Carleton College
Mathematics Department

A Bayesian Spatial Temporal Model of Ozone

Submitted in Fulfillment of Integrative Exercise 2008-2009

Christina Knudson
Edward Kuhn
Bassirou Sarr

Acknowledgements

We would like to thank Professor Katie St. Clair, our advisor for this project, for her guidance and thoughtful comments. We would also like to thank the Kari Palmer and Gregory Pratt of the Minnesota Pollution Control agency for providing ozone data and relevant information. Additionally, we are grateful to Wei-Hsin Fu for help with ArcGIS.

Abstract

We model and attempt to forcast low level atmospheric ozone concentration across Minnesota in a Bayesian setting. Our data include 2007 ozone measurements from the Minnesota Pollution Control Agency as well as meteorological data. Our model, based on McMillan et al., is hierarchical and incorporates nearest neighbor spatial and time series parameters. We compute the posterior distributions using Gibbs sampling and slice sampling, implemented with WinBUGS. We provide an overview of Bayesian hierarchical modeling and our data, and discuss the rationale for the model. Our results indicate that ozone concentration is primarily dependent on the previous day's ozone concentration in the same location and the concentration to the west, with temperature and pressure also significant factors. Our predictions are, at least for two days into the future, fairly accurate but miss isolated spikes in concentration, limiting the utility of the model for forecasting purposes.

1. Introduction

Tropospheric ozone O_3 is a pollutant of concern to environmental regulation agencies such as the Minnesota Pollution Control Agency which monitors ozone levels from the early spring until the autumn. Because it has toxic effects on humans [1] as well as vegetation through intereference with photosynthesis, the Minnesota Pollution Control Agency is required to implement air quality standards for stratospheric ozone O_3. These regulations stem from the National Air Quality Standards enforced by the U.S. EPA as mandated by the Clean Air Act. [2]

However, estimating and forecasting daily stratospheric ozone is difficult task due to its spatial and temporal variability as well as the influence of meteorlogical variables in its formation. In this paper, we attempt to capture all these effects through a Bayesian hierarchical model. In particular, we use a nearest neighbor spatial dependence framework to model the mouvement of stratospheric ozone as it is transported by the air mass in various directions. The nearest neighbor model also allows for the inclusion of a one day lagged ozone formation process accounting for the temporal variability of ozone. The persistence of global high pressure systems also increases the levels of stratospheric ozone. We account for these high pressure systems with the inclusion of a regime switching parameter that captures low and high pressure systems.

We implement the Bayesian hierarchical model with two statistical softwares, namely: Rand WinBUGS (Bayesian inference Using Gibbs Sampling, Spegelhgalter et al. 2003). WinBUGS facilitates the implementation of Markov Chain Monte Carlo (MCMC) sampling techniques for model parameters. WinBUGS requires the specification of a model, inital values and the data for a number of Markov chains.WinBUGS can be downloaded free of charge at http://www.mrc-bsu.cam.ac.uk/bugs/. R is an open source free software package that we use to manage the ozone and meteorlogical data and build matrices to be used by WinBUGS through the R2WinBUGS package (Sturtz, Uwe and Gelman 2005). In tackling the estimation and forecasting of daily ozone levels, we deal with large matrices not easily handled by WinBUGS. The R2WinBUGS package provides convenient functions to call WinBUGS from R after the data has been stored in matrices using R.

[1] There is evidence that exposure high concentrations of stratospheric ozone O_3 is associated with higher hospital admissions of people with poor cardiovascular conditions

[2] The Clean Air Act requires that the U.S. EPA revises the NAAQS every ten years to designate counties as attainment or non-attainment for pollutants such as O_3 and particulate matter

Section 1 provides a decription of the data sets as well as the use of ArcGIS for the initial steps of a spatial model. Section 2 is a general overview of Bayesian hierarchical models in the environmental sciences with a particular emphasis on ozone. Section 3 provides a general overview of hierarchical models, and section 4 discusses the spatial temporal ozone model. In section 5, we discuss the MCMC techniques that WinBUGS uses to obtain posterior distributions for model parameters. In sections 7, 8 and 9 we interpret the spatial temporal, meteorological and regime switching parameters, provide visuals for the predictive capabilties of our model and insights to improve the performance of the nearest neighbor model.

2. Data

The Minnesota Pollution Control Agency (MPCA) recorded ozone levels at fifteen locations in Minnesota from April 1 through September 30, 2007. These stations provide hourly ozone levels in parts per million. We use the maximum running eight hour average for each day. In accordance with the MPCA's primary mission of identifying dangerously high ozone levels, the stations were concentrated in the Minneapolis/St. Paul metropolitan area and areas likely to be downwind from the metropolitan area, although a few were in rural areas to provide background ozone levels. The MPCA collects data in compliance with EPA regulations and routinely checks to ensure that their stations are correctly calibrated.

The meteorological data was collected by the National Climatic Data Center. We select 18 weather stations scattered across Minnesota. Each station has hourly data for temperature in degrees Celsius, pressure in millibars, dewpoint, wind direction, and wind speed in miles per hour. For our modeling purposes, we use the daily maximum temperature, the average daily pressure, and the average daily wind speed. We treat wind direction as a categorical variable, indicating which of eight directions was most prominent for each day. The eight categories for wind direction are north-northeast, east-northeast, east-southeast, south-southeast, south-southwest, west-southwest, west-northwest, and north-northwest. We calculate the maximum temperature by finding the maximum of all eight-hour temperature averages. Based on the dewpoint and temperature, we calculate average relative humidity.

Occasionally, meteorological data are missing from the data set. The problem of missing data could arise for a variety of reasons, such as problems collecting the data due to faulty meteorological equipment or problems recording the data due to difficulties with the

computer or network. Because the software we use to create the model requires all meteorological data, we need to fill in the gaps. Most of the time that a station lacks data, only a few hourly readings were missing. In this instance, we calculate the daily averages and daily maxima by using the data that were collected and available. In rare cases, a station is missing data for an entire day. In this instance, we use the data from the last day data were recorded. For all eighteen stations, data are missing for April 25 and August 28. We solve this problem by using the previous day's data. In the extremely rare occasion that a station lacks several consecutive days of meteorlogical data, we fill in the holes using the last known data for that station.

In addition to the problem of missing meteorological data, ozone data are occasionally missing. However, WinBUGS does not require ozone data to be complete. Therefore, we do not fill in any missing data points in the ozone data set.

In order to create a spatial model, we superimpose a 3×4 rectangle onto the state of Minnesota. This rectangle is subdivided into square grids of side 150 km. Because Minnesota is not perfectly rectangular, some sqaures include small sections of surrounding states while small sections of Minnesota extend beyond the bounds of the grid. The inclusion of surrounding states does not interfere with our ability to model ozone levels for the portions of Minnesota contained in the boxes. The small sections of Minnesota that extend beyond the bounds of the grid are excluded for the sake of simplicity. If we instead created a rectangular grid that included all of Minnesota, many gridboxes would contain large areas of surrounding states and sections of Canada. Since our goal is to model ozone for Minnesota, this overinclusive grid would be excessive. If we utilized a non-rectangular grid designed to contain the entire state of Minnesota, our model would become complicated because some boxes would have only one neighboring box. For these reasons, we opt to use a rectangular grid that covers most of Minnesota.

To assign meteorological and ozone data to each box, we use ArcGIS to calculate the centroid of each box in the grid and assign it meteorological information from the nearest station. We also use ArcGIS to create a mapping matrix identifying the grid cell location of each ozone station. Figure 1 and Figure 2 show the state of Minnesota with the grid as well as the weather stations and ozone stations.

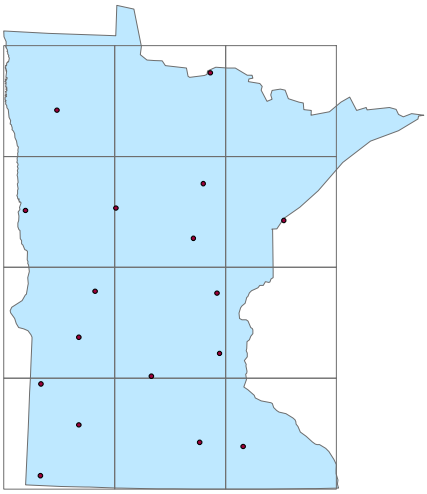

FIGURE 1. Meteorology stations plotted in Minnesota with a 150 km ×150 km grid overlaid.

