# High Dimensional Spatial Modeling of Extremes with Applications to United States Rainfalls

Jie Zhou

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2007

Approved by

Advisor: Dr. Richard L. Smith

Reader: Dr. Edward Carlstein

Reader: Dr. Gabriele Hegerl

Reader: Dr. Douglas G. Kelly

Reader: Dr. Vidyadhar G. Kulkarni

Reader: Dr. Zhengyuan Zhu

# ABSTRACT

JIE ZHOU: High Dimensional Spatial Modeling of Extremes with Applications to United States Rainfalls

(Under the direction of Dr. Richard L. Smith)

Spatial statistical models are used to predict unobserved variables based on observed variables and to estimate unknown model parameters. Extreme value theory(EVT) is used to study large or small observations from a random phenomenon. Both spatial statistics and extreme value theory have been studied in a lot of areas such as agriculture, finance, industry and environmental science.

This dissertation proposes two spatial statistical models which concentrate on non-Gaussian probability densities with general spatial covariance structures. The two models are also applied in analyzing United States Rainfalls and especially, rainfall extremes.

When the data set is not too large, the first model is used. The model constructs a generalized linear mixed model(GLMM) which can be considered as an extension of Diggle's model-based geostatistical approach(Diggle et al. 1998). The approach improves conventional kriging with a form of generalized linear mixed structure. As for high dimensional problems, two different methods are established to improve the computational efficiency of Markov Chain Monte Carlo(MCMC) implementation. The first method is based on spectral representation of spatial dependence structures which provides good approximations on each MCMC iteration. The other method embeds high dimensional covariance matrices in matrices with block circulant structures. The eigenvalues and eigenvectors of block circulant matrices can be calculated exactly by Fast Fourier Transforms(FFT). The computational efficiency is gained by transforming the posterior matrices into lower dimensional matrices. This method gives us exact update on each MCMC iteration. Future predictions are also made by keeping spatial dependence structures fixed and using the relationship between present days and future days provided by some Global Climate Model(GCM). The predictions are refined by sampling techniques. Both ways of handling high dimensional

covariance matrices are novel to analyze large data sets with extreme value distributions involved. One of the main outcomes of this model is for producing $N$-year return values and return years for a given value for precipitation at a single location given climate model projections based on a grid. This is very important, because in many applications, detailed precipitation information on pointwise locations is more important that predictions averaged over grids.

The second model can be applied to those large data sets and is based on transformed Gaussian processes. These processes are thresholded due to the emphasis on rainfall extremes.

**Keywords**: Block Circulant Matrix; Extreme value theory;Fast Fourier Transform; Generalized Linear Mixed Model; Kriging; Markov Chain Monte Carlo; Spectral Representation; Spatial statistics

***To my parents***

*... and numerous who helped*

# ACKNOWLEDGMENTS

I wish to express my most sincere gratitude to my advisor and mentor, Dr. Richard L. Smith, for leading me through my graduate study, and for his inspiring discussions and many valuable suggestions. Having him as an advisor has been the best part of my life at Chapel Hill.

I would also like to thank the rest of faculty members in Statistics Department, especially Dr.Zhengyuan Zhu and Dr.Chuanshu Ji, for their help in all aspects of my statistics study.

I have had a great time at UNC-Chapel Hill. UNC-Chapel Hill has provided me with a wonderful environment for both my life and study. The people on campus are all extremely nice.

Finally, I want to thank my family and friends for their support all along my pursuit of statistics.

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS AND SYMBOLS

| | |
|---|---|
| $A'$ | transpose of $A$ |
| a.s. | almost surely |
| DFFT | discrete fast Fourier transform |
| $\mathbb{E}$ | expected value |
| FFT | fast Fourier transform |
| GEV | generalized extreme value |
| GLMM | generalized linear mixed model |
| GPD | generalized Pareto distribution |
| i.i.d. | independent and identically distributed |
| MDA | maximum domain of attraction |
| MCMC | Markov chain Monte Carlo |
| $\mathbb{N}$ | set of natural numbers |
| $\Phi$ | standard Normal distribution function |
| $\phi$ | standard Normal density function |
| $\Re$ | set of real numbers |

CHAPTER 1

# Introduction

Extreme values are very important in a lot of areas such as agriculture, finance, industry, environmental science, etc. The Extreme Value Theory(EVT) is developed to study probabilistic and statistical properties of these unusual phenomena. The theory of univariate extremes has been well discussed and there is a large literature behind this subject.

Even though one may be able to collect a lot of data on many locations, it is still impossible to have the observations from all of the area. So it is necessary to use spatial statistics to study high dimensional spatial extreme problems.

Kriging, as one of the broadly used methods in spatial statistics, is an interpolation method to have spatial predictions based on least mean squared error under some assumptions of spatial dependent structures. The limitation of kriging is that it is a method of optimal spatial linear prediction and cannot be applied directly to those spatial data with complicated probability densities. In our case, we are focusing on the precipitation data which contain a lot of zeros and can never be a negative value. The conventional kriging does not reflect these constraints.

Model-based geostatistical approaches proposed by Diggle et al.(1998) discussed a generalized linear mixed model (GLMM) which improves on conventional kriging and is able to handle non-Gaussian cases. The model is solvable by Monte Carlo sampling but conventional Markov Chain Monte Carlo(MCMC) implementation is not practically implementable for high dimensional prediction problems. As an improvement, Wikle et al.(2002) proposed a spectral generalized linear prediction model which is based on spectral parameterized covariance structures. Diggle's model has the assumption that the parameters of the probability density distributions at observed locations are based on some conditional expectations,

while Wikle's model is based on exponential family distributions. It is possible to extend these two models to our extreme value cases, because essentially, the point of these two models is treating the parameters from distributions as Gaussian random fields with noises. The form of probability density distributions is not very important. So in fact, the spectral GLMM can be extended to any parametric distributions but it is an approximate approach.

Alternatively, it is possible to consider an exact approach without losing much of the computational efficiency. The essential point is to embed the covariance matrix in a block circulant matrix and therefore, the eigenvalues and eigenvectors can be obtained exactly by using FFT. If the embedded matrix is not too large, then one will still be able to generate a Gaussian random field quickly with this certain covariance matrix. This idea is proposed by Wood and Chan(1994). For our empirical computational problems, there are several critical adjustments need to be made. Some other linear algebraic techniques are also used to gain the computational efficiency.

Evidence shows that variations in global surface temperature cannot be explained by natural climate variability, especially, the unusual increase in global mean temperature during the latter half of the 20th century can be attributed to anthropogenic forcing. Furthermore, the increase in intense precipitation, more so than the changes in mean precipitation, are consistent with changes expected in a warmer atmosphere due to an acceleration of the hydrological cycle(e.g. Trenberth 1999). The observations of rainfall also support the relationship between extreme rainfall events and the climate changes.

In our work, we would like to focus on the United States due to the availability of a large amount of controlled station data. We will be using a comprehensive description of the statistics of extremes along with spatial interpolation and smoothing methods to simulate the United States rainfalls(extreme events) and make spatial(-temporal) predictions.

The rest of the dissertation is organized as follows. Chapter 2 is an overview of Extreme Value Theory and Spatial Statistics in the literature. It includes the basic concepts of EVT and some standard statistical approaches. Spatial statistics is also introduced in this part, including the conventional kriging method, the generalized linear mixed model proposed by Diggle *et al.*(1998), the spectral parametrization method developed by Wikle(2002) and simulation of stationary Gaussian processes by Wood and Chan(1994).

Chapter 3 introduces the background and motivation of our study on North American rainfall extremes.

Chapter 4 includes a high dimensional spatial modeling with EVT on United States rainfalls, two computational methods to solve the model and the statistical criteria and results for rainfall predictions in the future.

Chapter 5 describes an alternative statistical approach, transformed, thresholded Gaussian Process is considered to simulate the rainfall extremes.

Chapter 6 shows the outline of completion of current works and proposes some future possible extensions.

CHAPTER 2

# Extreme Value Theory and Spatial Statistics

## 2.1 Univariate Extreme Value Theory

Intuitively, Extreme Value Theory(EVT) is a theory concentrating on unusual extreme events and studying the probabilistic and statistical properties of these events. There is a large literature behind this subject going back to the 1920's, so it is impossible to include a complete list of references. There are some classic references including the books written by Leadbetter, Lindgren and Rootzén(1983) and Resnick(1987). The first chapter written by Smith(2003) in Finkenstadt and Rootzén's book *Extreme Values in Finance, Telecommunications and the Environment* is a comprehensive reference of EVT with the application in environment, insurance and finance.

In this section, we will briefly summarize some main results in extreme value theory. These results are mostly taken from the recent paper by Smith(2003), which is a good introduction to this subject.

### 2.1.1 The Extreme Value Distributions

Let $X_1, ..., X_n$ be i.i.d. random variables with common density function $F$. Let $M_n = max\{X_1, ..., X_n\}$, then

$$Pr\{M_n < x\} = F(x)^n \tag{2.1}$$

This won't give us anything other than the trivial result that for any fixed $x$ for which $F(x) < 1$, the probability will tend to 0 when $n$ tends to $+\infty$. Or in fact, the maximum of a sample will tend to the right endpoint of the distribution almost surely, regardless it is

finite or infinite. For instance, let $X_E$ be the right endpoint, since

$$\begin{aligned}\sum_{n=1}^{\infty} Pr\{|M_n - X_E| > \epsilon\} &= \sum_{n=1}^{\infty} Pr\{M_n < X_E - \epsilon\} \\ &= \sum_{n=1}^{\infty} Pr\{X_1 < X_E - \epsilon\}^n \\ &= \frac{Pr\{X_1 < X_E - \epsilon\}}{1 - Pr\{X_1 < X_E - \epsilon\}} < \infty.\end{aligned}$$

But after proper normalization and centralization, the form of limits

$$\lim_{n\to\infty} F^n(a_n x + b_n) = \lim_{n\to\infty} Pr\{\frac{M_n - b_n}{a_n} \leq x\} = H(x) \tag{2.2}$$

becomes meaningful and there is a limit law known as *Three Types Theorem* for this.

