Iranian Journal of Mathematical Sciences and Informatics
Vol.2, No.1 (2007),pp 55-70
DOI: 10.7508/ijmsi.2007.01.006

AN ADDITIVE MODEL FOR SPATIO-TEMPORAL SMOOTHING OF CANCER MORTALITY RATES

GENTRY WHITE, DONGCHU SUN AND MARIO SCHOOTMAN

NORTH CAROLINA STATE UNIVERSITY
UNIVERSITY OF MISSOUSI-COLUMBIA
WASHINGTON UNIVERSITY

E-MAIL: WHITE@STAT.NCSU.EDU
E-MAIL: SUND@MISSOURI.EDU
E-MAIL: MSCHOOTM@IM.WUSTL.EDU

Abstract. In this paper, a Bayesian hierarchical model is used to anaylze the female breast cancer mortality rates for the State of Missouri from 1969 through 2001. The logit transformations of the mortality rates are assumed to be linear over the time with additive spatial and age effects as intercepts and slopes. Objective priors of the hierarchical model are explored. The Bayesian estimates are quite robustness in terms change of the hyperparamaters. The spatial correlations are appeared in both intercepts and slopes.

Keywords: Bayesian analysis, Spatial statistics, Cancer mortality, Poisson distributions
2000 Mathematics subject classification: 62P10

1. Introduction

Over the past decades a great deal of effort has been expended in the collection and compilation of high quality data on cancer incidence and mortality in the United States. These data have largely been used in the creation and disbursement of descriptive statistics concerning the state of cancer in the U.S. The information available through these statistics present limited information concerning spatial or temporal trends in the course of cancer in the U.S. Recently, there have been more efforts made to investigate these trends, such as Jackson-Thompson et al. (2006). National data on mortality due to cancer has been examined using a variety of methods in work by Mungiole and Pickle

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

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

The data set considered here consists of the observed number of deaths in each county in the State of Missouri due to female breast cancer from 1969 through 2001. The data are stratified into eleven three year time periods, and into four ten year age-group periods, $40-49$, $50-59$, $60-69$ and $70+$ years of age. One of the goals of the analysis proposed here is to devise a suitable spatio-temporal smoother for this data set.

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

2. A Hierarchical Model for Mortality Rates

In order to begin the description of the additive model, first consider the likelihood of the data. Let y_{ijk} denote the number of cases of a given disease for the i^{th} county, j^{th} time period, and k^{th} age-group. Here $i = 1, \cdots, I = 115$, $j = 1, \cdots, J = 11$, and $k = 1, \cdots, K = 4$. Given the population size n_{ijk} and rate p_{ijk}, we assume that y_{ijk} follows an independent Poisson distribution,

$$(y_{ijk} \mid p_{ijk}) \overset{indep.}{\sim} Poisson(n_{ijk} p_{ijk}). \tag{1}$$

We consider the following hierarchical model for a monotonic transformation of the rate p_{ijk},

$$\nu_{ijk} \equiv \log(p_{ijk}) = z_i + \theta_k + (w_i + \mu_k)(t_j - \bar{t}) + \epsilon_{ijk}, \tag{2}$$

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

$$\epsilon_{ijk} \overset{iid}{\sim} N(0, \delta_0). \tag{3}$$

As a result, the prior for ν_{ijk} is

$$(\nu_{ijk} \mid z_i, \theta_k, w_i, \mu_k, \delta_0) \sim N(z_i + \theta_k + (w_i + \mu_k)(t_j - \bar{t}), \delta_0). \tag{4}$$

This form of spatio-temporal interaction, in which county slopes are allowed to have spatial correlation, is first suggested in Sun et al. (2001). In order to complete this model the priors for δ_0, θ_k and μ_k must be specified. z_i and w_i also have prior distributions with hyper-parameters that have prior distributions of their own.

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

$$C_{uv} = \begin{cases} 1, & \text{if counties } u \text{ and } v \text{ are adjacent,} \\ 0, & \text{otherwise, with } C_{uu} = 0. \end{cases} \tag{5}$$

If $\boldsymbol{z} = (z_1, \ldots, z_I)'$, then the CAR prior for $\boldsymbol{z}$ is then defined by the density

$$[\boldsymbol{z} \mid \rho_1, \delta_1] = \frac{|\boldsymbol{I}_I - \rho_1 \boldsymbol{C}|^{1/2}}{(2\pi\delta_1)^{I/2}} \exp\left\{-\frac{1}{2\delta_1}\boldsymbol{z}'(\boldsymbol{I}_I - \rho_1\boldsymbol{C})\boldsymbol{z}\right\}, \tag{6}$$

where $\delta_1 > 0$. In order for this to be a proper prior, the values for ρ_1 are constrained such that $\rho_1 \in (\lambda_I^{-1}, \lambda_1^{-1})$, where λ_1 and λ_I are, respectively, the minimum and the maximum eigenvalues of the adjacency matrix $\boldsymbol{C}$. Note that this interval contains 0 because $\lambda_1 < 0 < \lambda_I$. In the case $\rho_1 = 0$, the z_i are independent. As in the case of $\boldsymbol{z}$, $\boldsymbol{w} = (w_1, .., w_I)'$ follows a similar CAR prior.

$$[\boldsymbol{w} \mid \rho_2, \delta_2] = \frac{|\boldsymbol{I}_I - \rho_2 \boldsymbol{C}|^{1/2}}{(2\pi\delta_2)^{I/2}} \exp\left\{-\frac{1}{2\delta_2}\boldsymbol{w}'(\boldsymbol{I}_I - \rho_2\boldsymbol{C})\boldsymbol{w}\right\}, \tag{7}$$

where $\delta_2 > 0$ and $\rho_2 \in (\lambda_1^{-1}, \lambda_I^{-1})$. Note that the correlation coefficients ρ_1 and ρ_2 are assumed to be independent of age-group and time period. The CAR priors (6) and (7) chosen for both $\boldsymbol{z}$ and $\boldsymbol{w}$ as in He and Sun (2000) have the benefits of additional correlation parameters not in other priors such as that proposed by Besag et al. (1991).

