MASTER THESIS

Bayesian Spatial Quantile Regression: A Substudy of Tennessee Ground-level Ozone

Author:
Jen-Tse KAO

Supervisor:
Dr. Michael STEIN

A thesis submitted in fulfillment of the requirements
for the degree of

in the

May 7, 2018

Abstract

Bayesian Spatial Quantile Regression: A Substudy of Tennessee Ground-level Ozone

by Jen-Tse KAO

We analyze ground-level ozone concentration in Tennessee during the summer in 1999 using Bayesian spatial quantile regression developed by Reich et al. (2011). We modify their primitive (full Bayesian) model by employing piecewise polynomial (B splines) as basis functions.

In the simulation study, we find that B-Splines slightly outperforms Bernstein polynomials when the prior for updating centering distributions are correctly specified. Bernstein polynomials perform relatively better when the previous condition is violated. We provide possible explanations to this phenomenon in the context.

In the data analysis part, we find that the ozone distribution may behave differently when the cloud cover changes. To be more specific, when the cloud cover is 10 percent, east Tennessee tends to have higher ground-level ozones. When the cloud cover is 90 percent, Nashville tends to have higher concentrations of ozone.

This thesis is organized as follows. Chapter 1 introduces quantile regression models and literature review. Chapter 2 introduces an approach to Bayesian spatial quantile regression from Reich et al. (2011), and provides a modification to it. Chapter 3 applies these methods to ozone concentrations in Tennessee during the summer of 1999.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Regression is a widely used tool in various subjects for modeling the conditional mean of a response given covariates. In practice, the change in covariates might affect the entire distribution of the response, not just change the conditional mean. In light of this possibility, classical regression may not be well suited to explaining the distribution of a response variable. Quantile regression, on the other hand, allows researchers to explore this possibility.

Quantile regression has been widely studied in economics, finance, statistics and related quantitative areas. In climatology, P. Friederichs and A. Hense (2007) presented a censored quantile regression method to forecast the extreme precipitation in Europe. However, less attention [1] has been paid to quantile regression for analyzing spatial data.

Classical[2] quantile regression does not make distributional assumptions on error term. Econometricians usually require the error to be zero when conditioning on predictors, $E(\varepsilon_i|x_i) = 0$, to produce unbiased and consistent estimators, and further apply asymptotics on regression coefficient β to perform inference. That is, consider the conditional linear quantile regression

$$\hat{\beta}(\tau) = argmin_\beta \sum_{i=1}^{n} (\rho_\tau(y_i - X_i\beta)) \tag{1.1}$$

$$\rho_\tau(u) = u * (\tau - I(u < 0))$$

where $\rho_\tau(u)$ is the check function. It is known that the above minimization problem can be re-framed into a linear program, wherein the simplex or interior method is employed.

1.1 Literature Review

There is a large amount of literature on quantile regression since Koenker's (1978) seminal work. It is therefore a daunting task to keep track of the development of this method. Koenker (2017) provides a concise review of frequentist's approach. In this article, we focus attention on quantile regression in statistical literature. In particular, we review three components that are vital to quantile regression in statistics: Bayesian inference, non-crossing constraints, and spatial data.

[1] To our knowledge, relatively few papers consider using spatial quantile regression for research purpose before 2010. The earliest one might be M. Hallin et al. (2009). There exists a few quantile regression treatments for spatial data in econometrics, but since our attention is on statistical side, we choose to skip this part at this moment.

[2] For the model-free approach, we refer to "classical" or "frequentist" quantile regression.

1.1.1 Bayesian Quantile Regression

The frequentist's approach of quantile regression relies on asymptotic theory, while the model-based or Bayesian approach requires likelihood for inference. In order to introduce distribution into the quantile regression framework, Yu and Moyeed (2001) proposes a pseudo or auxiliary likelihood approach by modeling ε_p[3] as asymmetric Laplace distribution (ALD), where p is the p-th quantile for separate regression model. The term "auxiliary" comes from the fact that maximizing the log likelihood is equivalent to minimizing the check loss function. As a consequence, the spirit of Bayesian quantile regression adopted by Yu and Moyeed (2001) is very similar to the frequentist's approach. Finally, it is important to note that the data generating process is not distributed as ALD.

Other than using ALD for Bayesian inference, Reich et al. (2011) and Reich (2013) employ different schemes for constructing a likelihood, which will be discussed later.

1.1.2 Non-Crossing Constraint

For given covariates X_i, the conditional quantile function $\tau \longrightarrow q_{Y|X}(\tau|X_i)$ should be weakly non-decreasing in τ. However, this may not be the case for $\hat{q}_{Y|X}(\tau|X_i) = X_i^T\hat{\beta}(\tau)$, where $\hat{\beta}(\tau)$ is in (1.1). This problem may happen because, in practice, the quantile regression is implemented via simplex or interior optimizations, and because each quantile curve is estimated separately. Hence Bondell et al. (2010) proposes to estimate all conditional quantile function simultaneously under non-crossing constraints.

As for the Bayesian quantile regression, due to its embodying auxiliary likelihood, it is no surprise that the Bayesian approach inherits the same property as frequentist's approach, i.e. they estimate conditional quantiles separately, which may produce potential crossing quantiles. Tokdar and Kadane (2012) points out that this is not a serious issue in classical framework, as the asymptotic theory works whenever statistical inferences are needed. However, it is inappropriate to ignore the non-crossing constraint in the model-based approach.

1.1.3 Spatial Quantile Regression

The development of spatial quantile regression can be traced back to 2010. Reich et al. (2011) analyzes ground-level ozone in eastern United States from 1997 to 2005, for that ozone has been linked with several adverse health effects. They develop a model-based quantile regression which models the regression coefficients as Gaussian processes to take into account the spatial variations, and employs the Bernstein polynomial to fulfill the non-crossing constraints.

Reich (2013) develops a spatio-temporal quantile regression model for detecting the distributional changes in temperature in the southeast United States from 1931 to 2009. This paper is different from Reich et al. (2011) in several ways. Reich (2013)

[3]The readers may wonder the difference between ε_i and ϵ_p. The former focuses on i-th observation, while the latter stresses on modeling separate model according to p-th quantile. In the frequentist's approach and the earlier Bayesian approach, the correct way is to use ε_{ip}.

permits a closed-form density for the response distribution, and hence allows the Bayesian model to be fit larger spatio-temporal datasets. Also, Reich (2013) considers the Gaussian copula model to account for the residual correlation.

Besides the aforementioned literature, there are other advancements in this fields. Lum and Gelfand (2012) is an extension of Yu and Moyeed (2001) wherein they model the error distribution ε_p as a Gaussian process. Das and Ghosal (2017) is an application of Tokdar and Kadane (2012) into the spatial context. Contrary to Reich et al. (2011), wherein they set up different quantile functions at each spatial location, the quantile functions in Das and Ghosal (2017) share the same parameters at each spatial location, hence named "simultaneous" quantile regression. However, Das and Ghosal (2017) is restrictive in the sense that only single covariate is allowed in their approach. We finally decided to skip this part and consider more general works as in Lum and Gelfand (2010), Reich et al. (2011), Zhou et al. (2011), and Reich (2013).

Chapter 2

Bayesian Spatial Quantile Regression

We introduced the development of quantile regression by Reich et al. (2011) in Chapter 1. In the following, we adopt their full Bayesian approach[1], and make slight modifications to this model.

2.1 Statistical Model

Let $q(\tau|X_i, s) = \sum_{j=1}^{p} X_{ij}\beta_j(\tau, s)$ be the quantile function and X_{ij} be the covariates. The coefficients $\beta_j(\tau, s)$ are dependent on both τ and spatial locations $s = (s_{i1}, s_{i2})$. Reich et al. (2011) makes the following specification

$$\beta_j(\tau, s) = \sum_{m=1}^{M} B_m(\tau)\alpha_{jm}(s)$$

$$\alpha_{jm}(s) = \sum_{l=1}^{m} \delta_{jl}(s)$$

where $B_{m,M}(\tau) = \binom{M}{m}\tau^m(1-\tau)^{M-m}$ is the Bernstein polynomial, and M denotes the number of basis functions. The $\delta_{jl}(s)$ are defined as[2]

$$\delta_{jl}(s) = \begin{cases} \delta^*_{jl}(s), \text{ if } \delta^*_{jl}(s) + \sum_{j=2}^{p} I(\delta^*_{jl}(s) < 0)\delta^*_{jl}(s) \geq 0 \\ 0, \text{otherwise} \end{cases} \tag{2.1}$$

where $\delta^*_{jl}(s)$ are modeled as independent Gaussian processes with mean $E(\delta^*_{jl}(s)) = \bar{\delta}_{jl}(\Theta)$ and covariance $Cov(\delta^*_{jl}(s), \delta^*_{jl}(s')) = \sigma_j^2 exp(-||s - s'||/\rho_j)$.