**Theorem 2.1.1** (Three Types Theorem)**.** *If there exist normalizing constants* $a_n > 0$,$b_n \in \Re$*, and some non-degenerate df* $H$ *such that*

$$a_n^{-1}(M_n - b_n) \xrightarrow{d} H \tag{2.3}$$

*then* $H$ *belongs to the type of one of the following three dfs;*

*Gumbel:*

$$\Lambda(x) = e^{-e^{-x}}, x \in \Re. \tag{2.4}$$

*Fréchet:*

$$\Phi_\alpha(x) = \begin{cases} 0 & x \leq 0 \\ & \alpha > 0 \\ e^{-x^{-\alpha}} & x > 0 \end{cases} \tag{2.5}$$

*Weibull:*

$$\Psi_\alpha(x) = \begin{cases} e^{-|x|^\alpha} & x \leq 0 \\ & \alpha > 0 \\ 1 & x > 0 \end{cases} \tag{2.6}$$

The detailed proof of this theorem, issues about maximum domains of attraction and other related topics, can be found in Leadbetter et al.(1983) or Resnick(1987). And we say such $F$ (or $X$) belongs to the maximum domain of attraction of $H$ and write $F \in MDA(H)$.

The following theorems indicate the way to find such distribution functions.

**Theorem 2.1.2.** *Let $0 \leq \tau \leq \infty$ and suppose that for suitable normalizing constants $a_n > 0$ and $b_n$, $u_n(x) = a_n x + b_n$ such that*

$$n(1 - F(u_n)) \to \tau \qquad \text{as } n \to \infty \tag{2.7}$$

*then*

$$Pr\{M_n \leq u_n\} \to e^{-\tau} \qquad \text{as } n \to \infty \tag{2.8}$$

*Conversely, if (2.8) holds for some $\tau$, $0 \leq \tau \leq \infty$, then (2.7) holds.*

**Theorem 2.1.3.** *Necessary and sufficient conditions for the distribution $F$ belongs to the MDA of*

*(i) Gumbel: $\int_0^\infty (1 - F(u))du < \infty$*

$$\lim_{t \uparrow X_E} \frac{1 - F(t + xg(t))}{1 - F(t)} = e^{-x}$$

*for all real $x$, where*

$$g(t) = \frac{\int_t^{X_E} (1 - F(u))du}{1 - F(u)}$$

*for $t < X_E$.*

*(ii) Fréchet: $X_E = \infty$ and*

$$\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}$$

*$\alpha > 0$, for each $x > 0$.*

*(iii) Weibull: $X_E < \infty$ and*

$$\lim_{h \downarrow 0} \frac{1 - F(X_E - xh)}{1 - F(X_E - h)} = x^{\alpha}$$

*$\alpha > 0$, for each $x > 0$.*

There are some other results for finding $MDA(H)$ of $F$ and the normalizing constants in Leadbetter et al.(1983) and Resnick(1987). Consider a simple example here, suppose we

have the Pareto distribution

$$F(x) = 1 - \kappa x^{-\alpha}, \alpha > 0, \kappa > 0, x \geq \kappa^{1/\alpha}.$$

then we calculate

$$\frac{1 - F(tx)}{1 - F(t)} = \frac{(tx)^{-\alpha}}{t^{-\alpha}} = x^{-\alpha}$$

which indicates $F$ belongs to $MDA$ of a Fréchet type of extreme value distribution. By setting

$$n(1 - F(u_n)) = \tau$$

where $u_n(x) = a_n x + b_n$, we have

$$Pr\{(\kappa n)^{-1/\alpha} M_n \leq x\} \to \exp(-x^{-\alpha})$$

therefore,

$$a_n = (\kappa n)^{-1/\alpha}, \ b_n = 0. \tag{2.9}$$

The extreme value distributions lead to max-stable distributions, that is, distributions for which $H^n(a_n x + b_n) = H(x)$ holds for some constants $a_n > 0$, and $b_n$ for each $n \in \mathbb{N}$. Theorem 2.1.4(Leadbetter et al. 1983) shows the relationship between extreme value distributions and max-stable distributions.

**Theorem 2.1.4.** *Every max-stable distribution is of extreme value type, i.e. equal to $H(ax+b)$ for some $a > 0$ and $b$ and $H$ that is one of (1.4), (1.5) or (1.6); conversely, each distribution of extreme value type is max-stable.*

The *Generalized Extreme Value*(GEV) distribution is a combination of these three types:

$$H(x) = \exp\left\{-(1 + \xi\frac{x - \mu}{\psi})_+^{-1/\xi}\right\}, \tag{2.10}$$

where $\mu$ is a location parameter, $\psi > 0$ is a scale parameter and $\xi$ is a shape parameter. The limit $\xi \to 0$ corresponds to the Gumbel type, $\xi > 0$ and $\xi < 0$ correspond to Fréchet type and Weibull type with $\alpha = 1/\xi$ and $\alpha = -1/\xi$ respectively.

Furthermore, the case $\xi > 0$ refers to the "long-tailed" case for which $1 - H(x) \propto x^{-1/\xi}$, $\xi \to 0$ refers to the "medium-tailed" case for which $1-H(x)$ decreases exponentially for large $x$,i.e., $\exp\left(-\frac{x-\mu}{\psi}\right)$ and $\xi < 0$ refers to the "short-tailed" case for which the distribution has a finite endpoint at $x = \mu - \psi/\xi$.

Now suppose $\{X_i, i = 1, 2, ...\}$ are a stationary sequence with continuous marginal distribution function $F(x)$ and $\{\hat{X}_i, i = 1, 2, ...\}$ be the associated sequence of i.i.d. random variables with same marginal distribution function $F(x)$. Define

$$\begin{aligned} M_n &= max\{X_1, ..., X_n\} \\ \hat{M}_n &= max\{\hat{X}_1, ..., \hat{X}_n\}, \end{aligned}$$

then the limiting distribution of $M_n$ and $\hat{M}_n$ are associated with each other via a quantity $\theta$ defined as follows: $\theta$ is called the *extremal index*(Leadbetter, 1983) of a sequence $\{X_n\}$ if for every $\tau > 0$, there exists a sequence of thresholds $\{u_n\}$ such that

$$Pr\{\hat{M}_n \leq u_n\} \to e^{-\tau} \tag{2.11}$$

and under quite mild additional conditions,

$$Pr\{M_n \leq u_n\} \to e^{-\theta\tau}. \tag{2.12}$$

The index $\theta$ can be any real number between 0 and 1, and $\frac{1}{\theta}$ is the mean cluster size of exceedance over some threshold. For instance, if $\theta$ is close to 0, it corresponds to a strong dependence because the cluster size is very large. On the other hand, if $\theta$ tends to 1, then it indicates convergence towards asymptotic independence of extremes but meanwhile, the random variables in the original sequence do not have to be independent.

### 2.1.2 Exceedances Over Thresholds

Consider the excess distribution of $X$ over some high threshold $u$ and let $Y = X - u > 0$,

$$F_u(y) = Pr\{Y \leq y | Y > 0\} = \frac{F(u+y) - F(u)}{1 - F(u)}. \tag{2.13}$$

As $u \to \omega_F = \sup\{x : F(x) < 1\}$, we are able to get a limit

$$F_u(y) \approx G(y; \sigma_u, \xi), \tag{2.14}$$

G is called *Generalized Pareto Distribution*(GPD) and has the following form:

$$G(y; \sigma, \xi) = 1 - (1 + \xi \frac{y}{\sigma})_+^{-1/\xi}. \tag{2.15}$$

The case $\xi > 0$ corresponds to "long-tailed" distributions; the case $\xi \to 0$ is $1 - e^{-x/\sigma}$ and the case $\xi < 0$ has finite upper endpoint at $-\sigma/\xi$. The rigorous connection between classical extreme value theory and generalized Pareto distribution was established by Pickands(1975), who proves $F_u(y/\sigma_u) \to 1 - (1 + \xi y)^{-1/\xi}$ for all $y$ for which $1 + \xi y > 0$. As a result, the limit results of sample maxima and limit results for exceedances over thresholds are quite parallel.

### 2.1.3 Poisson-GPD model for exceedances

The Poisson-GPD model is another useful approach in extreme value theory. It is a limiting form of the joint process of exceedance times and excesses over the threshold. Let $N$ be the number of exceedances of the level $u$ in any one unit of time, then $N \sim Poisson(\lambda)$. Conditionally on $N \geq 1$, the excess values $Y_1, Y_2, ..., Y_N$ are i.i.d. from GPD with parameters $(\sigma, \xi)$. This is how the Poisson-GPD model is defined.