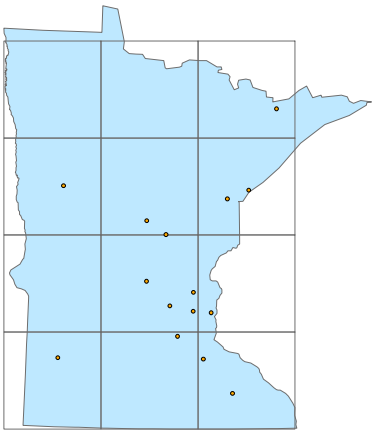

FIGURE 2. Ozone stations plotted in Minnesota with a 150 km ×150 km grid overlaid.

3. Overview of Bayesian Hierarchical models

Environmental processes such as the formation and dissipation of ozone are often complex. In particular, relationships between different variables may not be obvious, making it difficult to fit a model using classical techniques (Carlin, Clark and Gelfand 2006). In these situations, stochastic models are increasingly used to model the relationships between environmental variables. In particular, stochastic hierarchical models allow the inclusion of different stages for the dynamics environmental process as well as lagged variables and parameters. Each of these stages include unkown parameters and their corresponding uncertainties which help describe the overall uncertainty of the environmental process. Bayesian hierarchical models provide the adequate structure for the incoporation of these uncertainties.

Bayesian hierarchical models offer several advantages over classical frequentist ones . The structure of Bayesian models facilitates the inclusion of previously known information or expert opinion through the specification of the prior distributions. In addition, the interpretation of model estimates is simpler and therefore clearer [3]. However, Bayesian models are not flawless because the inclusion of uncertainty on stage processes and parameters through prior specifications can be difficult and often tricky. Consequentially, a poorly-chosen prior can produce results that are misleading or inconclusive. Furthermore, complex Bayesian models lead to integrals of high dimensions, requiring computationally intensive methods to obtain posterior estimates for model parameters. However, the recent advances in computing and the increase in computing power of most software packages has alleviated the burden of these intense computations.

3.1. **Bayes Theorem.** Suppose that $p(\mathbf{x}|\boldsymbol{\lambda})$ is the distribution of the observed data $\mathbf{x} = (x_1, x_2, ..., x_n)$ conditional on the vector of unknown parameters $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_n)$. Also asssume that $\boldsymbol{\lambda}$ is a random variable with prior distribution $p(\boldsymbol{\lambda})$. So by Baye's theorem:

$$p(\boldsymbol{\lambda}|\mathbf{x}) = \frac{p(\mathbf{x}, \boldsymbol{\lambda})}{p(\mathbf{x})} = \frac{p(\mathbf{x}|\boldsymbol{\lambda})p(\boldsymbol{\lambda})}{\int p(\mathbf{x}|\boldsymbol{\lambda})p(\boldsymbol{\lambda})d\boldsymbol{\lambda}} \tag{1}$$

[3]Bayesian credible intervals avoid the awkward interpretation of the confidence intervals of frequentist statistics. A 95% credible interval means that the probability of the parameter of interest falling within that interval is 0.95

As equation 1 suggests, Bayesian statistics treats both the data and the parameters as random variables, whereas frequentists statistics treats only the data as random variables. The purpose of frequentist statistics is to estimate the value of a fixed but unknown parameter by finding the parameter value that maximizes the likelihood of the given data. On the other hand, the goal of Bayesian statistics is to determine the distribution for the parameter given the data collected. In other words, the inference goal is to determine the posterior distribution from the likelihood and prior distributions. Furthermore, we can simplify the result above and rewrite as equation 2 because the denominator (the marginal distribution of $\mathbf{x}$) is a constant with respect to $\boldsymbol{\lambda}$. Therefore the posterior distribution of the parameter vector λ is proportional to the distribution of the data multiplied by the prior distribution $p(\boldsymbol{\lambda})$. The posterior is expressed as the product of the prior and the likelihood in equation 2:

$$p(\boldsymbol{\lambda}|\mathbf{x}) \propto p(\mathbf{x}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}) \tag{2}$$

3.2. **Hierarchical models: an illustration with ozone.** Let $\mathbf{Z_t}$ be the $S \times 1$ matrix of ozone measurements from all S stations on day t, conditional on $\mathbf{M_t}$, the corresponding vector storing meteorological information, the corresponding ozone process at the grid level $\mathbf{O_t}$, and $\boldsymbol{\xi_z}$: a vector storing error parameters. We can refer to this as the first stage or data process.

$$p(\boldsymbol{Z_t}|\mathbf{O_t}, \boldsymbol{\xi_z}) \tag{3}$$

Second in the hierarchical model is the description of the ozone process. Meteorology is scientifically known to impact ozone, and the hierarchical approach allows the incorporation of meteorological data into the model. Our inference goal is to estimate and predict daily ozone levels conditional on meteorology. We specify a distribution for the ozone process for a given day conditional on the observed meteorological data, previous ozone values, and other parameters which are stored in a vector $\boldsymbol{\xi_o}$. Then, the conditional distribution of the ozone process is given by:

$$p(\mathbf{O_t}|\mathbf{M_t}, \boldsymbol{\xi_o}) \tag{4}$$

The structure of the hierarchical model is completed with the specification of a prior distribution for the parameters stored in the vectors $\boldsymbol{\xi_z}$ and $\boldsymbol{\xi_o}$. That is, we assign prior distributions to each of them. We can include prior information or expert opinion in our model when we assign distributions to these parameters. In situations where no prior or expert opinion is available, we can assign vaguely informative priors. The use of strong, unjustified priors could potentially skew the posterior toward incorrect results; using vague priors gives the data more weight in the posterior, preventing misleading conclusions. With the distributions defined in the three stages, we can write posterior distributions model parameters in the form of equation 2. A sketch of the derivations of the posterior distributions for each of the model parameters is provided in the Appendix.

4. A Spatial Temporal Model for Ozone

4.1. **First Stage: Data Process.** Wikle et al. (1998) describe a general framework for the treatment of space time data in the environmental sciences. In particular, they formulate a Bayesian model capturing the spatial and temporal variability often displayed by environmental data. McMillan et al. (2005) apply this modeling strategy to estimate and forecast ozone through space and time for a domain over Lake Michigan during spring-summer 1999. Following this modeling strategy, we build a three stage process to capture the complex nature of daily ozone behavior. Ozone displays both spatial and temporal behavior which are often difficult to capture using classical statistical approaches. Furthermore, the Bayesian hierarchical model allows the inclusion of ozone measurement error in the first stage, the known interactions between meteorological information and ozone as well as the spatial-temporal structure of ozone. In the first stage, we specify a statistical measurement error model. Let $Z(s_j, t)$ denote the ozone measurements at station s_j which is possibly located at gridbox location i, where $j = 1, ..., S$ and $i = 1, ..., N$. Then we specify the model:

$$\mathbf{Z_t} = K\mathbf{O_t} + \boldsymbol{\epsilon_{z,t}} \tag{5}$$

We assume that the the spatial structure of each gridbox g_i and ozone measurements are independently and identically distributed (*i.i.d.*) Gaussian random variables with mean

$O(i,t)$ and variance $\sigma_{o,t}$. Therefore, $\mathbf{O_t}$ is an $N \times 1$ vector and K is an $S \times N$ matrix mapping the ozone values at grid locations to the observations with a one in the (i,j) position indicating that the ozone observation j is located in gridbox i. Finally, $\boldsymbol{\epsilon_{z,t}}$ is an $S \times 1$ vector for the sub-grid scale processes and measurement error. We assume $\boldsymbol{\epsilon_{z,t}} \sim N(0, \boldsymbol{\sigma^2_{z,t}} \boldsymbol{I})$.