Define $\theta_m(X_i, s) = \sum_{j=1}^{p} X_{ij}\alpha_{jm}(s)$ and adjust for the indices, we arrive at the following representation

$$q(\tau|X_i, s) = \sum_{j=1}^{p} X_{ij}\beta_j(\tau, s) = \sum_{j=1}^{p} X_{ij}\Big(\sum_{m=1}^{M} B_m(\tau)\alpha_{jm}(s)\Big) \tag{2.2}$$

$$= \sum_{m=1}^{M} B_m(\tau)\theta_m(X_i, s) \tag{2.3}$$

[1] We mention that Reich et al. (2011) proposes two approaches, namely, a full Bayesian method using Metropolis-Hastings algorithm, and an approximation method using Gibbs sampling.

[2] The condition 2.1 is shown in equaion (11) in Reich et al. (2011).

where we require that $\theta_m(X_i, s) \geq \theta_{m-1}(X_i, s)$, $m = 1, ..., M$ to ensure the non-crossing constraint. The monotonicity constraint is guaranteed by Chang et al. (2007). On the other hand, B-Splines possesses the same property whenever $\theta_m(X_i, s) \geq \theta_{m-1}(X_i, s)$ holds for $m = 1, ..., M$.

2.2 Update Centering distribution

When fitting the model using MCMC, a crucial step is to update the centering distribution f_0, that is, update the mean of $\delta_m^*(s)$ at each iteration. Let $q_0(\tau|\Theta)$ be the quantile function of f_0. We choose $\overline{\delta_{jl}}(\Theta)$ so that $q_0(\tau|\Theta) \approx \sum_{m=1}^{M} B_m(\tau)\overline{\alpha_{jm}}(\Theta)$, that is, $\overline{\delta_{jl}}(\Theta)$ are chosen to minimize the following ridge regression

$$\sum_{k=1}^{K}\left(q_0(\tau_k|\Theta) - \sum_{m=1}^{M} B_m(\tau_k)\left[\sum_{l=1}^{m} d_{jl}\right]\right)^2 + \lambda\sum_{m=1}^{M} d_m^2$$

The centering distribution f_0 is taken to be skew-normal, with location $\mu_0 \sim N(0, \sigma_\mu^2)$, scale $\sigma_0 \sim$ Uniform(a,b), and skewness $\phi_0 \sim N(0, \sigma_\phi^2)$. In section 2.8 the simulation study, we choose $\mu_0 \sim N(30, 100)$, $\sigma_0 \sim Uniform(0, 40)$, and $\phi_0 \sim N(0, 10)$. The choice of these prior distributions are based on density plot of simulated dataset as in Figure 3.1, and R package *sn*, in which we can estimate the above parameters by routine *sn.mple*.

2.3 Basis Functions

We illustrate the role of basis function expansions in the Bayesian spatial quantile regression, and present the close connection between Reich et al. (2011) and Reich (2013). Recall that the quantile functions in Reich et al. (2011) take the form

$$q(\tau|X_i, s) = X_i^T\beta(\tau, s) = X_i^T\left(\sum_{m=1}^{M} B_m(\tau)\alpha_{jm}(s)\right) = \sum_{m=1}^{M} B_m(\tau)\theta_m(X_i, s)$$

and observe that if we take $X_i^T = \begin{bmatrix} 1 & t_i \end{bmatrix}$, t_i is time variable

$$\begin{aligned} q(\tau|X_i, s) &= \beta_0(\tau, s) + t_i \cdot \beta_1(\tau, s) \text{ (Reich et al. (2011))} \\ &= q_0(\tau|s) + t \cdot q_1(\tau|s) \text{ (Reich (2013))} \end{aligned}$$

where $\beta_0(\tau, s)$, $\beta_1(\tau, s)$ and hence $q_0(\tau|s)$, $q_1(\tau|s)$ admit basis function expansions, e.g. Bernstein polynomial (Reich et al. 2011), or piecewise Gaussian basis function (Reich 2013), or I-Splines (Zhou 2011).

Although these models share the same building blocks, the non-crossing constraints will be deeply affected by the choice of basis function expansions. As a consequence, the non-crossing constraints have different forms:

(1) Reich et al. (2011) requires $\theta_{jm}(s) \geq \theta_{j-1,m}(s)$, no differentiation is involved.

(2) Reich (2013) requires $\theta_{0l}(s) + t \cdot \theta_{1l}(s) > 0$ after taking derivative $\frac{\partial}{\partial\tau}q(\tau|s, t)$.

(3) Zhou et al. (2011), the non-crossing constraints are simply non-negative coefficients $\alpha_{jm}(s) \geq 0$.

It is worth mentioning that the choice of basis function expansions also have impact on the likelihood computation:

(1) Reich et al. (2011) evaluates likelihood by $\mathbb{P}((q|X_i,s)) = \left[q(\tau_{j+1}|X_i,s) - q(\tau_j|X_i,s)\right]^{-1}$ as the use of Bernstein polynomial does not permit closed form likelihood.

(2) Reich (2013) permits closed-form density due to the use of Gaussian basis function

$$\mathbb{P}(q(\tau|s,t)) = \sum_{l=1}^{L} I\left(q(\tau_l) \leq q(\tau|s,t) \leq q(\tau_{l+1})\right) N\left(q(\tau|s,t)|a_l(s,t), b_l(s,t)^2\right)$$

2.4 B-Splines and Bernstein Polynomials

We generalize the Bernstein polynomial, which is defined on the entire interval $[0,1]$ to the piecewise polynomial B-Splines, and describe the relationship between Bernstein polynomial and B-Splines in the appendix. In fact, the Bernstein polynomial is a special case of B-Splines under certain conditions.[3] In this article we replace the Bernstein polynomial basis function in 2.2 with B-Splines. According to De Boor (2001) and Das and Ghosal (2017), the increasing coefficients $\theta_m(X_i,s) \geq \theta_{m-1}(X_i,s)$, $m = 1,...,M$ ensure the non-crossing constraint.

Figure 2.1 depicts B-Spline basis functions with different degrees. In this paper, we choose B-Spline with degree 3 and 9 equally spaced knots as our basis function, as in part (b). It is possible to choose non-equally spaced B-Splines basis function. We give a illustration of the special case in Figure 2.1 (c) and (d).

2.5 Metropolis-Hastings Algorithm

We review basic notions of the Metropolis-Hastings algorithm. Let $\pi(x)$ be the stationary probability density where we want to draw samples from. In particular, we assume $\pi(x) = \frac{1}{Z}p(x)$ where $p(x)$ is known and the normalizing constant Z is possibly unknown. Let $q(x,y)$ be the transition or proposal probability density function, and is often denoted by $q(y|x)$. Then the acceptance probability is

$$\alpha(x,y) = min\{\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\}$$

Metropolis - Hastings Algorithm Let $X^{(n)}$ be the random samples/chains drawn at time n. Suppose that Y is generated from the distribution $q(Y|X^{(n)})$ and that U is drawn from $Uniform(0,1)$. If $U \leq \alpha(X(n),Y)$ then we set $X^{(n+1)} = Y$. Otherwise we set $X^{(n+1)} = X^{(n)}$.

We describe the Metropolis-Hastings algorithm in Reich et al. (2011). In this context, the random samples $X^{(n)}$ are $\delta_{jl}^*(s)$ with dimensions $NS \times P \times M$, where NS denotes the number of spatial locations, P denotes the number of covariates, and M is the number of basis functions in $\beta_j(\tau,s) = \sum_{m=1}^{M} B_m(\tau)\alpha_{jm}(s)$ for $j = 1,...,P$, $m = 1,...,M$, and $s = 1,...,NS$.