Now we are able to find the relationship between GEV distribution and Poisson-GPD process by calculating

$$\begin{aligned} Pr\{\max_{1 \leq i \leq N} Y_i \leq x\} &= P\{N = 0\} + \sum_{n=1}^{\infty} Pr\{N = n, Y_1 \leq x, ..., Y_n \leq x\} \\ &= e^{-\lambda} + \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{n} \left[1 - (1 + \xi \frac{x-u}{\sigma})_+^{-1/\xi}\right]^n \\ &= \exp\left\{-\lambda \left(1 + \xi \frac{x-u}{\sigma}\right)_+^{-1/\xi}\right\}. \end{aligned} \tag{2.16}$$

By substituting

$$\sigma = \psi + \xi(u - \mu), \qquad \lambda = \left(1 + \xi\frac{u-\mu}{\psi}\right)^{-1/\xi}, \tag{2.17}$$

(2.16) reduces to the GEV form (2.10). This proves the GPD model and GEV model are consistent with each other above a certain threshold $u$, and (2.16) gives out an explicit relationship between the two set of parameters.

### 2.1.4 Statistical Approaches

#### Peaks Over Threshold

The *Peaks Over Threshold*(POT) model was originally developed by hydrologists. As currently applied, the basic idea of POT is to fix a high threshold $u$ and then fit the GPD to exceedances over the threshold. For datasets that with serial correlation, the threshold exceedances occur in clusters and this method can be directly applied to the peak values within each cluster.

There are also some extensions of the basic POT model including selecting the threshold, incorporating covariates and dependence in the time series. In environmental process, the probability of an extreme event typically varies over a year. In order to take such seasonality into account, one may try(Smith 2003):

1. Remove seasonal tread before threshold approach;
2. Use different Poisson-GPD model for different season;
3. Add some covariates to Poisson-GPD model.

#### Point Process Approach

The *Point Process Approach* of extreme value theory is to combine the times at which high threshold exceedances occur and the excess values over the threshold into one process. The estimation method was first proposed by Smith(1989) and it has subsequently been described in some detail by Coles (2001, chapter 7) as well as Smith (2003). This context theoretically relies on the point process presentation of extreme values which was developed by Leadbetter, Lindgren and Rootzén(1983).

In this approach, under suitable normalization, the process behaves like a *nonhomogeneous Poisson process*. In general, a nonhomogeneous Poisson process is defined as the following. Suppose $A$ is a measurable subset of some domain $D$, let $N(A)$ denote the number of points in A. With some intensity function $\lambda(x), x \in D$, if $N(A)$ has a Poisson distribution with mean

$$\Lambda(A) = \int_A \lambda(x)dx.$$

for every measurable set $A$, then $N$ is a nonhomogeneous Poisson process.

For the EVT application, assume $x$ is two-dimensional with time $t$ and excess value $y$ over high threshold $u$, $D = [0, T] \times [u, \infty]$. Write

$$\lambda(t, y) = \frac{1}{\psi}\left(1 + \xi\frac{y-\mu}{\psi}\right)_+^{-1/\xi-1}. \tag{2.18}$$

If $A = [t_1, t_2] \times [y, \infty]$, then

$$\Lambda(A) = (t_2 - t_1)\left(1 + \xi\frac{y-\mu}{\psi}\right)_+^{-1/\xi} \tag{2.19}$$

To fit this model, we note that if a nonhomogeneous process of intensity $\lambda(t, y)$ is observed on a domain $D$, and if $(T_1, Y_1), ..., (T_N, Y_N)$ are the $N$ observed points of the process, then the joint density is

$$\prod_{i=1}^{N} \lambda(T_i, Y_i) \cdot \exp\left\{-\int_D \lambda(t, y)dtdy\right\} \tag{2.20}$$

(2.20) may be treated as a likelihood function and maximized with respect to those unknown parameters.

As an extension of this approach, the parameters $\mu, \psi, \xi$ are allowed to be time-dependent. In homogenous case where $\mu, \psi, \xi$ are constants, the model is just the same as Poisson-GPD model with different parametrization.

## 2.2 Spatial Modeling

Spatial modeling is used to analyze data with spatial dependence. Based on current continuous or discrete observations on spatial locations, one will be able to make predictions on other locations by constructing spatial models under appropriate assumptions of spatially related structures. The literature of spatial statistics is very substantial, the standard reference include the books written by Ripley(1981,1988), Cressie(1993) and Stein(1999). There are also other specific references by Smith(2000) on environmental statistics and Yun and Smith(2001) on spatial extremes.

In this section, we will try to give an introductory review of spatial statistics, especially in the discipline of geostatistics.

### 2.2.1 Geostatistics and Kriging

Applications of spatial statistics cover many fields, but a lot of the original contexts of this subject are arisen from geostatistics. Among all kinds of modeling techniques, *Kriging*, as an optimal least square interpolation method over a random spatial field, is widely used. It was first proposed by Matheron(1963) and named after D.G. Krige, a South African mining engineer who developed the idea in the 1950's. The following results are mostly taken from "Spatial statistics in environmental science", a chapter written by Smith(2000) in Fitzgerald, Smith, Walden and Young's book "Nonlinear and Non-stationary Signal Processing".

#### Spatial Processes and Variogram

Consider a spatial process represented by $\{Z(s), s \in D\}$ where $D$ is a subset of $R^d$. Suppose we have a finite number of observations at $\{s_i, i = 1, ..., n\}$ and $z_i = Z(s_i)$. Given these $z_i, i = 1, ..., n$, predict $z_0 = Z(s_0)$ for some location $s_0$ other than $s_1, ...s_n$, this is how the classical "kriging" problem has been defined. If the problem can be solved, then it is easily extended to problems such as jointly making predictions on several unobserved sites or estimating other quantities like $\int_A Z(s)ds$ for $A \subseteq D$.

There are two traditional types of models for $z_i$:

$$z_i = \mu + \epsilon_i, \tag{2.21}$$

$$z_i = x_i^T \beta + \epsilon_i, \tag{2.22}$$

where in each case, $\{\epsilon_i\}$ are the "noise" which can be represented as some spatially correlated process with zero-mean. In (2.21), $\mu$ is an undetermined parameter supposed to be constant all over $D$ and it is described as the "ordinary kriging"(Matheron, 1971; Journel and Huijbregts, 1978). In (2.22), the mean at a specific location depends on a given set of covariates $x_i$ through a linear regression model with parameters $\beta$ to be determined. This is always recognized as "universal kriging"(Matheron, 1963).

The covariance structure of $\epsilon(\cdot)$ can be represented as

$$Cov\{\epsilon(s), \epsilon(s')\} = C(s, s'), \tag{2.23}$$

for some covariance function $C(\cdot, \cdot)$.

The process $Z$ is *strictly stationary* if the joint distribution of $(Z(s_1), Z(s_2), ..., Z(s_k))$ is the same as that of $(Z(s_1 + h), Z(s_2 + h), ..., Z(s_k + h))$ for any $k$ spatial locations $s_1, s_2, ..., s_k$ and any $h \in R^d$ with $s_1, s_2, ..., s_k, s_1 + h, ..., s_k + h \in D$. The process $Z$ is called *weakly stationary* if

$$Cov\{\epsilon(s), \epsilon(s')\} = C_0(s - s'), for\ all\ s, s' \in D. \tag{2.24}$$

Instead of focusing on covariance functions, it is more common in spatial statistics to work with *semivariogram function* $\gamma(\cdot)$, which is defined as following:

$$Var\{\epsilon(s) - \epsilon(s')\} = 2\gamma(s - s'). \tag{2.25}$$

The left hand side of (2.25) is defined as the *variogram*. If such $\gamma(\cdot)$ exists, the spatial process $Z$ is called *intrinsically stationary*. One reason for considering this variogram rather than the covariance function is because (2.25) is a weaker property comparing with (2.24).

If $\gamma(s - s')$ only depends on $||s - s'||$, i.e. the scalar distance between $s$ and $s'$, the

process is said to be *isotropic*. A spatial process is *homogeneous* if it is both stationary and isotropic.

**Kriging**

The model (2.22) can be rewritten as:

$$Z = X\beta + \epsilon \tag{2.26}$$

where $Z^T = (z_1\ z_2\ ...\ z_n)$, $X$ is $n \times p$ matrix of covariates, $\beta$ is a $p$-dimensional vector of unknown regression coefficients and $\epsilon$ is a vector of random errors with zero-mean and covariance matrix of the form $C = \alpha\Gamma$. Here $\alpha > 0$ is an unknown positive scalar while the matrix $\Gamma$ is supposed to be known or estimated. This is recognized as universal kriging, but ordinary kriging is just a special case of universal kriging with letting $X\beta = \mathbf{1}\mu$, where $\mathbf{1}$ is an $n$-dimensional vector of 1s and $\mu$ is a unknown constant.

Under this model, we would like to predict:

$$z_0 = x_0^T\beta + \epsilon_0, \tag{2.27}$$

where $x_0$ is a new location and $\epsilon_0$ is a random variable with zero-mean, variance $\alpha\gamma_0$ and covariance with $\epsilon$ given by $E\{\epsilon^T\epsilon_0\} = \alpha\omega^T$ with known scalar $\gamma_0$ and vector $\omega$. Now consider a linear predictor $\hat{z}_0 = \lambda^T Z$ where $\lambda$ satisfies the constraint

$$X^T\lambda = x_0, \tag{2.28}$$

The constraint (2.28)is used for justifying the reduction

$$\hat{z}_0 - z_0 = \lambda^T(X\beta + \epsilon) - x_0^T\beta - \epsilon_0 = \lambda^T\epsilon - \epsilon_0,$$

which means the prediction doesn't depend on the unknown regression coefficients $\beta$.