2.2. **Summary and Completion of the Hierarchical Model.** Evaluation of the model requires the likelihood and additional priors given here. The likelihood in (1) can be written in terms of $\nu_{ijk} = \log(p_{ijk})$,

$$[y_{ijk} \mid \nu_{ijk}] \propto \exp(\nu_{ijk} y_{ijk} - n_{ijk} e^{\nu_{ijk}}). \tag{8}$$

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

$$\theta_k \sim N(\xi_{mk}, \delta_{mk}), \tag{9}$$

$$\mu_k \sim N(\xi_{sk}, \delta_{sk}), \tag{10}$$

$$[\delta_l] \propto \frac{1}{\delta_l^{a_l - 1}} e^{(-b_l/\delta_l)},\ l = 0, 1, 2, \tag{11}$$

$$\rho_l \sim U(\lambda_1^{-1}, \lambda_I^{-1}),\ l = 1, 2. \tag{12}$$

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

2.3. **Propriety of Posterior Distribution.** To complete the hierarchical model, the hyper-parameters (ξ_{mj}, δ_{mj}), (ξ_{sj}, δ_{sj}) and (a_l, b_l) need to be specified. The commonly used objective prior for θ and μ are flat or constant priors. This can be a limiting case when δ_{mj} and $\delta_{sj} \to \infty$. Flat priors are naturally used for ρ_1 and ρ_2. Objective priors for δ_l can present problems. Traditionally the prior $1/\delta_l$ can be used for δ_l. The problem with this prior is that Sun et al. (2001) shows that the resulting joint posterior will be improper. In this model the priors $1/\sqrt{\delta_l}$ for the variance components are used. In this case, Sun et al. (2001) showed that the joint posterior is indeed proper.

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

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

Lemma 2.1.

a: *For given* $(z_i, \mu_k, w_i, \theta_k, \delta_0; y_{ijk})$, $(\nu_{111}, \ldots, \nu_{IJK})$ *are independent. Each* ν_{ijk} *depends only on* y_{ijk}, δ_0, *and* $a_{ijk} = z_i + \theta_k + (w_i + \mu_k)(t_j - \bar{t})$, *and has the following conditional density,*

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

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

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

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

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

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

g: $(\delta_0 \mid \boldsymbol{\nu}, \boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\theta}; \boldsymbol{y}) \sim IG\left(a_0 + \frac{IJK}{2}, b_0 + \frac{1}{2}\sum_{i,j,k}(\nu_{ijk} - a_{ijk})^2\right)$.

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

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

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

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

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

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

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

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

For part (k),

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

Part (m) is similar to part (k).

3. Results

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

3.1. **Objective and Data Dependent Priors.** The model is initially run using objective priors for δ_0, δ_1 and δ_2,

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

The resulting posterior means and variance are used to calculate a set of data-dependent priors for δ_0, δ_1 and δ_2 by inflating the mean and variances by some inflation factors (IF). Typically $IF = (2, 200)$ for the mean and variance respectively. Then we use the inflated mean and variance to calculate new values for the hyper-parameters of the priors. These new data-dependent priors are used to test the model for robustness in prior selection. Results in Table 1

for the objective priors (NI) and the inflated priors show that the model does indeed appear to be robust in terms of prior selection.

TABLE 1. Quantiles of (δ_0, δ_1, δ_2, ρ_1, ρ_2) for Objective (NI) and Inflation Factor (IF) Data-dependent Priors

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

3.2. **Age Effects θ_k and μ_k.** The range of posterior means of θ_k indicates a steady increase in mortality due to female breast cancer with age. Figure 1 shows that the rates for each age group appear to be increasing as age increases; the mortality rates are higher for older age-groups, regardless of location in space or time.

The posterior means for μ_k shown in Figure 1 demonstrate the change in rates with respect to age over time. The negative values for the two youngest age groups indicate a decrease in mortality rates over the time period for the those age groups, and the positive values for the two oldest age groups represent an increase in the mortality rates over time for those age groups. Looking at both plots indicates that the rates for the youngest age groups are less than the older age groups but also that there is a discrepancy in the rates over time; they are increasing for the older age groups but decreasing for the younger age

groups. This interaction between age and time is indication of a possible cohort effect.

3.3. **Variance Components δ_0, δ_1 and δ_2.** The relative importance of $\boldsymbol{z}$ and $\boldsymbol{w}$ and can be seen in their respective variances, δ_1, δ_2 and in δ_0, whose posterior distributions are shown in Figure 1. The mean of the posterior distribution of δ_2 is smaller than that of δ_0 and δ_1. In addition the posterior density of ρ_2 is quite diffuse and centered about 0. These results indicate that the $\boldsymbol{w}$ components are superfluous to the model, the small variance and density of ρ_2 indicates that there is little contribution by these terms.