[3] Please refer to appendix for the details.

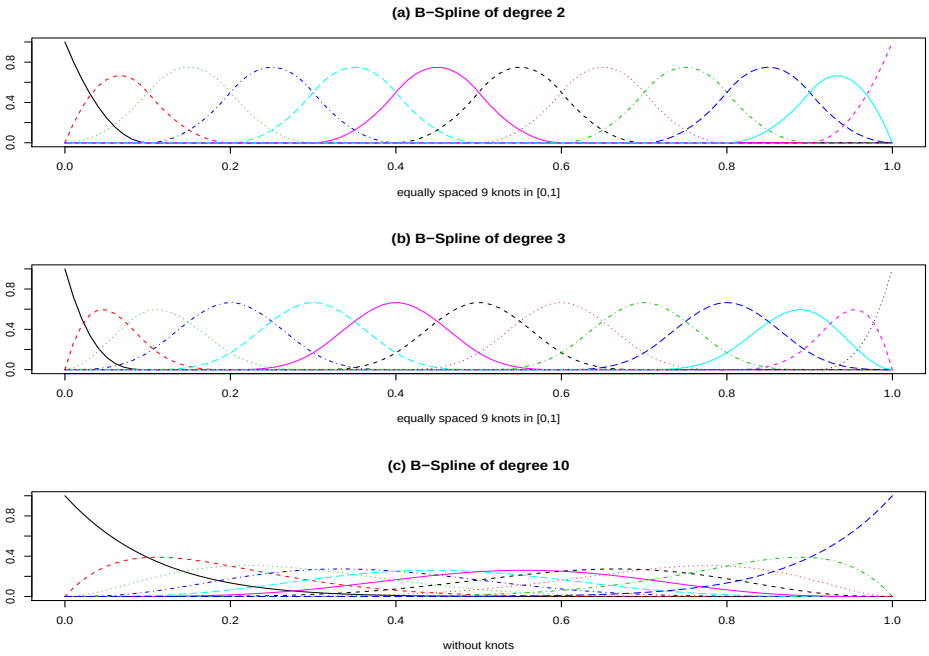

FIGURE 2.1: (a) and (b): B-Spline basis function of degree 2 and 3 with equally spaced 9 knots $\{0.1, 0.2, , ..., 0.9\}$ on unit interval. (c): B-Spline basis function of degree 10 without knots, which equals the Bernstein polynomial of degree 10.

Let Σ_j be the $NS \times NS$ covariance matrix indexed by $j = 1, ..., P$ and $\Lambda_j = \Sigma_j^{-1}$ be the corresponding precision matrix. Define

$$SD = \sqrt{\sigma_{jl}^{-2}\Lambda_j(s,s)}$$
$$MN = SD^2\sigma_{jl}^{-1}\left(\Lambda_j(s,s)\overline{\delta}_{jl} - \Lambda_j(s,-s)\left(\delta_{jl}^*(-s) - \overline{\delta}_{jl}\right)\right)$$

where $\delta_{jl}^* \sim N(\overline{\delta_{jl}}, \sigma_{jl}^2)$, $\Lambda_j(s,s)$ extracts the (s,s) position in precision matrix, $\Lambda_j(s,-s)$ denotes the precision matrix but remove the s-th column, and $\delta_{jl}^*(-s)$ denotes the vector of $\delta_{jl}^*(s)$ but remove the s-th component. We now are ready to present the Metropolis-Hastings algorithm for Bayesian spatial quantile regression.

Metropolis - Hastings Algorithm in Bayesian spatial quantile regression

$$(\text{For } j = 1, ..., P)(\text{For } s = 1, ..., NS)(\text{For } m = 1, ..., M)$$
$$\text{initialize } X^{(n)} = \delta_{jl}^*(s), n = 1, ..., P \times NS \times M$$
$$\text{draw } Y_{jl}(s) \text{ from } N(\delta_{jm}^*, \text{some variance})$$
$$\alpha(Y_{jl}(s), \delta_{jl}^*(s)) = \frac{N(Y_{jl}(s)|\mu = MN, \sigma = SD)}{N(\delta_{jl}^*(s)|\mu = MN, \sigma = SD)}\frac{f(Y_{jl}(s))}{f(\delta_{jl}^*(s))}$$
$$\text{If } U(0,1) < \alpha(Y_{jl}(s), \delta_{jl}^*(s)) \text{ then } \delta_{jl}^*(s) = Y_{jl}(s)$$
$$\text{Otherwise } X^{(n+1)} = \delta_{jl}^*(s)$$

where $f(\dot{})$ is the likelihood. It is worth mentioning that the strategy behind the algorithm is that they assume separate quantile functions at different locations but determine the chains $X^{(n)}$ jointly according to the stationary distribution $N(\mu = MN, \sigma = SD)$ and proposal density function $f()$.

2.6 MCMC Convergence

The convergence of Markov Chain Monte Carlo (MCMC) is an unsettled issue. Hence the purpose of this section is to provide a discussion of some of the issues related to convergence rather than a serious treatment of the topic.

The diagnostics of convergence may involve quantitative or graphical comparisons. Readers interested in this topic may reference Cowles and Carlin (1996) for early development. Recently, Goodman and Weare (2010) recommends using integrated autocorrelation time (IAT) to quantify sampling errors. The IAT, on the other hand, is related to the concept of effective sample size, which measures approximately how many independent samples we would obtain given the current correlated samples generated by a MCMC algorithm.

Let $\tau_{IAT}(f)$ be the integrated autocorrelation time

$$\tau_{IAT} = 1 + 2\sum_{n=1}^{\infty} corr(f(X^{(0)}), f(X^{(n)}))$$

where $X^{(n)}$ is the random samples drawn from stationary distribution at time n, $corr$ is the correlation between $f(X^{(0)})$ and $f(X^{(n)})$. The effective sample size (ESS) is $N/\tau_{IAT}(f)$, where N denotes the number of iterations.

We indicate that the estimation of IAT is a non-trivial problem, for instance, Thompson (2010) summarizes several methods for computing the IAT. In addition, the model presented in this thesis involves a huge number of coefficients and quantiles when compared with the mean regression. The current methods evaluating IAT may not be directly applied to this highly-structured system. In fact, it is possible to compute the IAT for regression coefficients $\beta(\tau, s)$ at different quantile levels. However, we may obtain a huge number of IATs for $\beta(\tau, s)$ due to interaction of the number of spatial locations and level of quantiles.

We next discuss how long the burn-in period should be. Geyer (1992) suggests that 1% or 2% of the run will suffice. However, early literature may not take into account the increasing complexity of current approaches. We also tried the method proposed by Sahlin (2011) in which the maximum effective sample size is determined via

$$ESS_{max} = max_{0 \leq n \leq N} ESS(n)$$

However, we may obtain a large range of ESS values for quantiles across different spatial locations. In the effort to compute IAT and ESS, we find that there are no universal criteria to follow in this thesis.

We finally decide to compute the RMSE under different lengths of burn-ins and from which to determine the length of burn-ins in the Bayesian spatial quantile regression.

2.7 Gaussian Copula

Modeling conditional densities via quantile functions $q(\tau|X_i, s) \equiv Q^{-1}(\tau)$ only specifies marginal distribution at each spatial location s and leaves the joint distribution undetermined. To introduce the spatial associations into spatial sites s, we may resort to the copula framework. The reader may refer to Nelson (2006) for a good reference of copulas.

A common definition for Gaussian copula is

$$C(u_1, ..., u_p) = \Phi_p\left(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_p)|0, \Theta\right)$$

where C is the copula, Φ_p is the multivariate normal distribution $MVN(\mathbf{0}, \Theta)$ with covariance matrix Θ, and Φ^{-1} is the inverse cdf of univariate normal, u_i is data input, usually taken as realizations of $U_i = F_X(X_i)$ with some random variable X_i and its cdf F_X.