The way to find the prediction is as follows: minimize $\mathfrak{Q} = E\{(\hat{z}_0 - z_0)^2\}$ under the constraint (2.28). The solution $\lambda$ only depends on the first and second order moments of $\hat{z}_0$ and is not related with any normality assumption. Using the method of Lagrange

multipliers, $\lambda$ can be explicitly written as:

$$\lambda^T = \{\left(x_0 - X^T\Gamma^{-1}\omega\right)^T \left(X^T\Gamma^{-1}X\right)^{-1} X^T + \omega^T\}\Gamma^{-1} \tag{2.29}$$

**Covariance matrix $\Gamma$**

Consider the most common situation that the spatial process is stationary and isotropic. There are typically several steps to estimate the covariance matrix.

1. Determine the shape of covariance function(or variogram).

2. According to the first step, choose from a family of positive-definite covariance functions(or variogram).

   In practice, there are a lot of standard families of covariance functions(or variograms). For example:

   - *Spherical model:*

$$\gamma_0(t) = \begin{cases} 0, & t = 0, \\ c_0 + c_1\{\frac{3t}{2R} - \frac{1}{2}\left(\frac{t}{R}\right)^3\}, & 0 < t \leq R, \\ c_0 + c_1, & t > R, \end{cases} \tag{2.30}$$

   - *Exponential model:*

$$\gamma_0(t) = \begin{cases} 0, & t = 0, \\ c_0 + c_1\left(1 - e^{-t/R}\right), & t > 0, \end{cases} \tag{2.31}$$

   - *Gaussian model:*

$$\gamma_0(t) = \begin{cases} 0, & t = 0, \\ c_0 + c_1\left(1 - e^{-(t/R)^2}\right), & t > 0, \end{cases} \tag{2.32}$$

- *Matérn model:*

$$C_0(t) = \frac{1}{2^{\nu-1}\Gamma_0(\nu)} \left(\frac{2\sqrt{\nu}t}{R}\right)^{\nu} K_\nu \left(\frac{2\sqrt{\nu}t}{R}\right), \tag{2.33}$$

where $\nu > 0$ is a shape parameter and $K_\nu$ is the modified Bessel function of the third kind of order $\nu$(Abramowitz and Stegun 1964). The special cases $\nu = \frac{1}{2}$ and $\nu \to \infty$ correspond to the exponential model and Gaussian model, respectively.

3. Estimate the parameters of the assumed covariance function(or variogram).

   Cressie(1993) proposed a simple but efficient way for this by using a weighted least squares technique. There are also other alternative methods based on likelihood procedure including maximum likelihood estimation(Kitanidis 1983, Mardia and Marshall 1984), restricted maximum likelihood estimation(Kitanidis 1983, Cressie 1993) and Bayesian methods(Le and Zidek 1992, Handcock and Stein 1993, Brown *et al.* 1994, Banerjee, S., Carlin, B.P. and Gelfand, A.E.(2004)).

4. Application to kriging.

   After the covariance matrix $\Gamma$ has been estimated, it is treated as known feature for solving the kriging problem. However, it is logical to adapt the kriging procedure with allowing unknown parameters in $\Gamma$. There are two possible approaches for this adaption:

   (a) *Corrections based on the delta method.* It is possible to use Taylor expansions of the covariance matrix and correct the prediction. The detailed method was proposed by Zimmerman and Cressie(1992).

   (b) *Fully Bayesian approach.* Treat $(\beta, \theta, z_0)$(where $\beta$ is the regression coefficient, $\theta$ are the unknown parameters of the covariance matrix, and $z_0$ is the predictor) as a random vector and solve the conditional density of $z_0$ given observed data $Z$ after integrating out $\beta$ and $\theta$. The standard kriging is just a special case of this when assuming $\theta$ are known and the prior distribution of $\beta$ is uniform.

### 2.2.2 Generalized Linear Mixed Modeling

The limitation of traditional kriging methods is that the predictions are always made under Gaussian assumptions. For non-Gaussian cases, the most widely implemented methodology is trans-Gaussian kriging(Cressie 1993) which applies the standard Gaussian methods after a marginal non-linear transformation, i.e. analyze the transformed data $\Phi(z_i), i = 1, ..., n$ for some specified function $\Pi(\cdot)$ instead of analyzing $z_i, i = 1, ..., n$ directly.

Diggle et al.(1998) embedded the linear kriging(and trans-Gaussian kriging) in a more general distributional framework, which is comparable with embedding the Gaussian linear model into a generalized linear mixed model.

Suppose $Z = \{Z(s) : s \in D\}$ denotes a spatial process, $D$ is a subset of $R^d$. Let $\{y_i, i = 1, ..., n\}$ be the observations at locations $(s_i, i = 1, ..., n)$,then $\{y_i, i = 1, ..., n\}$ can be treated as a function of $\{z_i = Z(s_i)\}$ plus some noise.

Now assume:

1. $Z$ is a stationary Gaussian process with $E[Z(s)] = 0$ and $cov\{Z(s), Z(s')\} = \alpha\Gamma(s-s')$;

2. Conditionally on Z, the observations $y_i, i = 1, ..., n$ are independent with distributions $f_i\{y|z_i\} = f(y; M_i)$ where $M_i = E[y_i|z_i]$;

3. $h(M) = KZ + \eta$, for some known link function h, vector of normal errors $\eta$ and location mapping matrix $K$.

Let $S^* = (s_1^* \ ... \ s_m^*)$ denote the locations we want to make predictions and $K_0$ the corresponding location mapping matrix, then $\hat{Z}(S^*) = E[Z(S^*)|y_1, ..., y_n]$ is the *generalized linear predictor* for $Z(S^*)$.

Suppose $g_n$ and $g_{m+n}$ denote multivariate normal distributions with dimensions $n$ and $m + n$,respectively. The unconditional distribution of $Y = (y_1 \ y_2 \ ... \ y_n)$ is given by the integral

$$f(Y) = \int f(Y|Z)g_n(KZ)dZ_1dZ_2...dZ_n \tag{2.34}$$

and the unconditional joint distribution of $(Y, K_0Z)$ is given by

$$f(Y, K_0Z) = \int f(Y|Z)g_{m+n}(KZ, K_0Z)dZ_1dZ_2...dZ_n \tag{2.35}$$

The conditional density of $K_0Z$ given $y_i, i = 1, ..., n$ is then the ratio of (2.35) and (2.34), so one will be able to produce predictions of Gaussian processes on $S^*$. The predictions $\hat{Z}(S^*)$ therefore can be reached after using some sampling technique.. Due to the fact that n is very large in many statistical problems, it is difficult to find the explicit form of $f(Y)$. A straightforward idea is to deploy Markov Chain Monte Carlo method. But since we have a high dimensional covariance structure, the convergence of MCMC is always very slow.

This approach was first discussed by Diggle, Tawn and Moyeed(1998).

### 2.2.3 Spectral Representation of Covariance Structures

To improve the computational efficiency, Wikle(2002) proposed an alternative basis-function representation of the Gaussian random field $Z$. Under his representation, one will be able to get a simpler form of covariance matrix(asymptotically diagonal), and therefore, have more easily computable MCMC processes to deal with.

Specifically, Wikle assumes we can write

$$Z = \Psi\alpha \tag{2.36}$$

where $\Psi$ (a $n \times p$ matrix) is some known set of $p$ basis functions evaluated at the $n$ spatial coordinates, and $\alpha$ is a set of coefficients, distributed as $\alpha \sim N[0, \Sigma_\alpha(\theta)]$ where $\Sigma_\alpha(\theta)$ is a known covariance matrix dependent on a finite-dimensional unknown parameter $\theta$.

The two representations are equivalent if

$$\Sigma_Z = \Psi\Sigma_\alpha(\theta)\Psi^T \tag{2.37}$$

and under some circumstances it might be appropriate to find $\Psi$ and $\Sigma_\alpha(\theta)$ so that (2.37) holds exactly. However this is also not feasible for large grids, and Wikle proposes approximations to (2.37).

In particular, if $Z$ is a stationary process on a regular lattice, it is possible to define $\Psi$ as an $n \times n$ array of Fourier coefficients so that $\alpha$ is the discrete Fourier transform of $Z$ and (2.36) is the inverse Fourier transform. This can greatly ease computation since if $n$ is a power of two, the calculation of (2.36) is an application of the FFT.

As described by Wikle, the matrix $\Sigma_\alpha(\theta)$ is asymptotically a diagonal matrix whose diagonal entries are determined by the spectral density of $Z$. For example, consider the Matérn class of covariance functions,

$$C(r) = \theta_3(\theta_1 r)^{\theta_2} K_{\theta_2}(\theta_1 r), \theta_1 > 0, \theta_2 > 0, \theta_3 > 0$$

where $K_{\theta_2}$ is modified Bessel function, $\theta_1$ and $\theta_2$ reflect the correlation range and the degree of smoothness of a spatial process, respectively, and $\theta_3$ is proportional to the variance of the process(Stein, 1999). This is equivalent to (2.33) under different notations. The corresponding spectral density function at frequency $\omega$ is then given by

$$f(\omega;\theta) = \frac{2^{\theta_2-1}\theta_3\Gamma(\theta_2+g)\theta_1^{2\theta_2}}{\pi^{g/2}(\theta_1^2+\omega^2)^{\theta_2+g/2}} \tag{2.38}$$

where $g$ is the dimension of the process. Thus, we can assume that $\Sigma_\alpha(\theta)$ is a diagonal matrix where the diagonal entries that correspond to Fourier components of frequency $\omega$ are given by (2.38).