3.4. **Spatial Correlation Parameters ρ_1 and ρ_2.** The plots of the posterior densities of ρ_1 and ρ_2 in Figure 1 show that the parameter ρ_1 for $\boldsymbol{z}$ is clearly non-zero, but the density for ρ_2, the similar parameter in the distribution of $\boldsymbol{w}$, is widely spread about 0. The implication in this is that the spatial structure between rates over time is not significant, even though the spatial effect overall is. In the CAR prior for $\boldsymbol{w}$ when ρ_2 is equal to zero, the w_i are in fact i.i.d, and would have the effect of adding random noise to the age group component of the temporal slope. This could lead to identifiability issues as the model would then have effectively two error terms.

3.5. **Disease Mapping.** The maps in Figures 2 and 3 compare the results of the additive model to the raw estimates of rates. As can be seen from these maps, the estimates from the additive model greatly smooth the raw data. Little if any spatial pattern evident in the data is visible from these maps. The smoothing that is taking place is due to the θ_k terms dominating the model estimates. As a result, the rate estimates are being smoothed toward a mean age-group effect. This result contradicts the maps of the raw rates, which show some indication of possible spatial patterns in the rates, though in reality interpretations of the raw rate maps could be misleading.

The second four sets of maps in Figures 4–7 show the rates for each age group through each time period. These clearly show the trend over time for the two youngest age groups ($k = 1, 2$) to be decreasing. The rates for the third age group ($k = 3$) are flat, and the rates for the oldest age group ($k = 4$) are increasing. These results verify what can be seen in the previous figures showing the posterior densities for μ_k.

4. Comments

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

terms of detecting age group differences in both mean rate and the temporal slope of the mean rate.

References

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

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

[3] J. Besag, J. York and A. Mollie, Bayesian image restoration with two applications in spatial statistcs, *Annals of the Institue of Statistical Mathematics*, **43**, (1991), 1-59.

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

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

[6] S. Devesa, D. Grauman, W. Blot and J. F. J. Fraumeni, Cancer Surveillance Se- ries: Changing Geographic Patterns of Lung Cancer Mortality in the United States, 1950 Through 1994, *Journal of the National Cancer Institute*, **91**, (1999),1040-1050.

[7] O. J. Devine, M. E. Halloran and T. A. Louis, Empirical Bayes Methods for Stabil- ising Incidence Rates Prior to Mapping, *Epidemiology* ,**5**,(1994), 622-630.

[8] O. Devine and T. A. Louis, A Constrained Empirical Bayes Esitmator for Incidence Rates in Areas with Small Populations, *Statistics in Medicine*, **13**, (1994),1119-1133.

[9] O. Devine, T. A. Louis and M. E. Halloran, Empirical Bayes Estimators for Spatially Correlated Incidence rates, *Environmetrics*, **5**, (1994),381-398.

[10] J. Ferrandiz, A. Lopez, A. Llopis, M. Morales and M. l. Tererizo, Spatial interaction between neighboring counties: Cancer mortality data in Valencia (Spain), *Biometrics*, **51**, (1995),665-678.

[11] A. E. Gelfan and A. F. M. Smith Sampling-Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, **85**, (1990), 398-409.

[12] W. Gilks and P. Wild, Adaptive rejection sampling for Gibbs sampling, *Applied Statistics*, **41**,(1992), 337- 348.

[13] Z. He and D. Sun,Hierarchical Bayesian estimation of hunting success rates with spatial correlations, *Biometrics*, **56**,(2000), 360-367.

[14] J. Jackson-Thompson, F. Ahmed, R. R. German, S. Lai and C. Friedman, *Descrip- tive epidemiologu of colorectal cancer in the united states*, 1998-2001, Cancer (2006), .

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

[16] K. Manton, M. Woodbury, E. Tallard, W. B. Riggan, J. P. Creason and A. Pellom, Empirical bayes procedures for stabilizing maps of u.s. cancer mortality rates, *Journal of the American Statistical Association*, **84**, (1989).

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

[18] L. Picle, E. Feuer and B. Edwards, U.S. Predicted Cancer Incidence, 1999: *Com- plete Maps by County and State From Spatial Projection Models, monograph 5*, National Cancer Institute, Bethesda, MD. NIH Publication No. 03-5435(2003).

[19] D. Sun, R. Tsutakawa and Z. He, Propriety of Posteriors with Improper Priors in Hierarchical Linear Mixed Models,, *Statistica Sinica*, **11**,(2001),77-95.

[20] D. Sun, R. Tsutsakawa, H. Kim, and Z. He, Spatio-temporal interaction with disease mapping, *Statistics in Medicine*, **19**,(2000), 2015-2035.

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

[22] S. Zhang, D. Sun, Z. He, and M. Schootman, A Bayesian semiparametric model for colorectal cancer incidence, *Statistics in Medicine*, **136**, (2006),2873-2897.

