β 's included for all regions
E2CS8	Model S8 with $e_i \sim Gam(1,1)$ and β 's included for all regions
E1CS9	Model S9 with $e_i = 1$ and β 's included for all regions
E2CS9	Model S9 with $e_i \sim Gam(1,1)$ and β 's included for all regions
E1PS1	$e_i = 1$ for scenario PS1
E2PS1	$e_i \sim Gam(1,1)$ for scenario PS1
E1PS2	$e_i = 1$ for scenario PS2
E2PS2	$e_i \sim Gam(1,1)$ for scenario PS2
E1PS3	$e_i = 1$ for scenario PS3
E2PS3	$e_i \sim Gam(1,1)$ for scenario PS3
E1PS4	$e_i = 1$ for scenario PS4
E2PS4	$e_i \sim Gam(1,1)$ for scenario PS4
E1PS5	$e_i = 1$ for scenario PS5
E2PS5	$e_i \sim Gam(1,1)$ for scenario PS5
E1PS6	$e_i = 1$ for scenario PS6

E2PS6	$e_i \sim Gam(1,1)$ for scenario PS6
E1PS7	$e_i = 1$ for scenario PS7
E2PS7	$e_i \sim Gam(1,1)$ for scenario PS7
E1PS8	$e_i = 1$ for scenario PS8
E2PS8	$e_i \sim Gam(1,1)$ for scenario PS8
E1PS9	$e_i = 1$ for scenario PS9
E2PS9	$e_i \sim Gam(1,1)$ for scenario PS9
E1CS1RE	Model S1 with $e_i = 1$, β 's included for all regions, and u_i is included
E1PS1RE	$e_i = 1$ for scenario PS1 and u_i is included
E1PS1CV	$e_i = 1$ for scenario PS1, u_i is included, and v_i is included

Table 2: Parameter Estimates for E1CS1 using BMS

Parameter	Mean (95% CI)	Parameter	Mean (95% CI)	Parameter	Mean (95% CI)
$\beta_{1,1}$	0.42 (-0.85,1.44)	$\beta_{2,1}$	0.22 (-1.60,1.78)	$\beta_{3,1}$	0.35 (-1.27,1.58)
$\beta_{1,2}$	0.07 (-0.91,1.03)	$\beta_{2,2}$	0.00 (-1.96,1.97)	$\beta_{3,2}$	0.03 (-1.32,1.40)
$\beta_{1,3}$	0.07 (-0.90,1.00)	$\beta_{2,3}$	0.00 (-1.96,1.96)	$\beta_{3,3}$	0.00 (-1.96,1.97)
$\beta_{1,4}$	0.06 (-1.00,1.06)	$\beta_{2,4}$	0.02 (-1.64,1.68)	$\beta_{3,4}$	0.00 (-1.96,1.96)
$\beta_{1,5}$	0.05 (-1.07,1.15)	$\beta_{2,5}$	0.04 (-1.67,1.69)	$\beta_{3,5}$	0.00 (-1.96,1.96)

Table 3: Model weights for non-spatial models fit with BMS

Model	p_1	p_2	p_3
E1CS1F1	0.706 (0.305)	0.360 (0.341)	0.536 (0.344)
E2CS1F1	0.633 (0.358)	0.310 (0.338)	0.600 (0.348)
E1CS2F2	0.712 (0.335)	0.555 (0.355)	0.229 (0.328)
E2CS2F2	0.661 (0.350)	0.582 (0.340)	0.183 (0.307)
E1CS3F3	0.873 (0.279)	0.387 (0.358)	0.397 (0.370)
E2CS3F3	0.810 (0.336)	0.422 (0.349)	0.389 (0.371)
E1CS4F4	0.360 (0.346)	0.577 (0.339)	0.410 (0.336)
E2CS4F4	0.312 (0.353)	0.459 (0.365)	0.412 (0.378)
E1CS5F5	0.660 (0.347)	0.578 (0.345)	0.089 (0.216)
E2CS5F5	0.661 (0.344)	0.537 (0.352)	0.080 (0.211)
E1CS6F6	0.338 (0.359)	0.810 (0.278)	0.071 (0.175)
E2CS6F6	0.361 (0.378)	0.797 (0.290)	0.073 (0.179)
E1CS7F7	0.907 (0.217)	0.373 (0.350)	0.339 (0.350)
E2CS7F7	0.943 (0.166)	0.324 (0.331)	0.318 (0.332)
E1CS8F8	0.806 (0.319)	0.441 (0.355)	0.346 (0.379)
E2CS8F8	0.769 (0.345)	0.427 (0.354)	0.238 (0.342)
E1CS9F9	0.886 (0.250)	0.526 (0.365)	0.021 (0.092)
E2CS9F9	0.885 (0.252)	0.531 (0.372)	0.042 (0.155)
E1CS1F1RE	0.214 (0.269)	0. 161 (0.240)	0. 693 (0.292)

Table 4: GoF measures for models fit with BMS.

	$\bar{D}$	DIC	$\frac{Var(\bar{D})}{2}$
E1CS1F1	603.42	608.84	5.41
E2CS1F1	516.11	521.23	5.12
E1CS2F2	589.86	619.20	29.34
E2CS2F2	502.85	523.14	20.30
E1CS3F3	590.29	636.90	46.61
E2CS3F3	504.63	544.59	39.96
E1CS4F4	584.00	648.26	64.26
E2CS4F4	501.47	537.60	36.13
E1CS5F5	597.41	600.68	3.27
E2CS5F5	507.01	510.24	3.23
E1CS6F6	603.40	605.02	1.62
E2CS6F6	511.63	513.36	1.73
E1CS7F7	585.76	590.79	5.03
E2CS7F7	477.80	482.92	5.12
E1CS8F8	586.38	627.13	40.77
E2CS8F8	497.38	516.05	18.67
E1CS9F9	627.74	630.17	2.43
E2CS9F9	530.14	532.55	2.41
E1CS1F1RE	578.19	645.47	67.28

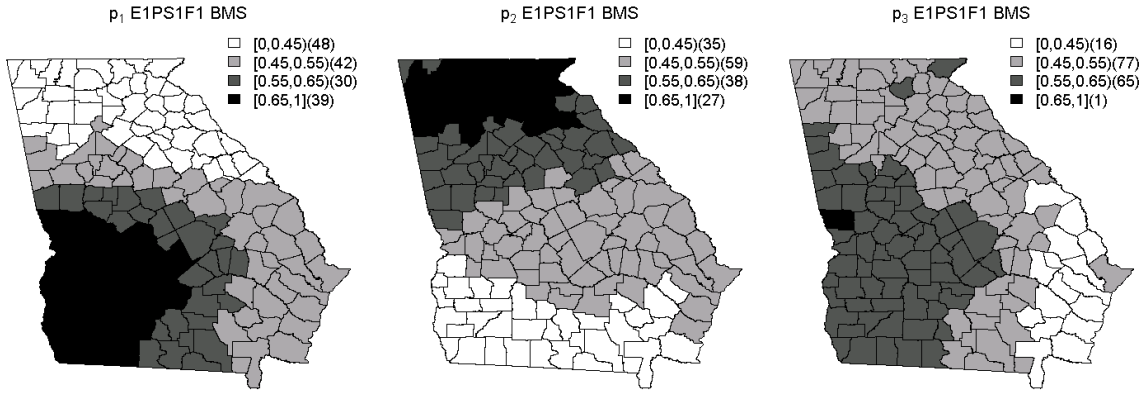
Table 5: Alternative model probabilities for complete models fit with BMA

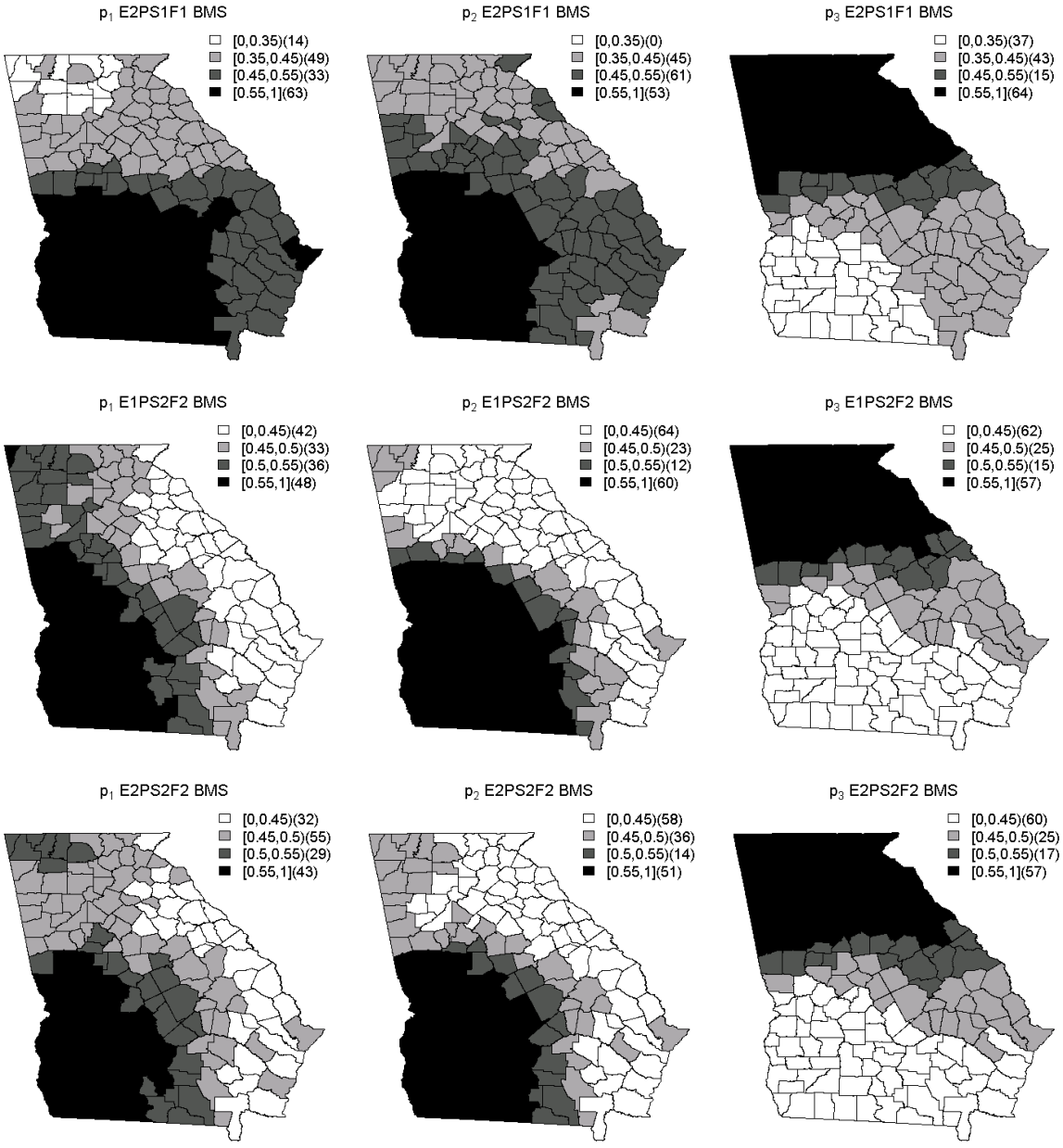
Model	p_1	p_2	p_3	p_1^{DIC1}	p_2^{DIC1}	p_3^{DIC1}	p_1^{DIC2}	p_2^{DIC2}	p_3^{DIC2}
E1CS1F1	0.406	0.283	0.298	1.000	0.000	0.000	0.044	0.932	0.024
E2CS1F1	0.405	0.283	0.300	1.000	0.000	0.000	0.000	0.660	0.338
E1CS2F2	0.283	0.283	0.419	0.000	0.000	1.000	0.000	0.000	1.000
E2CS2F2	0.283	0.284	0.418	0.000	0.000	1.000	0.000	0.000	1.000
E1CS3F3	0.283	0.283	0.419	0.000	0.000	1.000	0.000	0.000	1.000
E2CS3F3	0.283	0.284	0.418	0.000	0.001	0.999	0.000	0.000	1.000
E1CS4F4	0.283	0.284	0.419	0.000	0.000	1.000	0.003	0.004	0.993
E2CS4F4	0.283	0.283	0.418	0.000	0.000	1.000	0.000	0.000	1.000
E1CS5F5	0.392	0.315	0.283	0.999	0.001	0.000	0.009	0.003	0.988
E2CS5F5	0.394	0.314	0.283	0.970	0.030	0.000	0.000	0.000	1.000
E1CS6F6	0.345	0.320	0.332	0.165	0.003	0.833	0.571	0.193	0.236
E2CS6F6	0.349	0.317	0.330	0.009	0.001	0.989	0.969	0.021	0.010
E1CS7F7	0.418	0.284	0.284	1.000	0.000	0.000	1.000	0.000	0.000
E2CS7F7	0.418	0.285	0.283	1.000	0.000	0.000	0.000	0.000	1.000
E1CS8F8	0.351	0.287	0.348	0.000	0.000	1.000	0.000	0.000	1.000
E2CS8F8	0.284	0.283	0.418	0.00	0.000	1.000	0.000	0.255	0.745
E1CS9F9	0.409	0.288	0.290	0.999	0.000	0.001	0.200	0.261	0.538
E2CS9F9	0.410	0.287	0.290	0.982	0.000	0.018	0.000	0.833	0.167
E1CS1F1RE	0.388	0.284	0.319	1.000	0.000	0.000	0.042	0.899	0.059

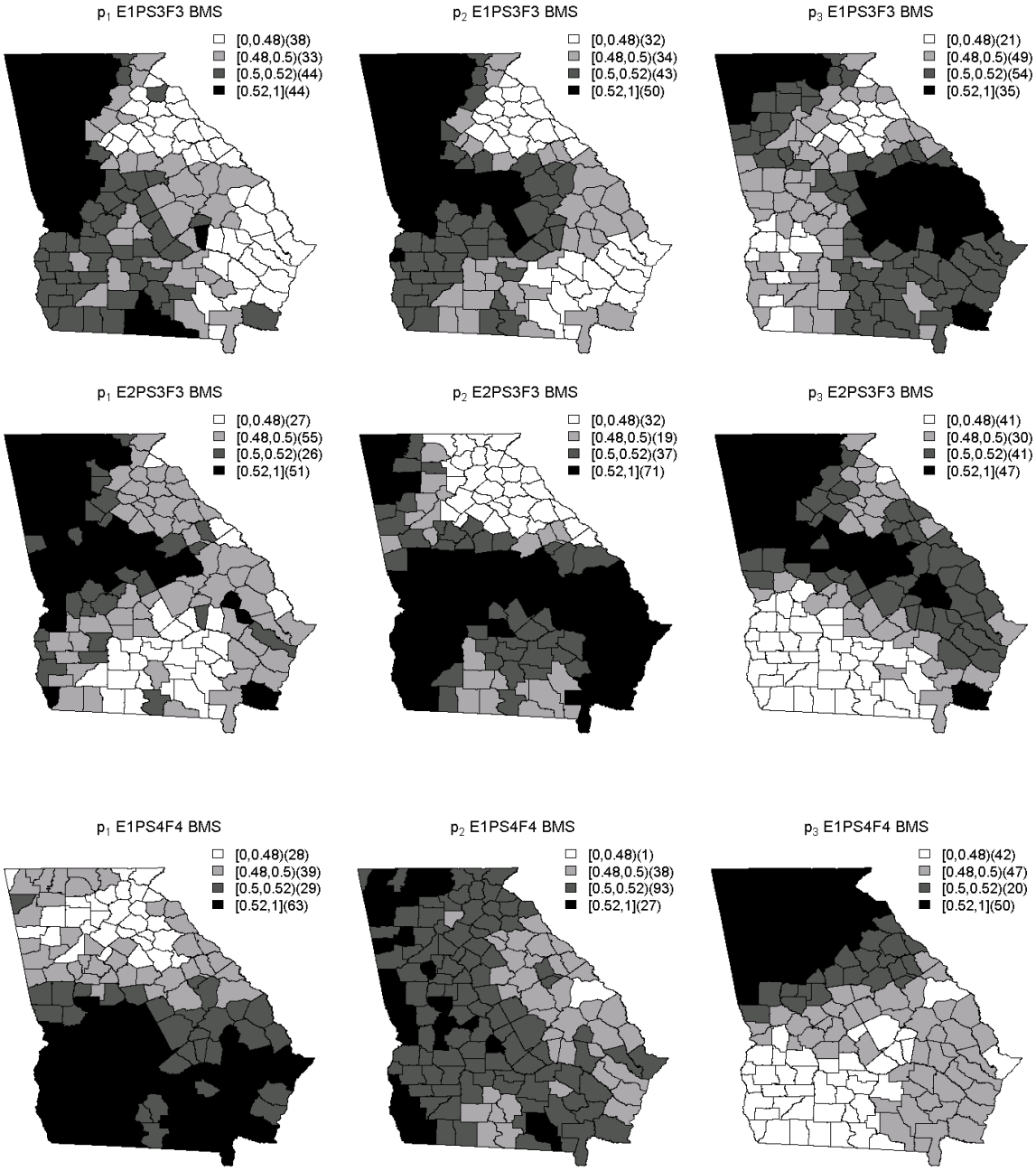
Table 6: GoF measures for complete models fit with BMA.

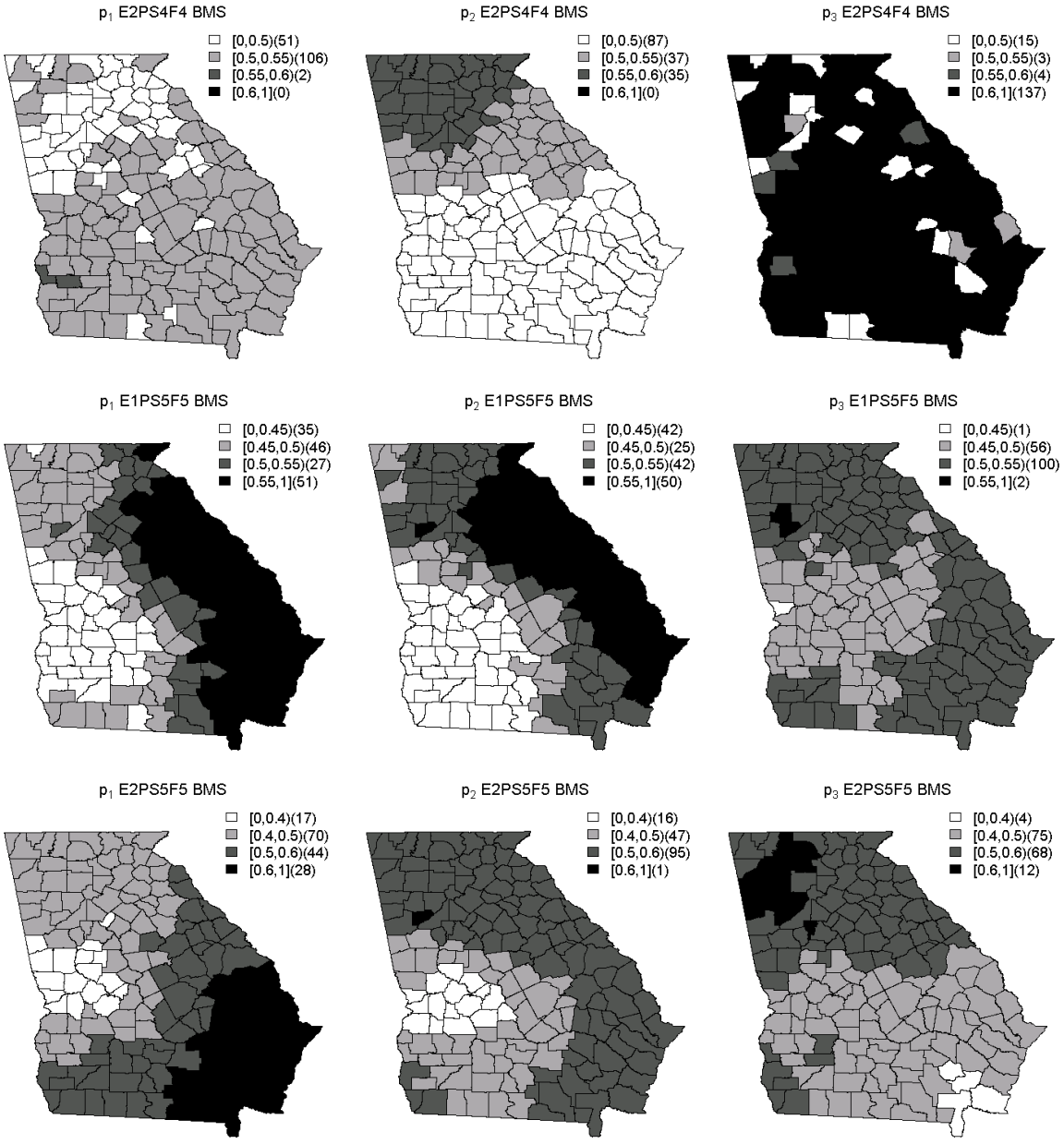
	$\bar{D}$	*DIC*	$\frac{Var(\bar{D})}{2}$
E1CS1F1	2441.81	2452.82	11.01
E2CS1F1	2096.86	2107.92	11.06
E1CS2F2	2228.79	2402.97	174.18
E2CS2F2	1903.82	2061.43	157.61
E1CS3F3	2276.39	2447.25	170.86
E2CS3F3	1956.93	2110.76	153.83
E1CS4F4	2227.32	2389.90	162.58
E2CS4F4	1905.62	2052.62	147.00
E1CS5F5	2417.5	2426.21	8.71
E2CS5F5	2062.73	2071.36	8.63
E1CS6F6	2413.35	2420.06	6.71
E2CS6F6	2046.98	2053.79	6.81
E1CS7F7	2392.73	2402.76	10.03
E2CS7F7	2062.94	2073.00	10.06
E1CS8F8	2294.05	2402.46	108.41
E2CS8F8	1902.00	2052.46	150.45
E1CS9F9	2520.96	2527.68	6.72
E2CS9F9	2143.05	2149.81	6.76
E1CS1F1RE	2229.25	2584.24	354.99

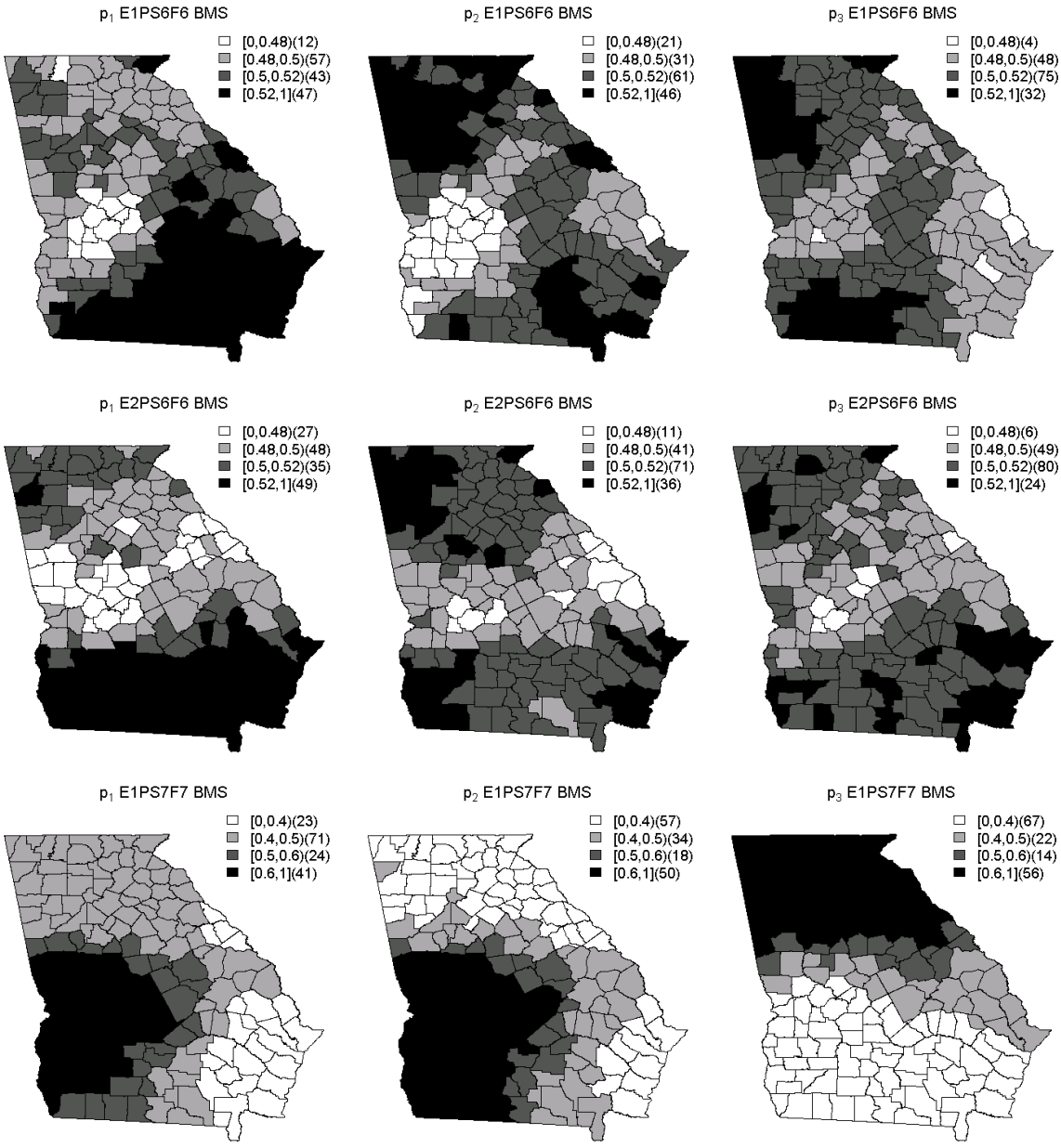
Figure 1: Model weights associated with the alternative linear predictors for BMS.