After the spectral decomposition, we have an approximated covariance matrix which can be easily updated through MCMC iterations. In fact, the covariance matrix of $\theta$ and $\alpha$ are both diagonal which let us avoid lots of computations on inverse of large dimensional matrices. And therefore, the computation is much more efficient than Diggle's model.

### 2.2.4 Exact Statistical Approach of Stationary Gaussian Processes

As for high dimensional covariance structures, the spectral method provides efficient approximations, but it is still possible to consider an exact approach without losing much of the computational efficiency. Wood and Chan(1994) first proposed a method for simulating stationary Gaussian processes on a rectangular grid in $[0,1]^d \subset \Re^d$. The following sections will briefly describe how the method works.

#### Toeplitz Matrix and Circulant Matrix

A Toeplitz matrix is a matrix in which each descending diagonal from left to right is constant. It can be defined as follows:

Suppose T is a $n \times n$ matrix with $T = [t_{k,j}; k, j = 0, 1, 2, \ldots, n-1]$ where $t_{k,j} = t_{k-j}$, i.e. a matrix of the form

$$T = \begin{pmatrix} t_0 & t_{-1} & t_{-2} & \cdots & t_{-(n-1)} \\ t_1 & t_0 & t_{-1} & & \\ t_2 & t_1 & t_0 & & \vdots \\ \vdots & & & \ddots & \\ t_{n-1} & & & \cdots & t_0 \end{pmatrix} \tag{2.39}$$

A Circulant matrix $C$ is a special Toeplitz matrix which having the form

$$C = \begin{pmatrix} c_0 & c_1 & c_2 & & \cdots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & c_2 & & \vdots \\ & c_{n-1} & c_0 & c_1 & \ddots & \\ \vdots & \ddots & \ddots & \ddots & & c_2 \\ & & & & & c_1 \\ c_1 & \cdots & & & c_{n-1} & c_0 \end{pmatrix} \tag{2.40}$$

And the eigen-decomposition of $C$ can be easily obtained by Fourier Transform. Define

$$F = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega^1 & \omega^2 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(n-1)} \\ \vdots & \cdots & \cdots & & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)(n-1)} \end{pmatrix}, \tag{2.41}$$

where $\omega = \exp -2\pi i/n$. Then $F$ is unitary($F^{-1} = F^*$) and we have the following theorem:

**Theorem 2.2.1.** *If $C$ is a circulant matrix, then $C$ is diagonalized by $F$. More specifically,*

$$C = F^* \Lambda F, \tag{2.42}$$

*where $\Lambda$ is the diagonal matrix consists of the eigenvalues of $C$.*

The eigenvalues of $C$ are therefore can be written as:

$$\lambda_k = \sum_{j=0}^{n-1} c_j e^{-2\pi\sqrt{-1}jk/n}, \quad \mathrm{k} = 0, 1, ..., n-1. \tag{2.43}$$

A general Toeplitz matrix can be easily embedded in a circulant matrix. For example we can extend $T$ as follows:

$$T \Rightarrow C_T = \begin{pmatrix} t_0 & t_{-1} & t_{-2} & \cdots & t_{-(n-1)} & t_{n-1} & \cdots & t_2 & t_1 \\ t_1 & t_0 & t_{-1} & \cdots & & t_{-(n-1)} & \cdots & & t_2 \\ t_2 & t_1 & t_0 & & & & & & \\ \vdots & & & & \ddots & & & \ddots & \\ t_{n-1} & t_{n-2} & t_{n-3} & \cdots & t_0 & t_{-1} & \cdots & & t_{-(n-1)} \\ \vdots & & & & & \ddots & \ddots & & \\ \vdots & & & & & & \ddots & \ddots & \\ t_{-2} & t_{-3} & t_{-4} & \cdots & t_{n-2} & & \cdots & t_0 & t_{-1} \\ t_{-1} & t_{-2} & t_{-3} & \cdots & t_{n-1} & t_{n-2} & \cdots & t_1 & t_0 \end{pmatrix} \tag{2.44}$$

where $C_T$ is now a circulant matrix. This is the way to extend a general Toeplitz matrix to a

circulant matrix with the least dimension, we can extend the matrix by adding more entries making the circulant matrix larger with better properties, for example, positive definite.

Now suppose $\{T_i, i = 0, \pm 1, \pm 2, \ldots, \pm(m-1)\}$ is a set of $n \times n$ Toeplitz matrices, then

$$\hat{T} = \begin{pmatrix} T_0 & T_{-1} & T_{-2} & \cdots & T_{-(m-1)} \\ T_1 & T_0 & T_{-1} & & \\ T_2 & T_1 & T_0 & & \vdots \\ \vdots & & & \ddots & \\ T_{m-1} & & & \cdots & T_0 \end{pmatrix} \tag{2.45}$$

is a matrix with block Toeplitz structure, and it can be embedded into a block circulant matrix. First, we can embed each of $T_i$ to a circulant matrix $C_i$ like (2.44), then a new block circulant matrix containing $\hat{T}$ can be obtained as follows:

$$\hat{T} \Rightarrow \hat{C} = \begin{pmatrix} T_0 & T_{-1} & T_{-2} & \cdots & T_{-(n-1)} & T_{n-1} & \cdots & T_2 & T_1 \\ T_1 & T_0 & T_{-1} & \cdots & & T_{-(n-1)} & \cdots & & T_2 \\ T_2 & T_1 & T_0 & & & & & & \\ \vdots & & & & \ddots & & & \ddots & \\ T_{n-1} & T_{n-2} & T_{n-3} & \cdots & T_0 & T_{-1} & \cdots & & T_{-(n-1)} \\ \vdots & & & & & \ddots & \ddots & & \\ \vdots & & & & & & \ddots & \ddots & \\ T_{-2} & T_{-3} & T_{-4} & \cdots & T_{n-2} & & \cdots & T_0 & T_{-1} \\ T_{-1} & T_{-2} & T_{-3} & \cdots & T_{n-1} & T_{n-2} & \cdots & T_1 & T_0 \end{pmatrix} \tag{2.46}$$

**The One-Dimensional Case**

Now consider a stationary Gaussian process $\{X(t) : t \in \Re\}$ with zero mean and covariance function $\gamma$. The purpose is to generate a random vector $X = \left(X(0), X\left(\frac{1}{n}\right), \cdots, X\left(\frac{n-1}{n}\right)\right)^T$.

We have $X \sim N_n[0, \Sigma]$, where

$$\Sigma = \begin{pmatrix} \gamma(0) & \gamma(\frac{1}{n}) & \cdots & \gamma(\frac{n-1}{n}) \\ \gamma(\frac{1}{n}) & \gamma(0) & \cdots & \gamma(\frac{n-2}{n}) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(\frac{n-1}{n}) & \gamma(\frac{n-2}{n}) & \cdots & \gamma(0) \end{pmatrix} \tag{2.47}$$

and $\Sigma$ has Toeplitz structure.

There are several steps to simulate the random vector $X$:

1. Embed $\Sigma$ in a positive definite circulant matrix $C$ with dimension $m = 2^g$ for some integer $g$ and $m \geq 2(n-1)$

2. Write $C = Q\Lambda Q^*$ where $\Lambda$ is the diagonal matrix of eigenvalues of $C$ which can be efficiently obtained by one-dimensional Discrete FFT, $Q$ is the corresponding unitary matrix and $Q^*$ is the conjugate transpose of $Q$.

3. Define $Y = Q\Lambda^{1/2}Q^*Z$, where $Z \sim N_m[0, I]$. Then $Y \sim N_m[0, C]$, and the random vector $X$ is a subvector of $Y$.

**The $n$-Dimensional Case**

The $n$-dimensional case is analogous to 1-dimensional case. After reordering the indices, the embedded positive definite matrix $C$ will have block circulant structure. For each dimension, a certain $m$ will be selected. Suppose $m[l]$ is corresponding to $l$th dimension and for the covariance function $\gamma$, if

$$\gamma(t[1], \ldots, t[k], \ldots, t[n]) = \gamma(t[1], \ldots, -t[k], \ldots, t[n]),$$

for all $(t[1], \ldots, t[n])^T \in \Re^n$, $\gamma$ is said to be even in $k$th coordinate. $\gamma$ has to be even in all coordinates in order to get the exact simulation. There is an alternative way proposed by Wood and Chan(1984)to handle the case $\gamma$ is not even in some coordinates, but it will be an approximation.

Other than the one dimensional case, we can use two-dimensional discrete Fourier trans-

form to get the superstructures of $\hat{C}$. And the eigenvalues of $\hat{C}$ turn to be:

$$\lambda(u') = \sum_{u \in I(m)} c(u) \exp\left\{-2\pi\sqrt{-1}u^T(\frac{u'}{m})\right\}, \quad u \in I(m) \tag{2.48}$$

where

$$I(m) = \{u = (i_1, i_2, ..., i_n) : 0 \leq i_1 \leq m[1] - 1, 0 \leq i_2 \leq m[2] - 1, ..., 0 \leq i_n \leq m[n] - 1\},$$

The Wood-Chan method cannot be directly used in Bayesian analysis because in Bayesian analysis, we are focusing on those posterior distributions and the covariance forms will be much more complicated. So an alternative approach will be discussed in later chapters.

CHAPTER 3

# North American Rainfall Extremes

## 3.1 Introduction

There is evidence that shows that variation in global surface temperature is significantly different from estimates of natural climate variability, and especially, that the unusual increase in global mean temperature during the latter half of the 20th century can be attributed to anthropogenic forcing. The early impacts of change on society and ecosystems are mostly felt via changes in climate variability and extreme climatic events. Therefore, it is very important to analyze changes in climatic extremes to determine whether these changes are statistically significant and to make predictions for future changes.