4.2. **Second Stage: Process Model.** In the second stage, we consider a model for the state process. Ozone fields have spatio-temporal dependence structures. To account for these spatial dependence structures, we follow the model described in McMillan et al. (2005) for a nearest neighbor dependence. In this setting, the ozone measurement at gridbox i is a function of the previous day's ozone measurement for the same site and the four closest ones. For the sites at the edges, we take a similar strategy as in McMillan et al. by extending the map through the creation of new boundary sites (see McMillan et al 2005, Wikle et al 2002). The second stage model is then:

$$\mathbf{O_t} = \boldsymbol{\mu_t}\mathbf{1} + H(\boldsymbol{\theta})\mathbf{O_{t-1}} + G(\boldsymbol{\theta})\mathbf{B_{t-1}} + \mathbf{M_t}\boldsymbol{\beta} + \boldsymbol{\epsilon_{o,t}} \tag{6}$$

Here $H(\boldsymbol{\theta})$ is an $N \times N$ matrix describing the nearest neighbor dependence structure for gridboxes whose boundaries are within the interior of Minnesota. We depart from the parametrization of the conditional mean ozone value for a certain gridbox as described in McMillan et al. They allow the mean ozone value of gridbox i for a given day to depend on the ozone reading of the same gridbox from the previous day and information based on the ozone readings of the neighboring gridboxes for the same day. Our approach is more plausible in describing the spatial dynamics of ozone. The observation at a particular station is likely to be affected by the previous day's ozone level of neighboring locations since ozone is transported by moving air masses. Therefore, borrowing strength across the nearest neighbors should provide better estimates for daily ozone levels. Hence, we consider a nearest neighbor array, with each entry in the array being an autoregressive parameter matrix in which the conditional of ozone for cell i on day t includes the previous day's information from the four closest cells. If we define $c_i = \{c_{i,1}, c_{i,2}, c_{i,3}, c_{i,4}\}$ as the four closest neighbors to cell i with $c_{i,1}$ as its neighbor to the north, $c_{i,2}$ its neighbor to the east, $c_{i,3}$ the neighbor to the south, and $c_{i,4}$ being on the west of cell i. Thus:

$$\begin{aligned} O(i,t) &= \theta_5 O(i,t-1) + \theta_1 O(c_{i,1},t-1) + \theta_2 O(c_{i,2},t-1) \\ &\quad + \theta_3 O(c_{i,3},t-1) + \theta_4 O(c_{i,4},t-1) \end{aligned}$$

A similar approach is adopted for the boundary sites. The boundary process is paramterized to allow nearest neighbor dependence for the cells at the edges. $G(\theta)B(i,t-1)$ will take a vector autoregressive form similar to the one described for the non-boundary cells (See Wikle et al. 2003). The extension of the map to include new neighbors for the gridboxes at the edge requires the specification of an ozone process for these new gridboxes. We believe that the mean ozone in these new gridboxes should not play a great role in the mean ozone of the gridboxes that are actually within Minnesota [4]. We understrengthen the mean ozone process in these new gridboxes by borrowing information from adjacent cells but with new parameters. The mean and standard error of these new parameters reflect this stance. Hence we write the mean ozone process for the newly added gridboxes as follow:

$$\begin{aligned} O(i,t) &= \psi_5 O(i,t-1) + \psi_1 O(c_{i,1},t-1) + \psi_2 O(c_{i,2},t-1) \\ &\quad + \psi_3 O(c_{i,3},t-1) + \psi_4 O(c_{i,4},t-1) \end{aligned}$$

In this second stage, $\boldsymbol{\mu_t}$ captures the arrival and departure of high pressure systems. The inclusion of this term stems from the possibility that high pressure systems are associated with extreme ozone measurements. While the active regime is high pressure, $\boldsymbol{\mu_t}$ and predicted ozone levels will be higher than while the active regime is normal pressure. Specifically, $\boldsymbol{\mu_t}$ is parametrized to increase the probability of higher ozone after the presence of a high pressure systems for several days. Therefore, $\boldsymbol{\mu_t}$ is modelled as an autoregressive times series of lag one whose mean and autoregression parameters are conditional to the pressure regime. If we assume that $\boldsymbol{\mu_t - \nu}$ is the deviation from the long run spatially averaged ozone level, then the mean of the t^{th} deviation as a function of all previous daily deviations is given by :

[4]Note that this would be less of an issue if data for border areas with neighboring states was considered

$$\boldsymbol{\varphi(t)} = \begin{cases} \boldsymbol{\alpha_1(\mu_{t-1} - \nu_1)} & \text{If } I_{t-1} = 0 \\ \boldsymbol{\alpha_2(\mu_{t-1} - \nu_2)} & \text{If } I_{t-1} = 1 \end{cases} \tag{7}$$

In equation 7, $\boldsymbol{\mu_t} \sim N(\boldsymbol{\varphi(t)}, \boldsymbol{\sigma^2_{\mu_t}})$. I_t is an indicator for the active regime for day t, with I_t=1 indicating a high pressure regime and I_t=0 indicating a normal pressure regime. Lu and Berliner (1999) describe a Markov switching time series model in which the runoff series are single-lag autoregressions. The regime indicators are modelled as a lag-one Markov chain in which the conditional probability for a given day's regime depend on the previous day's regime. Likewise, we assume that the pressure series has two states: a normal pressure regime and a high pressure regime. The probability transition matrix for this regime switching process is described in Table 4.2 where 0 denotes the low pressure regime while 1 indicates the high pressure regime. In table 4.2, for $i = 0, 1$,$p_{i,i}$ is the probability of the pressure regime staying constant between two consecutive periods. Otherwise, it is the probability of switching regimes.

Further, the transition probabilities depend on a recursively filtered and areally averaged pressure series. The pressure series are filtered by subtracting the lag 1 serial correlation coefficient of the pressure series from a given day's pressure to obtain the filtered value for that same day. The transition probabilties may be modeled using an arbitrarily chosen distribution function. We model them using the following probit function, although the logit function would produce similar results. The probit is defined as:

$$p_{ij}(t) = \phi(\delta_1 P_t + \delta_2 P_t I_{t-1} + \delta_3 + \delta_4 I_{t-1})$$

where $\phi(x)$ denotes the standard normal cumulative distribution function. P_t represents the recursively filtered and areally averaged pressure values.

Finally, we complete the specification of the second stage process with the daily meteorological data. This information is stored in the $m \times n$ matrix M_t which contains meteorological station data on daily maximum temperatures, atmospheric pressure, average humidity, and wind speed and direction. Here n is the number of variables in each of the meteorological information matrices. Therefore, β is an $n \times 1$ vector of parameters associated with

TABLE 1. Probability Transition Matrix

	I_t	
I_{t-1}	0	1
0	p_{00}	p_{01}
1	p_{10}	p_{11}

each of the meteorological variables. We assume that all error in this process are normally distributed with mean zero.

4.3. **Third Stage: Prior Specification.** We complete the Bayesian hierarchical model by selecting prior distributions for each of the parameters described in the second stage. We assume that the spatial parameters θ_1,θ_2,θ_4 and θ_5 for the non-boundary gridboxes are all independent Gaussian random variables. Likewise, the spatial parameters for the additional gridboxes at the edges, ψ_1, ψ_2,ψ_3,ψ_4 and ψ_5 are also Gaussian. Therefore for $i = 1, ..., 5$, we assign initial values for the mean and variance such that:

$$\theta[i] \sim N(\widetilde{\mu}_\theta[i], \sigma^2_\theta[i]) \tag{8}$$

and

$$\psi[i] \sim N(\widetilde{\mu}_\psi[i], \sigma^2_\psi[i]) \tag{9}$$

In addition, we also assume that the autoregression parameters in the daily mean shift are independent Gaussian random variables. Hence for $i = 1, 2$:

$$\alpha[i] \sim N(\widetilde{\mu}_\alpha[i], \sigma^2_\alpha[i]) \tag{10}$$

$$\nu[i] \sim N(\widetilde{\mu}_\nu[i], \sigma^2_\nu[i]) \tag{11}$$

The parameters specifying the regime states in the probit probability model are also independent Gaussian random variables. Therefore for $i = 1, ..., 4$:

$$\delta[i] \sim N(\widetilde{\mu}_\delta[i], \sigma^2_\delta[i]) \tag{12}$$

The variances of the error terms in the first and second stages follow Inverse Gamma distributions. Thus:

$$\sigma_{z,t} \sim IG(\zeta, \eta) \tag{13}$$

$$\sigma_{o,t} \sim IG(\rho, \varrho) \tag{14}$$

We assume flat priors for all the meteorlogical parameters stored in the $n \times 1$ vector $\boldsymbol{\beta}$. We also assume that the variances of the error terms in the mean ozone process for both boundary and non-boundary gridboxes follow normal distributions.