We resort to Gaussian copula model to introduce spatial dependence among locations s_i while preserving marginal distribution $q(\tau|X_i, s)$. By the Gaussian copula model, we refer to the models that are used in Sang and Gelfand (2009), and Fuentes et al. (2013). Hence in the following, we first introduce a latent process $\mathbf{Z}$

$$\begin{aligned} \mathbf{Z} &= (Z_1, ..., Z_p) \sim \Phi_p(\mathbf{0}, \Theta) = MVN(\mathbf{0}, \Theta) \\ q(\tau|X_i, s) &= Q^{-1}\Phi(Z_i|0, \theta_{ii}) \end{aligned}$$

where $\Theta = \text{diag}(\theta_{ii})$ $\Phi(Z_i) \sim$ uniform $(0, 1)$ by the probability integral transformation.

Let $\mathbf{Z}$ be a standardized Gaussian process with mean 0, variance 1, with correlation function $\rho(s, s')$, then the transformed Gaussian process is defined as $Y_i = Q^{-1}\Phi(Z_i)$. More precisely, let $(Y_1, ..., Y_p) = \left(Q^{-1}\Phi(Z_1), ..., Q^{-1}\Phi(Z_p)\right)$ then the joint distribution of $(Y_1, ..., Y_p)$ is given by the Gaussian copula

$$C_Z(u_1, ..., u_p) = \Phi_p\left(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_p)\right)$$

which leads to

$$\mathbb{P}(Y(s_1) \leq y_1, ..., Y(s_p) \leq y_p) = C_Z(Q(Y_1), ..., Q(Y_p)) = \Phi_p\left(\Phi^{-1}Q(Y_1), ..., \Phi^{-1}Q(Y_p)\right)$$

where $u_i = Q(Y_i) = \Phi(Z_i)$.

In the simulation study followed by Reich et al. (2011), we use the data generating process $Z(\tau|X_i, s)$ in equation 2.4. We will consider two cases in the simulation study, the first one is when the probabilities τ are generated independently from Uniform(0,1), and the second one is when the probabilities τ are generated via Gaussian copula.

2.8 Simulation Studies

We assume the following data generating process for the ground-level ozone[4][5] at each spatial location

$$Z(s) = 24s_2 + (T(s)+1)\Phi^{-1}(T(s)|\mu = 20, \sigma = 5) + (50s_1(T(s))^2)X_s \tag{2.4}$$

where $T(s) \sim$ Uniform(0,1), $s = (s_{i1}, s_{i2}) \in [0,1]^2$ denotes spatial locations, i.e. longitude and latitude, and $X(s) \in [0,1]$, and $\Phi^{-1}(\tau|\mu = 20, \sigma = 5)$ denotes the inverse distribution function of normal random variable with mean $\mu = 20$ and standard deviation $\sigma = 5$. The purpose of equation 2.4 is to provide a spatially varying quantile function that mimics the kind of patterns which might be observed in applications, as shown in Figure 3.1 and Figure 3.3.

We use 20 datasets, each dataset contains 20 stations, and station contains 80 observations for the simulation. To ensure reproducibility, we fix random seed starting from 244 and onwards. Besides generating independent probabilities $T(s) \sim$ Uniform(0,1) and substitute into the quantile function 2.4, we also generating correlated probabilities $T(s)$ by first drawing samples from Gaussian process with exponential covariance function $\exp(-||s_i - s_j||/0.5)$, and then apply probability integral transformation to obtain correlated observations from joint distributions whose marginal quantile function is given by 2.4.

To compare models, the following summary statistic[6] is used in the simulation study.

$$\sqrt{\frac{\sum_{s,i,k,n=1}^{D,L,Q,N}(\hat{q}(\tau_k|X(s),s) - q(\tau_k|X(s),s))^2}{DLQN}}$$

where D is the number of datasets, L is the number of stations, Q is the number of quantiles to be estimated, and N is the number of observations at each station. Figure 2.2 shows the RMSE under different choice of burn-in numbers. As we mentioned in section 2.6, there seems no consensus on the determination of burn-in numbers, we decide to use 5000 in the real data analysis in Chapter 3.

Figure 2.2 shows that the use of B-Splines performs slightly better than Bernstein polynomials whether the quantiles τ are independent or not.

We give possible explanations to this situation. We might think that there are two driving forces in this highly-structured system: approximation by (piecewise) polynomials, and approximation by MCMC samples. The former is relevant to numerical precisions, and the latter is relevant to likelihood inference. When the quantiles are independent, it is perhaps enough to approximate quantile with piecewise polynomials of degree 3, and higher order polynomials may result in noise in approximation.

[4] In this way, equation 2.4 also defines the quantile function of the process. We denote $q(\tau|X(s),s) = 24s_2 + (\tau+1)\Phi^{-1}(\tau|\mu = 20, \sigma = 5) + (50s_1\tau^2)X(s)$ as quantile function. Please refer to appendix 4.2 for the justification.

[5] Please refer to appendix 4.3 for the discussion.

[6] Reich et al. (2011) also takes into account MSE for regression coefficients. Since we do not compare the same model, we elect to consider RMSE only.

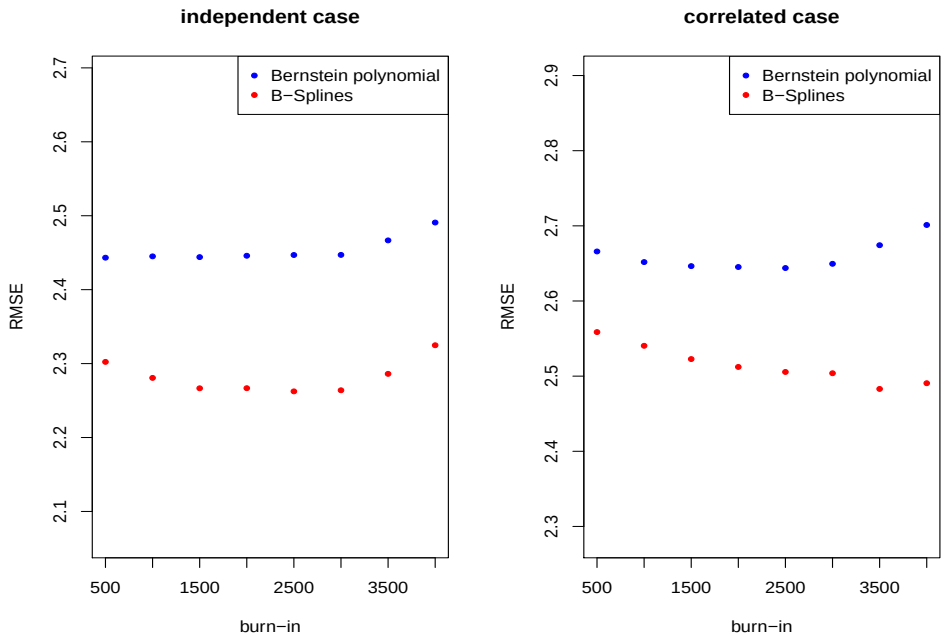

FIGURE 2.2: Left panel: The RMSE under different length of burn-ins when the probabilities τ are independent. Right panel: The RMSE under different length of burn-ins when the probabilities τ are correlated. In both cases the total number of runs are 5,000.

2.9 Conclusion

In this section we perform simulation study using Bernstein polynomials and B-Splines as basis functions. It is found that B-Splines performs generally better whether or not the system possess strong correlation, given that the prior distributions for updating centering distribution f_0 are correctly specified.

Chapter 3

Data Analysis

3.1 Introduction

Ground-level ozone is detrimental to human's respiratory system, especially for people with asthma, children, and older adults. Hence it is natural to ask whether there exists suitable statistical tools to anlayze the concentration of ozone to allow policy making.

Unlike some pollutants such as carbon monoxide and sulfur dioxide, such as PM2.5, ozone is not emitted directly into the air. Indeed, it is a by product of the chemical reaction among nitrogen oxides, volatile organic compounds, sunlight, and high temperature. Hence, we expect nitrogen oxides, temperature, sky cover, and wind speed play important roles in determining ground-level ozone levels. In the following context, we choose daily average temperature and cloud cover and their interaction terms as our covariates.