The increase in intense precipitation, rather than the change in mean precipitation, is consistent with changes expected in a warmer atmosphere due to an acceleration of the hydrological cycle(e.g. Trenberth 1999) and reflects the climate changes forced by scenarios of greenhouse gas and sulfate aerosol emissions. Recent discussion(Karl and Trenberth 2003) concludes even without any change in total precipitation, higher temperatures lead to a greater proportion of total precipitation in extreme events. Kharin and Zwiers(2000) found that North American rainfall extremes with 20-year return values are expected approximately every 10 years by the end of next century.

Observations of rainfall also support the relationship between extreme rainfall events and climate change. Karl and Knight (1998) found that the proportion of precipitation extremes has increased over the United States. Subsequent work has shown that changes in very heavy precipitation are positive in many regions(Groisman *et al.* 2004).

We focus on the United States because of the ready availability of quality-controlled

station data. We will be using three data sets in our analysis.The North American Rainfall Data collected by National Climatic Data Center(NCDC; Groisman et al.,2001) consists of rainfall values for the last 100 years. Data from United States are observed in about 6000 stations. The following two figures are from NCDC web site(`www.ncdc.noaa.gov`). Figure 3.1 shows the observational stations of NCDC all over the continental United States and Figure 3.2 shows the average annual precipitation for the period $1899 - 1999$.

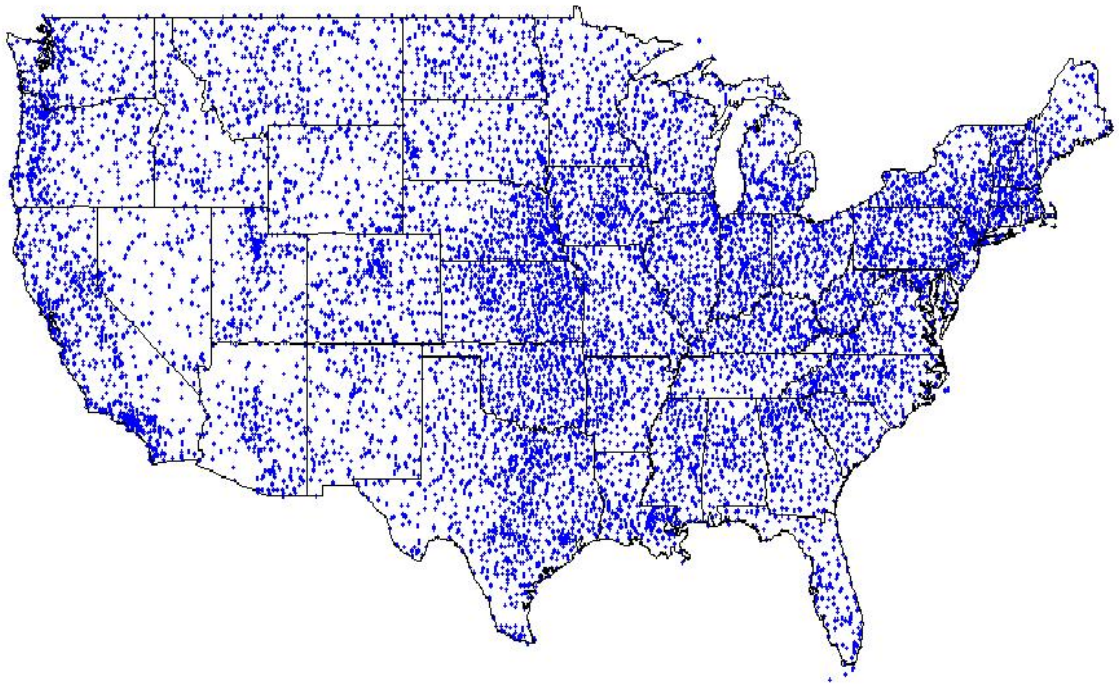


Figure 3.1: *Observational Stations of NCDC*

There is also a set of re-analysis data available which is provided by the National Center for Environmental Prediction(NCEP). A re-analysis is a reconstruction of the observed weather using forecasting models in which observational data define the boundary conditions. The third dataset we use is from the Community Climate System Model(CCSM), which is a global climate model (GCM). CCSM simply generates data that are representative of a particular set of climatic conditions - the output of CCSM is not correlated on a day to day basis with actual weather pattern, whereas the output of a re-analysis is. Re-analyses are useful in studies of this nature because they use similar grid cells to GCMs, but can

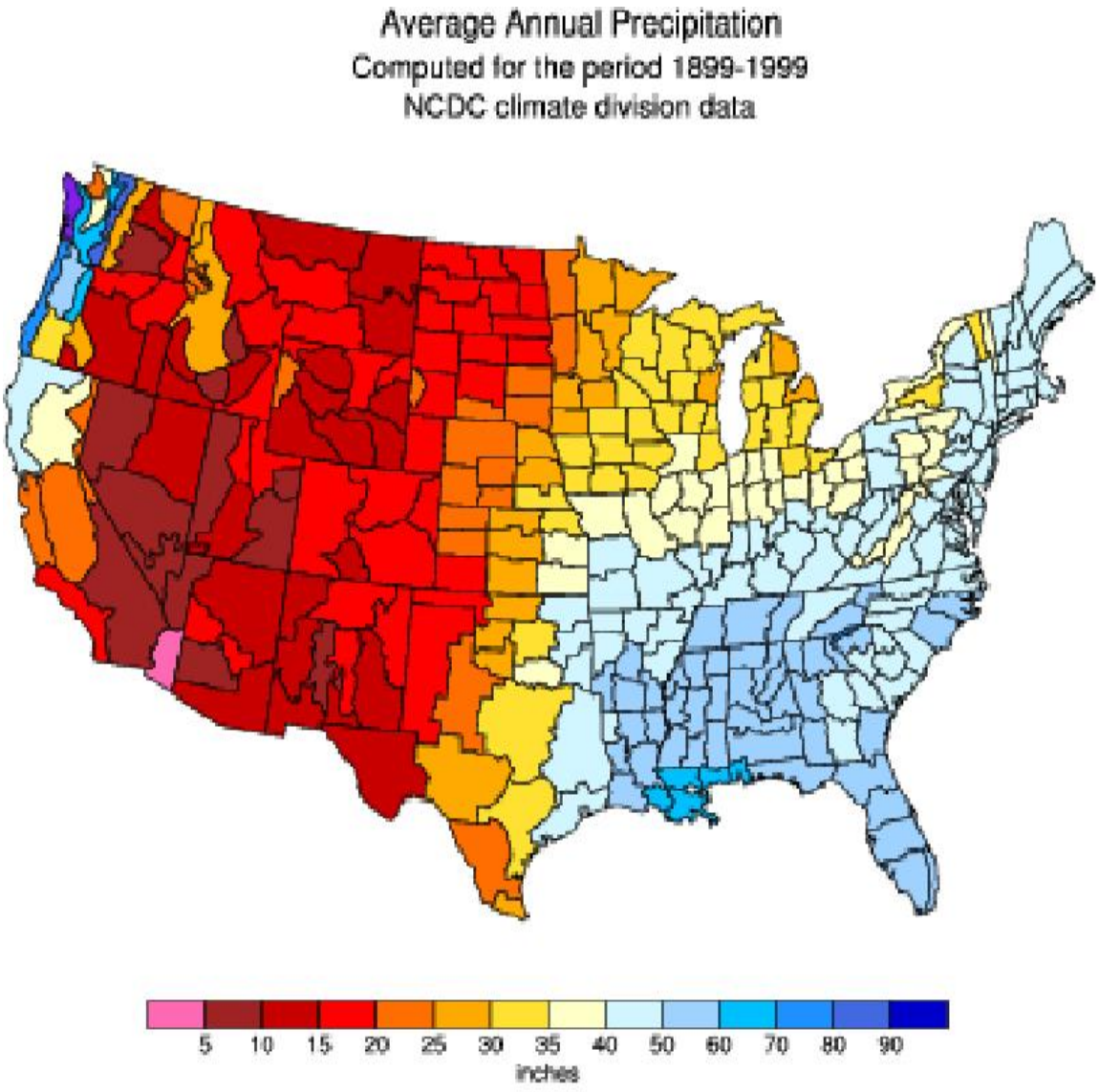



Figure 3.2: *Average Annual Precipitation(*1899 – 1999*, NCDC)*

also be compared directly with observational data, and therefore are useful for calibrating GCMs. In our study, we examine daily NCDC in one NCEP grid-cell and make predictions of rainfall amounts at locations other than the stations. These predictions are then aggregated to form predictions of the grid-cell average for each day, which can then be compared with the NCEP values. Figure 3.3 shows the grid cells used by NCEP(Smith, 2006, page 34). Having thus derived relationships between grid-cell and point-station values, we then apply these relationships to present and future CCSM grid-cell data.