5. Sampling Techniques: Gibbs Sampling and Slice Sampling

With the complexity of the ozone nearest neighbor model, the computation of full conditional distributions for model parameters involves the evaluation of high-dimensional integrals. Therefore calculating the exact posterior distribution can become impossibly time consuming. We use Markov Chain Monte Carlo simulation methods to evaluate the full conditional distributions for our model parameters. These methods sequentially sample parameter values from a Markov Chain whose stationary distribution is the full conditional distribution of interest. The full conditionals for most of our model parameters follow multivariate normal distributions, thus allowing easy computation with the Gibbs sampler (Geman and Geman 1984). Here we show the full conditional posterior distribition for the spatial parameters $\theta[i]$'s(See the appendix for the details of the derivation).

$$\begin{aligned}
p(\theta_i|\cdot) \quad \propto \quad & exp\Big\{-\frac{1}{2\sigma^2_{o,t}}\sum_{t=1}^{T}(O_t - \mu_t - \theta D_i O_{t-1} - H_i O_{t-1} - \theta E_i B_{t-1} - G_i B_{t-1} - M_t\beta)^T\Big\} \\
& \times exp\Big\{-\frac{1}{\sigma^2_\theta[i]}(\theta_i - \widetilde{\mu}_{\theta i})\Big\} \\
& \propto exp\Big\{-\frac{1}{2}\theta(\frac{1}{\sigma_{o,t}}\sum_{t=1}^{T} D_i O_{t-1} D_i O_{t-1} + (E_i B_(t-1))^T D_i O_{t-1} + (E_i B_{t-1})^T E_i B_{t-1} + \frac{1}{\sigma^2_\theta[i]})\theta \\
& -2\theta[(O_t - \mu_t - H_i O_{t-1} - G_i O_{t-1} - M_t\beta)^T(D_i O_{t-1} + E_i B_{t-1}) + \frac{\widetilde{\mu}_i}{\sigma^2_i}]\Big\}
\end{aligned}$$

As shown above, the full conditional posterior distribution of each of the spatial parameters is proportional to a multivariate normal distribution that we can write as follows.

$$
\begin{aligned}
p(\theta_i|\cdot) \quad \sim \quad & N\Bigg((\frac{1}{\sigma_{o,t}} \sum_{t=1}^{T} D_i O_{t-1} D_i O_{t-1} + B + (t-1)^T D_i O_{t-1} + E_i B_{t-1} E_i B_{t-1} + \frac{1}{\sigma_i^2})^{-1} \\
& [(O_t - \mu_t - H_i O_{t-1} - G_i O_{t-1} - M_t \beta)^T (D_i O_{t-1} + E_i B_{t-1}) + \frac{\widetilde{\mu}_i}{\sigma_i^2}]^T, \\
& (\frac{1}{\sigma_{o,t}} \sum_{t=1}^{T} D_i O_{t-1} D_i O_{t-1} + (E_i B_{t-1})^T D_i O_{t-1} + (E_i B_{t-1})^T E_i B_{t-1} + \frac{1}{\sigma_i})^{-1} \Bigg)
\end{aligned}
$$

The full conditional posterior distribution follows a known distribution, namely the multivariate one. In this case, WinBUGS uses Gibbs Sampling to sample parameter values. To avoid notational confusion, we will describe the Gibbs Sampling procedure with j parameters $\boldsymbol{\lambda} = \lambda_1, \lambda_2, ..., \lambda_j$ that are incorporated into a model with data x. The Gibbs sampler requires the full conditional distribution $p(\lambda_i|\lambda_{k \neq i}, x)$ for $i = 1, 2, ..., j$. Then for each iteration t of a chain of length T, the Gibbs Sampler follows these steps:

First Step: Draw λ_1^t from $p(\lambda_1|\lambda_2^{t-1}, \lambda_3^{t-1}, ..., \lambda_j^{t-1}, x)$
Second Step: Draw λ_2^t from $p(\lambda_2|\lambda_1^t, \lambda_3^{t-1}, ..., \lambda_j^{t-1}, x)$
$j - 1^{th}$ **Step:** Draw λ_{j-1}^t from $p(\lambda_{j-1}|\lambda_1^t, \lambda_2^t, ..., \lambda_{j-2}^t, \lambda_j^{t-1}, x)$
j^{th} **Step:** Finally draw λ_{j-1}^t from $p(\lambda_j|\lambda_1^t, \lambda_2^t, ..., \lambda_{j-1}^t, x)$

The samples from each iteration of the chain form a sample form the joint posterior distribution $p(\boldsymbol{\lambda}|x)$. When the chain is long enough, it approximates the posterior distribution. The first few samples are drawn during a burn-in period, a section of time before the samples have converged to the posterior $p(\boldsymbol{\lambda}|x)$. Because these samples were drawn before convergence, they are often not used to approximate the posterior distribution. Although the idea of excluding samples drawn during the burn-in period seems simple, determining the length of the burn-in period is tricky and often arbitrary since it is difficult to know when a chain has converged.

However for some model parameters, there are no close form full conditional posterior distributions. Whenever this situtation arises, WinBUGS computes the posterior distribution of model parameters using Slice Sampling, though other sampling methods would give similar outcome (Metropolis-Hastings algorithm, Metropolis et al. 1953; Hastings, 1970). Such is the case for the parameters specified in the probit model for transition probabilities, namely the δ_i's.

Again, assume that our parameter of interest is λ, the prior distribution is $\pi(\lambda)$, and the likelihood distribution is $f(x|\lambda)$. Bayes' Theorem indicates that the posterior distribution $p(\lambda|x) \propto \pi(\lambda) \times f(x|\lambda)$. Both $\pi(\lambda)$ and $f(x|\lambda)$ are known ans so is the unnormalized posterior $\pi(\lambda) \times f(x|\lambda)$. We use this product to approximate the conditional posterior distribution of λ.

Slice sampling involves taking "slices" of $\pi(\lambda) \times f(x|\lambda)$. As illustrated in Figure 3, we first introduce an auxiliary variable U with $U|\lambda \sim Unif(0, \pi(\lambda) \times f(x|\lambda))$. Then we draw one sample from this distribution and use the sample to create a vertical "slice" of the distribution $\pi(\lambda) \times f(x|\lambda)$. Next, we determine the set $S = \{\lambda|\pi(\lambda) \times f(x|\lambda) \geq U\}$ and draw a sample from this set. Continue drawing samples in this manner until you have a chain $\lambda_1, \lambda_2, \ldots, \lambda_t$. The collection of samples then give an approximation of $\pi(\lambda) \times f(x|\lambda)$, which means that the samples also given an approximation to the posterior distribution $p(\lambda|x)$.

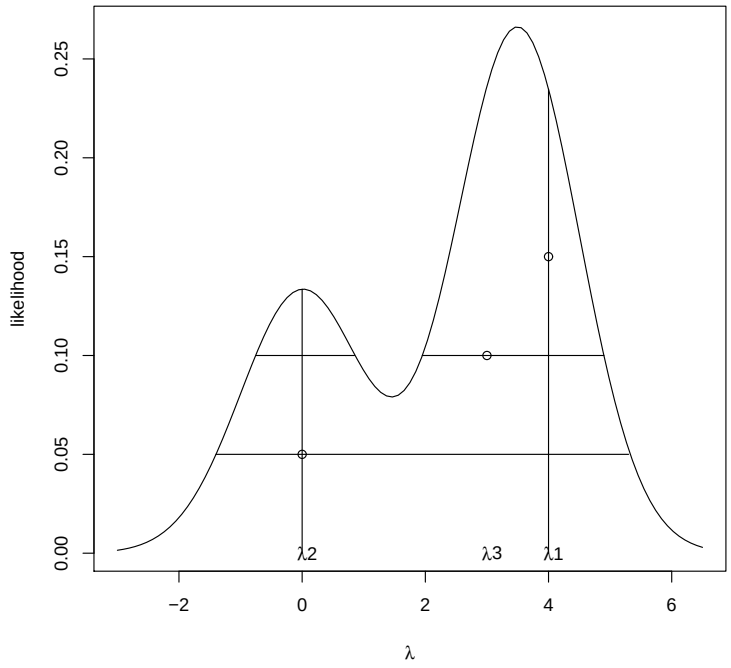

FIGURE 3. Illustration of Slice Sampling

We monitor the convergence of the chains using the Brooks-Gelman-Rubin(BGR) factor (Gelman and Rubin 1992, Brooks and Gelman 1998).We run 3 parallel chains with overdispered starting values with respect to the true posterior values. The BGR factor checks