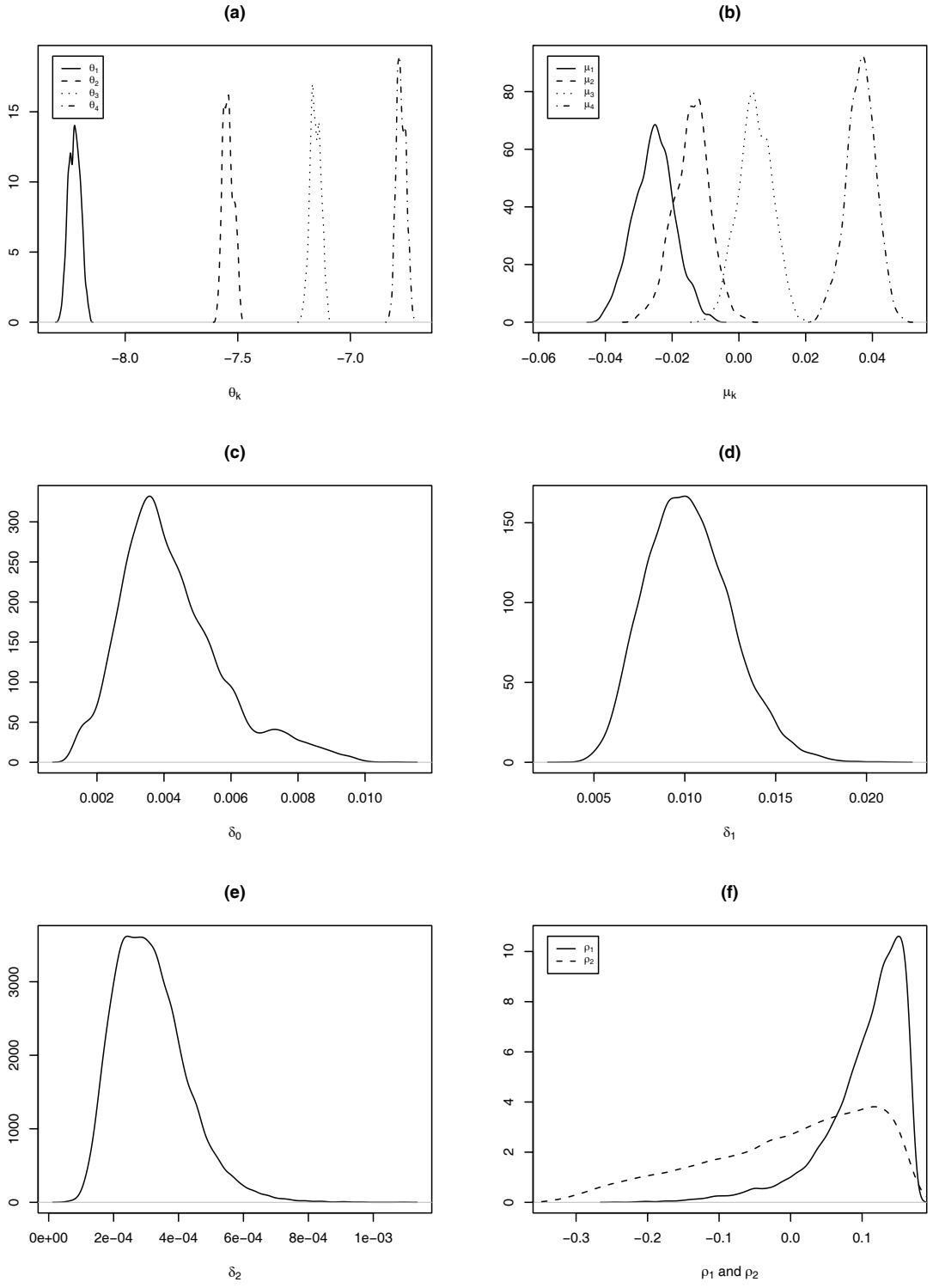

FIGURE 1. Posterior Densities of (a) θ_k, (b) μ_k, (c) δ_0, (d) δ_1, (e) δ_2, (f) ρ_1 and ρ_2 from Additive Model for Mortality Rates p_{ijk} for Female Breast Cancer in Missouri during 1969–2000.