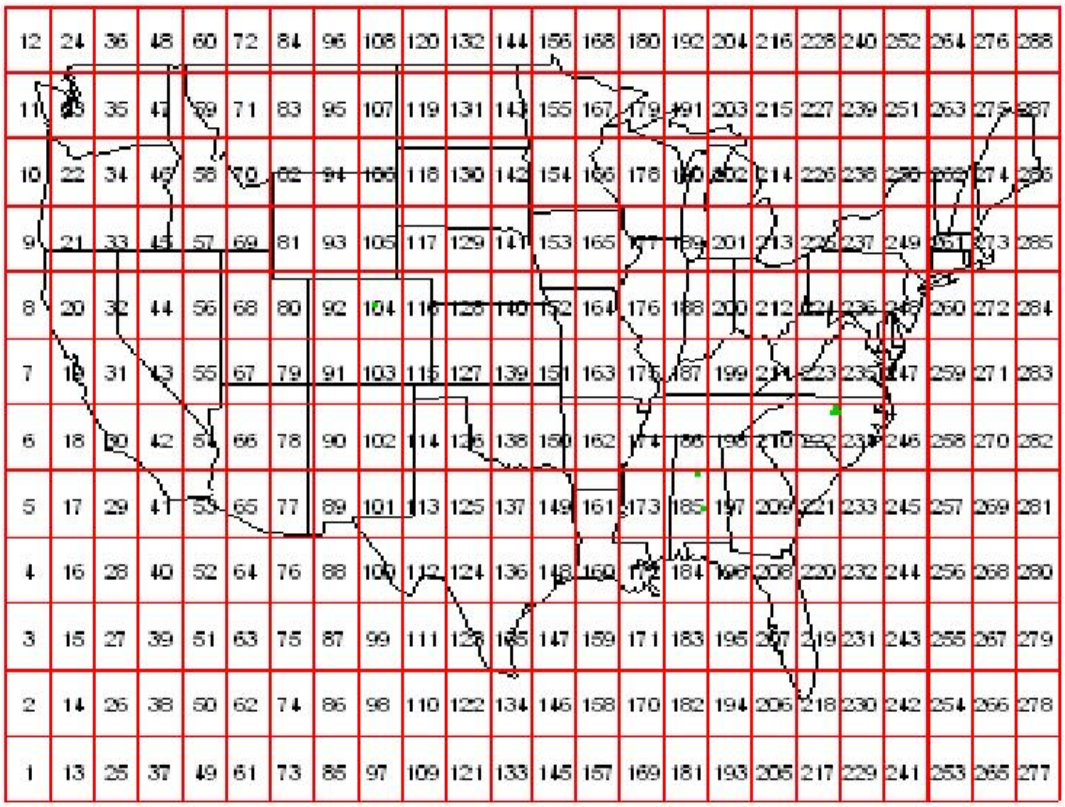



Figure 3.3: *NCEP Grid Cells(Continental United States)*

## 3.2 Statistics of Rainfall Extremes

There are several possible statistical approaches to analyze rainfall extremes, for example, estimating the distribution function for rainfall, using extreme value theory for

examining extreme events and indices for climate extremes. The first approach has a big disadvantage because the distribution fit is mostly determined to a large extent by the center of the distribution, which will not be able to reflect the extreme events very well. Meehl *et al.*(2000) and Frich *et al.*(2002) proposed and applied a large range of indices related to extreme events with "extreme" referring to events that can be observed at least once or several times every certain period(one year in general) on average. We will be using a more complete description of the statistics of extremes along with spatial interpolation and smoothing methods to simulate the North American rainfalls and make spatial(-temporal) predictions.

CHAPTER 4

# Generalized Linear Mixed Modeling on Rainfall Extremes

## 4.1 Generalized Linear Mixed Model

For United States rainfall data, we assume they follow a piecewise distribution. More specifically, at day n and location k,

- For the data exceed our threshold $u$, we assume $y$ follows POT approach, i.e.

$$f_{n,k}(y|\mu_{n,k}) \sim \frac{1}{\phi_{n,k}}\left(1+\xi_{n,k}\frac{y_{n,k}-\mu_{n,k}}{\phi_{n,k}}\right)_{+}^{-1/\xi_{n,k}-1}\exp\left\{-\left(1+\xi_{n,k}\frac{u-\mu_{n,k}}{\phi_{n,k}}\right)^{-1/\xi_{n,k}}\right\}$$

  where $\mu_{n,k}$ is a location parameter, $\xi_{n,k}$ is a shape parameter and $\phi_{n,k}$ is a scale parameter.

- Estimate $F_{n,k}(0)$ by the sample proportion of zeros.

- For $0 < y \leq u$(threshold), divide the range into $T$ equiprobable intervals(depending on the number of data we have in this interval) and assume $F_{n,k}(y)$ is piecewise linear within each interval.

Due to the huge observational data set we have from NCDC, it is more realistic to do one NCEP grid cell at a time. The NCEP grid cells are of dimensions 2.5° latitude × 2.5° longitude, centered on latitude-longitude coordinates that are multiples of 2.5°. We are going to take the grid cell that contains Raleigh-Durham airport(RDU) as our computational example. Raleigh-Durham airport, at latitude $35.87^{o}N$ and longitude $78.78^{o}W$, is in the grid cell centered at $35^{o}N$ and $80^{o}W$, which covers latitudes from $33.75^{o}N$ to $36.25^{o}N$ and

longitudes $78.75^oW$ to $81.25^oW$. Even within this one grid cell, there are 66 NCDC rainfall stations. We will only update the location parameters $\mu$ as the spatial parameter every day and use maximum likelihood estimations for the other two in one grid cell. Therefore $\mu_{n,k}$ is the only parameter considered as a noisy observation of a Gaussian process $Z_n$ on day n at location $k$.

Given that rainfall varies seasonally, it makes sense to break up the year into individual months, fitting a separate model to each month. We will be using 5 years of data at a time, that is approximately 150 days per analysis, which at 66 stations gives 9900 station-day combinations to be updated on each MCMC cycle. More than this, we would like to define a large value of $q = 30 \times 30$(number of grid points at which prediction is obtained) to ensure good spatial coverage on NCEP grid cell, which let us have $900 \times 150$ point-day combinations of Gaussian Processes to be updated at each MCMC iteration.

We are really talking about spatial-temporal data since we obviously don't want to analyze the data just one day at a time. But it is known with precipitation data that the temporal correlations are not very strong, and for this study, as a simplification, we are treating the days as independent so as to concentrate on the spatial dependence.

Now for the data exceeding threshold $u$, we have

$$f(y_{n,k}; \mu_{n,k}, \hat{\phi}, \hat{\xi}) \sim \frac{1}{\hat{\phi}} \left(1 + \hat{\xi}\frac{y_{n,k} - \mu_{n,k}}{\hat{\phi}}\right)_+^{-1/\hat{\xi}-1} \exp\left\{-\left(1 + \hat{\xi}\frac{u - \mu_{n,k}}{\hat{\phi}}\right)^{-1/\hat{\xi}}\right\} \quad (4.1)$$

Under this assumption, for each day $n$, two observations $y_{n,k}$ and $y_{n,k'}$ are independent conditional on location parameters $\mu_{n,k}$ and $\mu_{n,k'}$. The dependence among the data is then parameterized as a random effect. Now let us assume

$$\mu_{n,k} = \bar{\mu} + z_{n,k} + \eta_{n,k} \quad (4.2)$$

where $\bar{\mu}$ is the mean with some noise over grid cells. $z_{n,k}$ and $\eta_{n,k}$ are spatial related Gaussian processes and independent errors, respectively. The $\bar{\mu}$ could be written as a regression component such as $\bar{\mu}_{n,k} = \sum_{i=1}^{l} \beta_{n,i} x^i_{n,k}$ plus some noise with $x_{n,k}$ reflecting topography of the area and the model can be more generally used, but it is not our intention here.

## 4.2 Spectral Representation of Covariance Structure

To improve the computational efficiency, especially, to avoid calculating the inverse of large dimensional covariance matrices, we try to represent the covariance structure in a Fourier space. This method is based on the technique first described by Wikle (2002). He assumed the data are from exponential family distribution, but it is really not a necessary assumption for the model. We can write

$$Z_n = \Psi \alpha_n \tag{4.3}$$

where $\Psi$ is a $n \times n$ Fourier basis matrix and $\alpha_n$ is the corresponding vector of coefficients.

We will be using the Matérn class of covariance functions and the spectral density function at frequency $\omega$ is given by

$$f(\omega;\theta) = \frac{2^{\theta_2-1}\theta_3\Gamma(\theta_2+g)\theta_1^{2\theta_2}}{\pi^{g/2}(\theta_1^2+\omega^2)^{\theta_2+g/2}} \tag{4.4}$$

where $g$ is the dimension of the process. Thus, we can assume that $\Sigma_\alpha(\theta)$ is a diagonal matrix where the diagonal entries that correspond to Fourier components of frequency $\omega$ are given by (4.4). We assume $\theta_3 = 1$ and make up for the missing scale component by including a parameter $\gamma$ in the MCMC formulation that follows.

This leads us to the model

$$\begin{aligned}
\mu_{n,k}|Z &\sim N\left[\bar{\mu} + \gamma\sum_{j=1}^{q} k_{k,j} z_{n,j}, \frac{1}{\kappa_\eta}\right], \\
z_{n,j} &\sim N\left[\sum_{\ell=1}^{p}\psi_{j\ell}\alpha_{n,\ell}, \frac{1}{\kappa_\epsilon}\right], \\
\alpha_n &\sim N[0, \Sigma_\alpha(\theta)], \\
\bar{\mu} &\sim N\left[\bar{\mu}_0, \frac{1}{\kappa_\mu}\right], \\
\kappa_\eta &\sim Gamma[q_\eta, r_\eta], \\
\kappa_\epsilon &\sim Gamma[q_\epsilon, r_\epsilon], \\
\gamma &\sim \pi_\gamma(\gamma),
\end{aligned}$$

$$\theta \sim \pi_\theta(\theta).$$

where $n$ refers to day $n$, $K = [k_{i,j}]$ is the location matrix with entries 1 if there is a station at corresponding grid cell and entries 0 otherwise and $\Sigma_\alpha(\theta)$ is the asymptotic covariance matrix with (4.4) as diagonal entries .