TABLE 2. Prior Distributions

Parameter	Prior
θ_1	N(0.4, 0.2)
θ_2	N(0.4, 0.2)
θ_3	N(0.4, 0.2)
θ_4	N(0.4, 0.2)
θ_5	N(0.5, 0.2)
ψ_1	N(0.4, 0.2)
ψ_2	N(0.4, 0.2)
ψ_3	N(0.4, 0.2)
ψ_4	N(0.4, 0.2)
ψ_5	N(0.5, 0.2)
α_1	N(0, 100000)
α_2	N(0.013, 10000)
ν_1	N(0, 100000)
ν_1	N(0.02, 100000)
δ_1	N(0, 1000000)
δ_2	N(0,1000000)
δ_3	N(-1.4, 10)
δ_4	N(0.4, 10)
$\sigma_{z,t}$	Gamma(0.1, 0.1)
$\sigma_{o,t}$	Gamma(0.1, 0.1)
β_1	Flat
β_2	Flat
β_3	Flat
β_4	Flat
β_5	Flat
β_6	Flat
β_7	Flat
β_8	Flat
β_9	Flat
β_{10}	Flat
β_{11}	Flat

TABLE 3. Brooks-Gelman-Rubin Diagnostic

Parameter	BGR
θ_1 (North)	.9987
θ_2 (East)	.9992
θ_3 (South)	.9992
θ_4 (West)	.9998
θ_5 (Yesterday)	.9999
β_1 (Temperature)	1.002
β_2 (Pressure)	1.00008
β_3 (Relative Humidity)	1.0007
ν_1 (Normal Regime)	.9992
ν_2 (High Pressure Regime)	1.0008

whether the within chain variation is assymptotically equal to the variation accross chains. A BGR value of 1 supports this equality and indicates good convergence of the chains. Table 3 shows that the Brooks-Gelman-Rubin statistics are close to one for each parameter, indicating convergence. Visual inspection of chains yields a similar conclusion. Unfortunately, the posterior exhibits high intra-chain variability. Therefore, chains with different initial conditions converge to similar posteriors, but with enough variance that some barely significant parameters become not significant.

TABLE 4. Spatial and Temporal Parameters

Parameter	Mean	95% CI
θ_1 (North)	.0675	(.0189, .1392)
θ_2 (East)	-.02676	(-.0824,.0132)
θ_3 (South)	-.0291	(-.0689,-.0036)
θ_4 (West)	.1967	(.1255,.2451)
θ_5 (Yesterday)	.3421	(.2105,.4376)

6. RESULTS

6.1. **Spatial and Temporal Parameters.** As evidenced by Table 4, a considerable proportion of ozone (about 34%) remains from day to day. This parameter was significant, and had the largest effect among all spatial temporal parameters. Our model also indicates that about 20% of ozone each day is transported from the cell to the West. This West-to-East movement is the most significant of our spatial parameters. This was not unexpected, as the MPCA had previously known that weather systems tend to move West to East.

Paradoxically, the point estimate for transport from both the East and the South was negative. This is a counterintuitive result, since it essentially states that higher ozone in the cell to the south is correlated with lower ozone in the current cell the next day. However, higher ozone values in one cell are correlated with higher ozone values in nearby cells, so the model is in effect showing that higher ozone values in the cells to the East and South reduce the effect of the higher ozone from the West. Also, it is important to note that these two parameters are either not significant or barely significant at the 5% level; indicating that these negative correlations are not extremely strong. These results were still surprising, as it is known that much of Minnesota ozone is transported north from Iowa. We note that our model did not include any data about Iowa ozone, nor did we have any method to include information about Iowa ozone in our priors. Therefore, our model missed the effect of ozone transported North into Minnesota.

The value of these spatial parameters is dependent on the prior distributions and initial conditions of the MCMC chain. On any given day, there is high correlation between the average ozone concentration in neigboring cells. Unfortunately, this makes it difficult to separate out the effects of the various spatial parameters. Essentially, it is difficult to distinguish between ozone that remains in a cell and ozone that was transported to the cell from the west because of this high correlation. As a consequence, the prior distributions are important in the final form of the posterior. We run the model with different priors

TABLE 5. Meteorological Parameters

Parameter	Mean	95% CI
*Temperature (^{o}C)	3.89x10^{-4}	(2.96x10^{-4}, 5.11x10^{-4})
*Pressure (mb)	1.05x10^{-5}	(6.65x10^{-6},1.43x10^{-5})
Rel. Humidity	7.16x10^{-6}	(-3.14x10^{-5},4.94x10^{-5})
*Wind from ENE	-6.89x10^{-4}	(.-9.9x10^{-4},-3.9x10^{-4})
*Wind from ESE	-5.50x10^{-4}	(-8.64x10^{-4},-1.73x10^{-4})

emphasizing either θ_5 or the other four parameters, and found that the final parameter values depend on the prior chosen. That is, if we emphasize the impact of neighboring cells, then all four of those parameters become more significant, with larger values. We get the opposite result if we de-emphasize these parameters. This barely determined parameter space is an area of concern for the model.

6.2. **Meteorological Parameters.** As shown in Table 5, temperature and pressure have high impact on average grid cell ozone values. An increase in 10 degrees Celsius causes a .004 ppm increase in average ozone level. At this scale, given otherwise identical conditions, a hot summer day (30 degrees C) would have a .012 ppm higher ozone level than a frigid April day (0 degrees C), so temperature does have a large impact. Similarly, an increase of 100 mb in pressure causes a .001 ppm increase in ozone. Again given otherwise identical conditions, a high pressure day should have about .004 ppm higher ozone than a low pressure day. Relative Humidity had no discernable effect on average ozone process. Either relative humidity does not have an effect on ozone or the effect is too subtle for the model to pick up.

The parameters measuring wind speed and direction were only significant when the wind comes from the East-Northeast or East-Southeast. Even then, since the average wind speed is only a few miles per hour, wind does not cause a large change in ozone values. According to the MPCA, wind from the East is correlated with moving air systems and lower ozone values, so the negative sign makes sense. Weather parameters are robust against changes in the prior.

6.3. **Regime Switching Parameters.** The model was highly sensitive to initial conditions. In particular, small changes in initial values of our delta greatly affected the model. Essentially, the probability of the high state became zero after a few MCMC iterations, thereby preventing the regime switching parameter from having an effect at all. Additionally, increasing the value of either or both of the regime parameters destabilized the model, preventing the model from converging in a reasonable number of iterations.

6.4. **Ozone through Space and Time.** Figures 4 through 9 show estimated average grid cell ozone for each day of a given month. Each rectangle represents a different day, with the twelve colored squares each representing the level of ozone in one grid cell. Yellow represents lower levels of ozone, while red represents higher levels. Note that the scale changes from month to month, and that similar colors may represent different ozone levels in different months. The scale shows the ozone levels measured in parts per million.

These plots illustrate the spatial and temporal nature of ozone concentration. Our model shows high levels of ozone in a cell are generally accompanied by higher levels in surrounding cells. Moreover, these concentrations persist across time. The persistence of ozone across time indicates that θ_5 is significant. When ozone does dissipate, it tends to move east, confirming the higher value for θ_4. In general, the southern portion of the state has higher ozone levels on average. Finally, note the extended period of high ozone from June 10 through June 15. Ozone levels during this period were dangerously high each day. It is noteworthy that the model shows this high-ozone event.

7. Predictive Capabilities

Since the MPCA issues warnings when ozone readings are dangerously high, being able to predict high ozone levels would be very beneficial. We run the model given only the April data and attempt to predict ozone levels for the first week of May, given the weather data for that timespan. We experiment only with one prediction interval since we would have to rerun the model for each timespan we wish to predict. In that timespan, our prediction yielded mixed results. The credible intervals for ozone levels on a given day tend to be quite wide, ranging from .03 to .09 parts per million. However, the point predictions are much more informative. The mean square error is roughly .006 parts per million for the first week of a prediction .003 parts per million for the first day. Additionally, the correlation for the first day between the predicted and actual ozone level is .42, while the correlation for the first week between the predicted and actual ozone levels is .19. The mean square error increases and the correlation decreases as we increase the number of days we try to predict. In other words, the model is able to predict the first day more accurately than the subsequent days, since any error in the first day's prediction is compounded by the following days' predictions. Unfortunately, while the model is relatively close on average, it does not predict spikes isolated to one or two grid cells. Thus, the model is not effective for predicting dangerous ozone levels. Of course, a model is able to correctly predict only if the model has converged. If the model has not converged, then the predictions are extremely

biased.