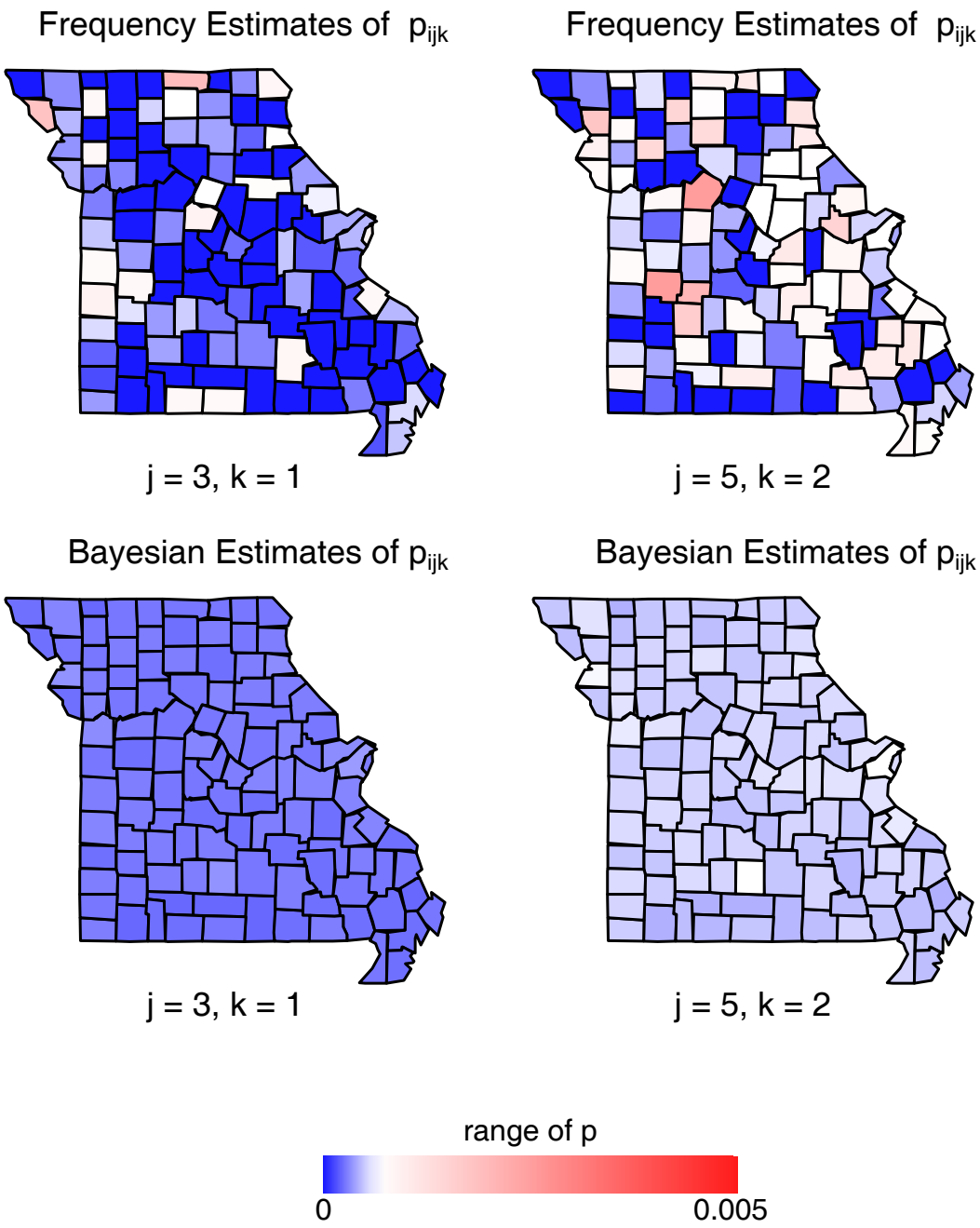

FIGURE 2. Maps of Frequency and Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri during 1969–2001 for $(j,k) = (3,1)$ and $(j,k) = (5,2)$.

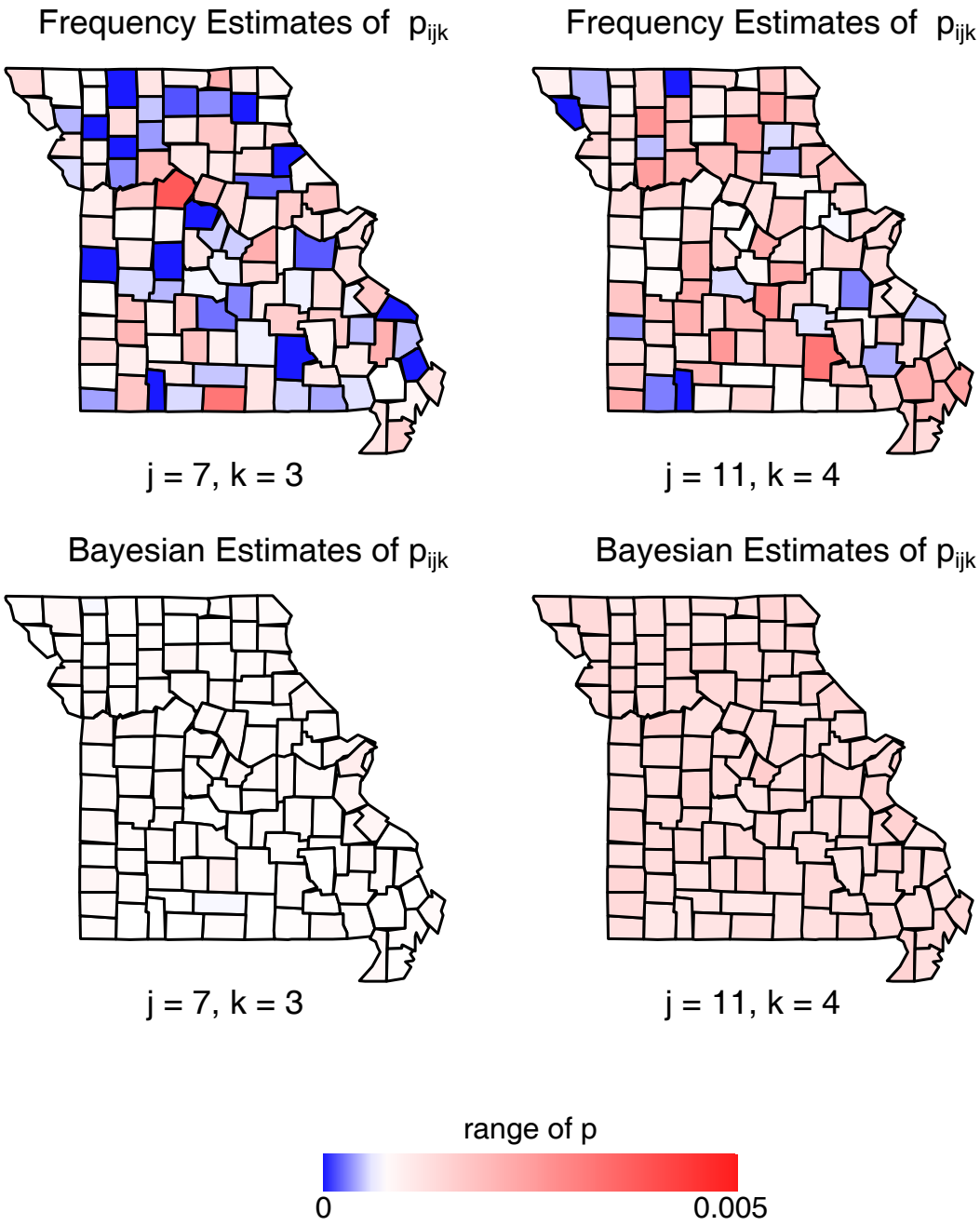

FIGURE 3. Maps of Frequency and Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri during 1969–2001 for $(j,k) = (7,3)$ and $(j,k) = (11,4)$.