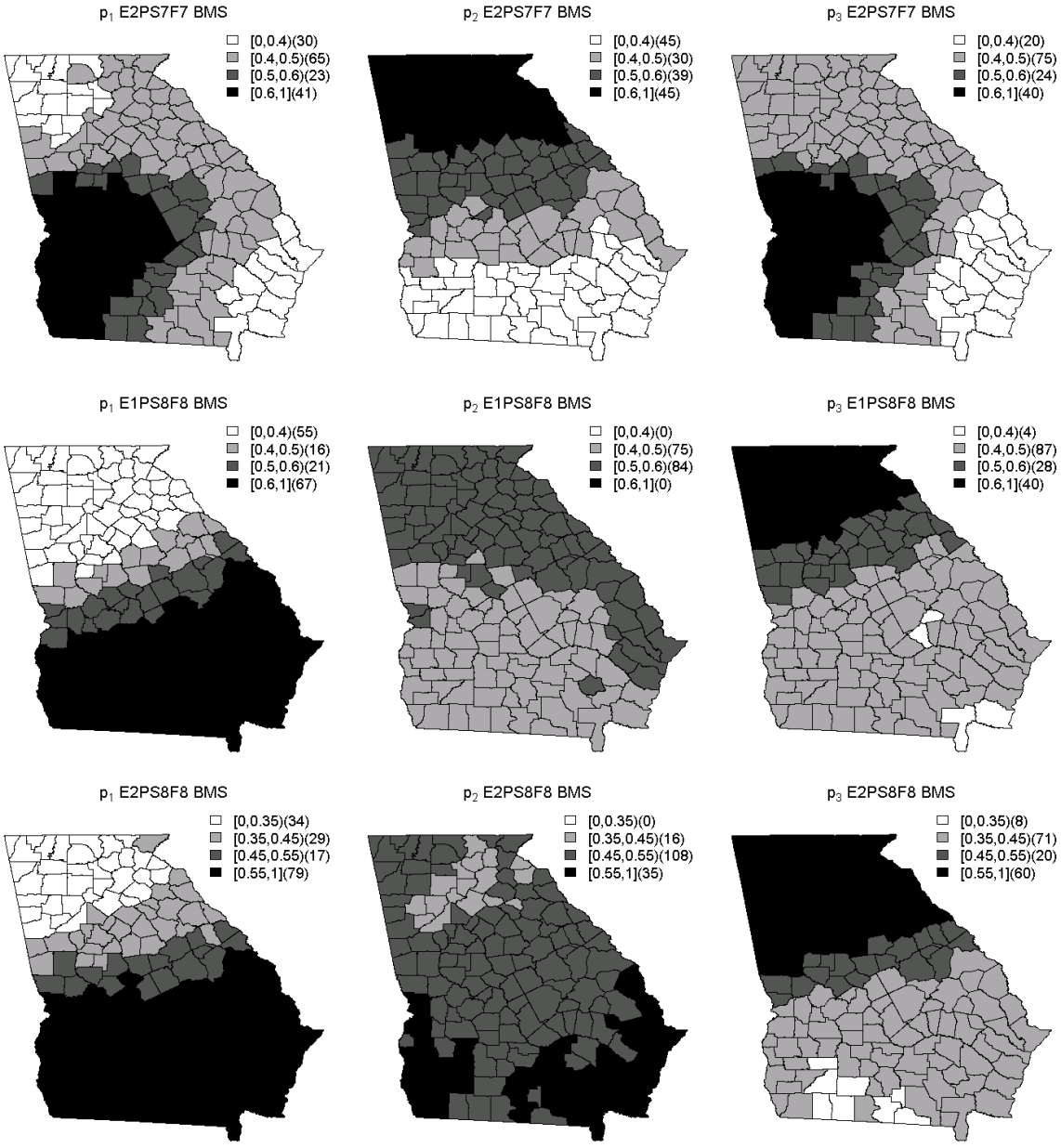


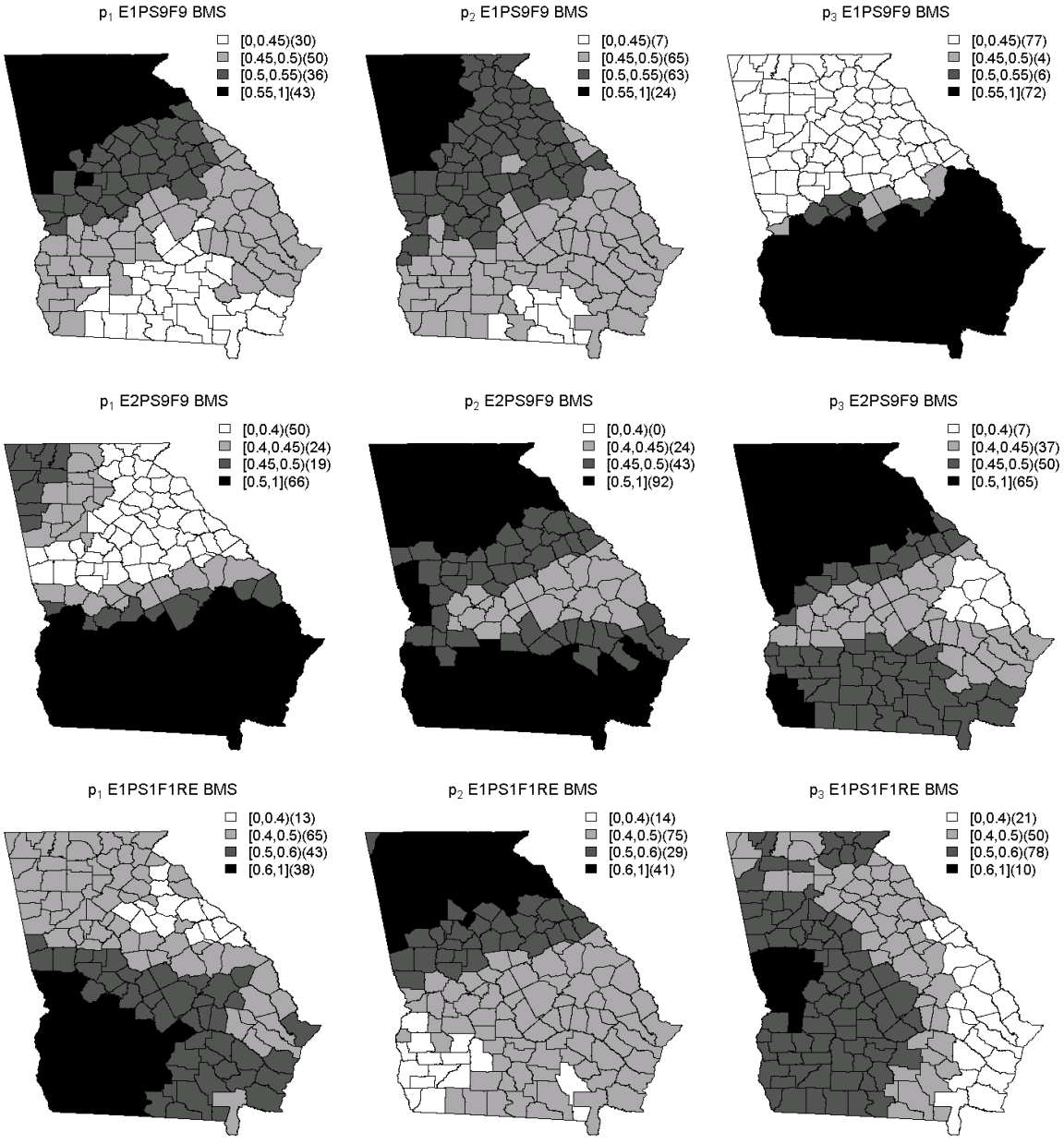

p_1 E2PS1F1 BMS
[0,0.35)(14)
[0.35,0.45)(49)
[0.45,0.55)(33)
[0.55,1](63)
p_2 E2PS1F1 BMS
[0,0.35)(0)
[0.35,0.45)(45)
[0.45,0.55)(61)
[0.55,1](53)
p_3 E2PS1F1 BMS
[0,0.35)(37)
[0.35,0.45)(43)
[0.45,0.55)(15)
[0.55,1](64)
p_1 E1PS2F2 BMS
[0,0.45)(42)
[0.45,0.5)(33)
[0.5,0.55)(36)
[0.55,1](48)
p_2 E1PS2F2 BMS
[0,0.45)(64)
[0.45,0.5)(23)
[0.5,0.55)(12)
[0.55,1](60)
p_3 E1PS2F2 BMS
[0,0.45)(62)
[0.45,0.5)(25)
[0.5,0.55)(15)
[0.55,1](57)
p_1 E2PS2F2 BMS
[0,0.45)(32)
[0.45,0.5)(55)
[0.5,0.55)(29)
[0.55,1](43)
p_2 E2PS2F2 BMS
[0,0.45)(58)
[0.45,0.5)(36)
[0.5,0.55)(14)
[0.55,1](51)
p_3 E2PS2F2 BMS
[0,0.45)(60)
[0.45,0.5)(25)
[0.5,0.55)(17)
[0.55,1](57)


p_1 E1PS3F3 BMS
[0,0.48)(38)
[0.48,0.5)(33)
[0.5,0.52)(44)
[0.52,1](44)
p_2 E1PS3F3 BMS
[0,0.48)(32)
[0.48,0.5)(34)
[0.5,0.52)(43)
[0.52,1](50)
p_3 E1PS3F3 BMS
[0,0.48)(21)
[0.48,0.5)(49)
[0.5,0.52)(54)
[0.52,1](35)
p_1 E2PS3F3 BMS
[0,0.48)(27)
[0.48,0.5)(55)
[0.5,0.52)(26)
[0.52,1](51)
p_2 E2PS3F3 BMS
[0,0.48)(32)
[0.48,0.5)(19)
[0.5,0.52)(37)
[0.52,1](71)
p_3 E2PS3F3 BMS
[0,0.48)(41)
[0.48,0.5)(30)
[0.5,0.52)(41)
[0.52,1](47)
p_1 E1PS4F4 BMS
[0,0.48)(28)
[0.48,0.5)(39)
[0.5,0.52)(29)
[0.52,1](63)
p_2 E1PS4F4 BMS
[0,0.48)(1)
[0.48,0.5)(38)
[0.5,0.52)(93)
[0.52,1](27)
p_3 E1PS4F4 BMS
[0,0.48)(42)
[0.48,0.5)(47)
[0.5,0.52)(20)
[0.52,1](50)


p1 E2PS4F4 BMS
[0,0.5)(51)
[0.5,0.55)(106)
[0.55,0.6)(2)
[0.6,1](0)
p2 E2PS4F4 BMS
[0,0.5)(87)
[0.5,0.55)(37)
[0.55,0.6)(35)
[0.6,1](0)
p3 E2PS4F4 BMS
[0,0.5)(15)
[0.5,0.55)(3)
[0.55,0.6)(4)
[0.6,1](137)
p1 E1PS5F5 BMS
[0,0.45)(35)
[0.45,0.5)(46)
[0.5,0.55)(27)
[0.55,1](51)
p2 E1PS5F5 BMS
[0,0.45)(42)
[0.45,0.5)(25)
[0.5,0.55)(42)
[0.55,1](50)
p3 E1PS5F5 BMS
[0,0.45)(1)
[0.45,0.5)(56)
[0.5,0.55)(100)
[0.55,1](2)
p1 E2PS5F5 BMS
[0,0.4)(17)
[0.4,0.5)(70)
[0.5,0.6)(44)
[0.6,1](28)
p2 E2PS5F5 BMS
[0,0.4)(16)
[0.4,0.5)(47)
[0.5,0.6)(95)
[0.6,1](1)
p3 E2PS5F5 BMS
[0,0.4)(4)
[0.4,0.5)(75)
[0.5,0.6)(68)
[0.6,1](12)


p1 E1PS6F6 BMS
[0,0.48)(12)
[0.48,0.5)(57)
[0.5,0.52)(43)
[0.52,1](47)
p2 E1PS6F6 BMS
[0,0.48)(21)
[0.48,0.5)(31)
[0.5,0.52)(61)
[0.52,1](46)
p3 E1PS6F6 BMS
[0,0.48)(4)
[0.48,0.5)(48)
[0.5,0.52)(75)
[0.52,1](32)
p1 E2PS6F6 BMS
[0,0.48)(27)
[0.48,0.5)(48)
[0.5,0.52)(35)
[0.52,1](49)
p2 E2PS6F6 BMS
[0,0.48)(11)
[0.48,0.5)(41)
[0.5,0.52)(71)
[0.52,1](36)
p3 E2PS6F6 BMS
[0,0.48)(6)
[0.48,0.5)(49)
[0.5,0.52)(80)
[0.52,1](24)
p1 E1PS7F7 BMS
[0,0.4)(23)
[0.4,0.5)(71)
[0.5,0.6)(24)
[0.6,1](41)
p2 E1PS7F7 BMS
[0,0.4)(57)
[0.4,0.5)(34)
[0.5,0.6)(18)
[0.6,1](50)
p3 E1PS7F7 BMS
[0,0.4)(67)
[0.4,0.5)(22)
[0.5,0.6)(14)
[0.6,1](56)


p_1 E2PS7F7 BMS
[0,0.4)(30)
[0.4,0.5)(65)
[0.5,0.6)(23)
[0.6,1](41)
p_2 E2PS7F7 BMS
[0,0.4)(45)
[0.4,0.5)(30)
[0.5,0.6)(39)
[0.6,1](45)
p_3 E2PS7F7 BMS
[0,0.4)(20)
[0.4,0.5)(75)
[0.5,0.6)(24)
[0.6,1](40)
p_1 E1PS8F8 BMS
[0,0.4)(55)
[0.4,0.5)(16)
[0.5,0.6)(21)
[0.6,1](67)
p_2 E1PS8F8 BMS
[0,0.4)(0)
[0.4,0.5)(75)
[0.5,0.6)(84)
[0.6,1](0)
p_3 E1PS8F8 BMS
[0,0.4)(4)
[0.4,0.5)(87)
[0.5,0.6)(28)
[0.6,1](40)
p_1 E2PS8F8 BMS
[0,0.35)(34)
[0.35,0.45)(29)
[0.45,0.55)(17)
[0.55,1](79)
p_2 E2PS8F8 BMS
[0,0.35)(0)
[0.35,0.45)(16)
[0.45,0.55)(108)
[0.55,1](35)
p_3 E2PS8F8 BMS
[0,0.35)(8)
[0.35,0.45)(71)
[0.45,0.55)(20)
[0.55,1](60)


p1 E1PS9F9 BMS
[0,0.45)(30)
[0.45,0.5)(50)
[0.5,0.55)(36)
[0.55,1](43)
p2 E1PS9F9 BMS
[0,0.45)(7)
[0.45,0.5)(65)
[0.5,0.55)(63)
[0.55,1](24)
p3 E1PS9F9 BMS
[0,0.45)(77)
[0.45,0.5)(4)
[0.5,0.55)(6)
[0.55,1](72)
p1 E2PS9F9 BMS
[0,0.4)(50)
[0.4,0.45)(24)
[0.45,0.5)(19)
[0.5,1](66)
p2 E2PS9F9 BMS
[0,0.4)(0)
[0.4,0.45)(24)
[0.45,0.5)(43)
[0.5,1](92)
p3 E2PS9F9 BMS
[0,0.4)(7)
[0.4,0.45)(37)
[0.45,0.5)(50)
[0.5,1](65)
p1 E1PS1F1RE BMS
[0,0.4)(13)
[0.4,0.5)(65)
[0.5,0.6)(43)
[0.6,1](38)
p2 E1PS1F1RE BMS
[0,0.4)(14)
[0.4,0.5)(75)
[0.5,0.6)(29)
[0.6,1](41)
p3 E1PS1F1RE BMS
[0,0.4)(21)
[0.4,0.5)(50)
[0.5,0.6)(78)
[0.6,1](10)

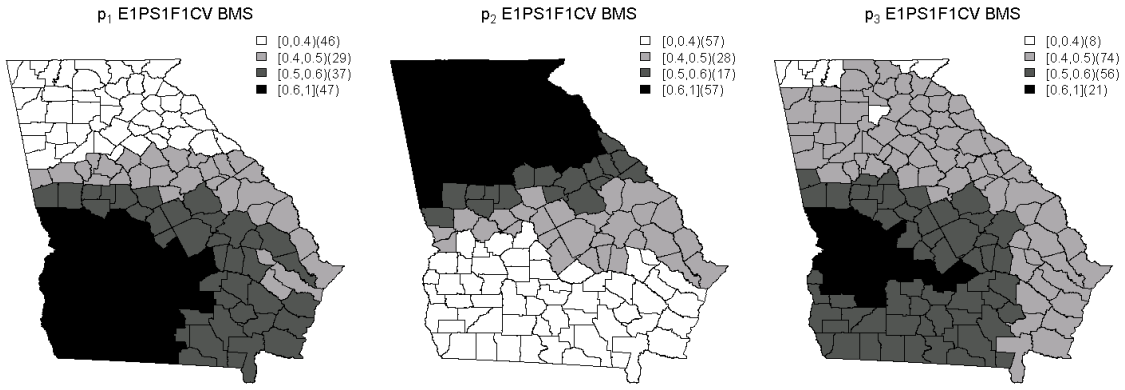
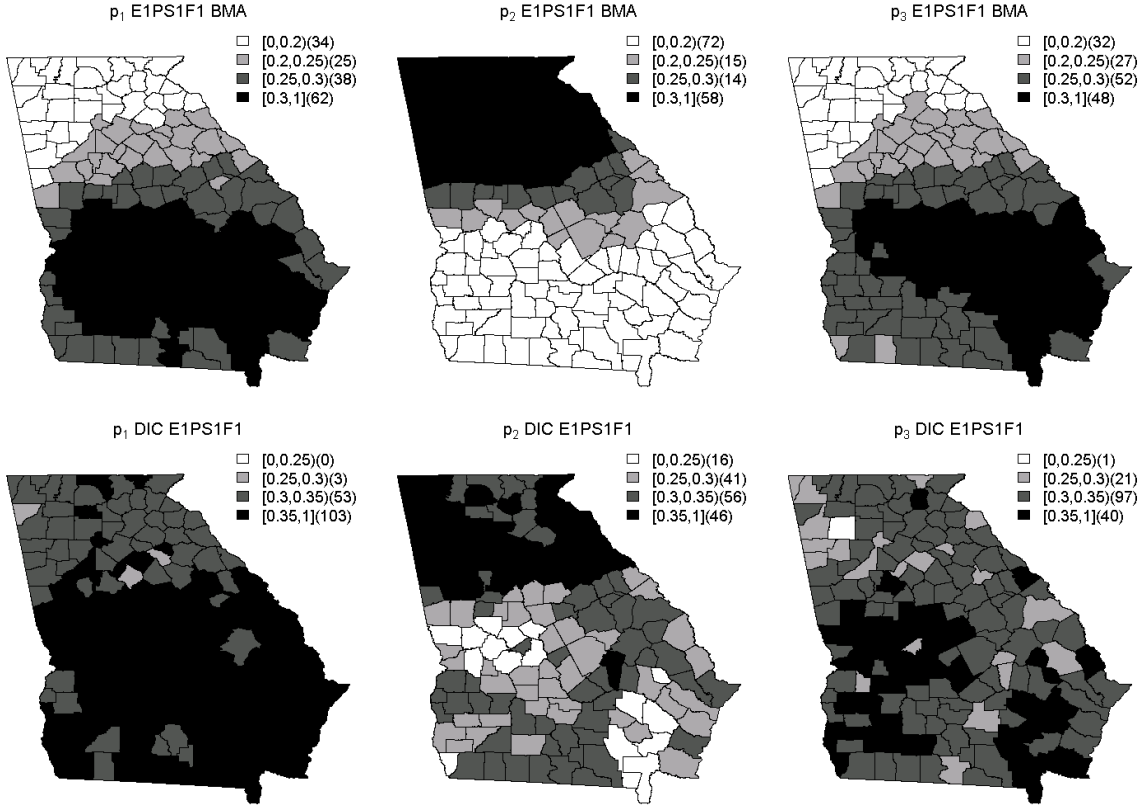
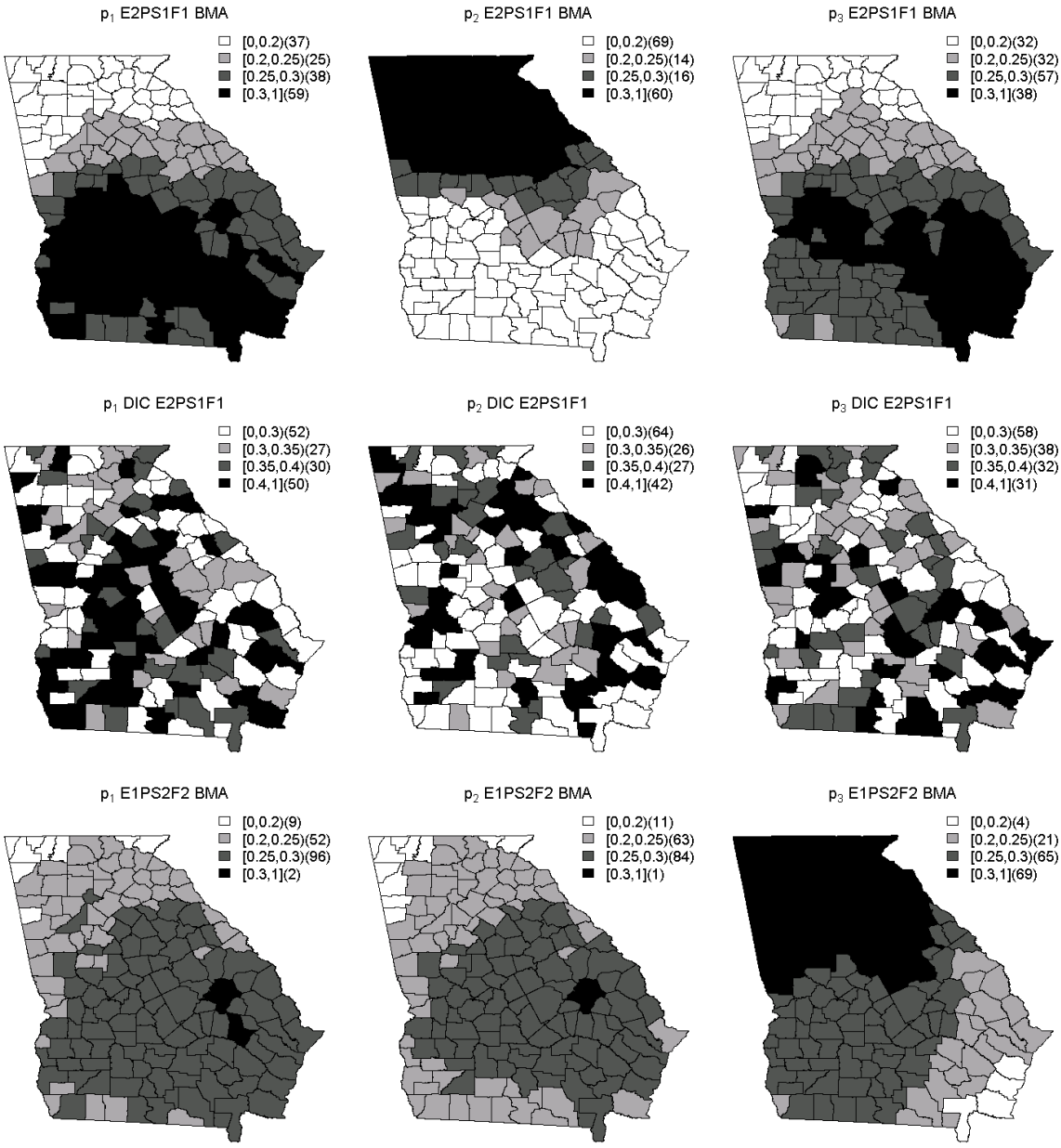
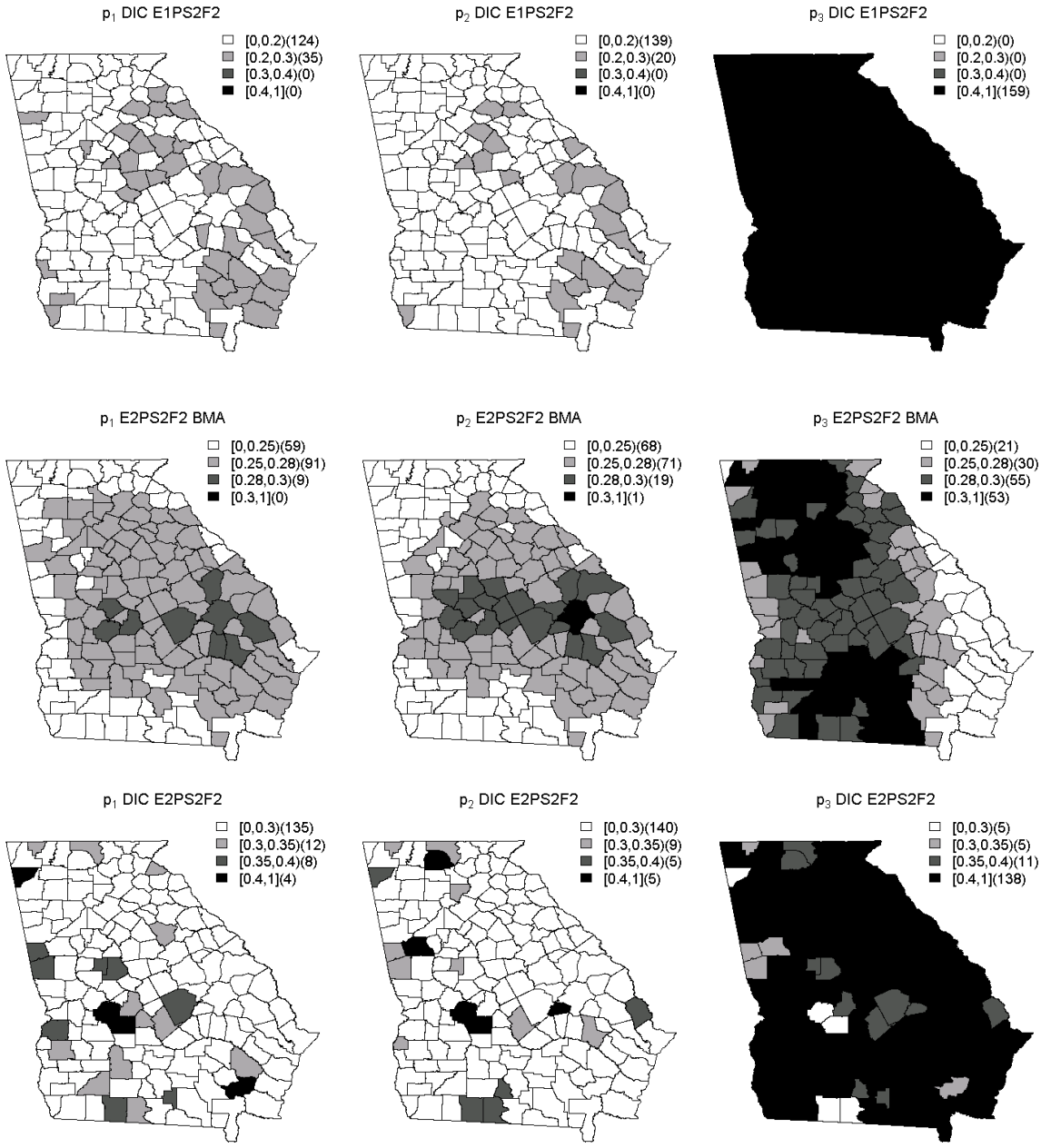