(need to change quantiles in the graph)

3.2 Data Collection Procedure

The data are collected from National Oceanic and Atmospheric Administration (NOAA), which contains several divisions, each division collects data in its respective regions. Meteorological data are obtained from Daily Summaries database and National Solar Radiation Database (NSRDB).

The daily average temperature (measured in $^{\circ}$F) are gathered from Daily Summaries database, which also contains daily maximum temperature, daily minimum temperature, average wind speed, precipitation, etc. NOAA has land-based stations across the United States to document hourly data, but the record years vary across sites. The daily maximum temperature is the highest value among hourly temperatures in a day, and vice versa for minimum temperature. The daily average temperature is the average of hourly temperature. We think that this is taken over 24 hours, although this is not stated explicitly in the official documentations.

Hourly cloud cover (in tenths) is from NSRDB, and we average it from 7:00 AM to 17:00 PM, because this time period might affect the amount of sunlight, and hence the production of ozone. The unit 1tenth means it is measured by the amount of sky covered by the clouds at specific a time.

The ozone measurements are from the Environmental Protection Agency (EPA). It is measured by taking consecutive eight-hour averages starting from 7:00 AM until

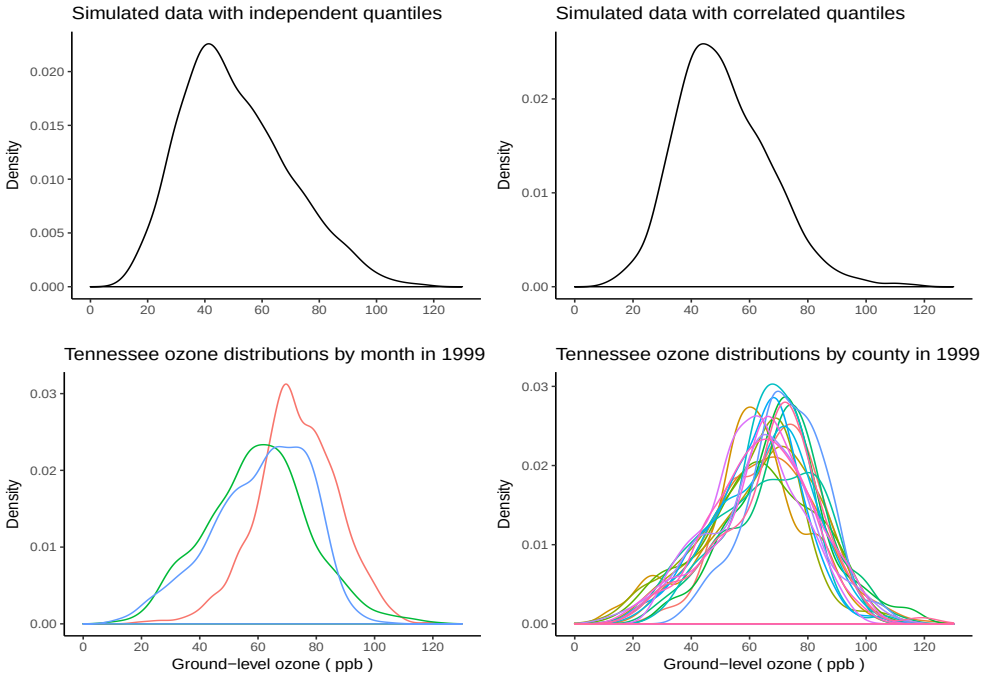

FIGURE 3.1: (a) Density of $Z(\tau|X_i, s)$ via independent probabilitiesτ) and (b) Density of $Z(\tau|X_i, s)$ via correlated probabilities τ, both aim to mimic the marginal distributions of ground-level ozone at spatial locations. (c) Blue: June; Green: July; Red: August.

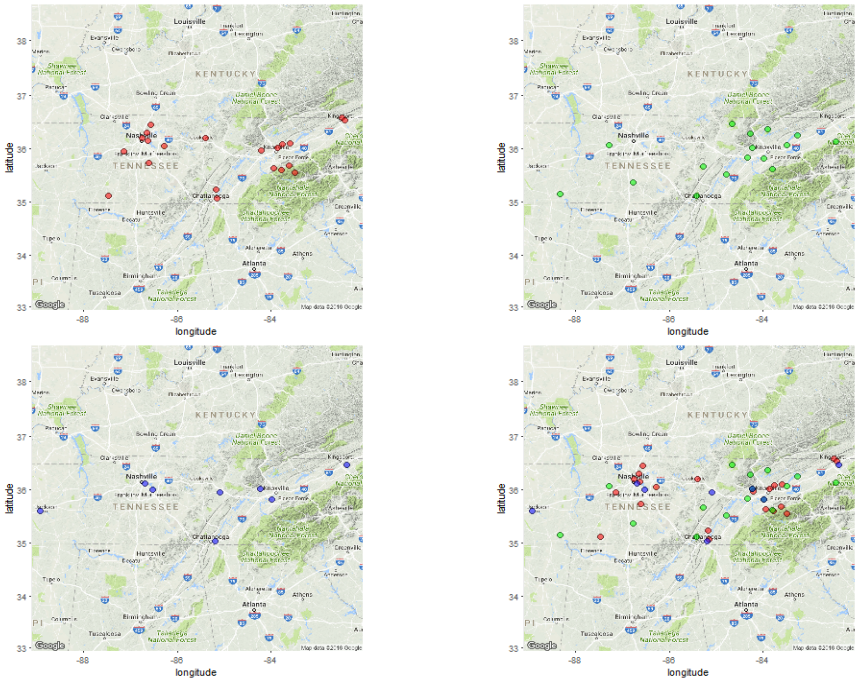

FIGURE 3.2: Meteorological stations in Tennessee: Red: ozone; Green: temperature; Blue: cloud cover. The last panel is a combination of three panels.

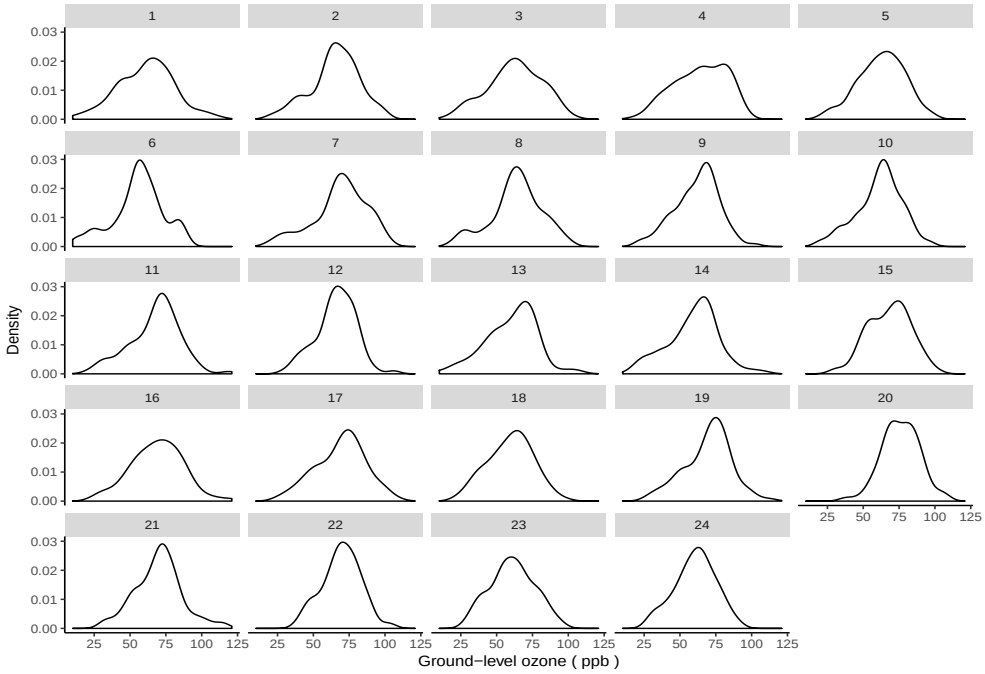

FIGURE 3.3: Ozone distributions for 24 spatial locations in Tennessee in the summer of 1999. Since the name of stations are unavailable, we label them with numbers.