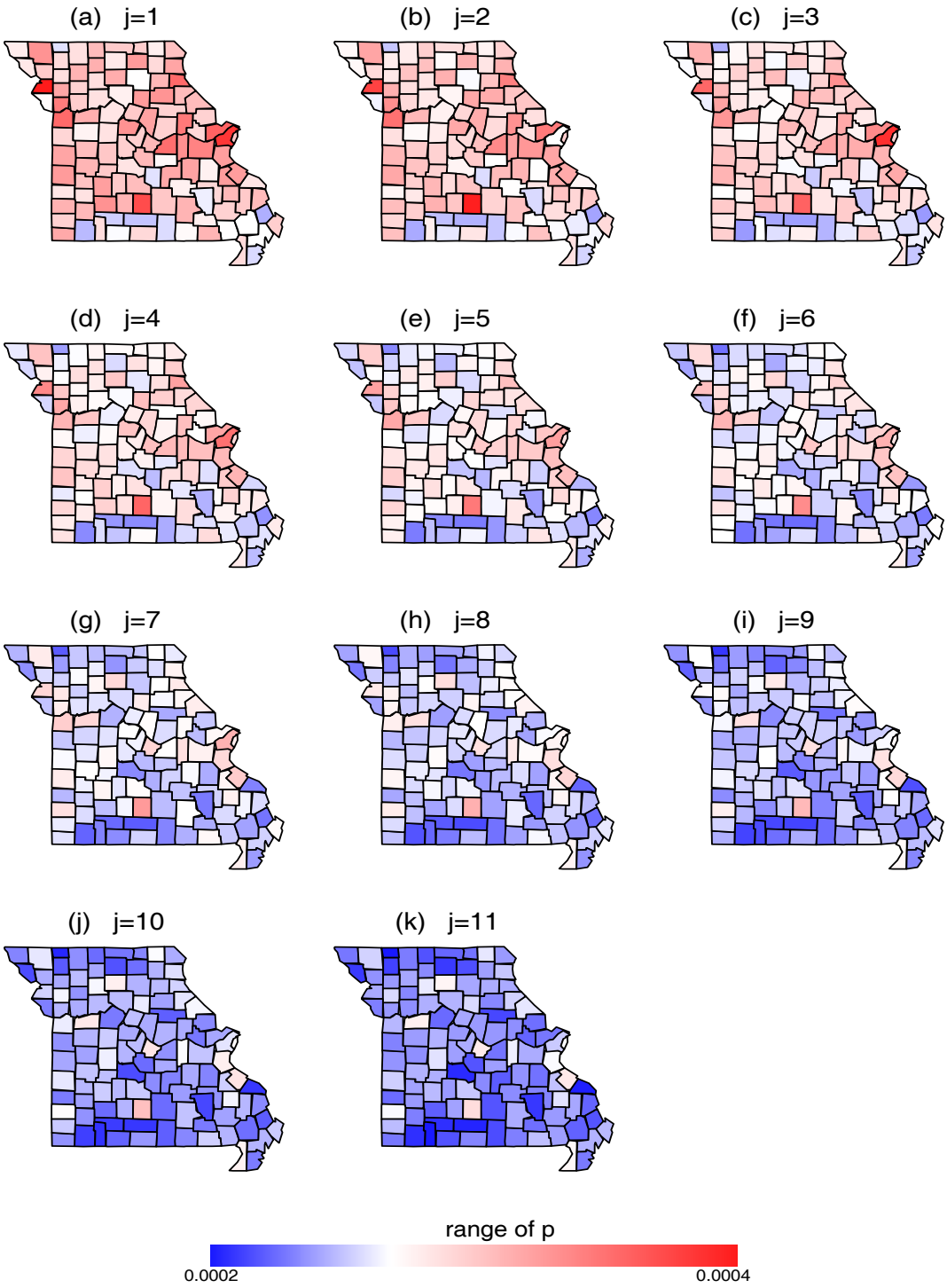

FIGURE 4. Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri during 1969-2001 for $j = 1, \ldots, 11$ and $k = 1$.