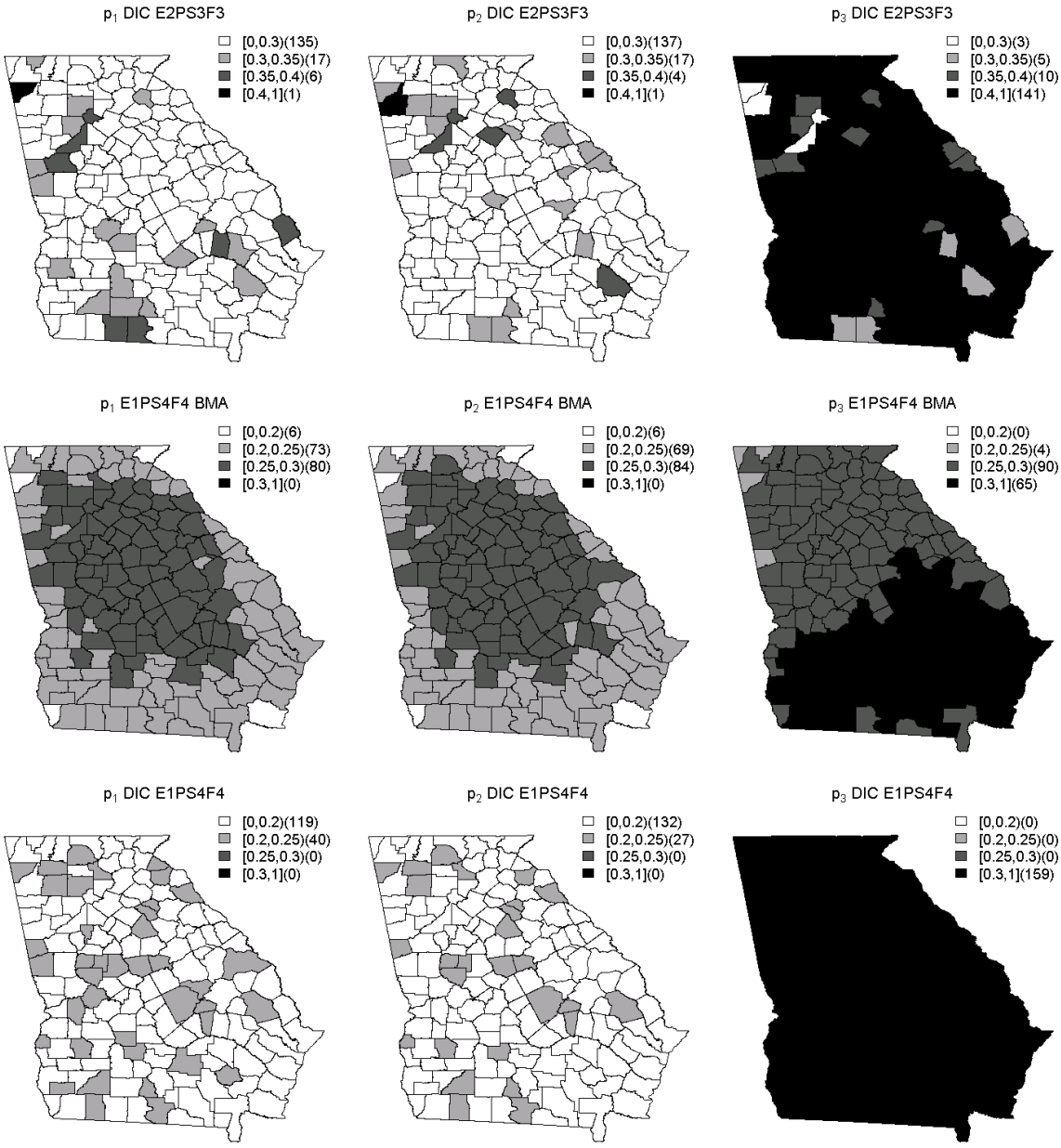
Figure 2: Model probabilities associated with the alternative linear predictors for BMA.

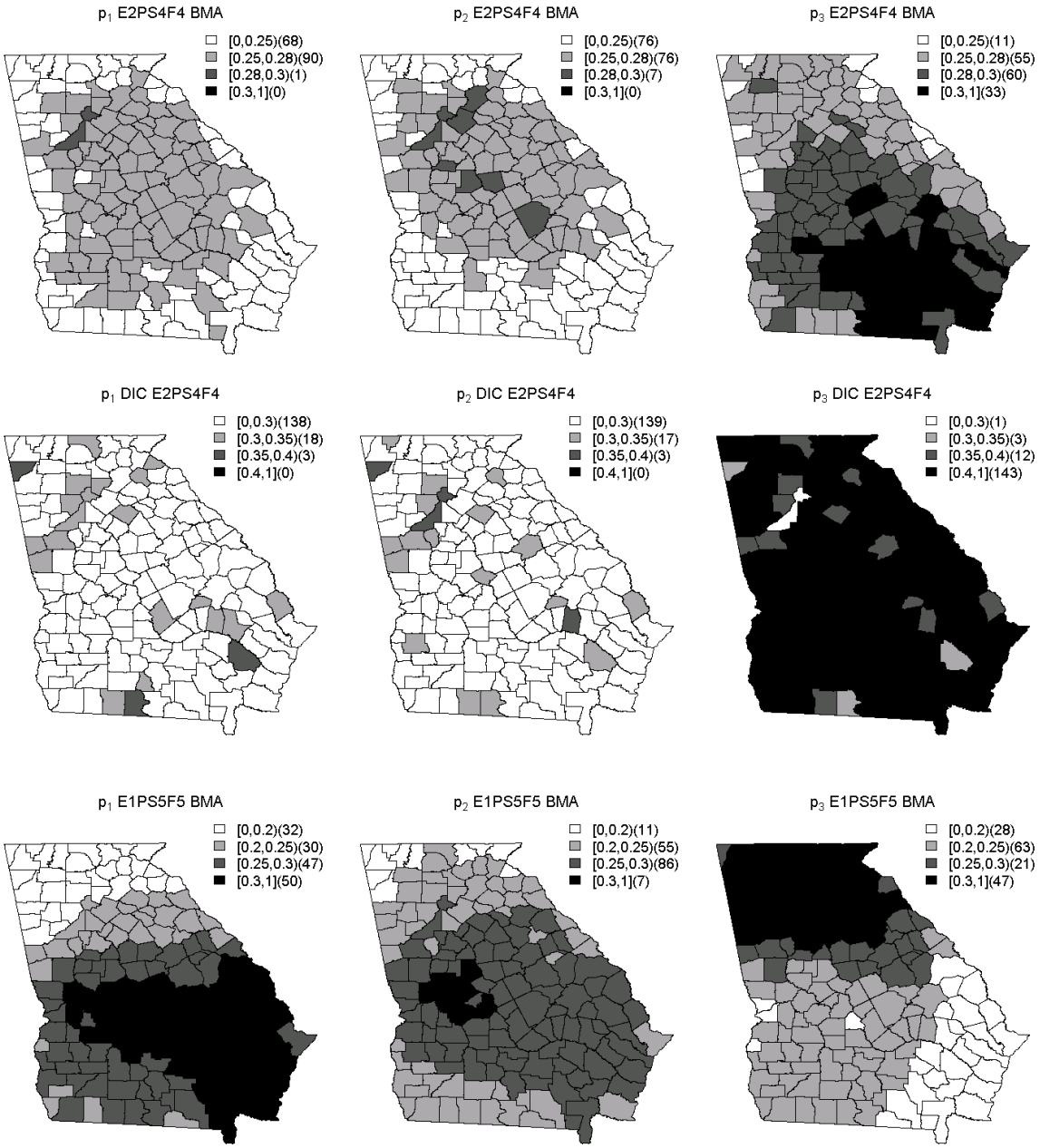


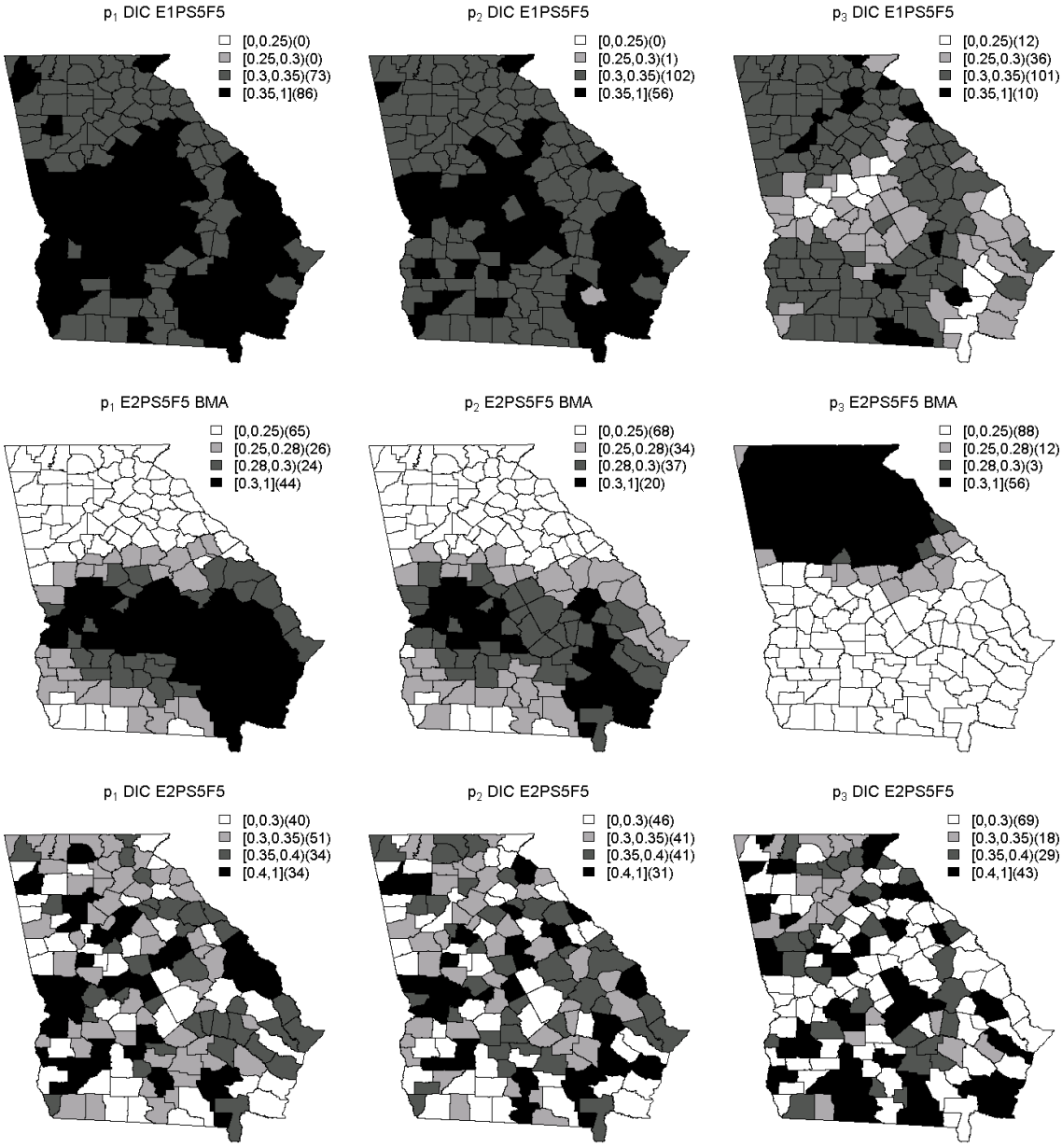

p_1 E2PS1F1 BMA
[0,0.2)(37)
[0.2,0.25)(25)
[0.25,0.3)(38)
[0.3,1](59)
p_2 E2PS1F1 BMA
[0,0.2)(69)
[0.2,0.25)(14)
[0.25,0.3)(16)
[0.3,1](60)
p_3 E2PS1F1 BMA
[0,0.2)(32)
[0.2,0.25)(32)
[0.25,0.3)(57)
[0.3,1](38)
p_1 DIC E2PS1F1
[0,0.3)(52)
[0.3,0.35)(27)
[0.35,0.4)(30)
[0.4,1](50)
p_2 DIC E2PS1F1
[0,0.3)(64)
[0.3,0.35)(26)
[0.35,0.4)(27)
[0.4,1](42)
p_3 DIC E2PS1F1
[0,0.3)(58)
[0.3,0.35)(38)
[0.35,0.4)(32)
[0.4,1](31)
p_1 E1PS2F2 BMA
[0,0.2)(9)
[0.2,0.25)(52)
[0.25,0.3)(96)
[0.3,1](2)
p_2 E1PS2F2 BMA
[0,0.2)(11)
[0.2,0.25)(63)
[0.25,0.3)(84)
[0.3,1](1)
p_3 E1PS2F2 BMA
[0,0.2)(4)
[0.2,0.25)(21)
[0.25,0.3)(65)
[0.3,1](69)

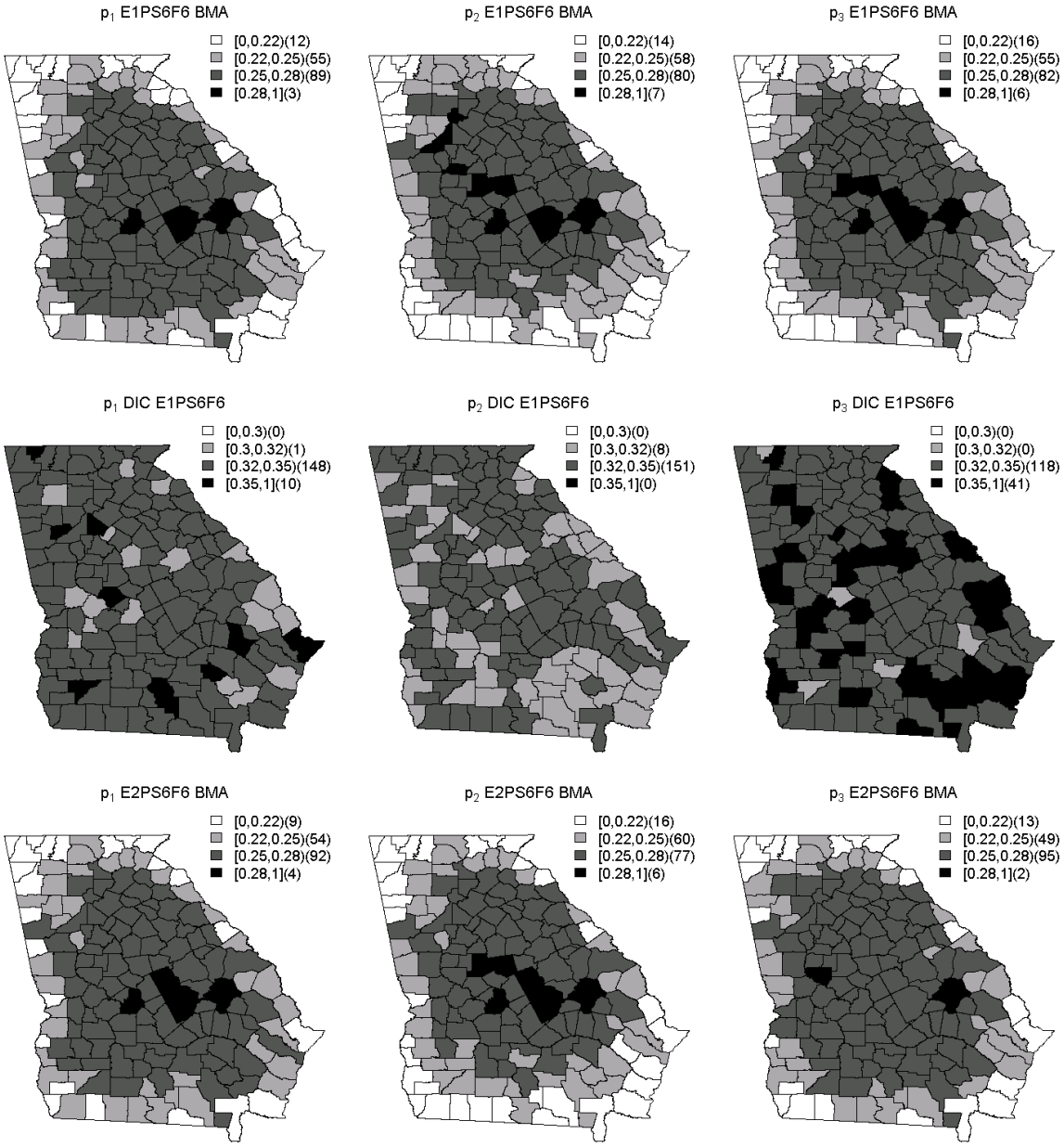

p1 DIC E1PS2F2
[0,0.2)(124)
[0.2,0.3)(35)
[0.3,0.4)(0)
[0.4,1](0)
p2 DIC E1PS2F2
[0,0.2)(139)
[0.2,0.3)(20)
[0.3,0.4)(0)
[0.4,1](0)
p3 DIC E1PS2F2
[0,0.2)(0)
[0.2,0.3)(0)
[0.3,0.4)(0)
[0.4,1](159)
p1 E2PS2F2 BMA
[0,0.25)(59)
[0.25,0.28)(91)
[0.28,0.3)(9)
[0.3,1](0)
p2 E2PS2F2 BMA
[0,0.25)(68)
[0.25,0.28)(71)
[0.28,0.3)(19)
[0.3,1](1)
p3 E2PS2F2 BMA
[0,0.25)(21)
[0.25,0.28)(30)
[0.28,0.3)(55)
[0.3,1](53)
p1 DIC E2PS2F2
[0,0.3)(135)
[0.3,0.35)(12)
[0.35,0.4)(8)
[0.4,1](4)
p2 DIC E2PS2F2
[0,0.3)(140)
[0.3,0.35)(9)
[0.35,0.4)(5)
[0.4,1](5)
p3 DIC E2PS2F2
[0,0.3)(5)
[0.3,0.35)(5)
[0.35,0.4)(11)
[0.4,1](138)

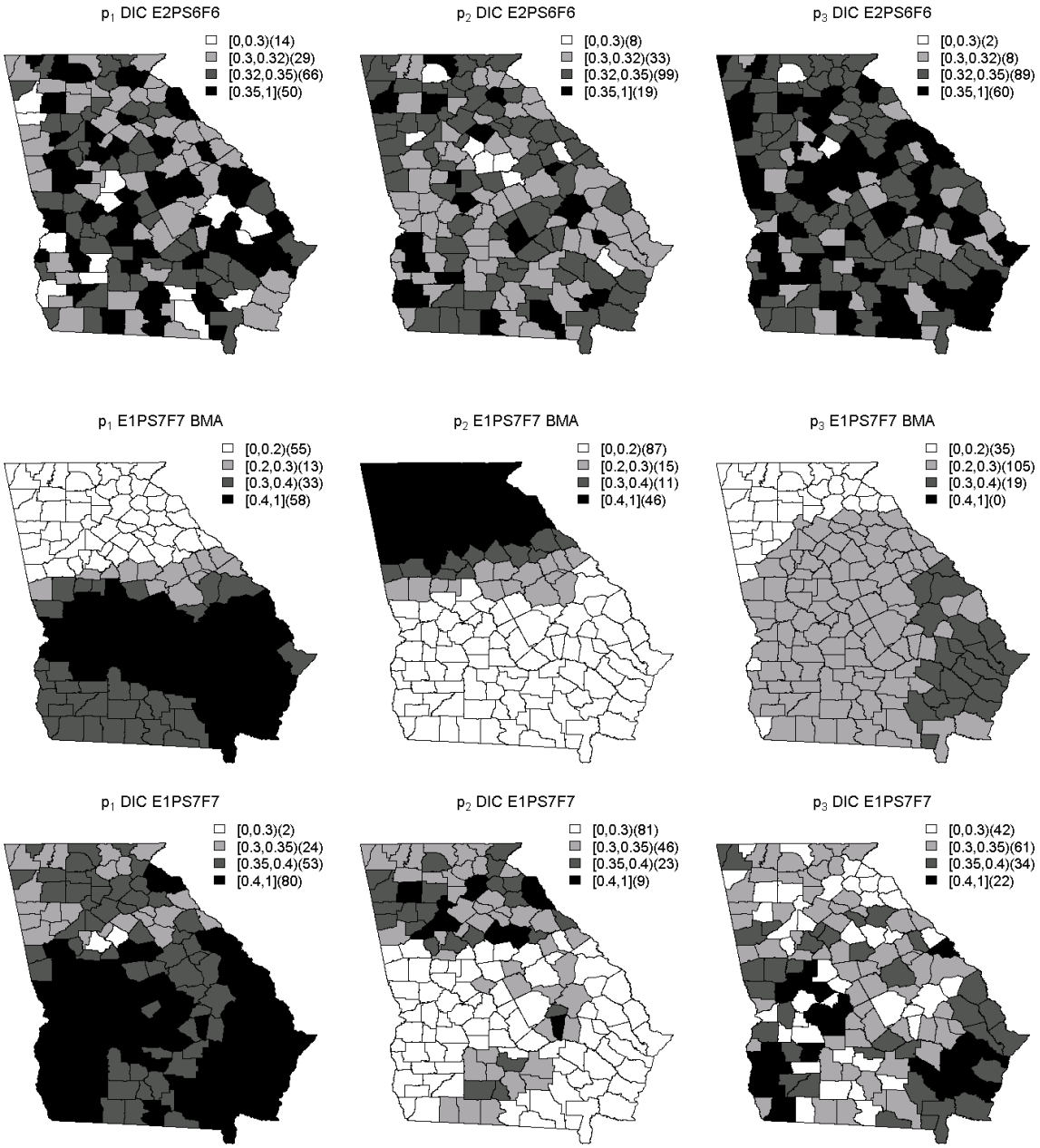

p1 E1PS3F3 BMA
[0,0.2)(4)
[0.2,0.25)(58)
[0.25,0.3)(97)
[0.3,1](0)
p2 E1PS3F3 BMA
[0,0.2)(4)
[0.2,0.25)(63)
[0.25,0.3)(92)
[0.3,1](0)
p3 E1PS3F3 BMA
[0,0.2)(0)
[0.2,0.25)(1)
[0.25,0.3)(99)
[0.3,1](59)
p1 DIC E1PS3F3
[0,0.15)(6)
[0.15,0.2)(126)
[0.2,0.25)(27)
[0.25,1](0)
p2 DIC E1PS3F3
[0,0.15)(15)
[0.15,0.2)(128)
[0.2,0.25)(16)
[0.25,1](0)
p3 DIC E1PS3F3
[0,0.15)(0)
[0.15,0.2)(0)
[0.2,0.25)(0)
[0.25,1](159)
p1 E2PS3F3 BMA
[0,0.25)(68)
[0.25,0.28)(86)
[0.28,0.3)(5)
[0.3,1](0)
p2 E2PS3F3 BMA
[0,0.25)(58)
[0.25,0.28)(97)
[0.28,0.3)(4)
[0.3,1](0)
p3 E2PS3F3 BMA
[0,0.25)(9)
[0.25,0.28)(40)
[0.28,0.3)(84)
[0.3,1](26)