We also analyze model residuals, comparing our data with the corresponding average grid cell values for each station and day. Most stations have acceptable residuals randomly distributed and centered around zero. However, a few stations consistently under-predict or over-predict ozone. In particular, our model consistently predicts higher ozone concentrations than observed for Mille Lac than were observed, and lower concentrations than observed for Brainerd. This occurs because Mille Lac and Brainerd are located in the same grid cell, and therefore have the same predicted concentrations. However, Mille Lac is rural and Brainerd is urban (in fact, the Brainerd monitor is located in an airport). This inability to differentiate between different areas within a grid cell is a weakness of the model. Because we have time series data, we also look at autocorrelation between residuals. There is not any significant autocorrelation, indicating that we don't have problems compensating for errors in previous days.

7.1. **Ozone across Time and Space.** ⊤

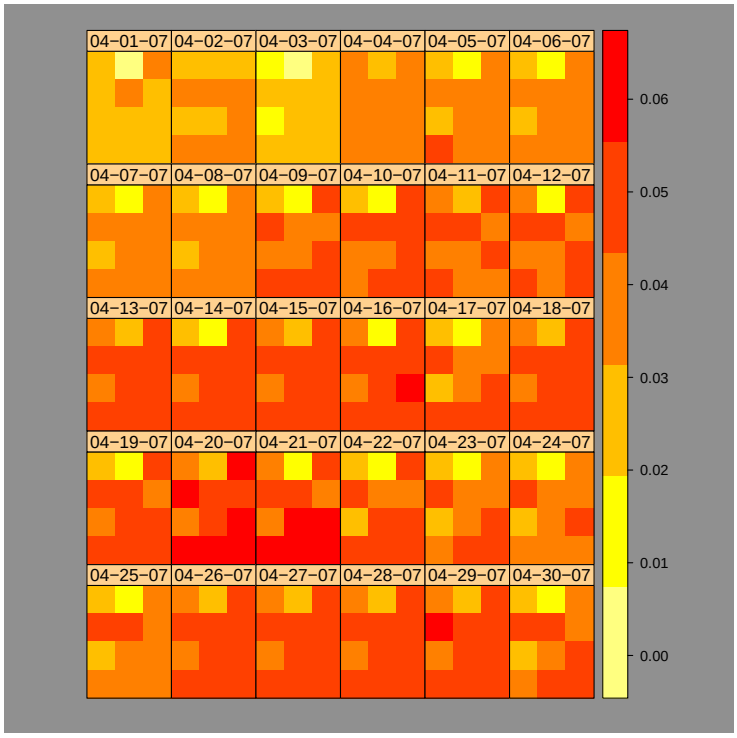

FIGURE 4. Mean grid level ozone, April 2007

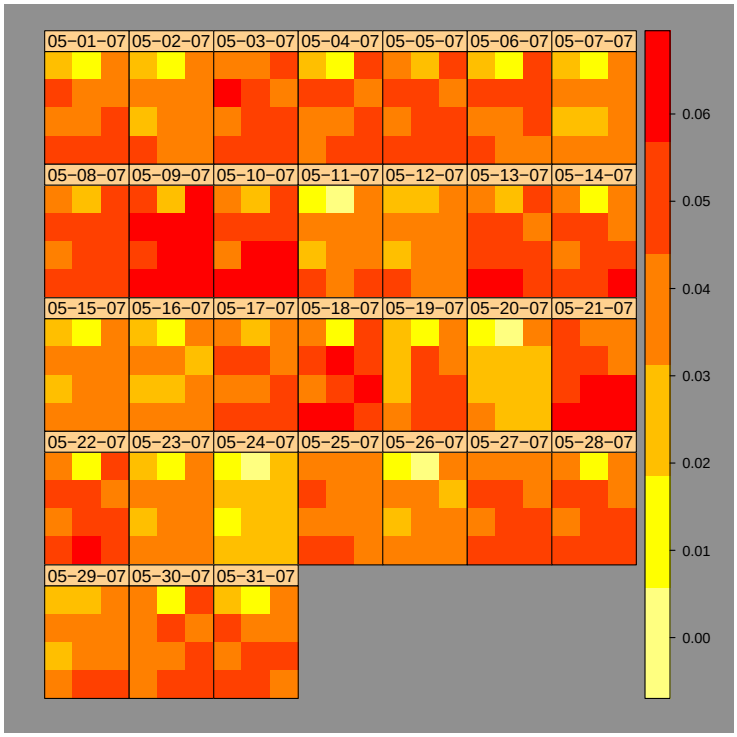

FIGURE 5. Mean grid level ozone, May 2007

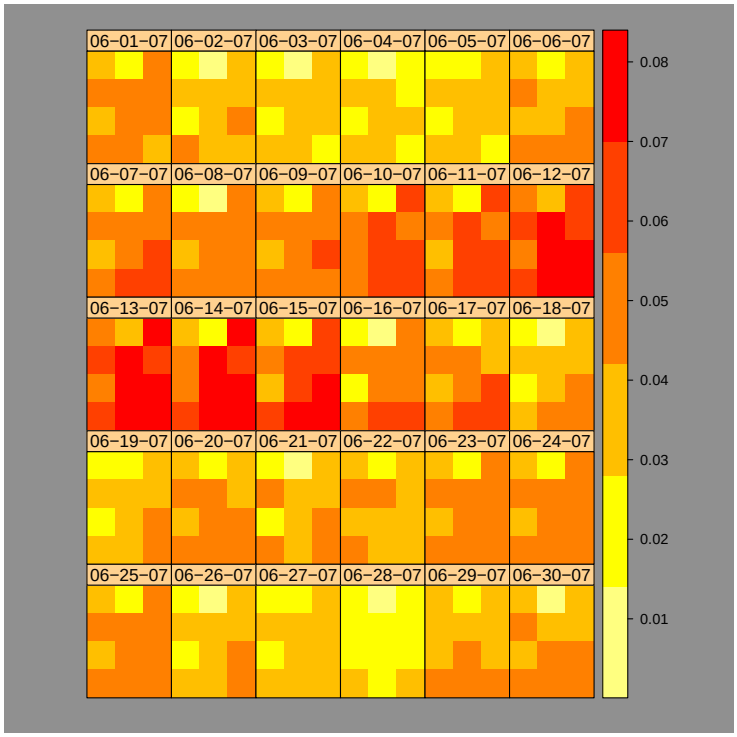

FIGURE 6. Mean grid level ozone, June 2007

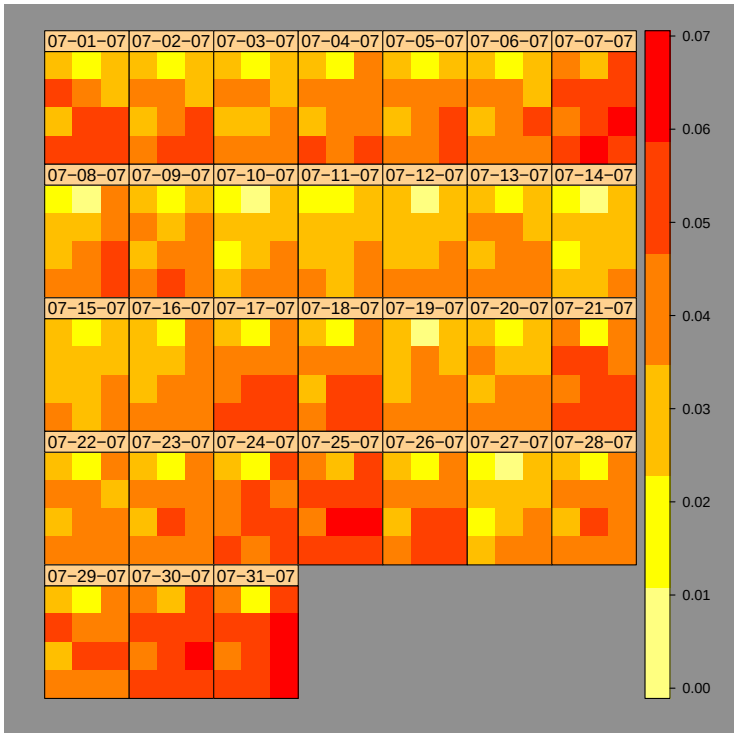

FIGURE 7. Mean grid level ozone, July 2007

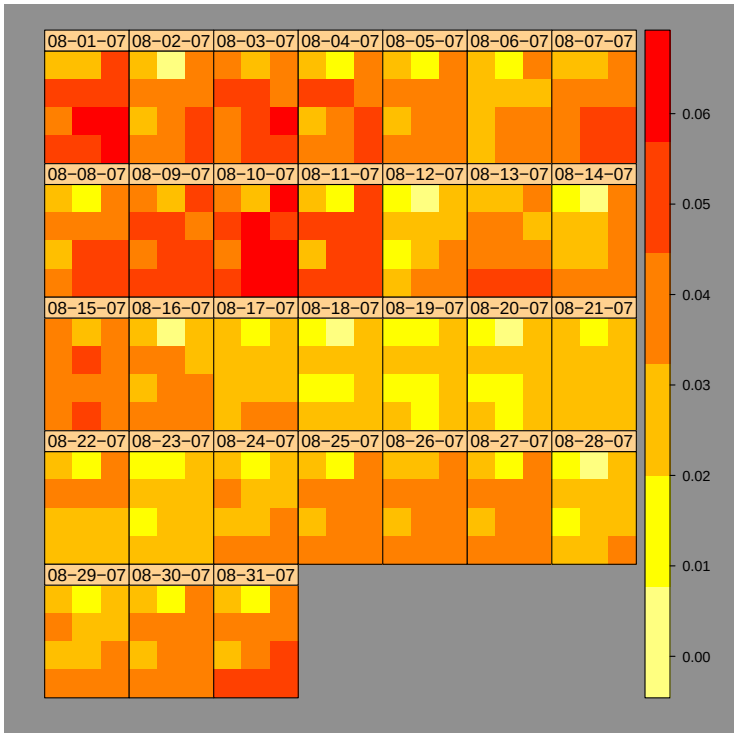

FIGURE 8. Mean grid level ozone, August 2007

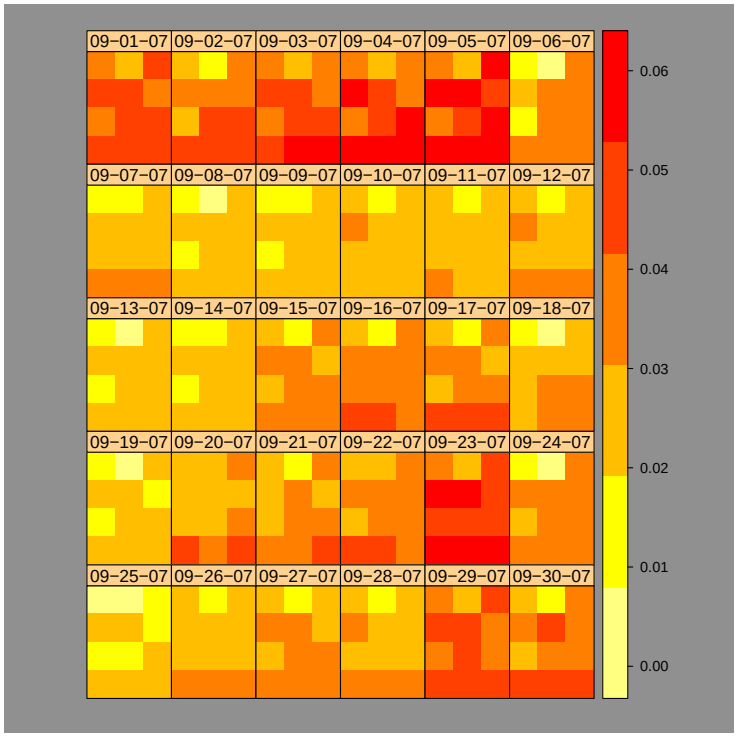

FIGURE 9. Mean grid level ozone, September 2007

8. Model Improvement

Many opportunities exist for model improvement. First of all, creating a finer grid of cells would put stations that are currently in the same box into separate boxes. This is because a finer grid would decrease the measurement error due to ozone variability within a cell. Another way to reduce the measurement error due to intracellular variability is to differentiate between rural ozone monitor sites and urban ozone monitor sites.

Another way to improve the model would involve placing ozone monitoring sites more evenly throughout the state. Some cells in our current model do not contain ozone monitoring stations, so these cells are assigned information from a monitoring site in a different cell. This means that the ozone value assigned to a cell is actually a value collected from many miles away, so the error associated with this assignment has the potential to be quite large. Placing ozone monitoring sites more evenly throughout Minnesota would lead to fewer gridboxes depending on monitoring sites contained in other gridboxes. Therefore, the error in the ozone level would be lower. This is not practical, as the MPCA places their monitors to maximize the probability of detecting dangerous conditions as they are occuring, so it would be a waste of their resources to allocate sensors to locations that are unlikely to have dangerous ozone concentrations.

Additionally, incorporating the HYSPLIT model would improve our model. Utilizing the HYSPLIT model would eliminate our dependence on $\boldsymbol{\theta}$ to capture the effect of moving masses of ozone. This is because the HYSPLIT model (http://www.arl.noaa.gov/HYSPLIT.php) calculates where a given air mass was a certain number of hours ago and where it will be in a given number of hours. Incorporation of the HYSPLIT model would be better capture the spatial-temporal nature of ozone than $\boldsymbol{\theta}$.