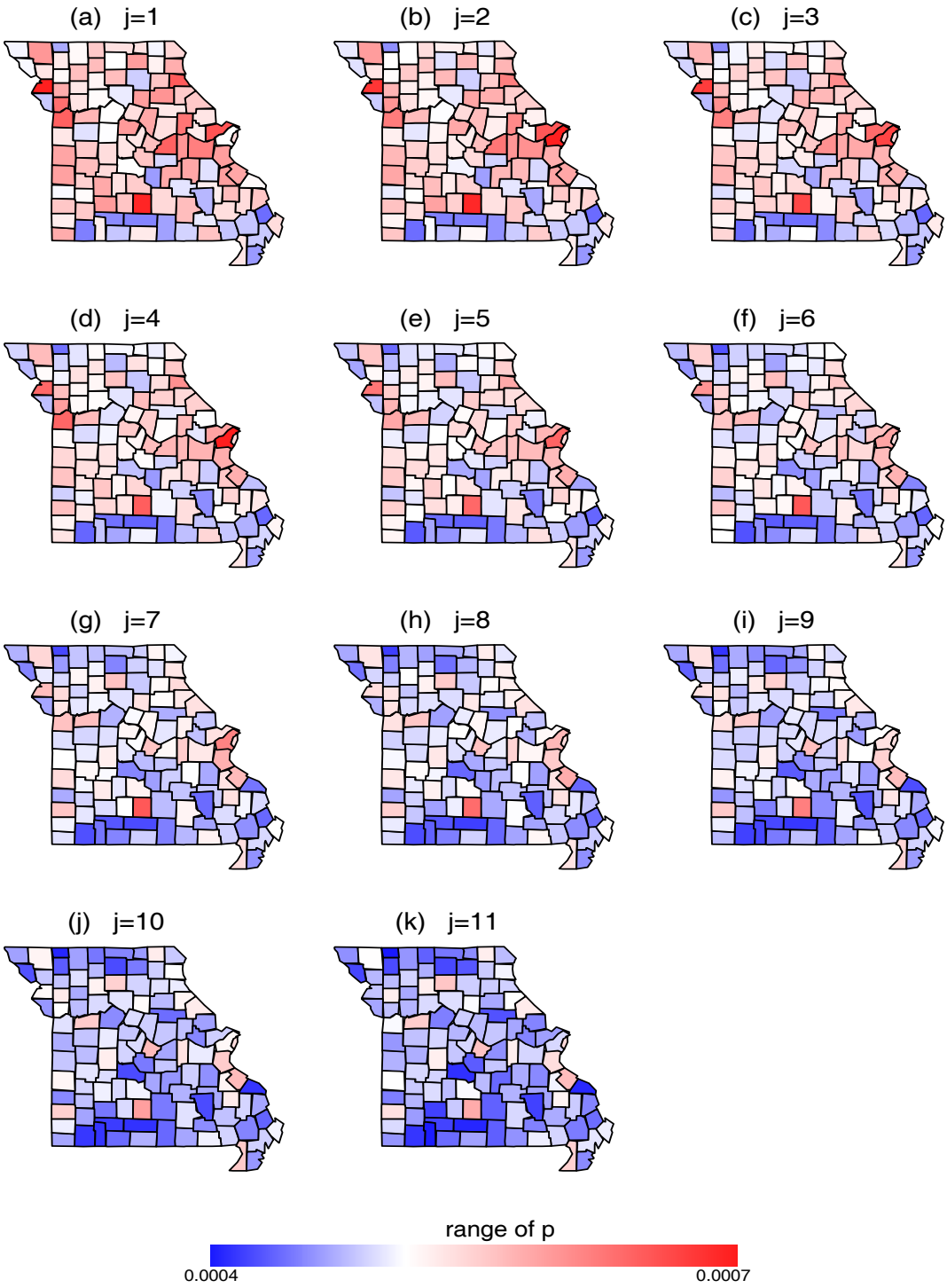

FIGURE 5. Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri during 1969-2001 for $j = 1, \ldots, 11$ and $k = 2$.