23:00 PM, and select the maximum of these 17 eight-hour averages. The unit is ppm, and we may convert it to ppb in the analysis.

Notice that most meteorological sites are well separated from each other. In order to obtain predictions of temperature and cloud cover at ozone stations, we choose to use spatial kriging with an exponential covariance function to impute values at these sites. To be more specific, we use all neighboring states of Tennessee when imputing average temperature and cloud cover. We hope this could improve of the accuracy of spatial kriging.

The spatial kriging is implemented by R package *gstat* and assume exponential covariance function $C(h) = \beta \exp(-|h|/\alpha), |h| \geq 0$ across space, where $|h|$ is the relative distance. We illustrate this procedure as follows.

- use $SpatioPoints$ to set correct distance
- compute sample variogram via $variogram$
- estimate parameters α and β via $fit.variogram$ and choose $vgm(\text{"Exp"})$
- employ ordinary kriging using $gstat$
- predict temperature and cloud cover at ozone locations using $predict$

It is also possible to perform spatio-temporal kriging by R package $spacetime$ and package $gstat$, which may obtain more accurate estimates for temperature and cloud cover.

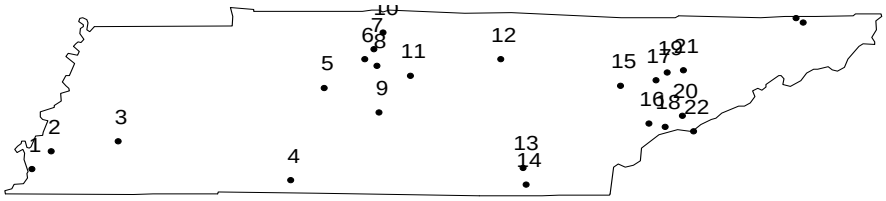

FIGURE 3.4: Ozone distributions for 24 spatial locations in Tennessee in the summer of 1999 corresponding to Figure 3.3.

3.3 Substudy of Tennessee

In this study, we focus on Tennessee and restrict the time span to June to August in 1999. The substudy includes 24 ozone monitoring stations and 92 days.

Before proceeding to statistical analysis, we first notice that the upper two panels in figure 3.1 presents the simulated ozone distributions by independent and correlated probabilities τ. From the lower panel in figure 3.1, we find that the ozone distribution has similar shape across counties in Tennessee, and tends to be higher concentration in August. Also, figure 3.3 shows the ozone distributions at 24 different spatial locations in the summer of 1999.

We next apply the model in Chapter 2, in which B-splines are used as basis functions for Bayesian spatial quantile regression. We implement the following model for 10,000 iterations and use length 5,000 as burn-ins[1]. Notice that we slightly change the subscript for j and k to correspond to Reich et al. (2011) paper.

$$ozone_{ik} = \beta_0(s) + \beta_1(s)\text{Temp}_{ik} + \beta_2(s)\text{Cover}_{ik} + \beta_3(s)(\text{Temp*Cover})_{ik} + \epsilon_{ik}$$

$$\beta_j(s) = \sum_{m=1}^{M} B_m(\tau)\alpha_{jm}(s)$$

[1] As discussed in section 2.6 and section 2.8, there are no apparent criteria for the length of burn-ins. We follow Reich et al. (2011) to choose half iterations as length of burn-ins. We also provide plots with burn-in length 2,000 and 8,000 in appendix 4.4.

where i is month, j is covariate, k is day, $B_m(\alpha)$ is B-Splines basis functions, $\alpha_{jm}(s)$ $N(\bar{\alpha}, \sigma_\alpha)$ are independent (over j and m) spatial (Gaussian) processes.

The prior for centering distributions $\mu_0 \sim N(70, 60)$, $\sigma_0 \sim$ Uniform(0,30), and ϕ_0 $\sim N(0, 5)$, which are obtained from the R routine $sn.mple$ in package sn and corresponding density plot.

In figure 3.5 and figure 3.6 we examine the conditional quantiles given average temperature at 68 °F with cloud cover 10%, and temperature at 68 °F with cloud cover 90%, respectively. The range of the image plots is fixed from 10 to 130 (ppb) so we can compare ground-level ozone at different quantiles and covariates.

For instance, when cloud cover is 10%, the change of ozone levels at $\tau = 0.2$ is less radical than the change of ozone levels at quantile levels greater than $\tau = 0.2$. Interestingly, when the cloud cover is 10%, eastern Tennessee usually has higher ground-level ozone concentrations than the rest of Tennessee at almost all quantile levels, whereas when cloud cover is 90%, the higher ozone concentrations occur around Nashville, the central part of Tennessee.

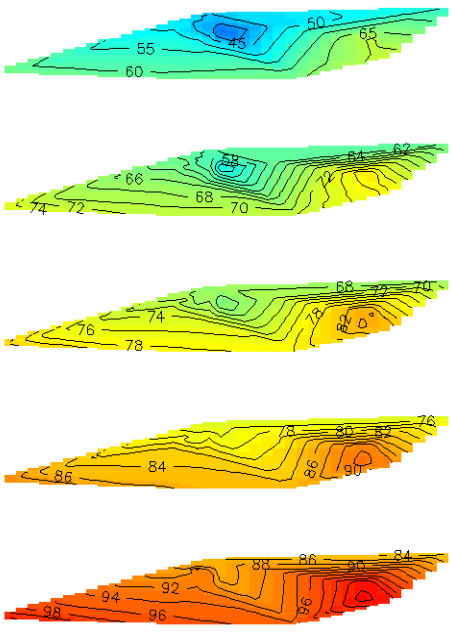

FIGURE 3.5: Temperature at 68 ° F with cloud cover 10% for quantile levels 0.1, 0.3, 0.5, 0.7, 0.9.

3.4 Conclusion

In this section we employ Bayesian spatial quantile regression to analyze conditional ozone distributions in Tennessee from June to August in 1999. We find that when the cloud cover is 10%, eastern Tennessee usually has higher ozone concentrations. But when the cloud cover is 90%, the Nashville tends to produce higher ozone concentrations.

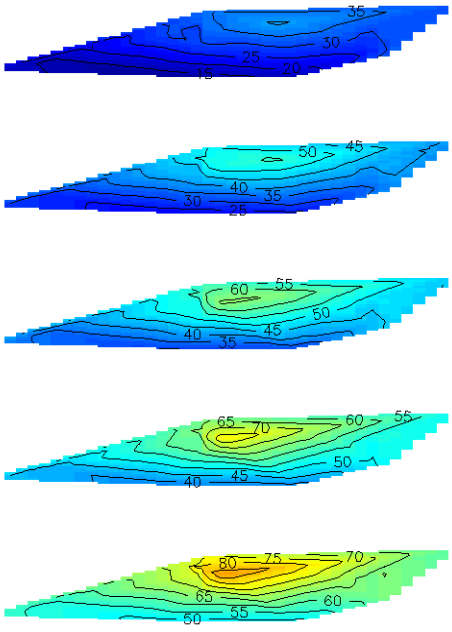

FIGURE 3.6: Temperature at 68 ° F with cloud cover 90 percent for quantile levels from 0.1, 0.3, 0.5, 0.7, 0.9.

Lastly, we conclude that the choice of updating centering distribution and its priors play important roles in this approach. We may not blindly use this model for other kinds of meteorological variables if the response variable does not follow skew-normal distribution. For future research direction, we hope to design faster Metropolis-Hastings algorithm for large spatial datasets.

Chapter 4

Appendix

4.1 Bernstein polynomial and B-Splines

We follow Floater's (2007) lecture notes, and explain that Bernstein polynomial is a special case of B-Splines by providing a simple example.