Also, the model would benefit from more informative prior distributions. Incorporating expert opinions or prior information into the prior distributions would allow the Markov Chains to converge more efficiently. Thus, the burn-in period would be shorter, so the same total number of iterations would produce a better approximation of the posterior distribution.

References

[1] Albert, J. *Baysian Computation with R*, Spring Science+Business Media, New York-U.S.A., 2008.

[2] Banerjee,Sudipto, Carlin, Bradley P. amd Gelfand, Alan E. *Hierarchical Modeling and Analysis for Spatial Data*, Chapman and Hall/CRC Boca Raton-U.S.A., 2004.

[3] Bivand, R.S., Pebesma, E.J., Gomez-Rubio, V., *Applied Spatial Data Analysis with R*, Springer Science+Business Media, New York-U.S.A., 2008.

[4] Camalier, L., Cox, W., Dolwick, P.: *The Effects of Meteorology on Ozone in Urban Areas and Their Use in Assessing Ozone Trends*, Atmospheric Environment, **41**, (2007), 7127-7137.

[5] Carlin, B., Louis, T.A., *Bayesian Methods for Data Analysis*, Chapman and Hall/CRC Press, Boca Raton-U.S.A., 2009.

[6] Chatfield, C. *The Analysis of Time Series: An Introduction*, Chapman and Hall, London-England, 5th edition, 1997.

[7] Chinkin, L.R., Crews, J.M., MacDonald, C.P., Funk, T.H., Wheeler, N.J.M., Dye, T.S. *Preliminary Assessment of Ozone Air Quality Issues in the Minneapolis/St. Paul Region*, Minnesota Pollution Control Agency, St. Paul-U.S.A., 2002.

[8] Clark, J.S., Gelfand, A.E.: *Hierarchical Modelling for the Environmental Sciences*, Oxford University Press.

[9] Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London, UK, 1996.

[10] Lu, Z.Q., Berliner, L.M.: *Markov Switching Time Series Models with Application to a Daily Runoff Series*, Water Resources Research, **35**, (1999), 523-534.

[11] McMillan, N., Bortnick, S.M., Irwin, M.E., Berliner, M.: *A Hierarchical Bayesian Model to Estimate and Forecast Ozone Through Space and Time*, Atmospheric Environment, **39**, (2005), 1373-1382.

[12] Ramsey, F.L., Schafer D.W, *The Statistical Sleuth*, Duxbury Press, Pacific Grove-U.S.A., 2003.

[13] Shumway, R.H. *Applied Statistical Time Series Analysis*, Prentice Hall, Englewood Cliffs-U.S.A., 1988.

[14] Wikle, C., Berliner, M., Cressie, N.: *Hierarchical Bayesian Space-Time models*, Environmental and Ecological Statistics, **5**, (1998), 117-154.

[15] Wikle, C.K., Berliner, L.M., Milliff, R.F.: *Hierarchical Bayesian Approach to Boundary Value Problems with Stochastic Boundary Conditions*, Monthly Weather Review, **131**, (2003), 1051-1062.

9. Appendix

In this section, we sketch the derivation of the full conditional posterior distributions of each of the model parameters. The derivation of these posterior conditional distributions is done using the principle outlined by equation 2. For most model parameters, the assumption of Gaussian random variables and Gaussian prior distributions implies posterior conditional distributions that are proportional to Gaussian distributions.As an illustration, we give details for the full conditional posterior distribution of the ozone process O_t. The remaining ones follow a fairly similar pattern.