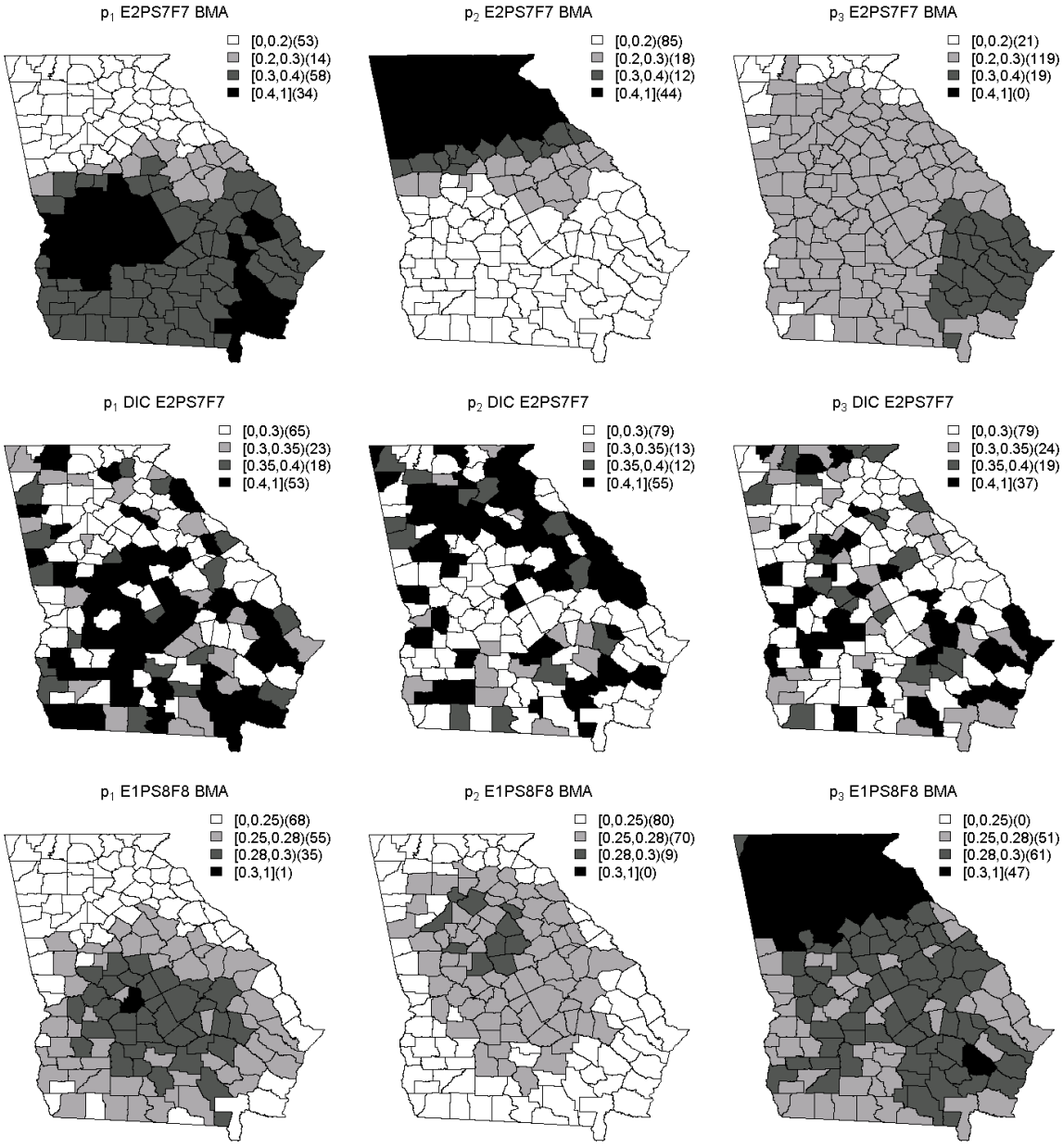

p_1 DIC E2PS3F3
[0,0.3)(135)
[0.3,0.35)(17)
[0.35,0.4)(6)
[0.4,1](1)
p_2 DIC E2PS3F3
[0,0.3)(137)
[0.3,0.35)(17)
[0.35,0.4)(4)
[0.4,1](1)
p_3 DIC E2PS3F3
[0,0.3)(3)
[0.3,0.35)(5)
[0.35,0.4)(10)
[0.4,1](141)
p_1 E1PS4F4 BMA
[0,0.2)(6)
[0.2,0.25)(73)
[0.25,0.3)(80)
[0.3,1](0)
p_2 E1PS4F4 BMA
[0,0.2)(6)
[0.2,0.25)(69)
[0.25,0.3)(84)
[0.3,1](0)
p_3 E1PS4F4 BMA
[0,0.2)(0)
[0.2,0.25)(4)
[0.25,0.3)(90)
[0.3,1](65)
p_1 DIC E1PS4F4
[0,0.2)(119)
[0.2,0.25)(40)
[0.25,0.3)(0)
[0.3,1](0)
p_2 DIC E1PS4F4
[0,0.2)(132)
[0.2,0.25)(27)
[0.25,0.3)(0)
[0.3,1](0)
p_3 DIC E1PS4F4
[0,0.2)(0)
[0.2,0.25)(0)
[0.25,0.3)(0)
[0.3,1](159)

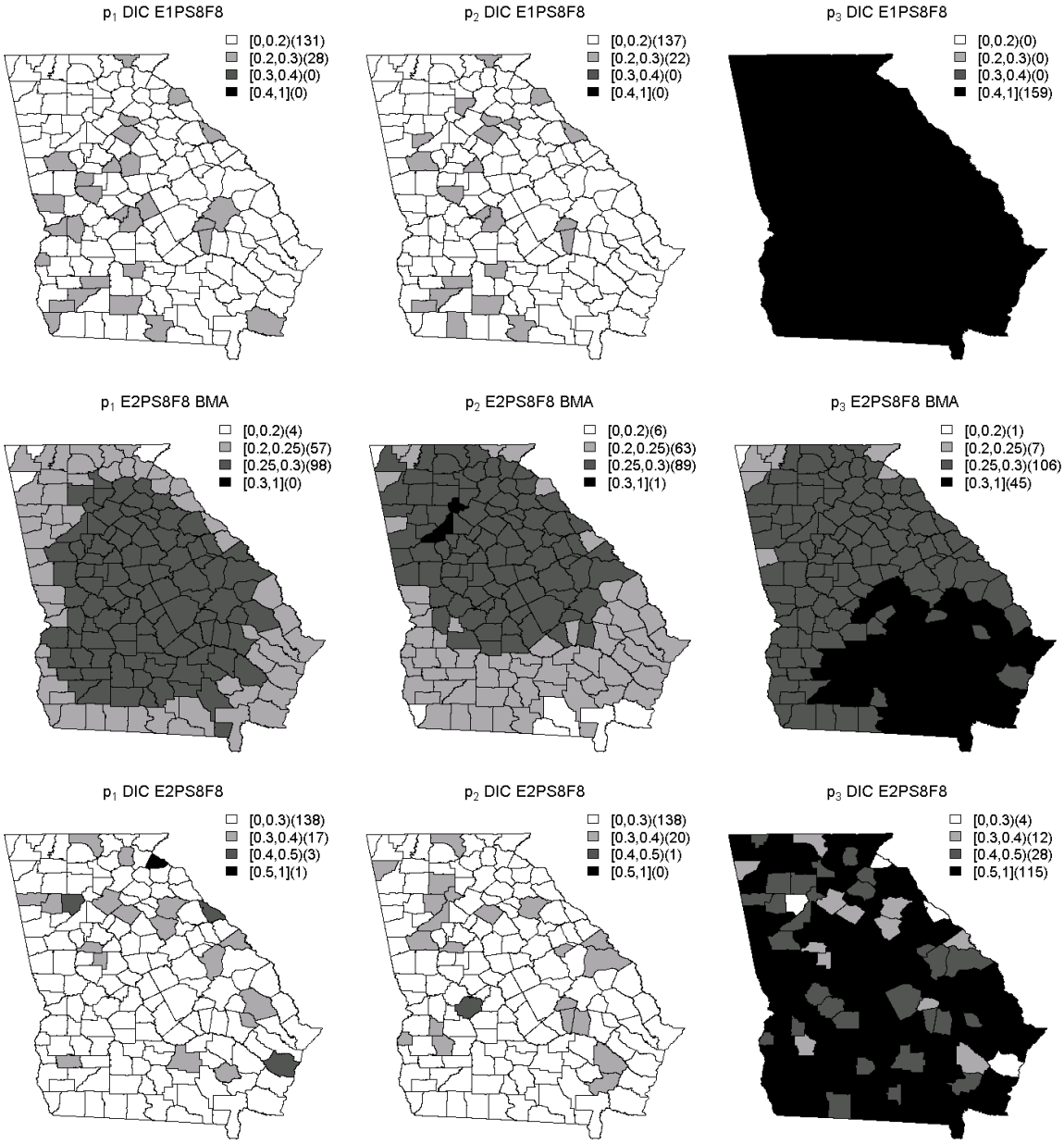

p1 E2PS4F4 BMA
[0,0.25)(68)
[0.25,0.28)(90)
[0.28,0.3)(1)
[0.3,1](0)
p2 E2PS4F4 BMA
[0,0.25)(76)
[0.25,0.28)(76)
[0.28,0.3)(7)
[0.3,1](0)
p3 E2PS4F4 BMA
[0,0.25)(11)
[0.25,0.28)(55)
[0.28,0.3)(60)
[0.3,1](33)
p1 DIC E2PS4F4
[0,0.3)(138)
[0.3,0.35)(18)
[0.35,0.4)(3)
[0.4,1](0)
p2 DIC E2PS4F4
[0,0.3)(139)
[0.3,0.35)(17)
[0.35,0.4)(3)
[0.4,1](0)
p3 DIC E2PS4F4
[0,0.3)(1)
[0.3,0.35)(3)
[0.35,0.4)(12)
[0.4,1](143)
p1 E1PS5F5 BMA
[0,0.2)(32)
[0.2,0.25)(30)
[0.25,0.3)(47)
[0.3,1](50)
p2 E1PS5F5 BMA
[0,0.2)(11)
[0.2,0.25)(55)
[0.25,0.3)(86)
[0.3,1](7)
p3 E1PS5F5 BMA
[0,0.2)(28)
[0.2,0.25)(63)
[0.25,0.3)(21)
[0.3,1](47)

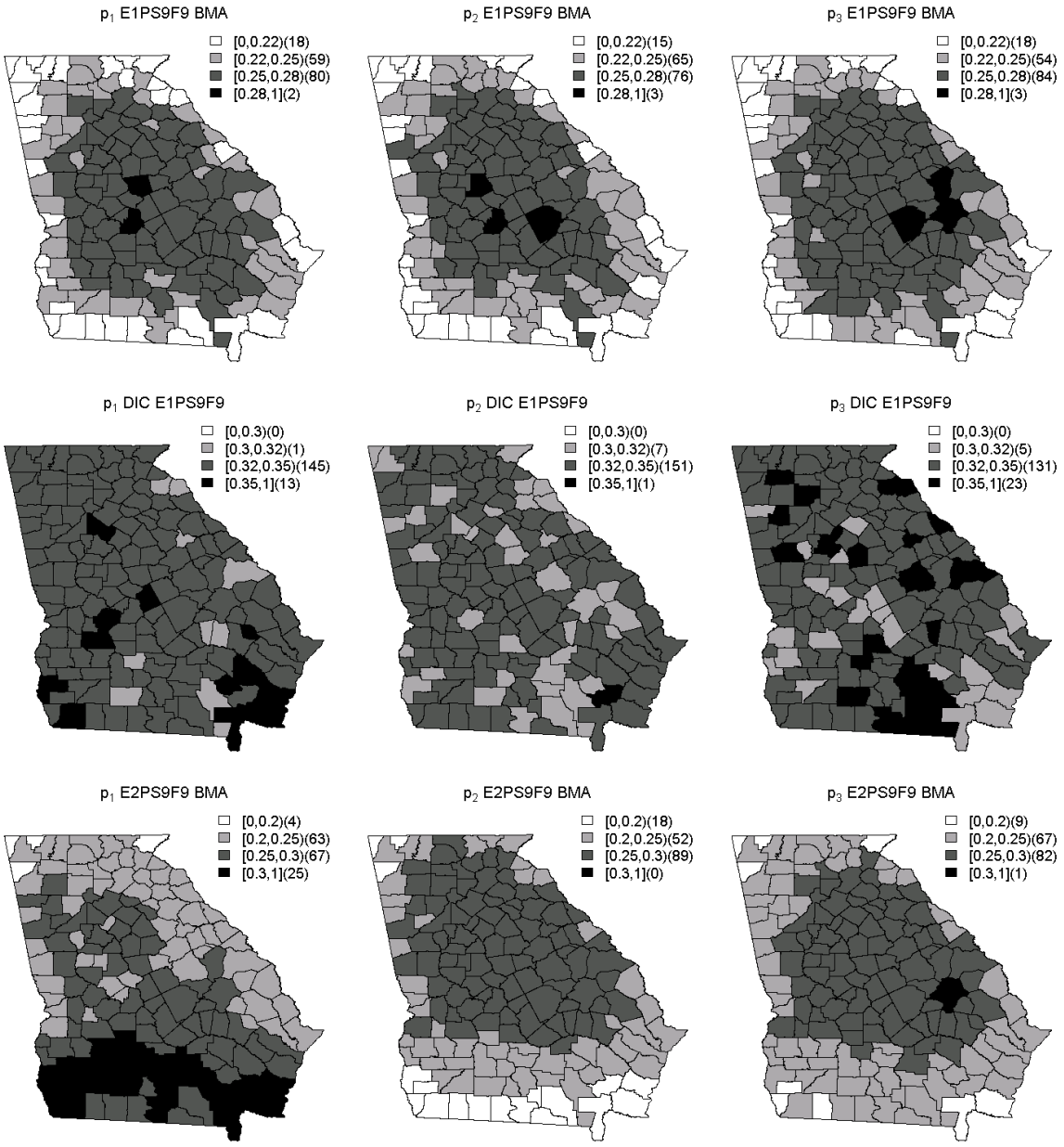

p1 DIC E1PS5F5
[0,0.25)(0)
[0.25,0.3)(0)
[0.3,0.35)(73)
[0.35,1](86)
p2 DIC E1PS5F5
[0,0.25)(0)
[0.25,0.3)(1)
[0.3,0.35)(102)
[0.35,1](56)
p3 DIC E1PS5F5
[0,0.25)(12)
[0.25,0.3)(36)
[0.3,0.35)(101)
[0.35,1](10)
p1 E2PS5F5 BMA
[0,0.25)(65)
[0.25,0.28)(26)
[0.28,0.3)(24)
[0.3,1](44)
p2 E2PS5F5 BMA
[0,0.25)(68)
[0.25,0.28)(34)
[0.28,0.3)(37)
[0.3,1](20)
p3 E2PS5F5 BMA
[0,0.25)(88)
[0.25,0.28)(12)
[0.28,0.3)(3)
[0.3,1](56)
p1 DIC E2PS5F5
[0,0.3)(40)
[0.3,0.35)(51)
[0.35,0.4)(34)
[0.4,1](34)
p2 DIC E2PS5F5
[0,0.3)(46)
[0.3,0.35)(41)
[0.35,0.4)(41)
[0.4,1](31)
p3 DIC E2PS5F5
[0,0.3)(69)
[0.3,0.35)(18)
[0.35,0.4)(29)
[0.4,1](43)

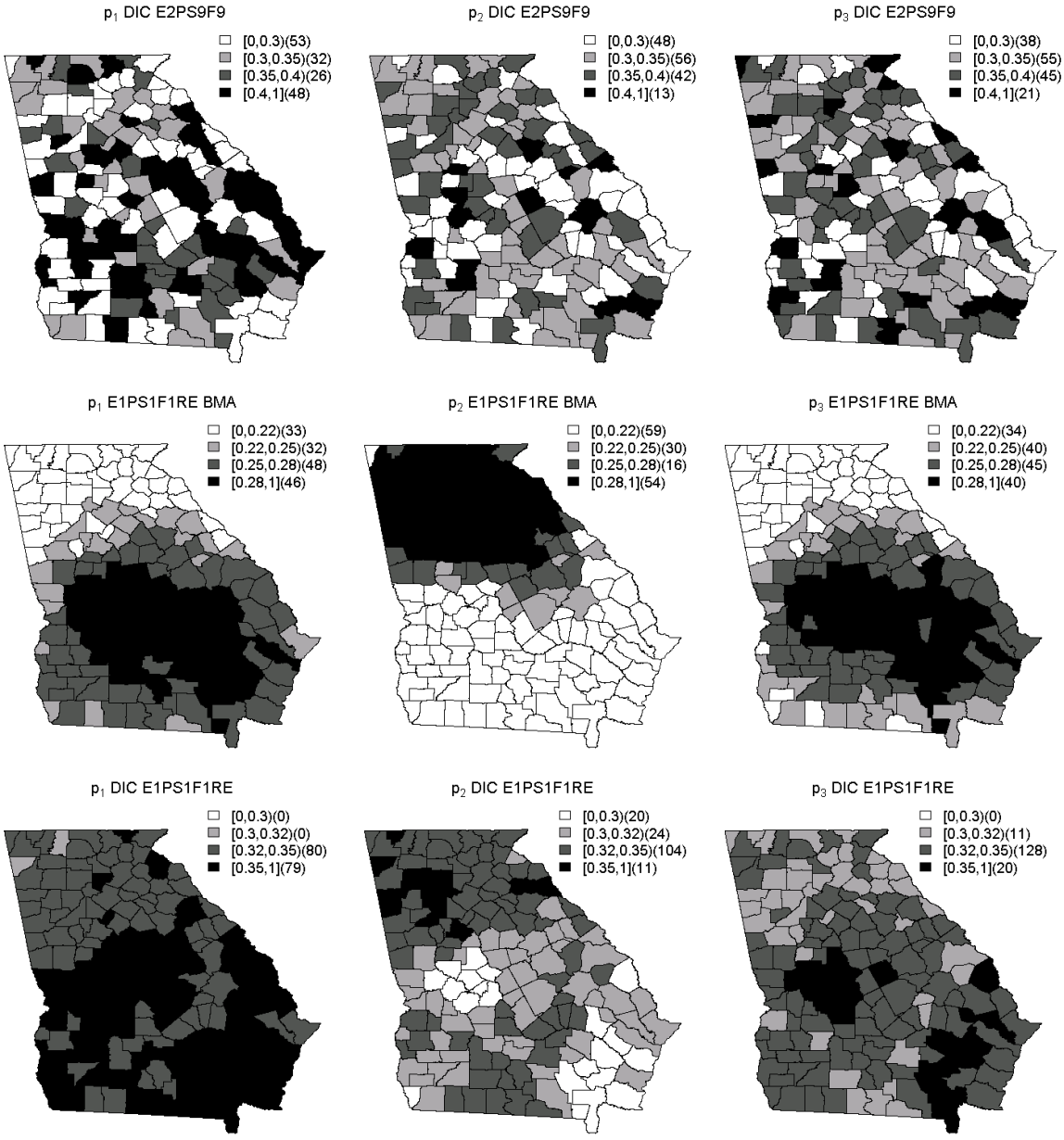

p1 E1PS6F6 BMA
[0,0.22)(12)
[0.22,0.25)(55)
[0.25,0.28)(89)
[0.28,1](3)
p2 E1PS6F6 BMA
[0,0.22)(14)
[0.22,0.25)(58)
[0.25,0.28)(80)
[0.28,1](7)
p3 E1PS6F6 BMA
[0,0.22)(16)
[0.22,0.25)(55)
[0.25,0.28)(82)
[0.28,1](6)
p1 DIC E1PS6F6
[0,0.3)(0)
[0.3,0.32)(1)
[0.32,0.35)(148)
[0.35,1](10)
p2 DIC E1PS6F6
[0,0.3)(0)
[0.3,0.32)(8)
[0.32,0.35)(151)
[0.35,1](0)
p3 DIC E1PS6F6
[0,0.3)(0)
[0.3,0.32)(0)
[0.32,0.35)(118)
[0.35,1](41)
p1 E2PS6F6 BMA
[0,0.22)(9)
[0.22,0.25)(54)
[0.25,0.28)(92)
[0.28,1](4)
p2 E2PS6F6 BMA
[0,0.22)(16)
[0.22,0.25)(60)
[0.25,0.28)(77)
[0.28,1](6)
p3 E2PS6F6 BMA
[0,0.22)(13)
[0.22,0.25)(49)
[0.25,0.28)(95)
[0.28,1](2)


p_1 DIC E2PS6F6
[0,0.3)(14)
[0.3,0.32)(29)
[0.32,0.35)(66)
[0.35,1](50)
p_2 DIC E2PS6F6
[0,0.3)(8)
[0.3,0.32)(33)
[0.32,0.35)(99)
[0.35,1](19)
p_3 DIC E2PS6F6
[0,0.3)(2)
[0.3,0.32)(8)
[0.32,0.35)(89)
[0.35,1](60)
p_1 E1PS7F7 BMA
[0,0.2)(55)
[0.2,0.3)(13)
[0.3,0.4)(33)
[0.4,1](58)
p_2 E1PS7F7 BMA
[0,0.2)(87)
[0.2,0.3)(15)
[0.3,0.4)(11)
[0.4,1](46)
p_3 E1PS7F7 BMA
[0,0.2)(35)
[0.2,0.3)(105)
[0.3,0.4)(19)
[0.4,1](0)
p_1 DIC E1PS7F7
[0,0.3)(2)
[0.3,0.35)(24)
[0.35,0.4)(53)
[0.4,1](80)
p_2 DIC E1PS7F7
[0,0.3)(81)
[0.3,0.35)(46)
[0.35,0.4)(23)
[0.4,1](9)
p_3 DIC E1PS7F7
[0,0.3)(42)
[0.3,0.35)(61)
[0.35,0.4)(34)
[0.4,1](22)


p1 E2PS7F7 BMA
[0,0.2)(53)
[0.2,0.3)(14)
[0.3,0.4)(58)
[0.4,1](34)
p2 E2PS7F7 BMA
[0,0.2)(85)
[0.2,0.3)(18)
[0.3,0.4)(12)
[0.4,1](44)
p3 E2PS7F7 BMA
[0,0.2)(21)
[0.2,0.3)(119)
[0.3,0.4)(19)
[0.4,1](0)
p1 DIC E2PS7F7
[0,0.3)(65)
[0.3,0.35)(23)
[0.35,0.4)(18)
[0.4,1](53)
p2 DIC E2PS7F7
[0,0.3)(79)
[0.3,0.35)(13)
[0.35,0.4)(12)
[0.4,1](55)
p3 DIC E2PS7F7
[0,0.3)(79)
[0.3,0.35)(24)
[0.35,0.4)(19)
[0.4,1](37)
p1 E1PS8F8 BMA
[0,0.25)(68)
[0.25,0.28)(55)
[0.28,0.3)(35)
[0.3,1](1)
p2 E1PS8F8 BMA
[0,0.25)(80)
[0.25,0.28)(70)
[0.28,0.3)(9)
[0.3,1](0)
p3 E1PS8F8 BMA
[0,0.25)(0)
[0.25,0.28)(51)
[0.28,0.3)(61)
[0.3,1](47)


p1 DIC E1PS8F8
[0,0.2)(131)
[0.2,0.3)(28)
[0.3,0.4)(0)
[0.4,1](0)
p2 DIC E1PS8F8
[0,0.2)(137)
[0.2,0.3)(22)
[0.3,0.4)(0)
[0.4,1](0)
p3 DIC E1PS8F8
[0,0.2)(0)
[0.2,0.3)(0)
[0.3,0.4)(0)
[0.4,1](159)
p1 E2PS8F8 BMA
[0,0.2)(4)
[0.2,0.25)(57)
[0.25,0.3)(98)
[0.3,1](0)
p2 E2PS8F8 BMA
[0,0.2)(6)
[0.2,0.25)(63)
[0.25,0.3)(89)
[0.3,1](1)
p3 E2PS8F8 BMA
[0,0.2)(1)
[0.2,0.25)(7)
[0.25,0.3)(106)
[0.3,1](45)
p1 DIC E2PS8F8
[0,0.3)(138)
[0.3,0.4)(17)
[0.4,0.5)(3)
[0.5,1](1)
p2 DIC E2PS8F8
[0,0.3)(138)
[0.3,0.4)(20)
[0.4,0.5)(1)
[0.5,1](0)
p3 DIC E2PS8F8
[0,0.3)(4)
[0.3,0.4)(12)
[0.4,0.5)(28)
[0.5,1](115)


p_1 E1PS9F9 BMA
[0,0.22)(18)
[0.22,0.25)(59)
[0.25,0.28)(80)
[0.28,1](2)
p_2 E1PS9F9 BMA
[0,0.22)(15)
[0.22,0.25)(65)
[0.25,0.28)(76)
[0.28,1](3)
p_3 E1PS9F9 BMA
[0,0.22)(18)
[0.22,0.25)(54)
[0.25,0.28)(84)
[0.28,1](3)
p_1 DIC E1PS9F9
[0,0.3)(0)
[0.3,0.32)(1)
[0.32,0.35)(145)
[0.35,1](13)
p_2 DIC E1PS9F9
[0,0.3)(0)
[0.3,0.32)(7)
[0.32,0.35)(151)
[0.35,1](1)
p_3 DIC E1PS9F9
[0,0.3)(0)
[0.3,0.32)(5)
[0.32,0.35)(131)
[0.35,1](23)
p_1 E2PS9F9 BMA
[0,0.2)(4)
[0.2,0.25)(63)
[0.25,0.3)(67)
[0.3,1](25)
p_2 E2PS9F9 BMA
[0,0.2)(18)
[0.2,0.25)(52)
[0.25,0.3)(89)
[0.3,1](0)
p_3 E2PS9F9 BMA
[0,0.2)(9)
[0.2,0.25)(67)
[0.25,0.3)(82)
[0.3,1](1)


p_1 DIC E2PS9F9
[0,0.3)(53)
[0.3,0.35)(32)
[0.35,0.4)(26)
[0.4,1](48)
p_2 DIC E2PS9F9
[0,0.3)(48)
[0.3,0.35)(56)
[0.35,0.4)(42)
[0.4,1](13)
p_3 DIC E2PS9F9
[0,0.3)(38)
[0.3,0.35)(55)
[0.35,0.4)(45)
[0.4,1](21)
p_1 E1PS1F1RE BMA
[0,0.22)(33)
[0.22,0.25)(32)
[0.25,0.28)(48)
[0.28,1](46)
p_2 E1PS1F1RE BMA
[0,0.22)(59)
[0.22,0.25)(30)
[0.25,0.28)(16)
[0.28,1](54)
p_3 E1PS1F1RE BMA
[0,0.22)(34)
[0.22,0.25)(40)
[0.25,0.28)(45)
[0.28,1](40)
p_1 DIC E1PS1F1RE
[0,0.3)(0)
[0.3,0.32)(0)
[0.32,0.35)(80)
[0.35,1](79)
p_2 DIC E1PS1F1RE
[0,0.3)(20)
[0.3,0.32)(24)
[0.32,0.35)(104)
[0.35,1](11)
p_3 DIC E1PS1F1RE
[0,0.3)(0)
[0.3,0.32)(11)
[0.32,0.35)(128)
[0.35,1](20)