Let d be an non-negative integer, and **t** be a sequence of knots in $[0,1]$ denoted by $\mathbf{t} = \{0 = t_0 < t_1 < ... < t_{n-1} < t_n = 1\}$, then the jth B-spline of degree d is defined as

$$B_{j,d}(x) = \frac{x - t_j}{t_{j+d} - t_j} B_{j,d-1}(x) + \frac{t_{j+d+1} - x}{t_{j+d+1} - t_{j+1}} B_{j+1,d-1}(x)$$

$$B_{j,0}(x) = \begin{cases} 1, \text{if } t_j \leq x < t_{j+1} \\ 0, \text{otherwise} \end{cases}$$

We note that $B_{j,d}$ depends only on the knots $\{t_k\}_{k=j}^{j+d+1}$ on interval $[a,b]$ and use the notation $B(x|t_j, t_{j+1}, ..., t_{j+d}, t_{j+d+1})$. The Bernstein polynomial, on the other hand, can be denoted by $B_{j,d}(x)$ for the jth Bernstein polynomial of degree d, and write $B(x|\overbrace{a,...,a}^{\text{d+1-j}}, \overbrace{b,...,b}^{\text{j+1}})$. According to the recursion formla

$$\begin{aligned} B_{j,d}(x|\overbrace{a,...,a}^{\text{d+1-j}}, \overbrace{b,...,b}^{\text{j+1}}) &= \frac{x-a}{b-a} B(x|\underbrace{a,...,a}_{\text{(remove one a)}}, \overbrace{b,...,b}^{\text{j+1}}) + \frac{b-x}{b-a} B(x|\overbrace{a,...,a}^{\text{d+1-j}}, \underbrace{b,...,b}_{\text{(remove one b)}}) \\ &= \binom{d}{j} \left(\frac{x-a}{b-a}\right)^j \left(\frac{b-x}{b-a}\right)^{d-j} B(x|a,b) \end{aligned}$$

For instance, the Bernstein polynomial of degree 2 on interval $[a,b]$ is composed of three basis functions: $B_{0,2}(x) = B(x|a,a,a,b)$, $B_{1,2}(x) = B(x|a,a,b,b)$, and $B_{2,2}(x) = B(x|a,b,b,b)$. Notice that the knots consist only of endpoints of the interval. By the previous arguments, we note that the recursion formula removes endpoints successively either from the right or from the left, we therefore define $B(x|a,...,a) = 0$ or $B(x|b,...,b) = 0$ whenever case occurs.

$$B_{0,2}(x) = \frac{x-a}{b-a}B(x|a,a,b) + \frac{b-x}{b-a}B(x|b,b,b) = \frac{x-a}{b-a}\Big[\frac{x-a}{b-a}B(x|a,b) + \frac{b-x}{b-a}B(x|a,a)\Big]$$
$$= \Big(\frac{x-a}{b-a}\Big)^2 B(x|a,b)$$

$$B_{1,2}(x) = \frac{x-a}{b-a}B(x|a,b,b) + \frac{b-x}{b-a}B(x|a,a,b) = \frac{x-a}{b-a}\Big[\frac{x-a}{b-a}B(x|b,b) + \frac{b-x}{b-a}B(x|a,b)\Big]$$
$$+ \frac{b-x}{b-a}\Big[\frac{x-a}{b-a}B(x|a,b) + \frac{b-x}{b-a}B(x|a,a)\Big] = 2\Big(\frac{x-a}{b-a}\Big)\Big(\frac{b-x}{b-a}\Big)B(x|a,b)$$

$$B_{2,2}(x) = \frac{x-a}{b-a}B(x|b,b,b) + \frac{b-x}{b-a}B(x|a,b,b) = \frac{b-x}{b-a}\Big[\frac{x-a}{b-a}B(x|b,b) + \frac{b-x}{b-a}B(x|a,b)\Big]$$
$$= \Big(\frac{b-x}{b-a}\Big)^2 B(x|a,b)$$

We finally note that $B(x|a,b) = 1$ if $a \leq x \leq b$ and $B(x|a,b) = 0$ otherwise.

4.2 Representing function and quantile function

We follow Wichura's (2001) lecture notes and provide the following theorems.

Definition 1. Let X be a random variable with distribution function F. A mapping R: $(0,1) \to \mathbb{R}$ is called a representing function if (1) R is nondecreasing, and (2) $R(U) \sim X$ whenever $U \sim$ Uniform(0,1).

Theorem 1. Suppose R is a representing function for X and $Y = T(X)$ is a transformation of X. If T is nondecreasing, then the mapping $u \to T(R(u))$ is a representing function for Y.

Theorem 2. Let X be a random variable. A mapping Q from $(0,1)$ to $\mathbb{R}$ is a quantile function for X if and only if Q is a representing function of X.

Let N, not necessarily normally distributed, be a random variable with distribution function Φ. According to the inverse probability transformation, the left-continuous inverse Φ^{-1} is a representing function. Furthermore, suppose $T(N)$ is a transformation of N and T is nondecreasing, then $T(\Phi^{-1})$ is a representing function of $T(N)$. Therefore, we know that the equation 2.4, $Z(\tau|X(s),s)$, is also a quantile function for some underlying process.

4.3 The Simulated Datasets

This section aims to explain the data generating process used in section 2.8. The data generating process in Reich et al. (2011) is

$$Z(\tau|X(s),s) = 2s_{i2} + (T(s)+1)\Phi^{-1}(T(s)) + (5s_{i1}T(s)^2)X(s) \tag{4.1}$$

where $\Phi^{-1}(\tau)$ is the inverse CDF of standard normal distribution $N(0,1)$, and s_{i1}, s_{i2}, $X(s)$ are defined in section 2.8. However, there are concerns about this process.

First, it may not represent the actual ozone distribution. When plotting the density function, we find that the simulated data ranges from -2 to 10, as shown in Figure 1. (a) in Reich et al. (2011), which is far from the real ozone data when converted into unit ppb. As displayed in Figure 2. (a) in Reich et al. (2011), they use unit ppb which ranges from 30 to 60 for median ozone. Second, we experiment with 30 datasets with each 5,000 iterations[1], and find that the MSE for regression coefficients β_i, $i = 1, 2, 3$ are much smaller (around 0.0055) than what are reported (around 0.21) in Table 1. in Reich et al. (2011). We think they might use a different data generating process in their article rather than equation 4.1.

Our initial attempt was to employ the following process

$$Z(\tau|X_i, s) = 20s_{i2} + 10(T(s) + 1)\Phi^{-1}(T(s)) + (50s_{i1}T(s)^2)X(s) \tag{4.2}$$

where $\Phi^{-1}(\tau)$ is the inverse CDF of standard normal distribution $N(0, 1)$, and s_{i1}, s_{i2}, $X(s)$ are defined in section 2.8. We tried 20 datasets with each 10,000 iterations, and find that the RMSE for B-Splines is smaller than the RMSE for Bernstein polynomial, as shown in Figure 4.1. However, this simulated dataset is unrealistic because the ozone levels are less likely to be negative, and this gives us motivation for the data generating process defined in section 2.8. The purpose of defining process 2.4 is to make the marginal density of ground-level ozone closer to the authentic dataset. Finally, the priors for updating centering distributions in Figure 4.1 are chosen to be $\mu_0 \sim N(0, 100)$, $\sigma_0 \sim$ Uniform(0,40), and $\phi_0 \sim N(0, 10)$.

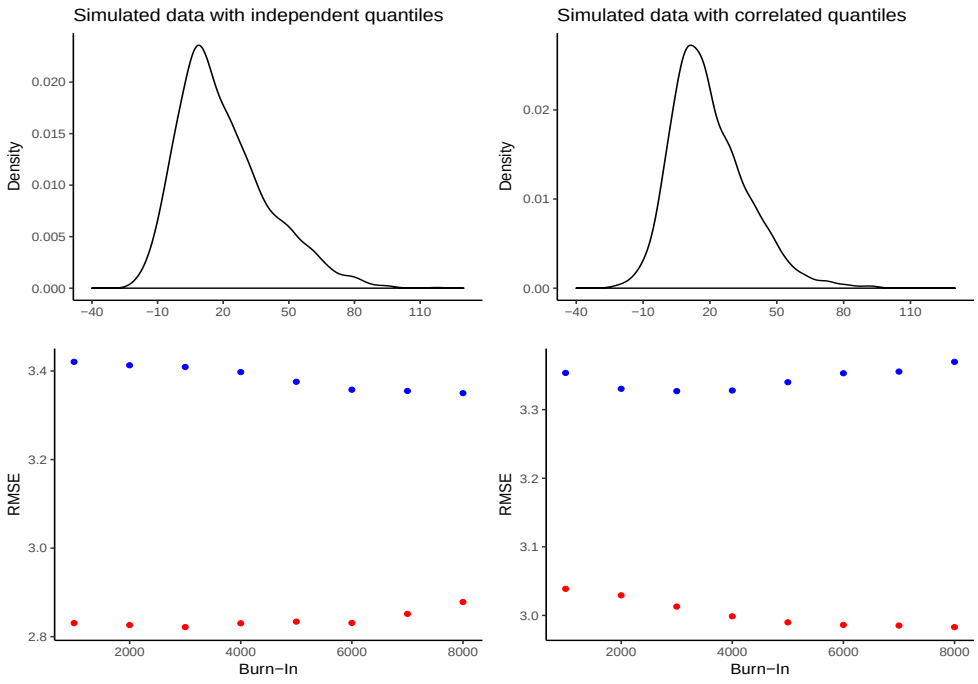

FIGURE 4.1: Simulated datasets generated from the process 4.2 and their respective RMSE in the lower panels.