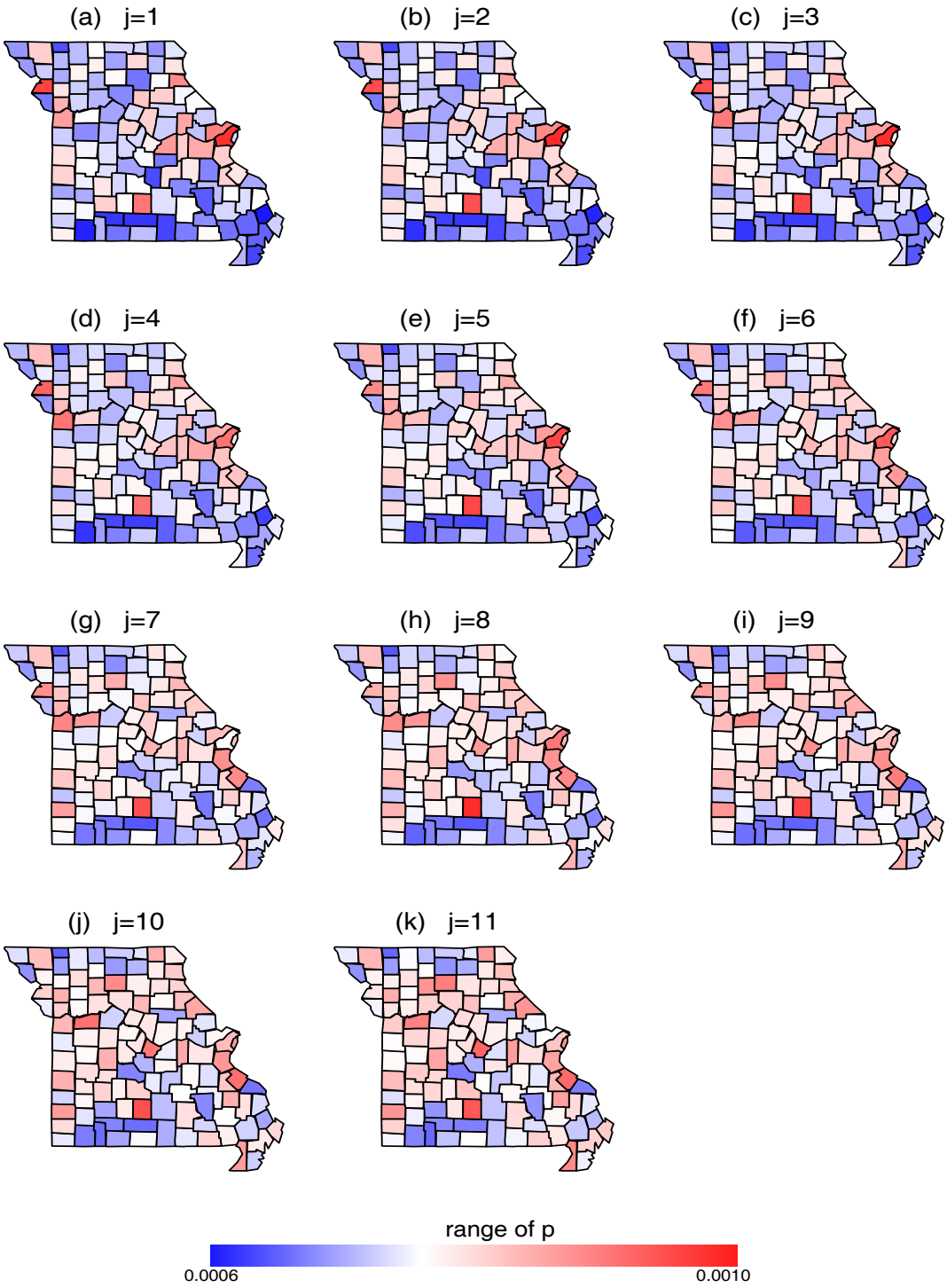

FIGURE 6. Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri during 1969-2001 for $j = 1, \ldots, 11$ and $k = 3$.

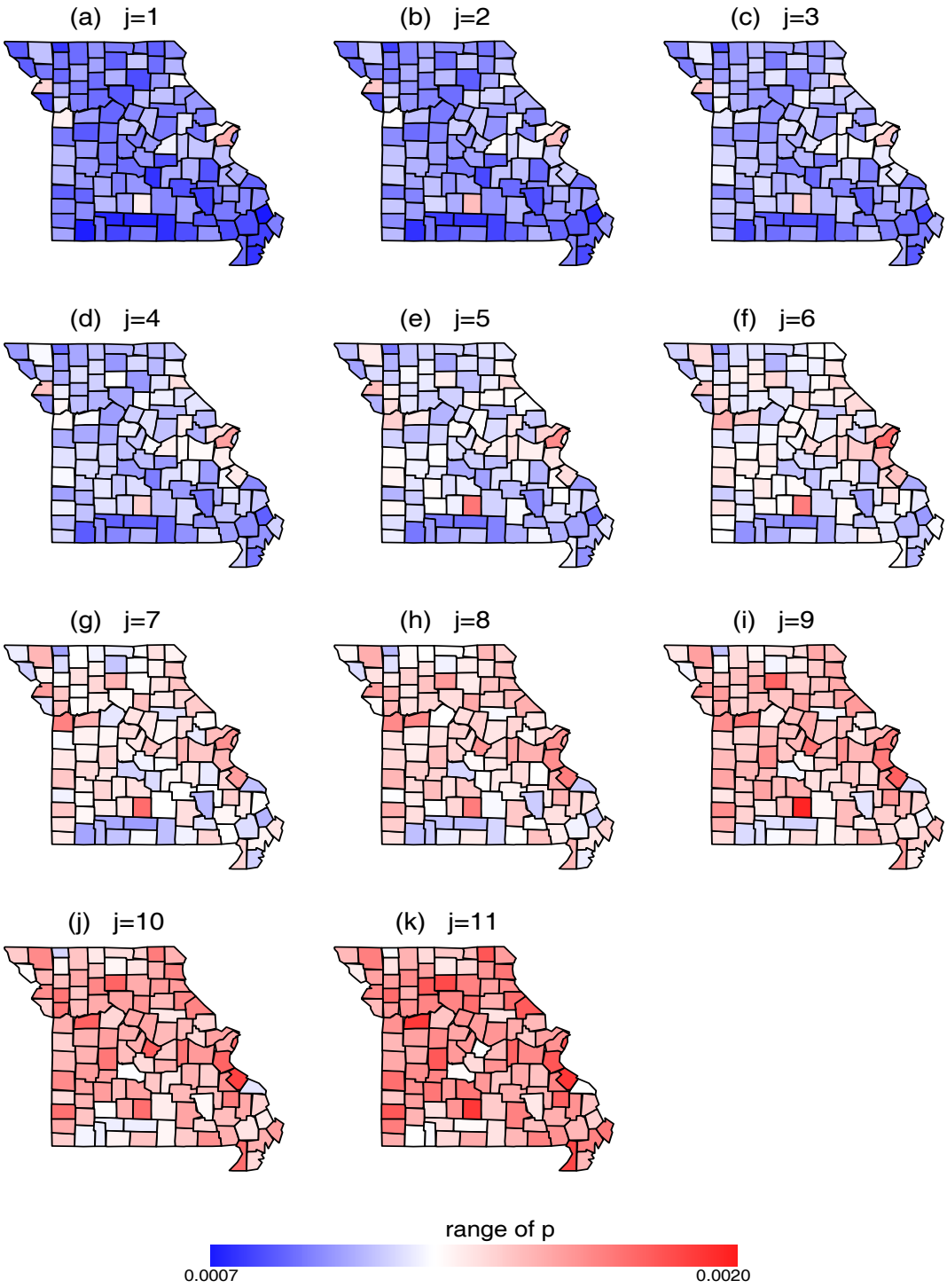

FIGURE 7. Maps of Bayesian Estimates of Mortality Rates p_{ijk} for Female Breast Cancer in Missouri during 1969-2001 for $j = 1, \ldots, 11$ and $k = 4$.