1. Download the google trends flu data corresponding to the first 12 weeks of 2013 from http://www.google.org/flutrends/us/data.txt for the contiguous continental states. Record a binary variable that is equal to 1 if the index for a given state in a given week is greater than 7,500, and 0 otherwise. These data correspond to binary data on a lattice. Write a spatial model that allows for known covariates, spatially structured variability, unstructured variability and uses a link function based on the distribution of a student random variable with fixed number of degrees of freedom. Assume that time can be included in the model as a covariate. Are there any other covariates that you should consider? Assume appropriate priors for the model parameters and write the full conditionals. Use the full conditionals to obtain samples from the model parameters and make inferences about the probability that the Google flu index will be higher than 7,500. Show maps for such probabilities.

Solution:

There are 48 contiguous states, and so the number of total data points is $48 \times 12 = 576$. The data were transformed to a binary data whose value is 1 if the index for a given state in a given week is greater than 7,500, and 0 otherwise.

The spatial model being considered is built as follows.

$$y_i = \begin{cases} 1 & \text{if } \omega_i \geq 0 \\ 0 & \text{otherwise} \end{cases}, \quad i = 1, \ldots, m = 576$$

$$\omega_i = \alpha + \beta t_i + \gamma_k(i) + \epsilon_i$$

$$\gamma_k = u_k + v_k, \quad k = 1, \ldots, n = 48$$

$$\pi(u) \propto \kappa_u^{0.5(n-1)} \exp \left\{ -\frac{\kappa_u}{2} \sum_{i \sim j} (u_i - u_j)^2 \right\}$$

$$v \sim N(0, \kappa_v^{-1}I)$$

$$\epsilon_i \sim N(0, 1/\lambda_i)$$

$$\lambda_i \sim \text{iid } Ga(d/2, d/2)$$

$$\pi(\alpha, \beta) \propto 1$$

$$\kappa_u \sim Ga(a_u, b_u)$$

$$\kappa_v \sim Ga(a_v, b_v)$$

We introduce auxiliary variable $\lambda_i$ that follows gamma distribution to have a scale mixture of normal that generate Student-t distribution with degrees of freedom $d$. The covariates of latent variables include intercept, time (week), and spatial variations. $v_k$ corresponds unstructured variation or nugget effect, and $u_k$ represents the first-order GMRF. Here we define two states as neighbors if they share the
same border. We use flat prior on the regression coefficients $\alpha$ and gamma prior on precision coefficients $\kappa_u$ and $\kappa_v$. The model parameters are $\alpha, \beta, u, v, \kappa_u, \kappa_v, \omega, \lambda$. The posterior density can be written as

$$
\pi(\alpha, \beta, u, v, \kappa_u, \kappa_v, \omega, \lambda|y) \propto \pi(y|\omega)\pi(\omega|\alpha, \beta, u, v, \lambda)\pi(u|\kappa_u)\pi(v|\kappa_v)\pi(\alpha, \beta)\pi(\kappa_u, \kappa_v)\pi(\lambda)
$$

We then derive the full conditionals for the model parameters. From our model, we have for $i = 1, \ldots, m$,

$$(\omega_i | \cdots) \sim TN(\alpha + \beta t_i + \gamma_k(i), 1/\lambda_i),$$

meaning that $(\omega_i | \cdots)$ is normally distributed but truncated to be positive if $y_i = 1$ and negative if $y_i = 0$.

For $k = 1, \ldots, n$,

$$\pi(v_k | \cdots) \propto \exp\left\{-\frac{\kappa_v}{2} v_k^2\right\} \exp\left\{-\frac{1}{2} \sum_{i:k(i)=k} \lambda_i (v_{k(i)}^2 - 2(\omega_i - (\alpha + \beta t_i + u_{k(i)}))v_{k(i)})\right\}$$

$$\propto \exp\left\{-\frac{1}{2} \left(\kappa_v + \sum_{i:k(i)=k} \lambda_i\right)v_k^2 - 2 \sum_{i:k(i)=k} \lambda_i(\omega_i - (\alpha + \beta t_i + u_{k(i)}))v_{k(i)}\right\}$$

$$\sim N\left(\left(\kappa_v + \sum_{i:k(i)=k} \lambda_i\right)^{-1}\left[\sum_{i:k(i)=k} \lambda_i(\omega_i - (\alpha + \beta t_i + u_{k(i)}))\right], \left(\kappa_v + \sum_{i:k(i)=k} \lambda_i\right)^{-1}\right)$$