[1]We mention that they experimented with 50 datasets with each 20,000 iterations. However, because the data generated by equation 4.1 are close to each other, it is unlikely for the regression coefficients estimates to deviate much from the true coefficients. (The difference between estimates and true coefficients are probably less than 1, and by taking square of it.....)

4.4 MCMC Results under Different Burn-ins

This section provides the approximation results in section 3.3 for burn-in numbers 2,000 and 8,000 respectively. Although there are discrepancies for different burn-in, we argue that the overall patterns are quiet similar for burn-in numbers 2,000, 5,000, and 8,000.

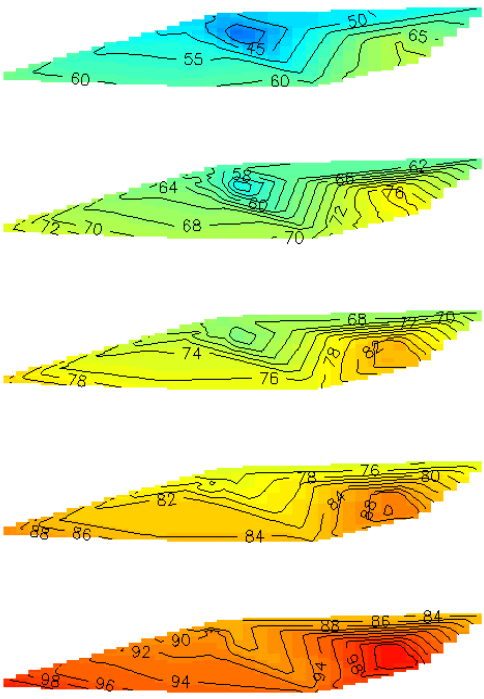

FIGURE 4.2: Temperature at 68 °F with cloud cover 90% for different quantiles with burn-in 2,000 for quantile levels 0.1, 0.3, 0.5, 0.7, 0.9.

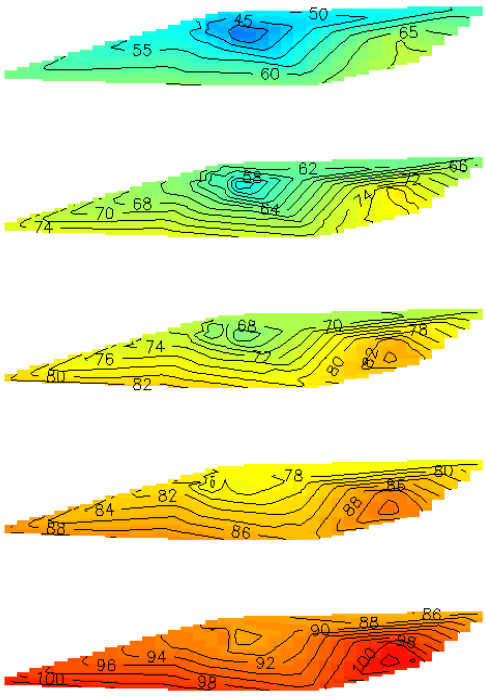

FIGURE 4.3: Temperature at 68 °F with cloud cover 90% for different quantiles with burn-in 8,000 for quantile levels 0.1, 0.3, 0.5, 0.7, 0.9.

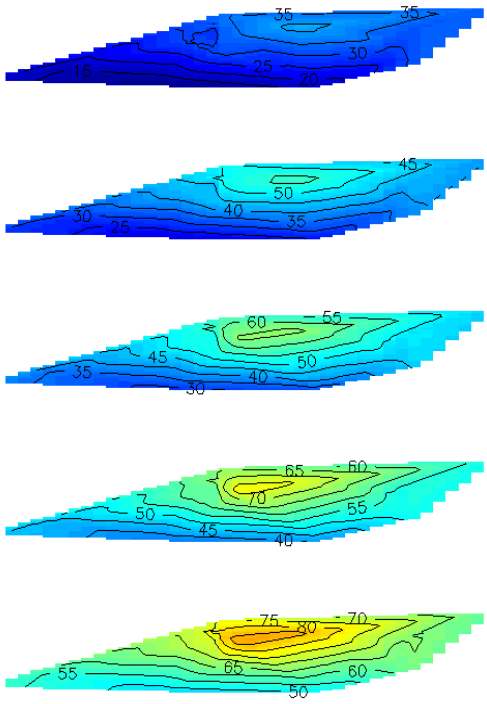

FIGURE 4.4: Temperature at 68 °F with cloud cover 90% for different quantiles with burn-in 2,000 for quantile levels 0.1, 0.3, 0.5, 0.7, 0.9.

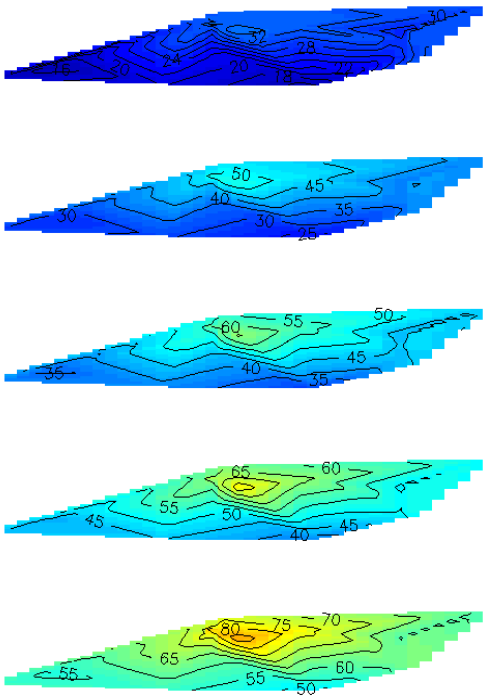

FIGURE 4.5: Temperature at 68 °F with cloud cover 90% for different quantiles with burn-in 8,000 for quantile levels 0.1, 0.3, 0.5, 0.7, 0.9.

Chapter 5

Reference

Das and Ghosal (2017), "Analyzing ozone concentration by Bayesian spatio-temporal quantile regression", Environmetrics

Floater (2007), lecture notes, http://www.uio.no/studier/emner/matnat/ifi/INF-MAT5340/v07/undervisningsmateriale/

Geyer (1992), "Practical Markov Chain Monte Carlo", Statistical Science

Lum and Gelfand (2012), "Spatial Quantile Multiple Regression Using the Asymmetric Laplace Process", Bayesian Analysis

Reich, Fuentes, Dunson (2011), "Bayesian Spatial Quantile Regression", Journal of American Statistical Association

Reich (2013) "Spatiotemporal quantile regression for detecting distributional changes in environmental processes", Applied Statistics

Sahlin (2011), "Estimating convergence of Markov chain Monte Carlo simulations", Master Thesis

Thompson (2010), "A comparison of methods for computing autocorrelation time", Technical Report

Tokdar and Kadane (2012), "Simultaneous linear quantile regression: a semiparametric Bayesian approach", Bayesian Analysis

Wichura (2001), lecture notes, https://galton.uchicago.edu/ wichura/Stat304/handouts.html

Zhou, Fuentes, and Davis (2011), "Calibration of numerical model output using nonparametric spatial density functions", Journal of Agricultural, Biological, and Environmental Statistics