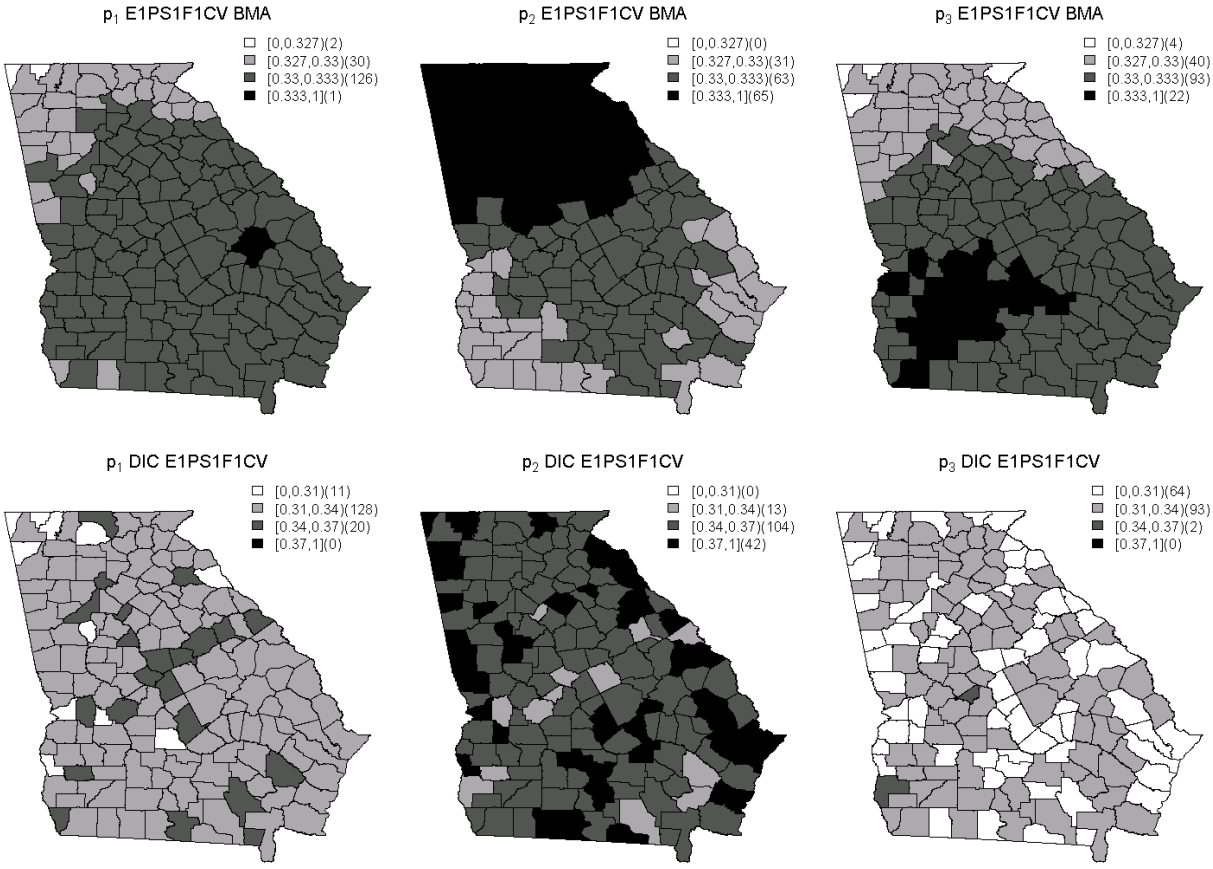

p1 E1PS1F1CV BMA
[0,0.327)(2)
[0.327,0.33)(30)
[0.33,0.333)(126)
[0.333,1](1)
p2 E1PS1F1CV BMA
[0,0.327)(0)
[0.327,0.33)(31)
[0.33,0.333)(63)
[0.333,1](65)
p3 E1PS1F1CV BMA
[0,0.327)(4)
[0.327,0.33)(40)
[0.33,0.333)(93)
[0.333,1](22)
p1 DIC E1PS1F1CV
[0,0.31)(11)
[0.31,0.34)(128)
[0.34,0.37)(20)
[0.37,1](0)
p2 DIC E1PS1F1CV
[0,0.31)(0)
[0.31,0.34)(13)
[0.34,0.37)(104)
[0.37,1](42)
p3 DIC E1PS1F1CV
[0,0.31)(64)
[0.31,0.34)(93)
[0.34,0.37)(2)
[0.37,1](0)

Table 7: GoF measures associated with partial models fit with BMA.

	$\bar{D}$	DIC	$\frac{Var(\bar{D})}{2}$	MSE_{A1}	MSE_{A2}	MSE_{A3}	$MSPE_{A1}$	$MSPE_{A2}$	$MSPE_{A3}$
E1PS1F1	2455.38	2476.58	21.20	0.97	1.50	1.18	1.02	1.44	1.22
E2PS1F1	2097.75	2117.27	19.52	4.84	5.01	4.46	4.82	4.93	4.42
E1PS2F2	2298.06	2487.52	189.45	1.13	1.52	1.27	1.06	1.44	1.24
E2PS2F2	1963.62	2117.66	154.04	5.04	4.66	3.98	4.98	4.65	3.96
E1PS3F3	2287.83	2460.36	172.52	1.24	1.26	1.17	1.18	1.20	1.12
E2PS3F3	1954.24	2088.57	134.32	5.62	5.59	4.24	4.19	3.98	3.95
E1PS4F3	2282.17	2451.06	168.88	1.06	1.10	1.11	1.02	1.05	1.05
E2PS4F3	1957.81	2086.93	129.13	4.70	3.55	4.16	4.67	3.47	4.10
E1PS5F5	2406.87	2424.65	17.78	1.06	1.18	1.14	1.10	1.26	1.15
E2PS5F5	2051.09	2067.67	16.59	4.96	4.01	5.49	4.97	4.04	5.49
E1PS6F6	2394.56	2402.73	8.17	0.96	0.93	0.98	1.00	0.98	1.04
E2PS6F6	2049.45	2057.78	8.33	4.84	3.53	4.39	4.83	3.50	4.40
E1PS7F7	2445.02	2481.95	35.93	1.05	2.59	1.23	1.15	2.77	1.34
E2PS7F7	2093.38	2125.59	32.21	3.68	7.89	3.98	3.83	8.32	4.00
E1PS8F8	2291.10	2465.30	174.20	1.46	1.39	1.43	1.40	1.35	1.38
E2PS8F8	1949.31	2079.34	133.03	4.32	4.61	5.25	4.29	4.61	5.15
E1PS9F9	2513.67	2525.15	11.48	1.08	2.88	1.46	1.08	3.02	1.45
E2PS9F9	2126.08	2137.42	11.34	4.11	4.58	6.30	4.08	4.55	6.32
E1PS1F1RE	2240.84	2602.07	361.23	1.10	1.71	1.34	1.14	1.63	1.37
E1PS1F1CV	2240.70	2585.74	345.03	1.84	3.06	1.85	2.04	3.30	2.03

A.3. Aim 3

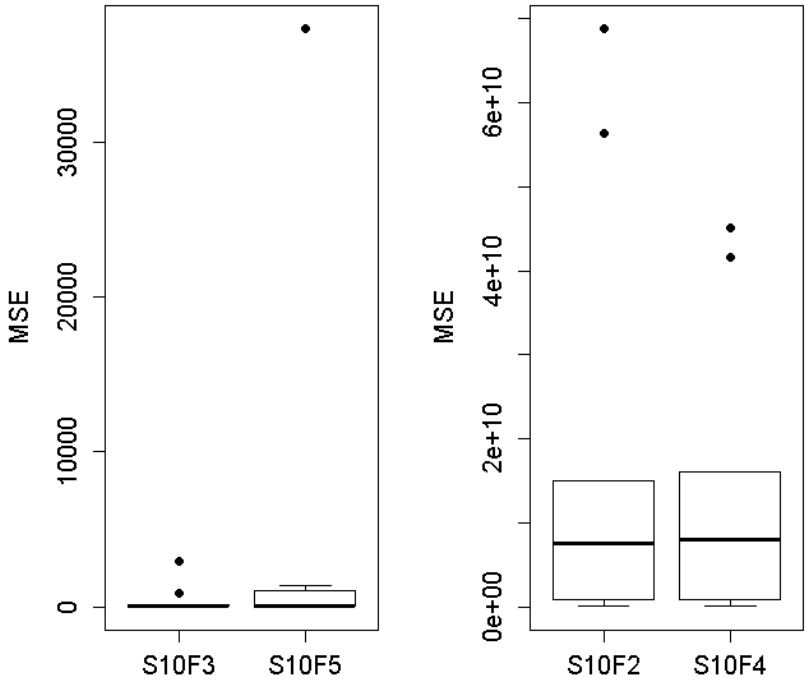
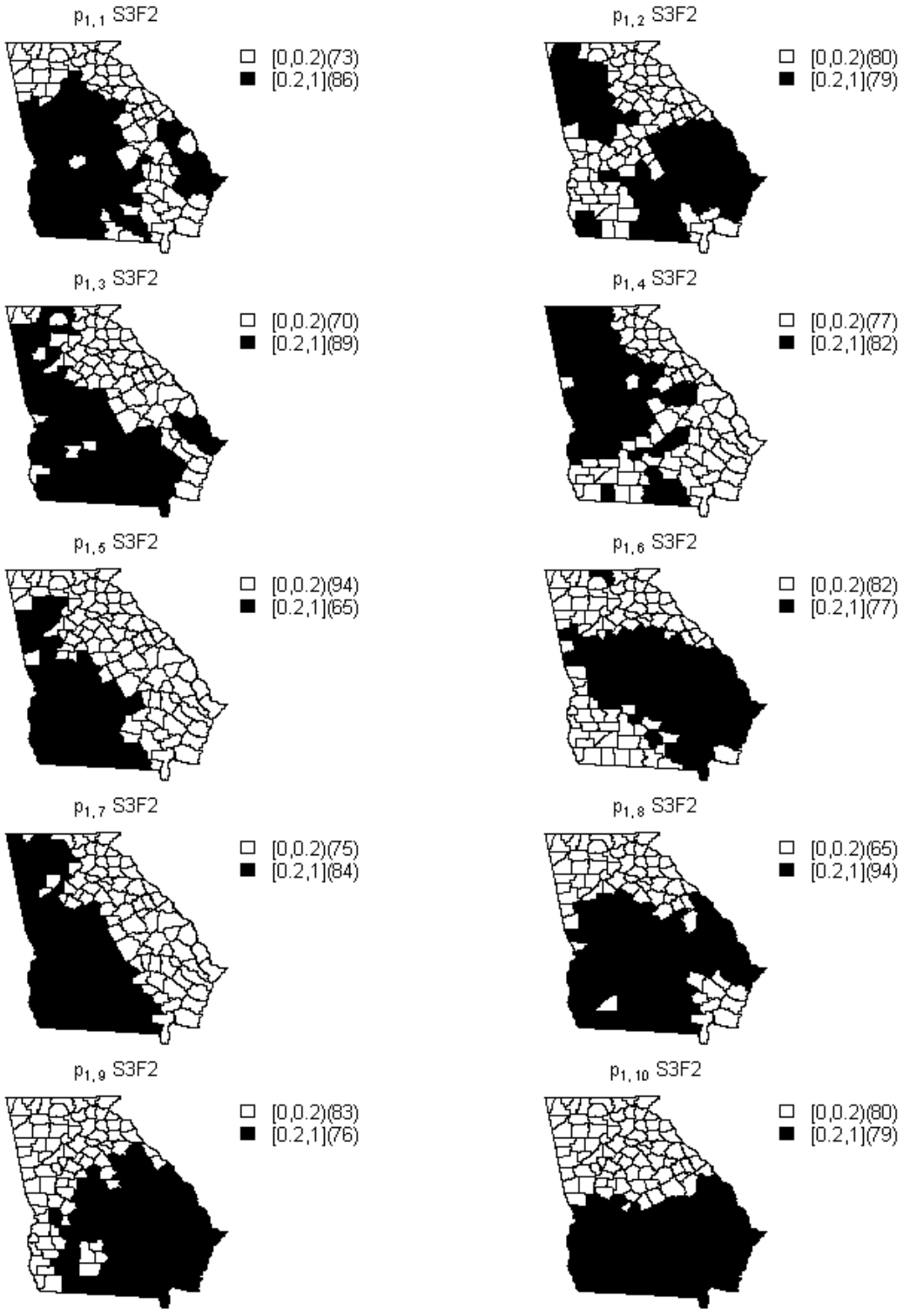
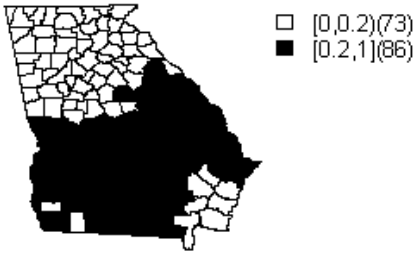


Figure 1: MSE measures for simulated data model M10.

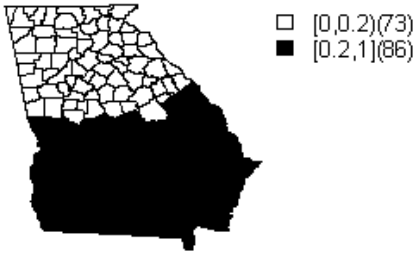
Figure 2: Model probabilities associated with fitted model F2.