Next, for $k = 1, \ldots, n$,

$$\pi(u_k | \cdots) \propto \exp\left\{-\frac{\kappa_u}{2} (\sum_{k\sim j} (u_k^2 - 2u_ku_j))\right\} \exp\left\{-\frac{1}{2} \sum_{i:k(i)=k} \lambda_i (u_k^2 - 2(\omega_i - (\alpha + \beta t_i + v_{k(i)}))u_k)\right\}$$

$$\propto \exp\left\{-\frac{1}{2} \left(\kappa_u n_k + \sum_{i:k(i)=k} \lambda_i\right)u_k^2 - 2 \left(\kappa_u \sum_{j \sim k} u_j + \sum_{i:k(i)=k} \lambda_i(\omega_i - (\alpha + \beta t_i + v_{k(i)}))\right)u_k\right\}$$

$$\sim N(\mu_{u_k}, \sigma_{u_k}^2),$$

where $\sigma_{u_k}^2 = (\kappa_u n_k + \sum_{i:k(i)=k} \lambda_i)^{-1}$ and $\mu_{u_k} = \sigma_{u_k}^2 \left(\kappa_u \sum_{j \sim k} u_j + \sum_{i:k(i)=k} \lambda_i(\omega_i - (\alpha + \beta t_i + v_{k(i)}))\right)$ and $n_k$ is the number of neighbors of $u_k$.

Define $\beta = (\alpha, \beta)^T$, $x_i^t = (1, t_i)$, $X = (x_1^T, \ldots, x_m^T)^T$, and $\Sigma^{-1}_\omega = \text{diag}(\lambda_1, \ldots, \lambda_m)$. Then

$$\pi(\beta | \cdots) \propto \exp\left\{-\frac{1}{2} (\omega - X\beta - \gamma_k)^T \Sigma^{-1}_\omega (\omega - X\beta - \gamma_k)\right\}$$

$$\propto \left\{-\frac{1}{2} \left[\beta^T X^T \Sigma^{-1}_\omega X \beta - 2 \beta^T X^T \Sigma^{-1}_\omega (\omega - \gamma_k)\right]\right\}$$

$$\sim N(\mu_{\beta}, \Sigma_{\beta}),$$

where $\Sigma_{\beta} = (X^T \Sigma^{-1}_\omega X)^{-1}$, and $\mu_{\beta} = \Sigma_{\beta}[X^T \Sigma^{-1}_\omega (\omega - \gamma_k)]$. 


The followings show the full conditionals for the precision parameter and $\lambda$.

\[
\pi(\kappa_u \cdots) \propto \kappa_u^{((0.5)(n-1)+a_u)-1} \exp \left\{ -((0.5) \sum_{i \sim j} (u_i - u_j)^2 + b_u)\kappa_u \right\}
\sim Ga((0.5)(n - 1) + a_u, (0.5) \sum_{i \sim j} (u_i - u_j)^2 + b_u)
\]

\[
\pi(\kappa_v \cdots) \propto \kappa_v^{((0.5)(n-1)+a_v)-1} \exp \left\{ -((0.5) \sum_{k=1}^n v_k^2 + b_v)\kappa_v \right\}
\sim Ga((0.5)(n - 1) + a_v, (0.5) \sum_{k=1}^n v_k^2 + b_v)
\]

For $i = 1, \ldots, m$,

\[
\pi(\lambda_i \cdots) \propto (\lambda_i)^{(0.5)(d+1)-1} \exp \left\{ -((0.5)[(\omega_i - (\alpha + \beta t_i + \gamma k(i))]^2 + d))\lambda_i \right\}
\sim Ga((0.5)(d + 1), (0.5)[(\omega_i - (\alpha + \beta t_i + \gamma k(i))]^2 + d))
\]

Given the full conditionals above, we implement Gibbs sampler with hyperparameters $a_u = a_v = 1$, $b_u = b_v = 0.001$ and $d = 1$ (Cauchy distribution). I found that the algorithm is poor at mixing, so I draw 50000 times and keep 1000 draws after burning and thinning. To get the probability of $y = 1$, for each posterior draw, I sample $\omega_i$ using $N(\alpha + \beta t_i + \gamma k(i),1/\lambda_i)$ 1000 times and the probability is the number of times that $\omega > 0$ divided by 1000. The original data and the probability map are shown in Figure 1 and Figure 2 respectively. We can see that the prediction is quite good. Note that Oregon state in week 1 has high probability but is labelled as 0 from the data. One possible reason is that its neighbors are all labelled as 1, which is not same for the states Utah and Colorado and states Iowa and Missouri.
Figure 1: Binary data map
Figure 2: Binary data map