First, with the assumption of normal distribution on both the daily ozone measurement and process, we write:

$$p(Z_t|K, O_t, \sigma^2_{z,t}) \sim N(KO_t, \sigma^2_{z,t})$$

$$p(O_t|\mu_t, H(\theta), O_{t-1}, G(\theta), B_{t-1}, M_t, \beta, \sigma^2_{o,t}) \sim N(\mu_t + H(\theta)O_{t-1} + G(\theta)B_{t-1} + M_t\beta, \sigma^2_{o,t})$$

Now, we derive the full conditional posterior distribution of O_t. Note that we have to include an O_{t+1} conditional to illustrate the predictive capacity of the model. So for $t = 1, 2, ..., T-1$:

$$\begin{aligned}
p(O_t|\cdot) \quad \propto \quad & p(Z_t|K, O_t, \sigma^2_{z,t}) \times p(O_t|\mu_t, H(\theta), O_{t-1}, G(\theta), B_{t-1}, M_t, \beta, \sigma^2_{o,t}) \\
& \times p(O_{t+1}|\mu_{t+1}, H(\theta), O_t, G(\theta), B_t, M_{t+1}, \beta, \sigma^2_{o,t}) \\
& \propto exp\{-\frac{1}{2\sigma^2_{z,t}}(Z_t - KO_t)^T(Z_t - KO_t)\} \\
& \times exp\{-\frac{1}{2\sigma^2_{o,t}}(O_t - \mu_t - H(\theta)O_{t-1} - G(\theta)B_{t-1} - M_t)^T \\
& (O_t - \mu_t - H(\theta)O_{t-1} - G(\theta)B_{t-1} - M_t)\} \\
& \times exp\{-\frac{1}{2\sigma^2_{o,t}}(O_{t+1} - \mu_{t+1} - H(\theta)O_t - G(\theta)B_t - M_{t+1})^T \\
& (O_{t+1} - \mu_{t+1} - H(\theta)O_t - G(\theta)B_t - M_{t+1})\}
\end{aligned}$$

Thus we can factorize the terms that are related to $p(O_t|\cdot)$ and omit the other ones as they are constants. This simplification leads to the grouping of term whose common factor is either $O_t^T O_t$ or O_t which leads to the following form:

$$
\begin{aligned}
p(O_t|\cdot) \quad \propto \quad & exp\{-\frac{1}{2}[(O_t(\frac{1}{\sigma^2_{z,t}}K^TK + \frac{1}{\sigma^2_{o,t}} + \frac{1}{\sigma^2_{o,t}}H(\theta)^TH(\theta)]O_t \\
& -2(\frac{1}{\sigma^2_{o,t}}Z_tK + \frac{1}{\sigma^2_{o,t}}(\mu_t^TI^T + O_{t-1}^TH^T + B_{t-1}^TG^T + \beta^TM_t^T + \\
& O_{t+1}^TH(\theta) - \mu_{t+1}^TI^TH(\theta) - B_t^TG(\theta)^TH(\theta) - \beta_tM_{t+1}^TH(\theta))O_t\}
\end{aligned}
$$

Now we apply the completing the square approach to the latter result to conclude that $p(O_t|\cdot)$ is multivariate normal distribution whose mean and covariance matrix is decribed below:

$$
\begin{aligned}
p(O_t|\cdot) \quad \sim \quad & N([O_t(\frac{1}{\sigma^2_{z,t}}K^TK + \frac{1}{\sigma^2_{o,t}} + \frac{1}{\sigma^2_{o,t}}H(\theta)^TH(\theta))O_t]^{-1} \\
& [\frac{1}{\sigma^2_{o,t}}Z_tK + \frac{1}{\sigma^2_{o,t}}(\mu_t^TI^T + O_{t-1}^TH^T + B_{t-1}^TG^T + \beta^TM_t^T \\
& +O_{t+1}^TH(\theta) - \mu_{t+1}^TI^TH(\theta) - B_t^TG(\theta)^TH(\theta) - \beta_tM_{t+1}^TH(\theta))O_t]^{-1}, \\
& [O_t(\frac{1}{\sigma^2_{z,t}}K^TK + \frac{1}{\sigma^2_{o,t}} + \frac{1}{\sigma^2_{o,t}}H(\theta)^TH(\theta))O_t]^{-1})
\end{aligned}
$$

9.1. **Spatial Parameters.**

$$
\begin{aligned}
p(\theta_i|\cdot) \quad \propto \quad & \prod_{t=1}^{T} p(O_t|\mu_t, H(\theta), O_{t-1}, G(\theta), B_{t-1}, M_t, \beta, \sigma^2_{o,t}) \times p(\theta_i|\widetilde{\mu}_i, \sigma_i^2) \\
& \propto exp\{-\frac{1}{2\sigma^2_{o,t}}\sum_{t=1}^{T}(O_t - \mu_t - H(\theta)O_{t-1} - G(\theta)B_{t-1} - M_t\beta)^T\} \\
& \times exp\{-\frac{1}{\sigma_i^2}(\theta_i - \widetilde{\mu}_i)\}
\end{aligned}
$$

To complete the full posterior distribution for $\theta[i]$, we decompose the matrices $H(\theta)O_{t-1}$ and $G(\theta)B_{t-1}$. Let $H(\theta)O_{t-1} = \theta D_iO_{t-1} + H_iO_{t-1}$ and $G(\theta)B_{t-1} = \theta E_iB_{t-1} + G_iB_{t-1}$. Here we define H_i as the $H(\theta)$ matrix with the diagonal corresponding to θ_i replaced by zero's. D_i is an $S \times S$ matrix with $1's$ in the diagonal corresponding to the diagonal in $H(\theta)$ with $\theta_i's$, this matrix is completed by $0's$ in all of its other entries. Similarly, G_i is the $G(\theta)$ matrix but with the diagonal corresponding to θ_i replaced by $0's$. Finally, E_i is also an $S \times S$ matrix with $1's$ in the diagonal whose corresponding entries in $G(\theta)$ are θ_i.

$$
\begin{aligned}
p(\theta_i|\cdot) \quad \propto \quad & exp\{-\frac{1}{2\sigma^2_{o,t}}\sum_{t=1}^{T}(O_t-\mu_t-\theta D_iO_{t-1}-H_iO_{t-1}-\theta E_iB_{t-1}-G_iB_{t-1}-M_t\beta)^T\} \\
& \times exp\{-\frac{1}{\sigma_i^2}(\theta_i-\widetilde{\mu}_i)\} \\
& \propto exp\{-\frac{1}{2}\theta(\frac{1}{\sigma_{o,t}}\sum_{t=1}^{T}D_iO_{t-1}D_iO_{t-1}+(E_iB_{t-1})^TD_iO_{t-1}+(E_iB_{t-1})^TE_iB_{t-1}+\frac{1}{\sigma_i^2})\theta \\
& -2\theta[(O_t-\mu_t-H_iO_{t-1}-G_iO_{t-1}-M_t\beta)^T(D_iO_{t-1}+E_iB_{t-1})+\frac{\widetilde{\mu}_i}{\sigma_i^2}]\}
\end{aligned}
$$

9.2.

$$
\begin{aligned}
p(\theta_i|\cdot) \quad \sim \quad & N((\frac{1}{\sigma_{o,t}}\sum_{t=1}^{T}D_iO_{t-1}D_iO_{t-1}+B+(t-1)^TD_iO_{t-1}+E_iB_{t-1}E_iB_{t-1}+\frac{1}{\sigma_i^2})^{-1} \\
& [(O_t-\mu_t-H_iO_{t-1}-G_iO_{t-1}-M_t\beta)^T(D_iO_{t-1}+E_iB_{t-1})+\frac{\widetilde{\mu}_i}{\sigma_i^2}]^T, \\
& (\frac{1}{\sigma_{o,t}}\sum_{t=1}^{T}D_iO_{t-1}D_iO_{t-1}+(E_iB_{t-1})^TD_iO_{t-1}+(E_iB_{t-1})^TE_iB_{t-1}+\frac{1}{\sigma_i})^{-1})
\end{aligned}
$$

$$
p(\sigma^2_{z,t}|\cdot) \propto \prod_{t=1}^{T}p(Z_t|O_t,\sigma^2_{z,t})p(\sigma^2_{z,t}|\zeta,\eta)
$$