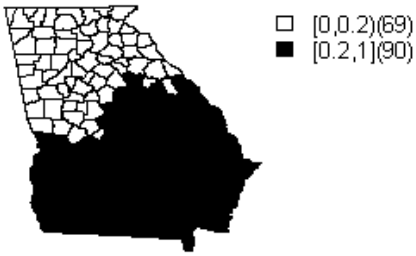
$p_{2,1}$ S3F2



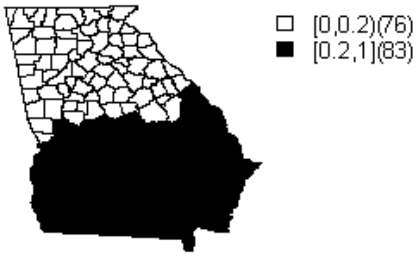
$p_{2,2}$ S3F2



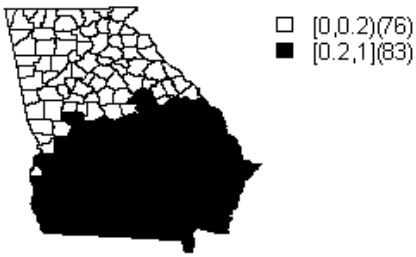
$p_{2,3}$ S3F2



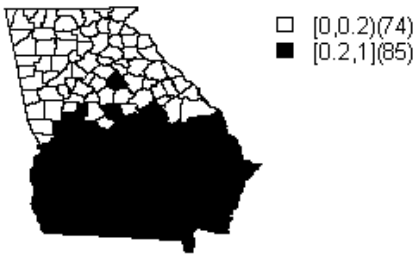
$p_{2,4}$ S3F2



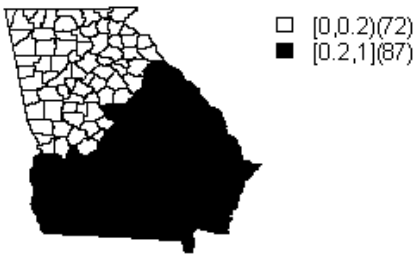
$p_{2,5}$ S3F2



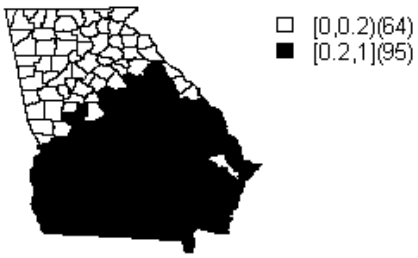
$p_{2,6}$ S3F2



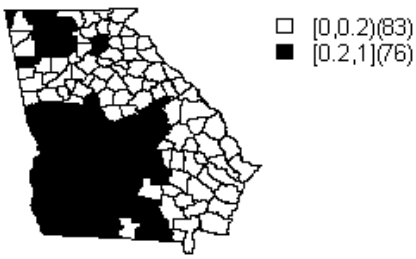
$p_{2,7}$ S3F2



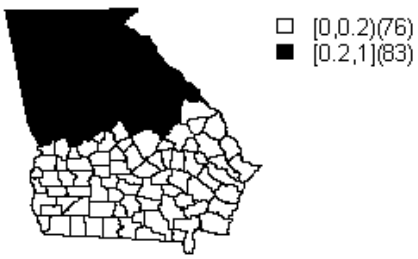
$p_{2,8}$ S3F2



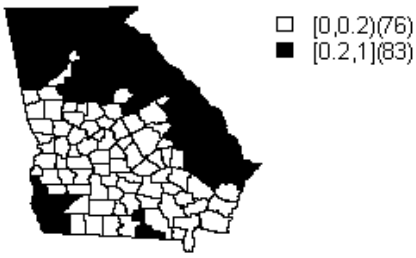
$p_{2,9}$ S3F2



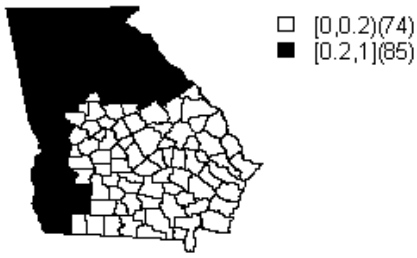
$p_{2,10}$ S3F2



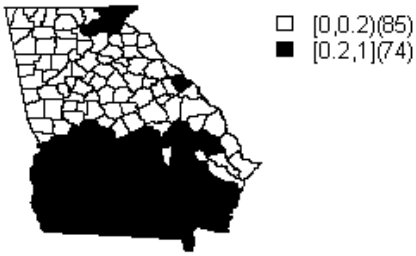
$p_{3,1}$ S3F2



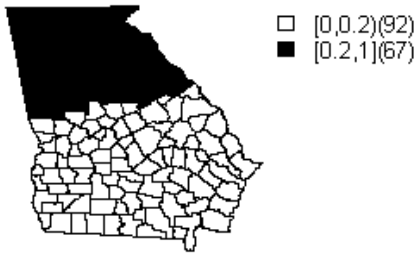
$p_{3,2}$ S3F2



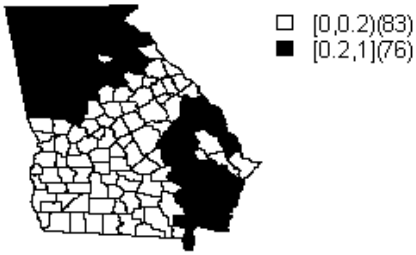
$p_{3,3}$ S3F2



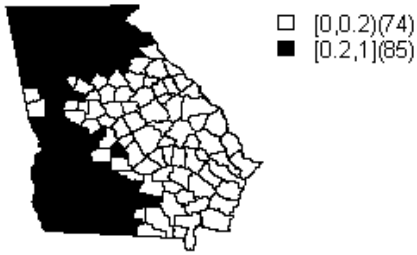
$p_{3,4}$ S3F2



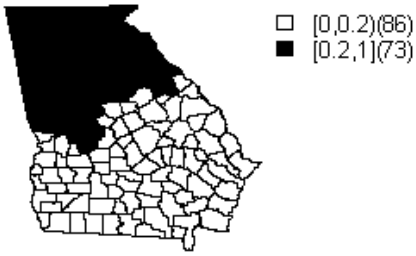
$p_{3,5}$ S3F2



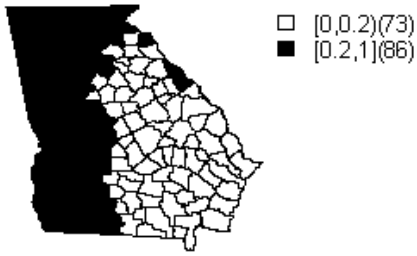
$p_{3,6}$ S3F2



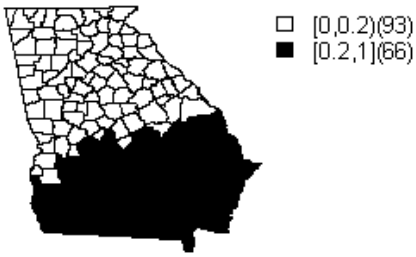
$p_{3,7}$ S3F2



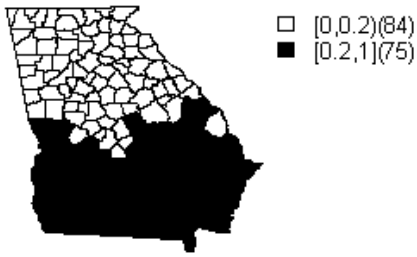
$p_{3,8}$ S3F2



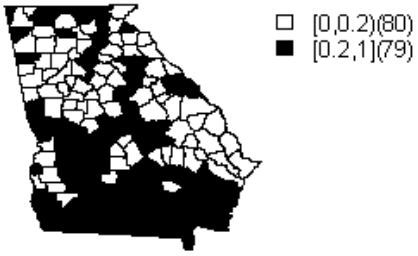
$p_{3,9}$ S3F2



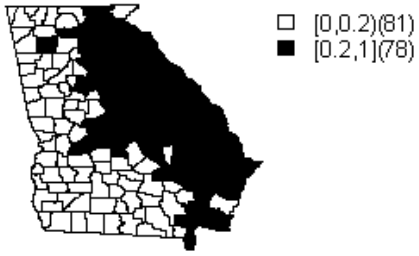
$p_{3,10}$ S3F2



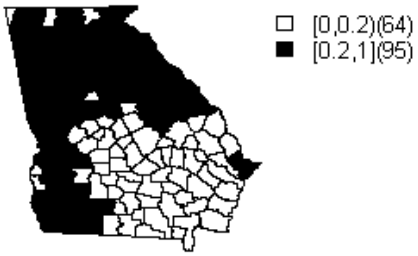
$p_{4,1}$ S3F2



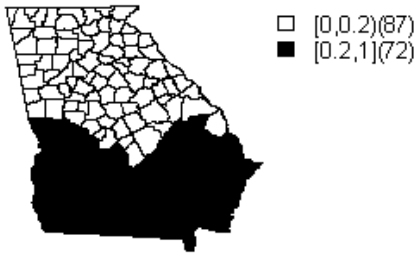
$p_{4,2}$ S3F2



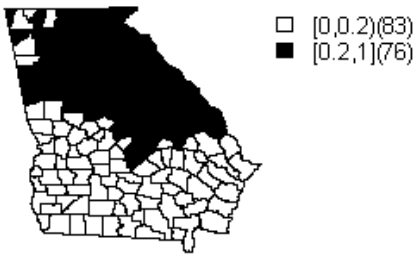
$p_{4,3}$ S3F2



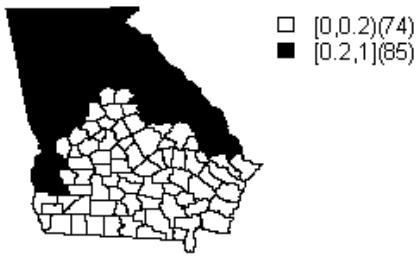
$p_{4,4}$ S3F2



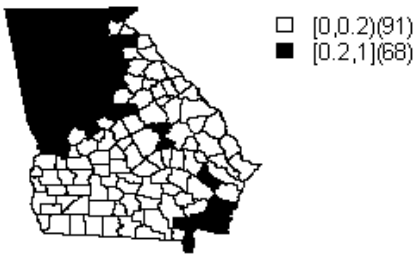
$p_{4,5}$ S3F2



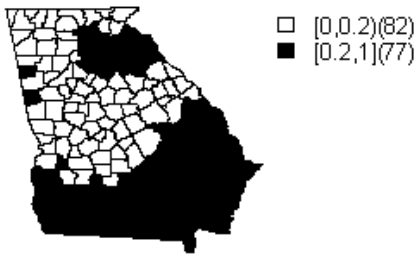
$p_{4,6}$ S3F2



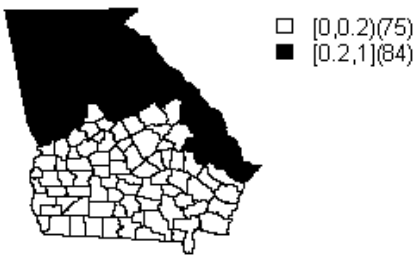
$p_{4,7}$ S3F2



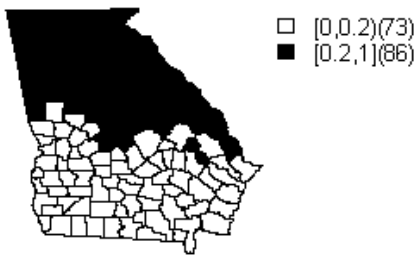
$p_{4,8}$ S3F2



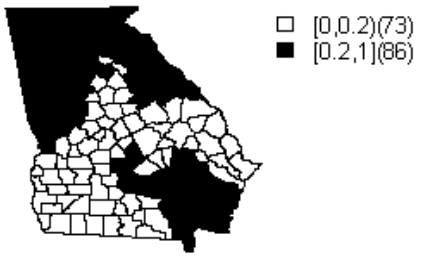
$p_{4,9}$ S3F2



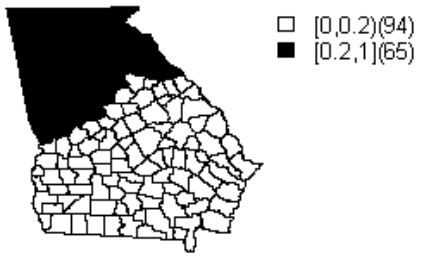
$p_{4,10}$ S3F2



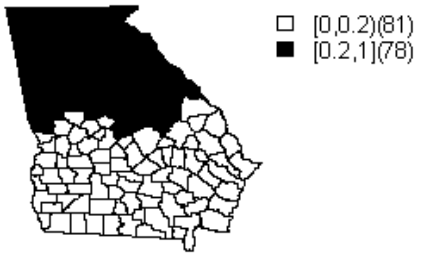
$p_{5,1}$ S3F2



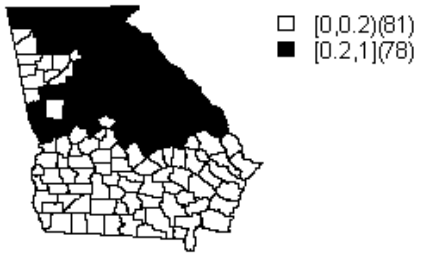
$p_{5,2}$ S3F2



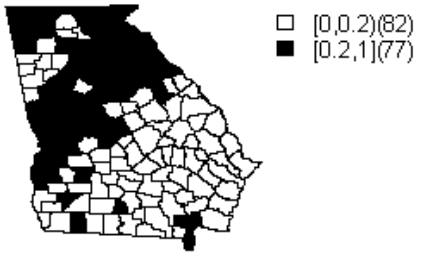
$p_{5,3}$ S3F2



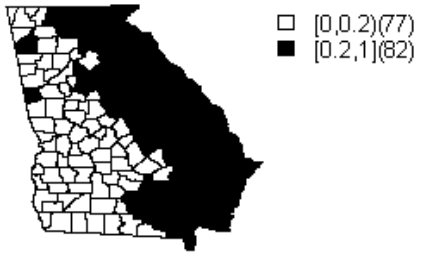
$p_{5,4}$ S3F2



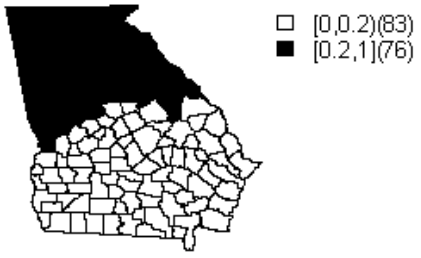
$p_{5,5}$ S3F2



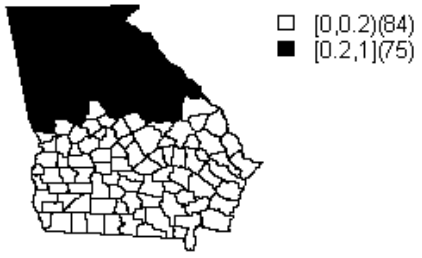
$p_{5,6}$ S3F2



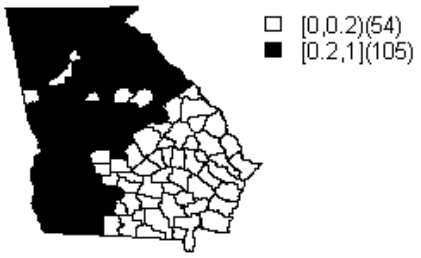
$p_{5,7}$ S3F2



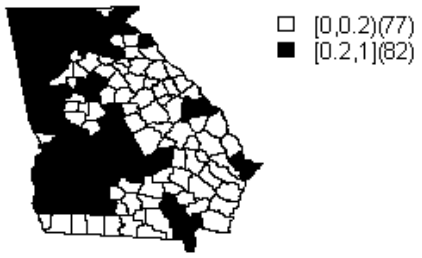
$p_{5,8}$ S3F2



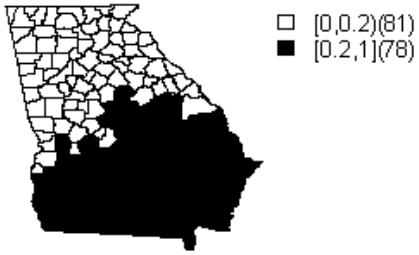
$p_{5,9}$ S3F2



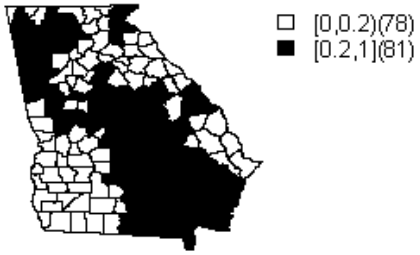
$p_{5,10}$ S3F2



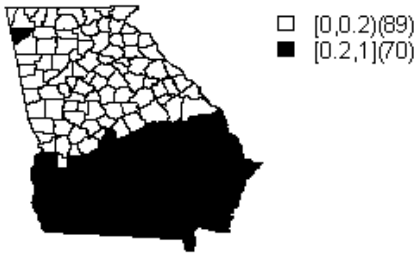
$p_{1,1}$ S4F2



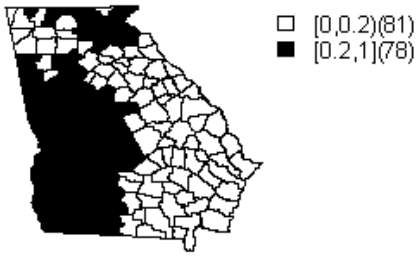
$p_{1,2}$ S4F2



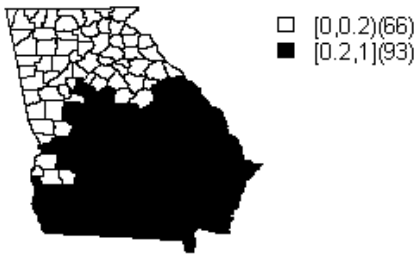
$p_{1,3}$ S4F2



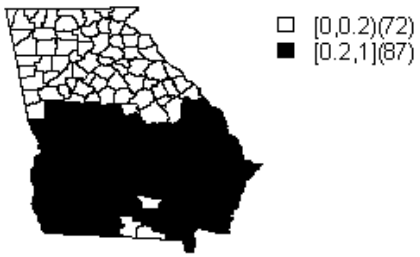
$p_{1,4}$ S4F2



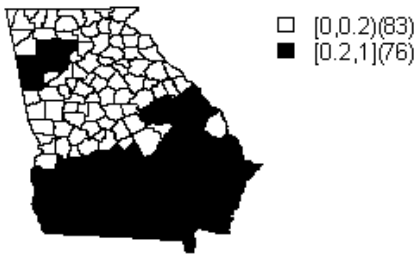
$p_{1,5}$ S4F2



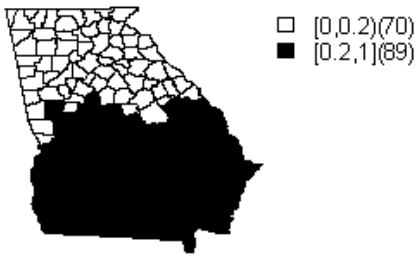
$p_{1,6}$ S4F2



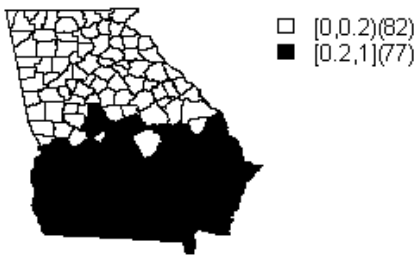
$p_{1,7}$ S4F2



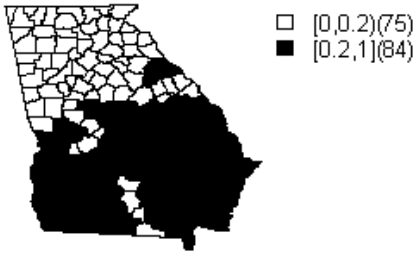
$p_{1,8}$ S4F2



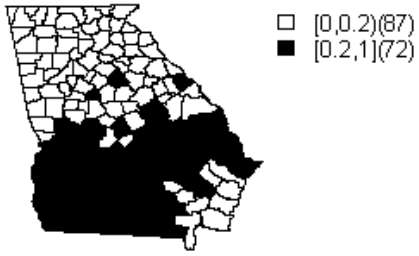
$p_{1,9}$ S4F2



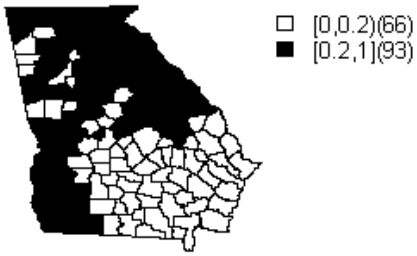
$p_{2,1}$ S4F2



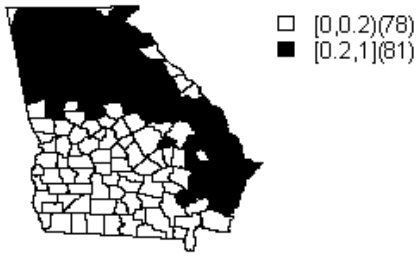
$p_{2,2}$ S4F2



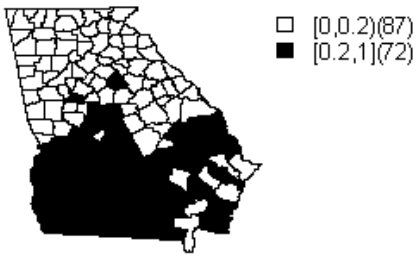
$p_{2,3}$ S4F2



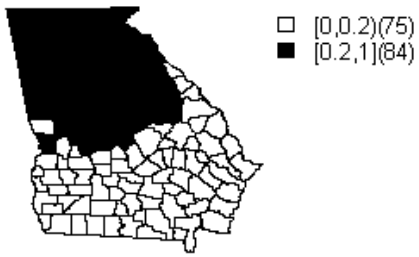
$p_{2,4}$ S4F2



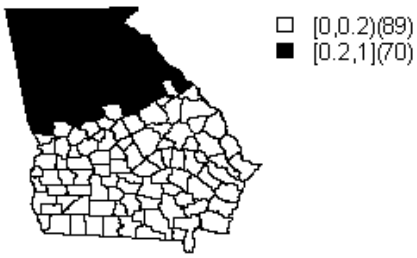
$p_{2,5}$ S4F2



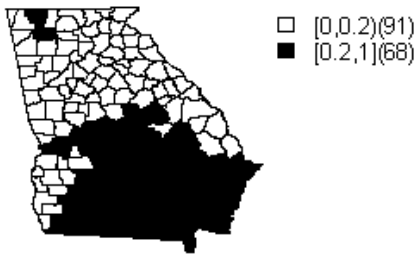
$p_{2,6}$ S4F2



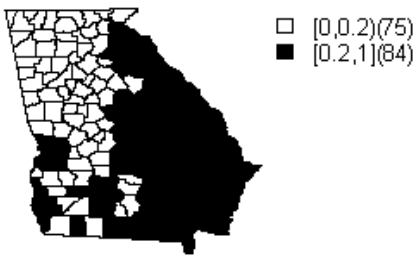
$p_{2,7}$ S4F2



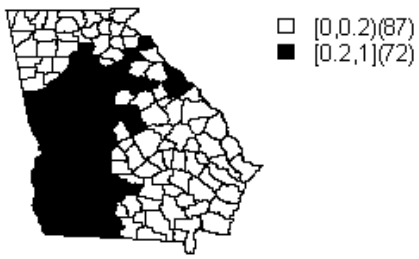
$p_{2,8}$ S4F2



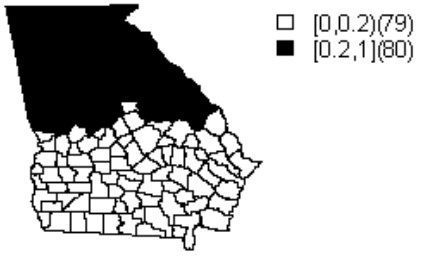
$p_{2,9}$ S4F2



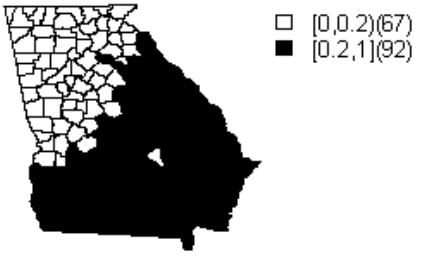
$p_{2,10}$ S4F2



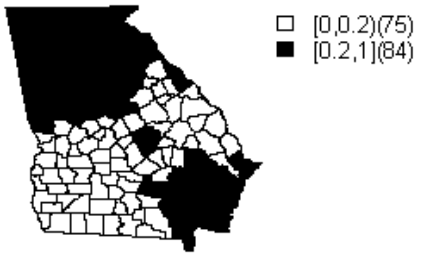
$p_{3,1}$ S4F2



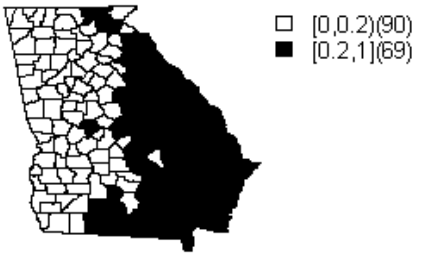
$p_{3,2}$ S4F2



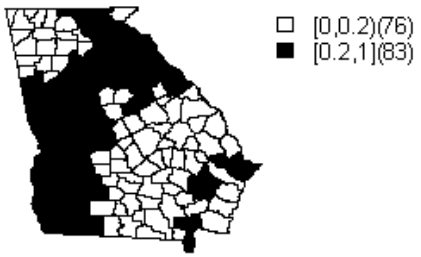
$p_{3,3}$ S4F2



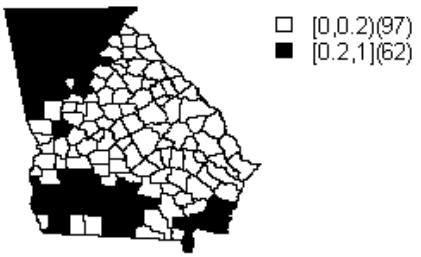
$p_{3,4}$ S4F2



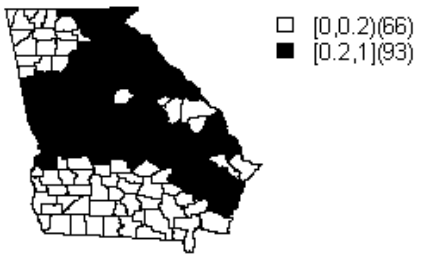
$p_{3,5}$ S4F2



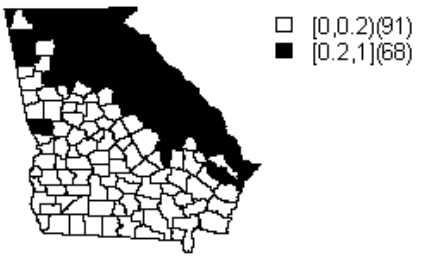
$p_{3,6}$ S4F2



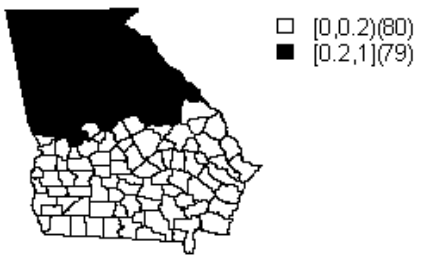
$p_{3,7}$ S4F2



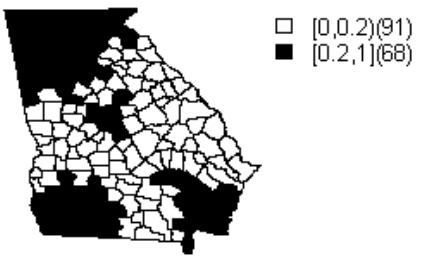
$p_{3,8}$ S4F2



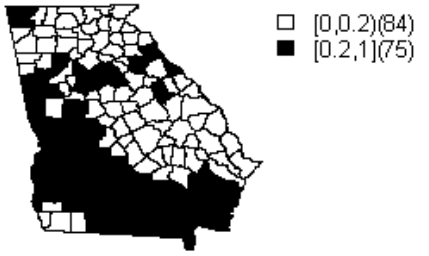
$p_{3,9}$ S4F2



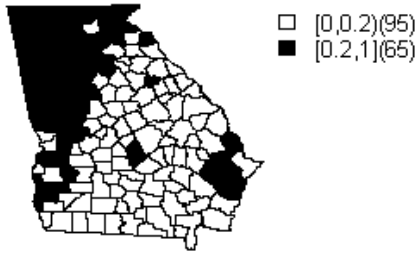
$p_{3,10}$ S4F2



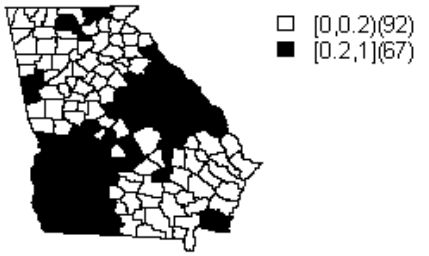
$p_{4,1}$ S4F2



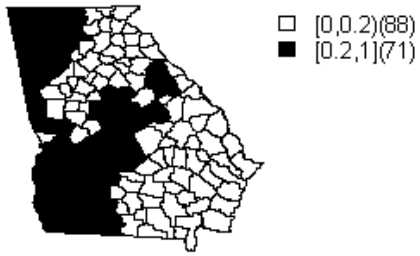
$p_{4,2}$ S4F2



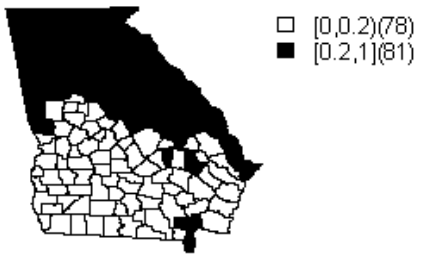
$p_{4,3}$ S4F2



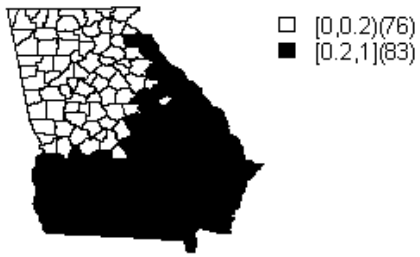
$p_{4,4}$ S4F2



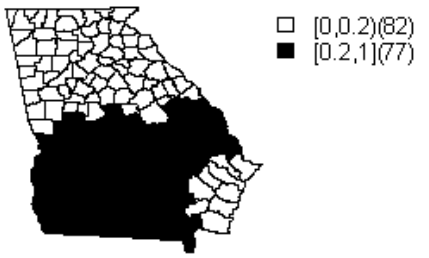
$p_{4,5}$ S4F2



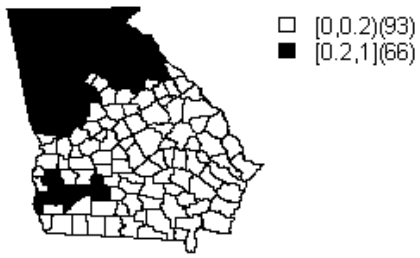
$p_{4,6}$ S4F2



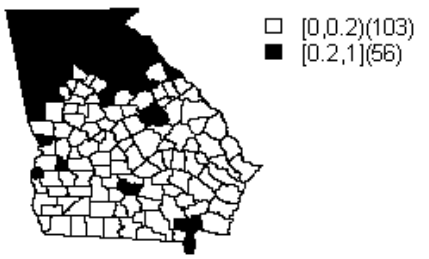
$p_{4,7}$ S4F2



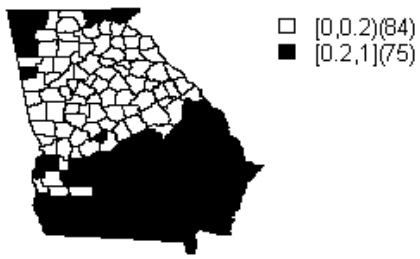
$p_{4,8}$ S4F2



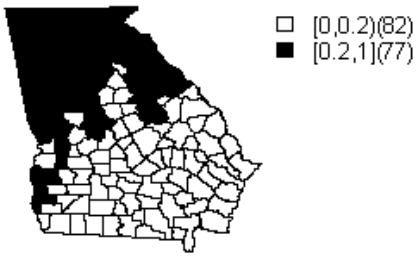
$p_{4,9}$ S4F2



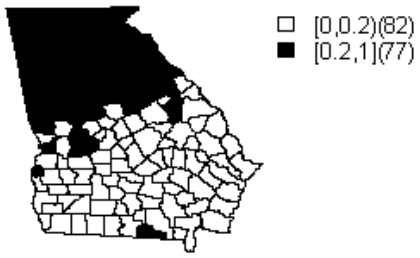
$p_{4,10}$ S4F2



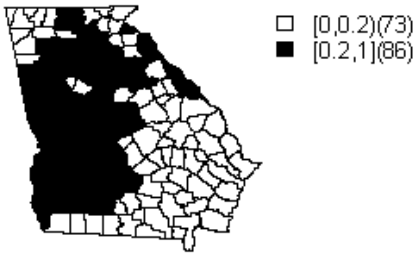
$p_{5,1}$ S4F2



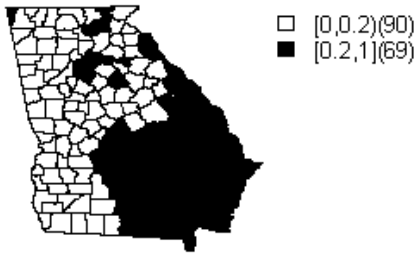
$p_{5,2}$ S4F2



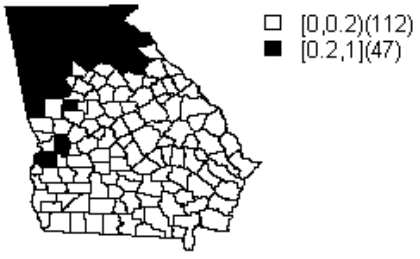
$p_{5,3}$ S4F2



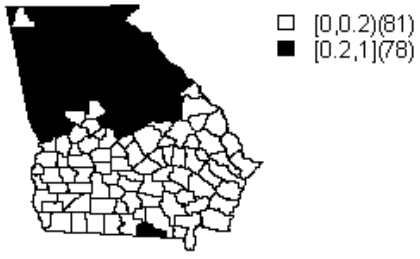
$p_{5,4}$ S4F2



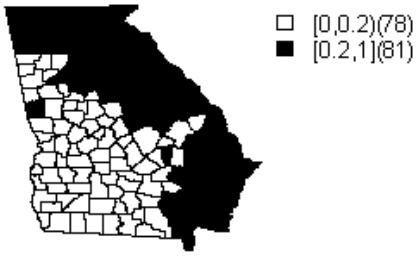
$p_{5,5}$ S4F2



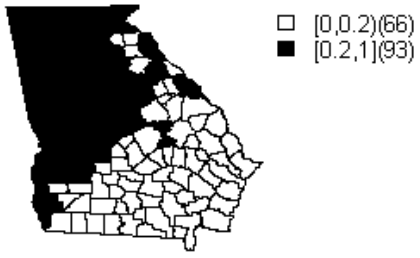
$p_{5,6}$ S4F2



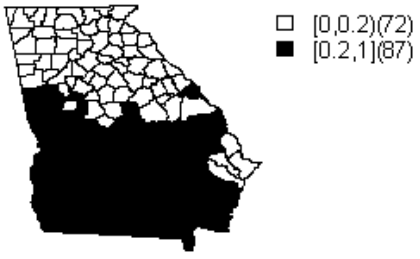
$p_{5,7}$ S4F2



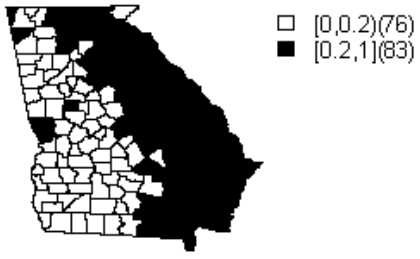
$p_{5,8}$ S4F2

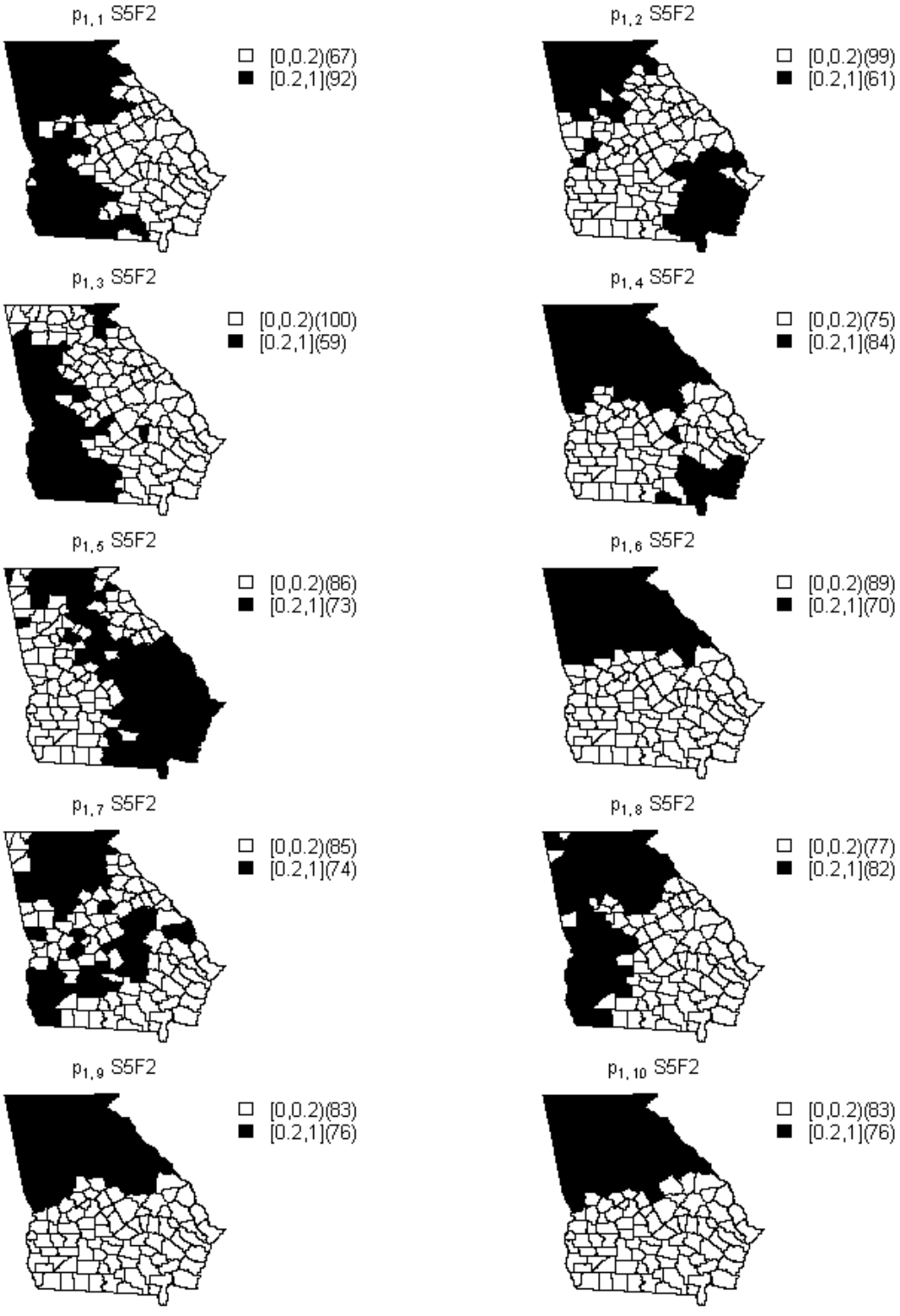


$p_{5,9}$ S4F2

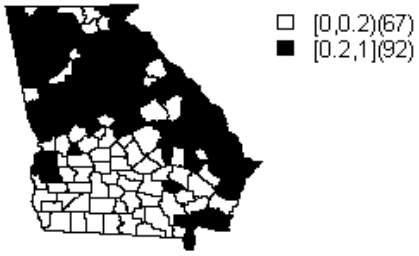


$p_{5,10}$ S4F2

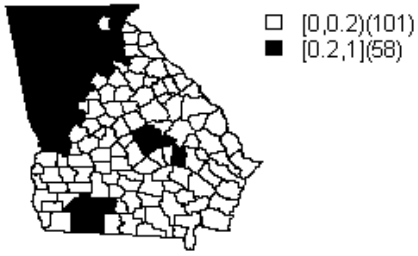



$p_{1,1}$ S5F2
[0,0.2)(67)
[0.2,1](92)
$p_{1,2}$ S5F2
[0,0.2)(99)
[0.2,1](61)
$p_{1,3}$ S5F2
[0,0.2)(100)
[0.2,1](59)
$p_{1,4}$ S5F2
[0,0.2)(75)
[0.2,1](84)
$p_{1,5}$ S5F2
[0,0.2)(86)
[0.2,1](73)
$p_{1,6}$ S5F2
[0,0.2)(89)
[0.2,1](70)
$p_{1,7}$ S5F2
[0,0.2)(85)
[0.2,1](74)
$p_{1,8}$ S5F2
[0,0.2)(77)
[0.2,1](82)
$p_{1,9}$ S5F2
[0,0.2)(83)
[0.2,1](76)
$p_{1,10}$ S5F2
[0,0.2)(83)
[0.2,1](76)

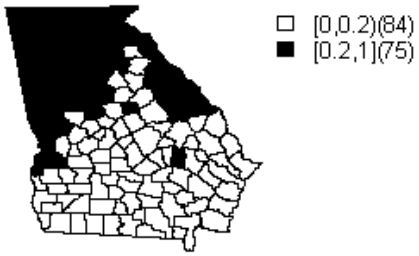
$p_{2,1}$ S5F2



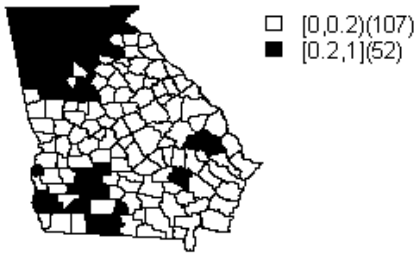
$p_{2,2}$ S5F2



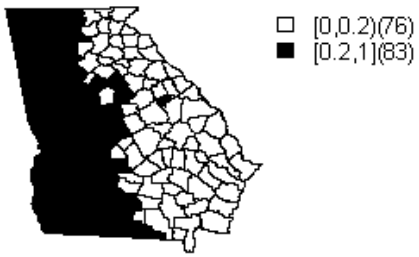
$p_{2,3}$ S5F2



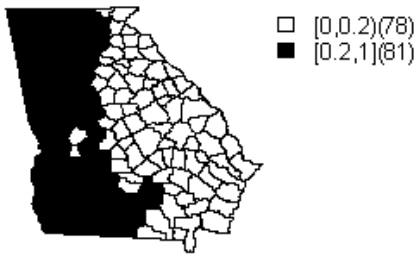
$p_{2,4}$ S5F2



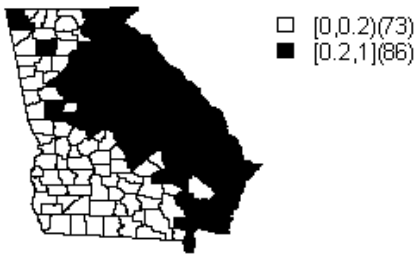
$p_{2,5}$ S5F2



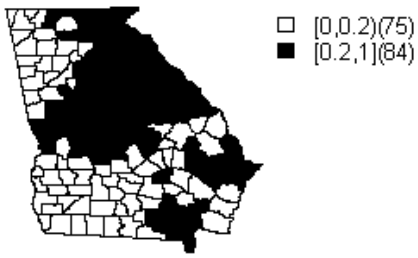
$p_{2,6}$ S5F2



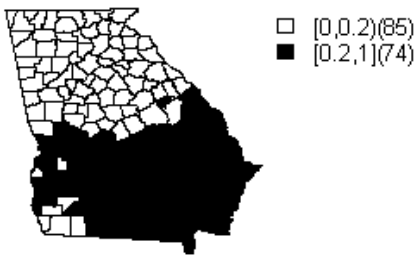
$p_{2,7}$ S5F2



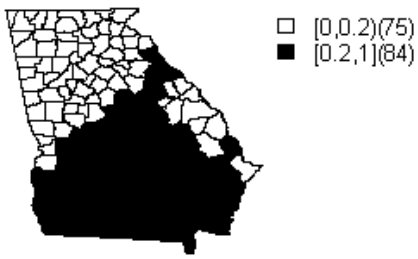
$p_{2,8}$ S5F2



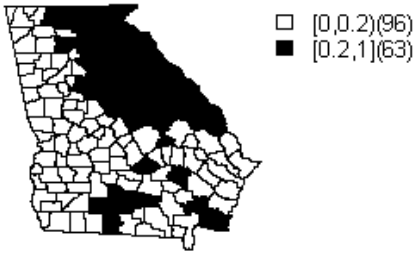
$p_{2,9}$ S5F2



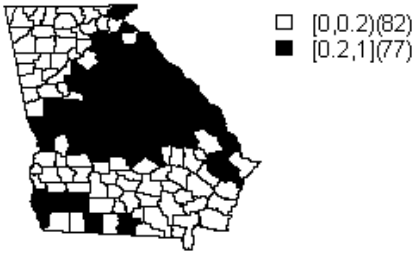
$p_{2,10}$ S5F2



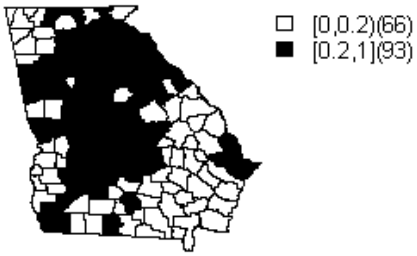
$p_{3,1}$ S5F2



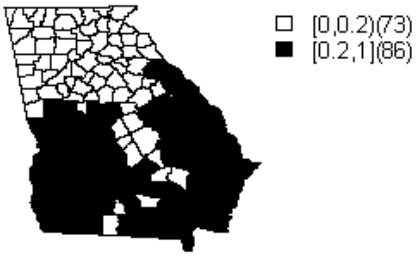
$p_{3,2}$ S5F2



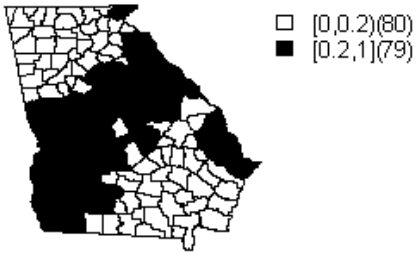
$p_{3,3}$ S5F2



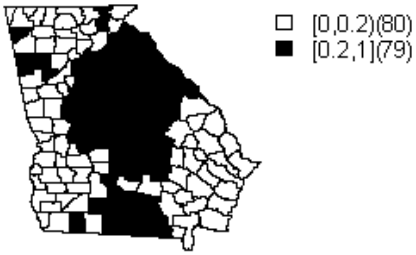
$p_{3,4}$ S5F2



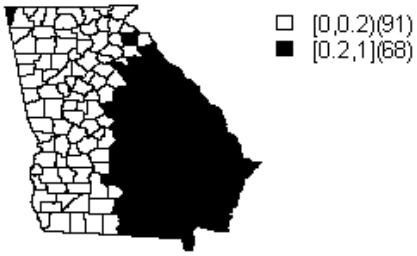
$p_{3,5}$ S5F2



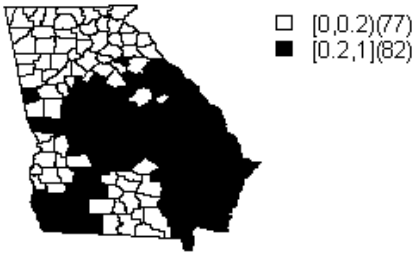
$p_{3,6}$ S5F2



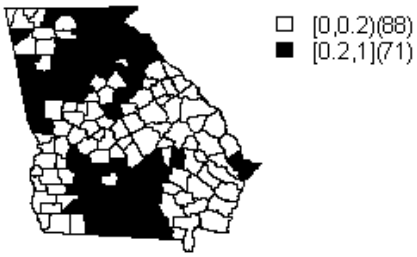
$p_{3,7}$ S5F2



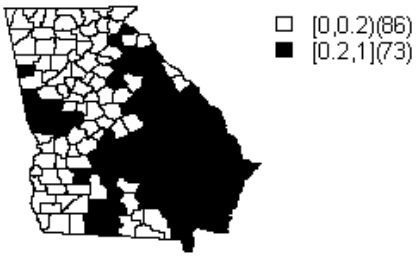
$p_{3,8}$ S5F2



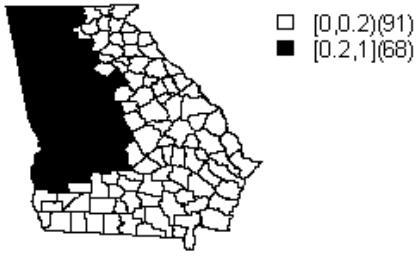
$p_{3,9}$ S5F2



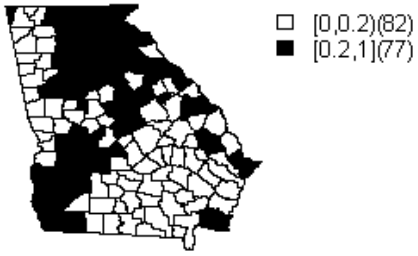
$p_{3,10}$ S5F2



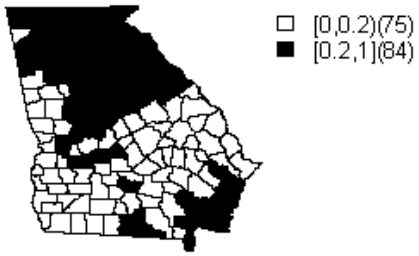
$p_{4,1}$ S5F2



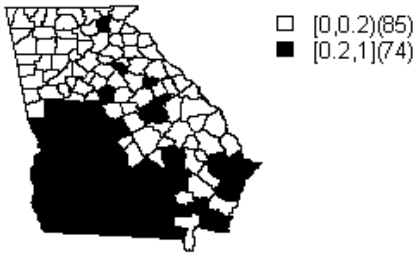
$p_{4,2}$ S5F2



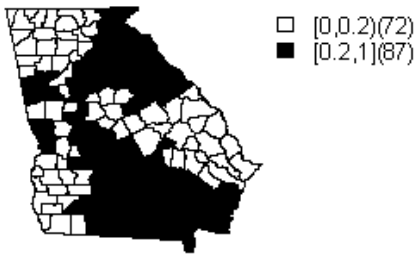
$p_{4,3}$ S5F2



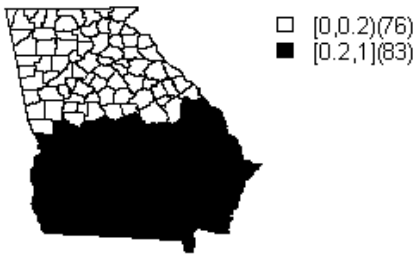
$p_{4,4}$ S5F2



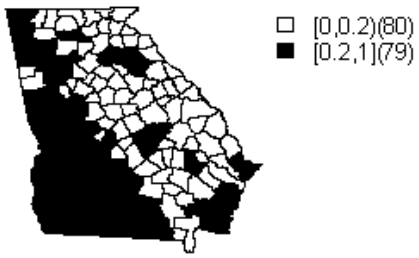
$p_{4,5}$ S5F2



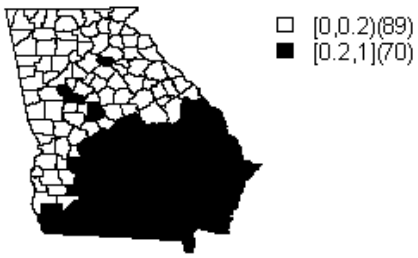
$p_{4,6}$ S5F2



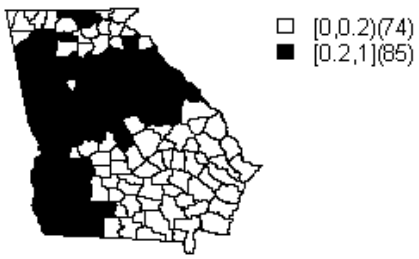
$p_{4,7}$ S5F2



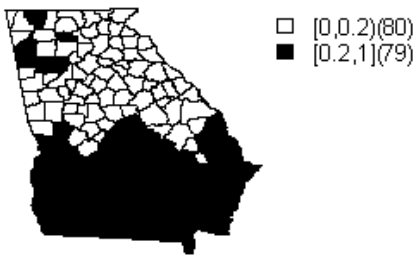
$p_{4,8}$ S5F2



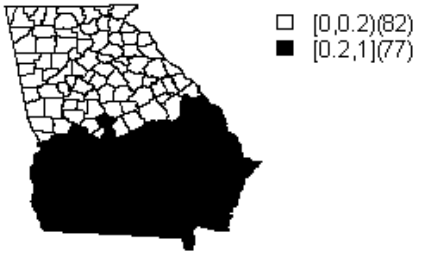
$p_{4,9}$ S5F2



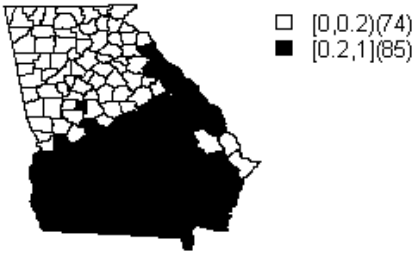
$p_{4,10}$ S5F2



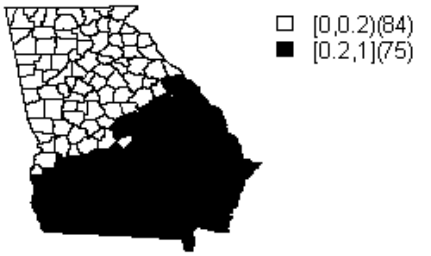
$p_{5,1}$ S5F2



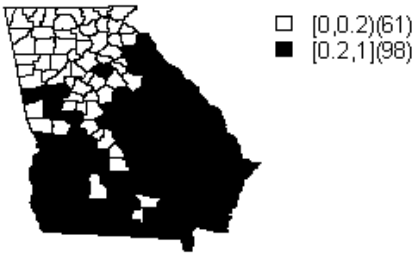
$p_{5,2}$ S5F2



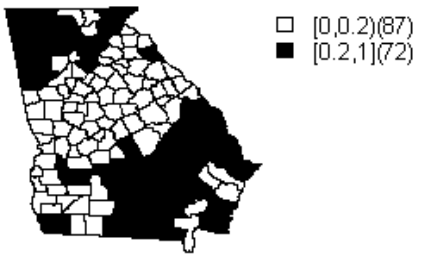
$p_{5,3}$ S5F2



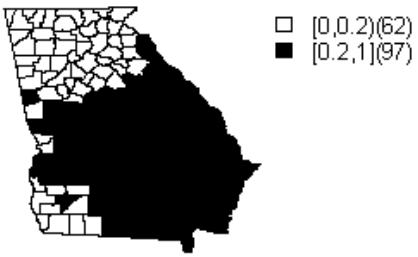
$p_{5,4}$ S5F2



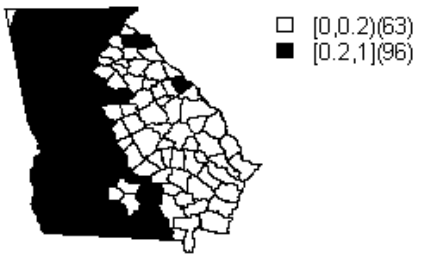
$p_{5,5}$ S5F2



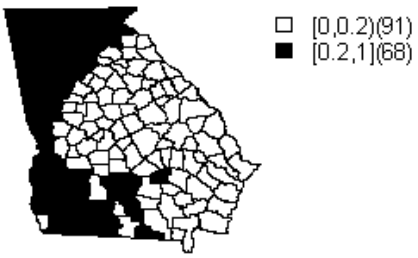
$p_{5,6}$ S5F2



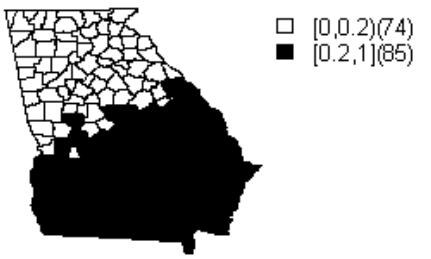
$p_{5,7}$ S5F2



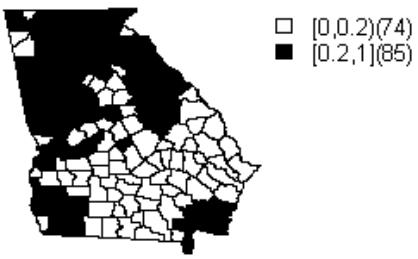
$p_{5,8}$ S5F2



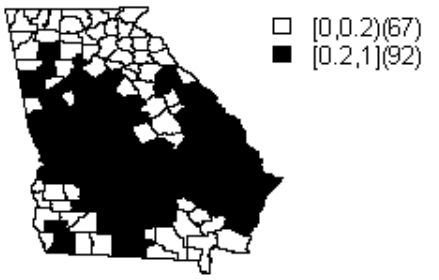
$p_{5,9}$ S5F2



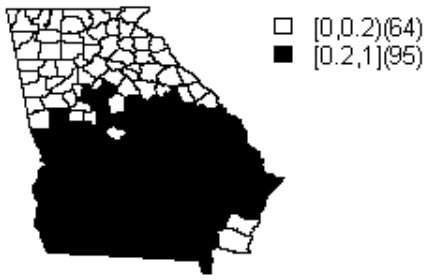
$p_{5,10}$ S5F2



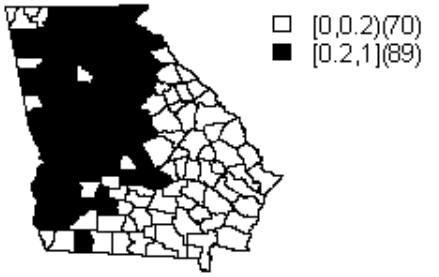
$p_{1,1}$ S10F2



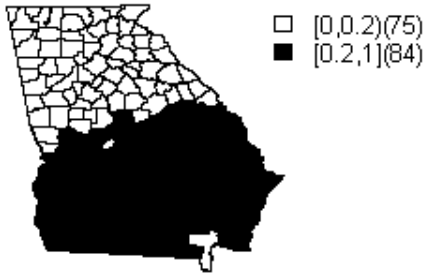
$p_{1,2}$ S10F2



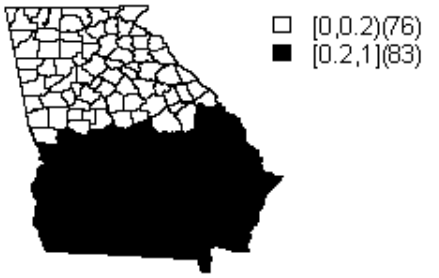
$p_{1,3}$ S10F2



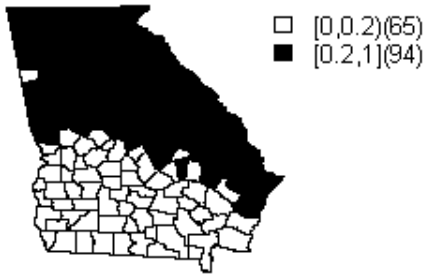
$p_{1,4}$ S10F2



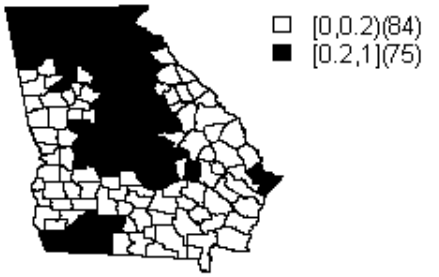
$p_{1,5}$ S10F2



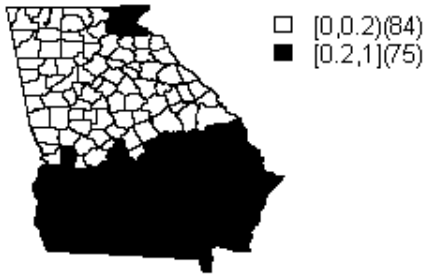
$p_{1,6}$ S10F2



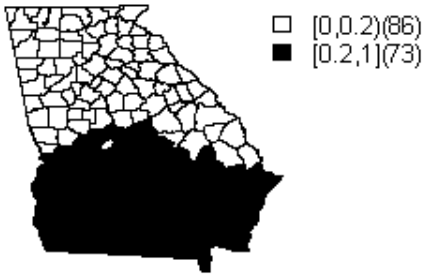
$p_{1,7}$ S10F2



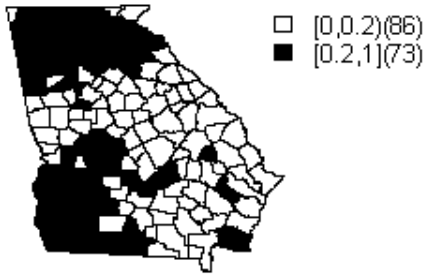
$p_{1,8}$ S10F2



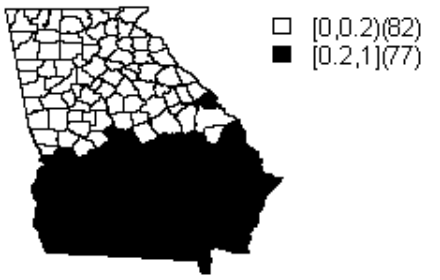
$p_{1,9}$ S10F2



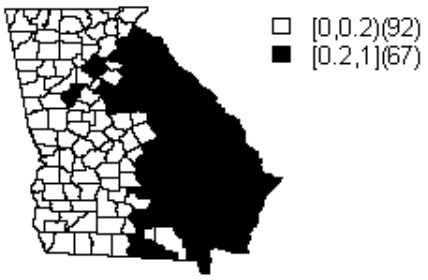
$p_{1,10}$ S10F2



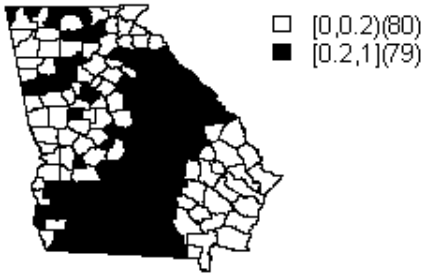
$p_{2,1}$ S10F2



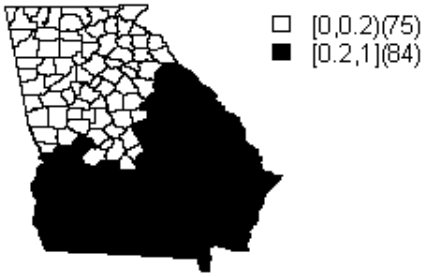
$p_{2,2}$ S10F2



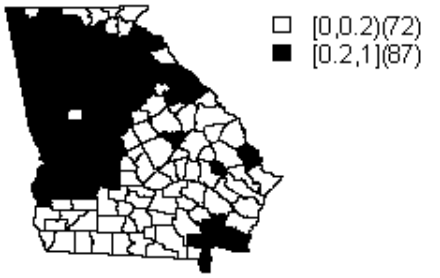
$p_{2,3}$ S10F2



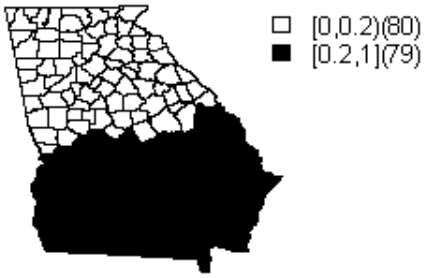
$p_{2,4}$ S10F2



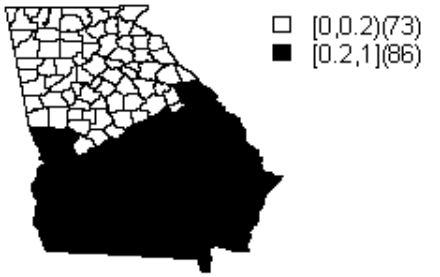
$p_{2,5}$ S10F2



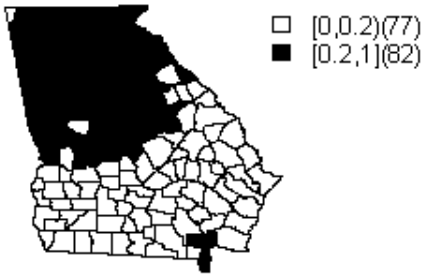
$p_{2,6}$ S10F2



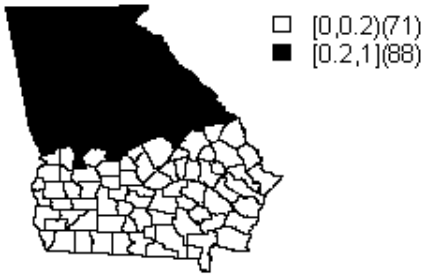
$p_{2,7}$ S10F2



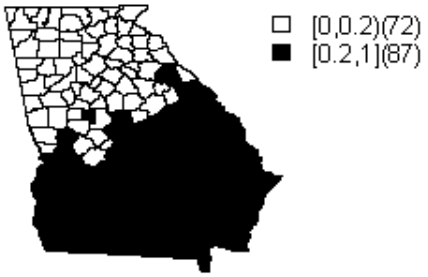
$p_{2,8}$ S10F2



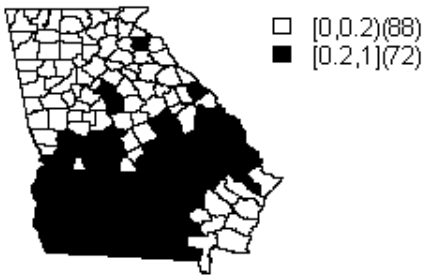
$p_{2,9}$ S10F2



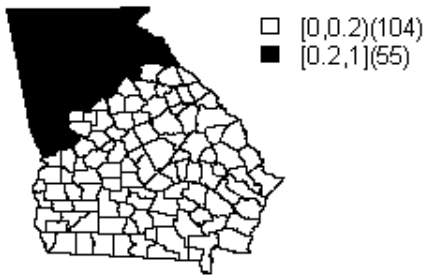
$p_{2,10}$ S10F2



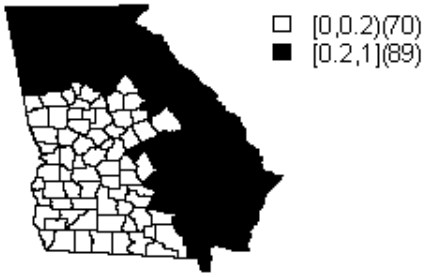
$p_{3,1}$ S10F2



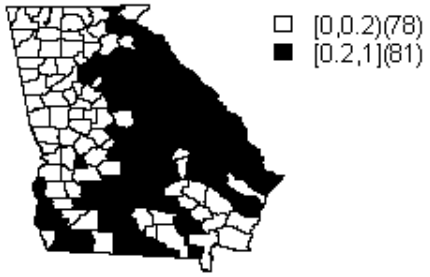
$p_{3,2}$ S10F2



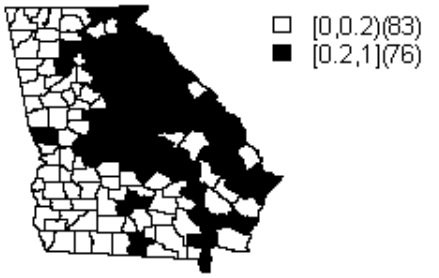
$p_{3,3}$ S10F2



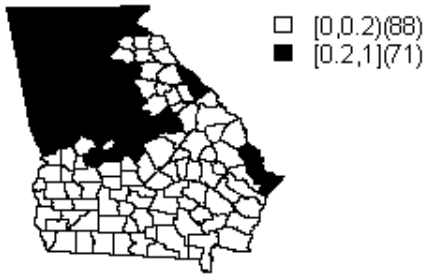
$p_{3,4}$ S10F2



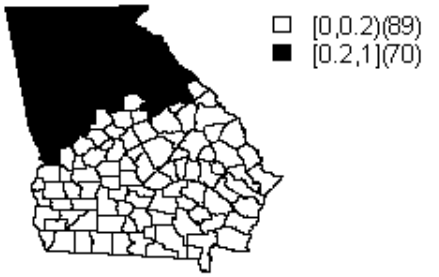
$p_{3,5}$ S10F2



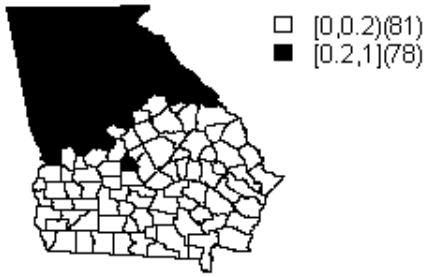
$p_{3,6}$ S10F2



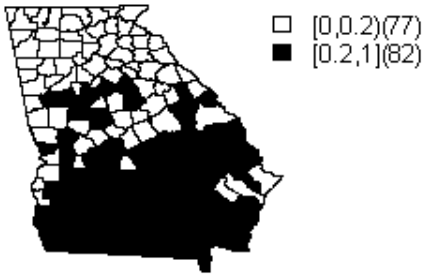
$p_{3,7}$ S10F2



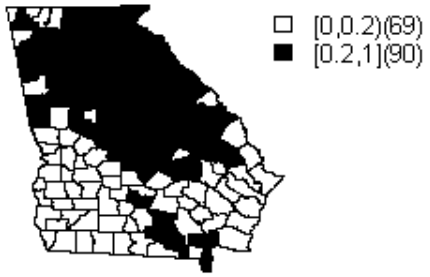
$p_{3,8}$ S10F2



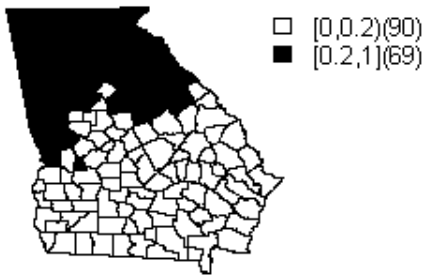
$p_{3,9}$ S10F2



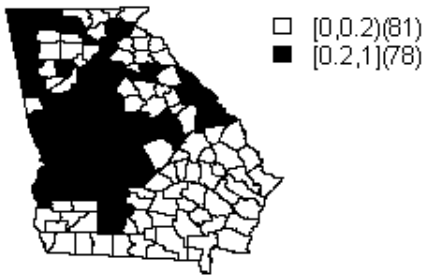
$p_{3,10}$ S10F2



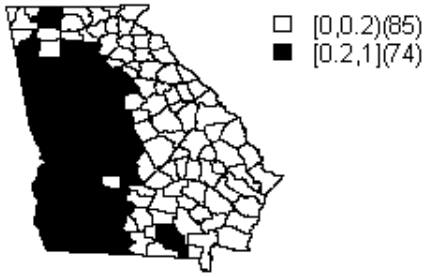
$p_{4,1}$ S10F2



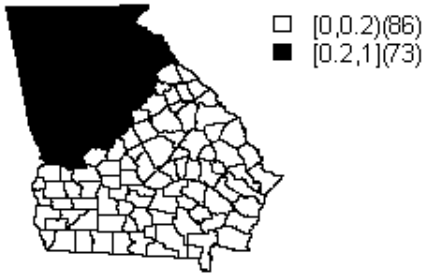
$p_{4,2}$ S10F2



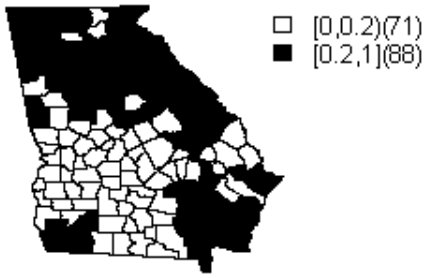
$p_{4,3}$ S10F2



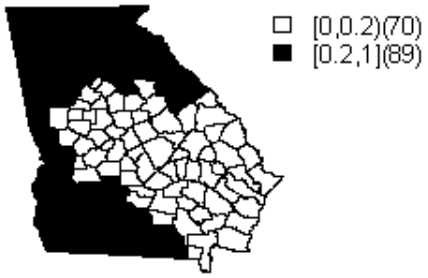
$p_{4,4}$ S10F2



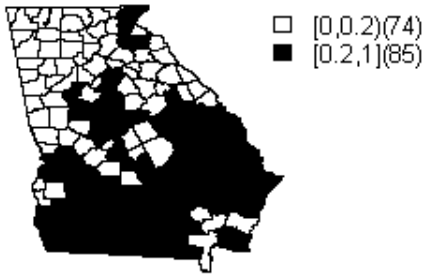
$p_{4,5}$ S10F2



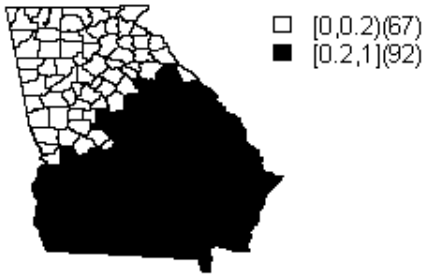
$p_{4,6}$ S10F2



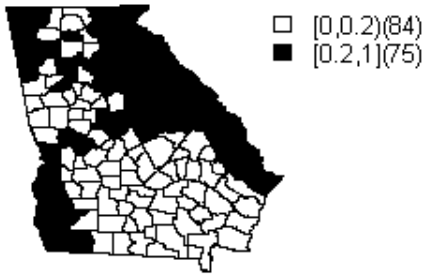
$p_{4,7}$ S10F2



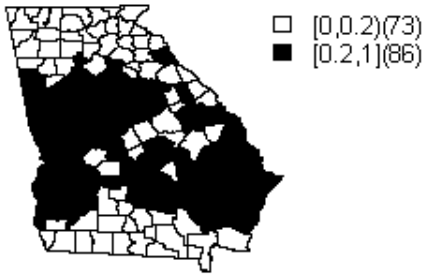
$p_{4,8}$ S10F2



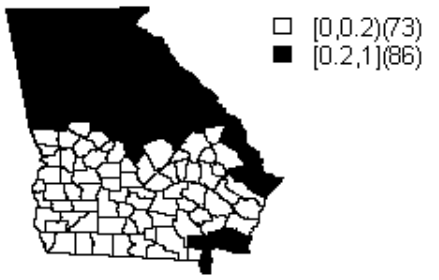
$p_{4,9}$ S10F2



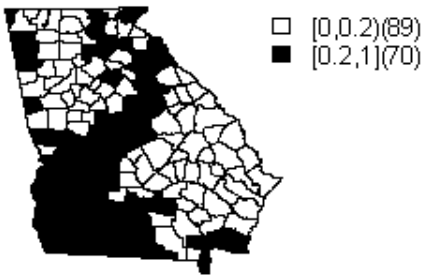
$p_{4,10}$ S10F2



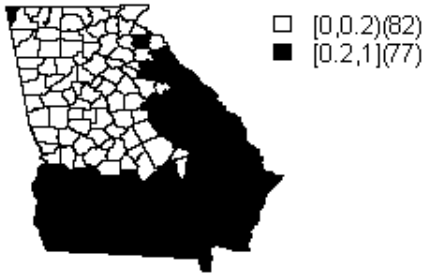
$p_{5,1}$ S10F2



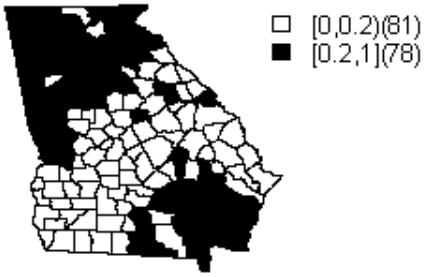
$p_{5,2}$ S10F2



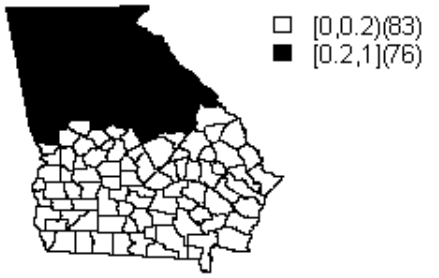
$p_{5,3}$ S10F2



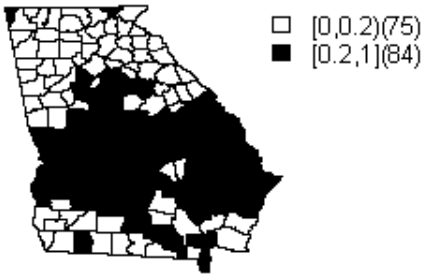
$p_{5,4}$ S10F2



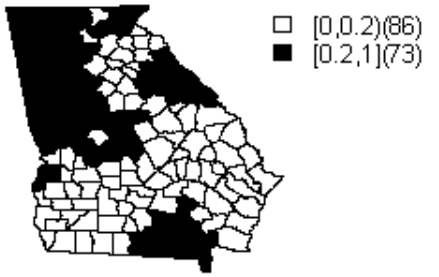
$p_{5,5}$ S10F2



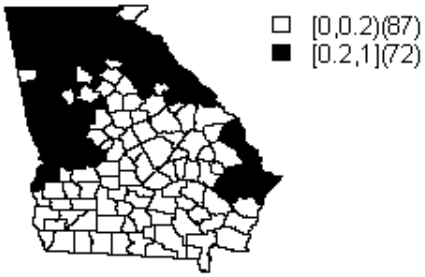
$p_{5,6}$ S10F2



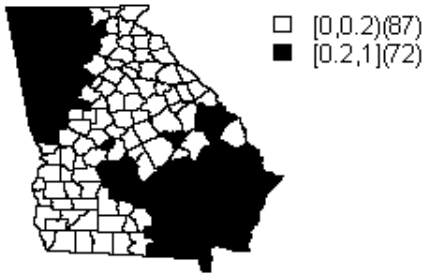
$p_{5,7}$ S10F2



$p_{5,8}$ S10F2



$p_{5,9}$ S10F2



$p_{5,10}$ S10F2

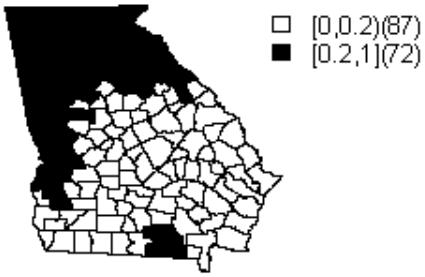


Table 1: Random walk estimates for the appropriate selected linear predictors.

Model	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7
F4	-0.0001	0.0004	0.0009	0.001	0.0005	0.0006	0.00009
	0.0003	0.0006	0.002	0.002	0.003	0.003	0.003
F5	-0.002	-0.003	-0.005	-0.003	-0.002	-0.004	-0.001

Figure 2: Theta estimates for fitted model F3 in the real data example.

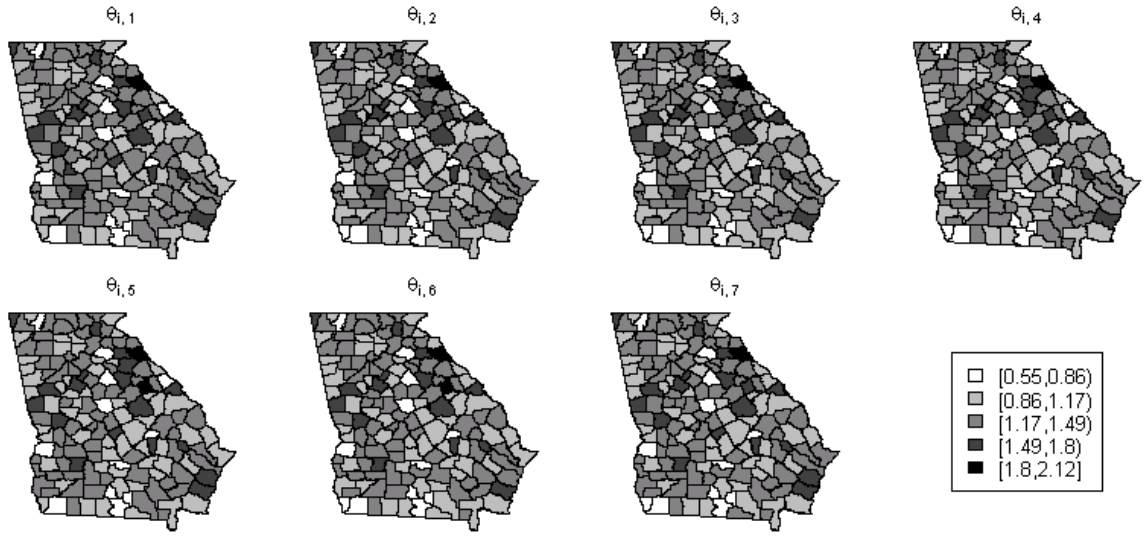


Figure 3: UH and CH estimates for fitted model F3 in the real data example.

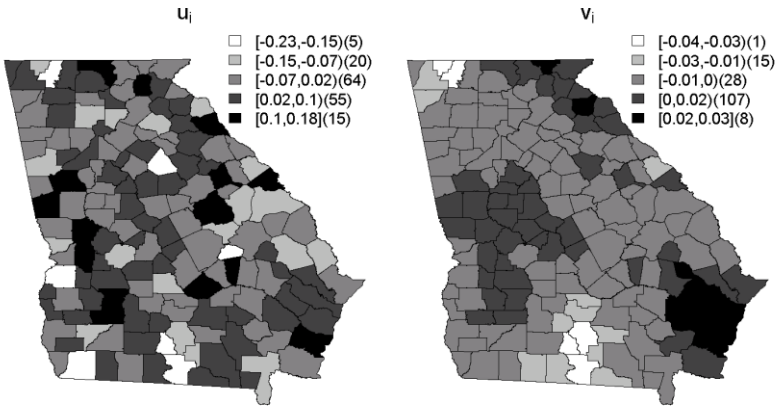


Figure 4: Theta estimates for fitted model F4 in the real data example.

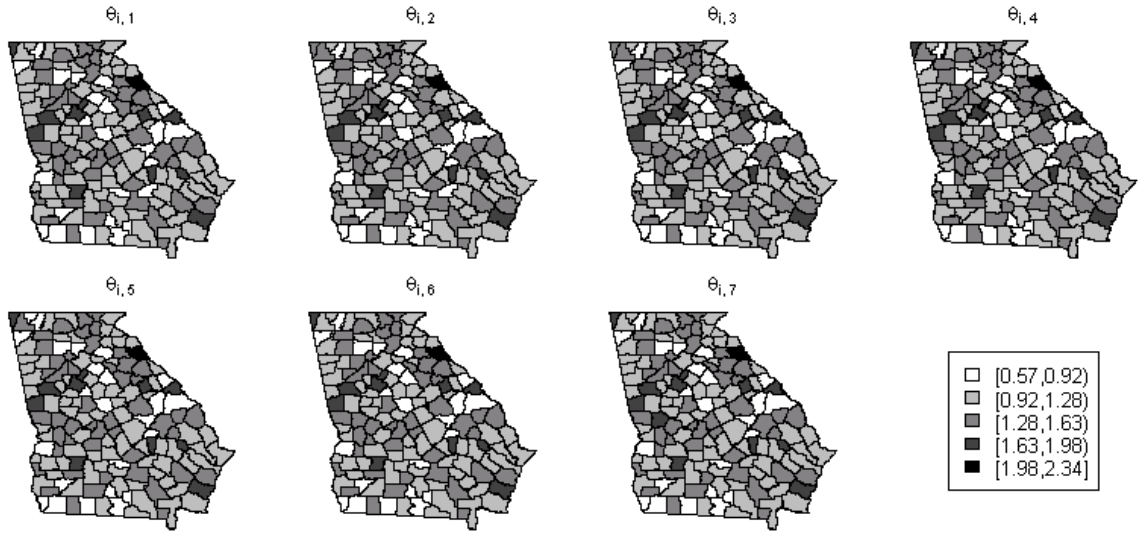


Figure 5: UH and CH estimates for fitted model F4 in the real data example.

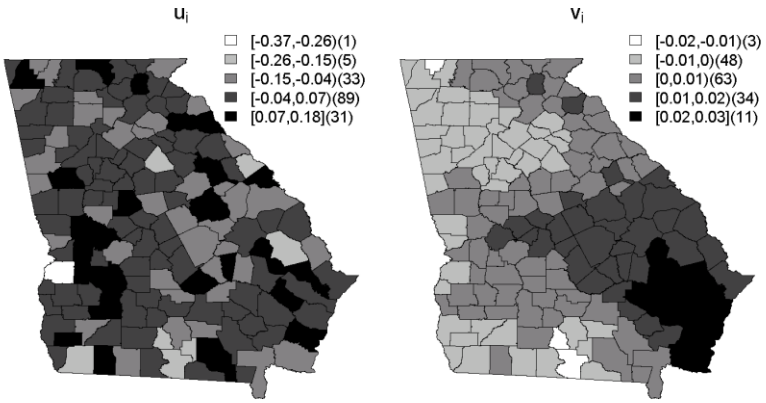


Figure 6: Theta estimates for fitted model F5 in the real data example.

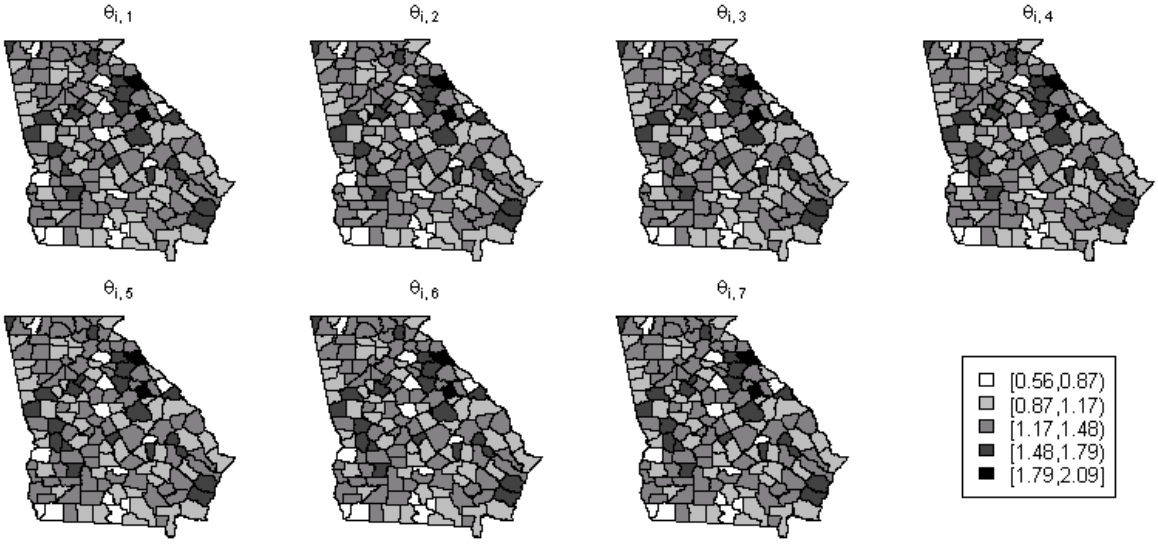


Figure 7: UH and CH estimates for fitted model F5 in the real data example.

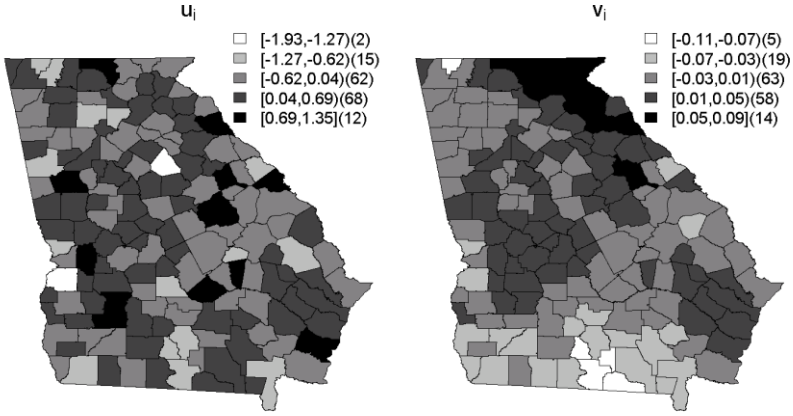


Table 2: Colon cancer data example model re-fit DIC measures.

Model	*DIC*	*pD*	$\bar{D}$
F3: UH+CH	5071.31	348.84	4722.47
F3: UH	4934.36	201.33	4733.03
F5	4903.09	160.72	4742.36

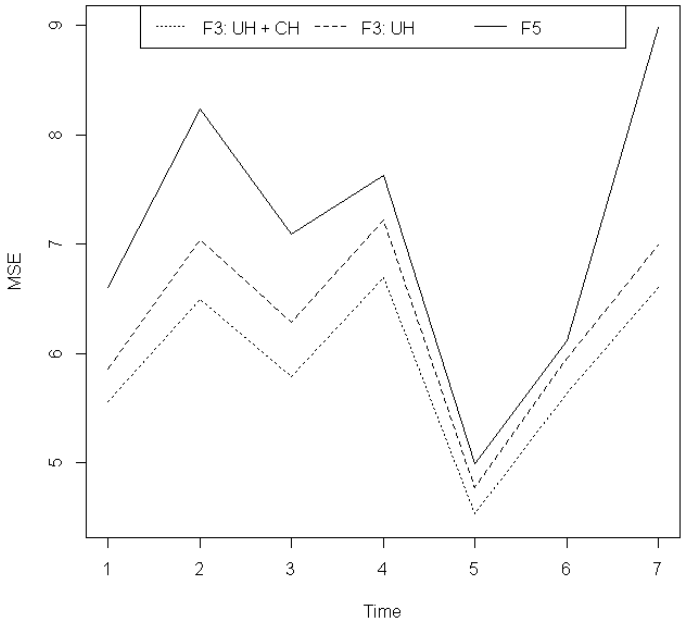


Figure 8: Colon cancer data example model re-fit MSE values.

Table 3: Parameter Estimates for the colon cancer data example.

Parameter	F3: UH	F3: UH+CH	F5
α_{01}	0.373*	0.391*	-0.295*
α_{02}	0.373*	0.379*	
α_{03}	0.390*	0.397*	
α_{04}	0.391*	0.406*	
α_{05}	0.419*	0.424*	
α_{06}	0.428*	0.445*	
α_{07}	0.458*	0.465*	
α_{1}	-0.0004	-0.002	-0.010*
α_{2}	0.005	0.011	0.160*
α_{31}	0.009	0.011	0.000
α_{32}			0.000
α_{33}			0.000
α_{34}			0.000
α_{35}			0.000
α_{36}			0.000
α_{37}			0.000
α_{41}	-0.011*	-0.009*	0.000
α_{42}			0.000
α_{43}			0.000
α_{44}			0.000
α_{45}			0.000
α_{46}			0.000
α_{47}			0.000

*indicates that the parameter is well estimated.