Under this model, the joint density of $(\mu, \kappa_\eta, \kappa_\epsilon, \gamma, \theta, \alpha, Z, \phi, y)$ beyond threshold $u$ at day $n$ is proportional to

$$\begin{aligned}
&\exp\left\{-\frac{\kappa_\mu}{2}(\bar{\mu}-\bar{\mu}_0)^2\right\}\cdot\kappa_\eta^{q_\eta-1}e^{-r_\eta\kappa_\eta}\cdot\kappa_\epsilon^{q_\epsilon-1}e^{-r_\epsilon\kappa_\epsilon}\cdot\pi_\gamma(\gamma)\cdot\pi_\theta(\theta)\\
&\cdot|\Sigma_\alpha(\theta)|^{-1/2}\exp\left\{-\frac{1}{2}\alpha^T\Sigma_\alpha(\theta)^{-1}\alpha\right\}\cdot\kappa_\epsilon^{q/2}\exp\left\{-\frac{\kappa_\epsilon}{2}(Z-\Psi\alpha)^T(Z-\Psi\alpha)\right\}\\
&\cdot\kappa_\eta^{m/2}\exp\left\{-\frac{\kappa_\eta}{2}(\mu-\bar{\mu}\mathbf{1}-\gamma K^T Z)^T(\mu-\bar{\mu}\mathbf{1}-\gamma K^T Z)\right\}\\
&\cdot\prod_{k=1}^{m}\frac{1}{\hat{\phi}}\left(1+\hat{\xi}\frac{y_{n,k}-\mu_{n,k}}{\hat{\phi}}\right)_+^{-1/\hat{\xi}-1}\exp\left\{-\left(1+\hat{\xi}\frac{u-\mu_{n,k}}{\hat{\phi}}\right)^{-1/\hat{\xi}}\right\}. \qquad (4.5)
\end{aligned}$$

And for the data below threshold $u$, we have

$$Pr\{Y<y_{n,k}\}=F(y_{n,k})=\begin{cases} C_1\cdot\left(\frac{t-1}{T}+\frac{y_{n,k}-u_{t-1}}{u_t-u_{t-1}}\right), & 0<y_{n,k}\le u \text{ and } y_{n,k}\in(u_{t-1},u_t];\\ C_0, & y_{n,k}=0. \end{cases} \qquad (4.6)$$

where $C_0$ is the sample proportion of zeros, $C_1$ is the sample proportion of numbers between 0 and $u$ and $u_{t-1}$ and $u_t$ are the boundaries of the $t^{th}$ interval.

These implies the following sequence of conditional distributions:

$$\pi(\mu_{n,k} \mid \text{rest}) \propto \exp\left\{-\frac{\kappa_\eta}{2}\left(\mu_{n,k}-\bar{\mu}-\gamma\sum_j k_{kj}z_{n,j}\right)^2\right\}\cdot f(y_{n,k}|\mu_{n,k}), \qquad (4.7)$$

$$Z_n \mid \text{rest} \sim N\left[(\kappa_\epsilon I_q+\kappa_\eta\gamma^2K^TK)^{-1}(\kappa_\epsilon\Psi\alpha_n+\kappa_\eta\gamma K(\mu_n-\bar{\mu}\mathbf{1}),(\kappa_\epsilon I_q+\kappa_\eta\gamma^2K^TK)^{-1}\right], \qquad (4.8)$$

$$\alpha_n \mid \text{rest} \sim N\left[\left(\Sigma_\alpha(\theta)^{-1}+\kappa_\epsilon\Psi^T\Psi\right)^{-1}\kappa_\epsilon\Psi^TZ_n,\left(\Sigma_\alpha(\theta)^{-1}+\kappa_\epsilon\Psi^T\Psi\right)^{-1}\right], \qquad (4.9)$$

$$\bar{\mu} \mid \text{rest} \sim N\left[\frac{\kappa_\mu\bar{\mu}_0+\kappa_\eta\mathbf{1}^T(\mu-\gamma K^TZ)}{\kappa_\mu+m\kappa_\eta},\frac{1}{\kappa_\mu+m\kappa_\eta}\right], \qquad (4.10)$$

$$\kappa_\eta \mid \text{rest} \sim Gamma\left[q_\eta+\frac{m}{2},r_\eta+\frac{1}{2}(\mu-\bar{\mu}\mathbf{1}-\gamma K^TZ)^T(\mu-\bar{\mu}\mathbf{1}-\gamma K^TZ)\right], \qquad (4.11)$$

$$\kappa_\epsilon|\text{ rest} \sim Gamma\left[q_\epsilon + \frac{q}{2}, r_\epsilon + \frac{1}{2}(Z-\Psi\alpha)^T(Z-\Psi\alpha)\right], \tag{4.12}$$

$$\gamma|\text{ rest} \sim \pi_\gamma(\gamma)\exp\left\{-\frac{\kappa_\eta}{2}(\mu-\bar{\mu}\mathbf{1}-\gamma K^T Z)^T(\mu-\bar{\mu}\mathbf{1}-\gamma K^T Z)\right\}, \tag{4.13}$$

$$\theta|\text{ rest} \sim \pi_\theta(\theta)|\Sigma_\alpha(\theta)|^{-1/2}\exp\left\{-\frac{1}{2}\alpha^T\Sigma_\alpha(\theta)^{-1}\alpha\right\}. \tag{4.14}$$

Because we are considering different days as independent, the "global" parameters, $\gamma, \theta, \mu, \kappa_\epsilon$ and $\kappa_\eta$ will be providing us the link between different days and will update processes (4.10)–(4.14) combining data from all days. This won't increase the computational difficulty because our day-independent assumption gives us that the combined log likelihood from all days will simply be a sum of log likelihoods from the individual days.

In this case, the $Z$ and $\alpha$ processes are defined each day on every grid point and they are the main computational obstacle we want to conquer. It is really not practically implementable to deal with such high dimensional($q = 900$) covariance structures of $Z$ directly as in Diggle et al. (1998). This is because we need to invert these matrices at each MCMC iteration and it will take too long to get the whole MCMC process convergent. But after the spectral parameterizations, the covariance matrices of conditional densities of both the Gaussian process,$Z_n$, and the random vector of coefficients,$\alpha_n$, become very simple. More specifically, $K^TK$ and $\Psi^T\Psi$ are obviously diagonal. The matrix $\Sigma_\alpha(\theta)$ derived from Matérn's covariance matrix is asymptotically diagonal which has been presented before. Therefore,$\kappa_\epsilon I_q + \kappa_\eta\gamma^2 K^TK$, the covariance matrix of $Z_n$, and $\Sigma_\alpha(\theta)^{-1} + \kappa_\epsilon\Psi^T\Psi$, the covariance matrix of $\alpha_n$ are both very easy to deal with. At this point, the MCMC sampling can proceed quite efficiently.

The simulated data generated by this method therefore are daily based on each unit of NCEP grid cell. In order to compare with the NCEP data, we can take the grid-cell average based on the simulated data and then it is possible to see on each day, how good the NCEP data are. Even though the method is very efficient, it is still a large amount of work to generate simulated data within every NCEP grid cell on each day for all over the last 100 years. But nevertheless, this approach gives us a possibility to provide detailed simulated extreme precipitation under a much smaller scale compared with other climate models.

By this method, for each day we obtain a sample of predicted precipitation values at

each point of the grid-cell. We can use this to calculate a sample of predicted grid-cell averages for eahc day, which may be directly compared with the NCEP value.

## 4.3 Embedded Block Circulant Approach to Spatial Dependence Structures

The spectral method provides us an approximation of spatial dependence structures for generating the MCMC iterations. But it is also possible to consider an alternative approach which is based on embedding our current covariance structures from stationary Gaussian processes in large dimensional matrices with block circulant structures. That is to say, we can expand the spatial related area to a much bigger area with simpler spatial dependence structure(block circulant). Then after some transformations, the MCMC iterations can still proceed very quickly.

We will still use the generalized linear mixed model and assume the rainfall data follow piecewise distribution, and only consider it at one certain day $n$, then

- For the data exceed our threshold $u$, we assume $y$ follows POT approach, i.e.

$$f_{i,j}(y_{i,j}|\mu_{i,j}) \sim \frac{1}{\phi_{i,j}}\left(1+\xi_{i,j}\frac{y_{i,j}-\mu_{i,j}}{\phi_{i,j}}\right)_{+}^{-1/\xi_{i,j}-1}\exp\left\{-\left(1+\xi_{i,j}\frac{u-\mu_{i,j}}{\phi_{i,j}}\right)^{-1/\xi_{i,j}}\right\}$$

  where $\mu_{i,j}$ is a location parameter, $\xi_{i,j}$ is a shape parameter and $\phi_{i,j}$ is a scale parameter.

- Estimate $F_{i,j}(0)$ by the sample portion of zeros.

- For $0 < y \leq u$(threshold), divide the range into $T$ equiprobable intervals(depending on the number of data we have in this interval) and assume $F_{n,k}(y)$ is piecewise linear within each interval.

Where $(i, j)$ corresponds to the indices of locations in a $30 \times 30$ grid cell.

According to our generalized linear mixed model approach, we have

$$\mu_{i,j} = \bar{\mu} + z_{i,j} + \eta_{i,j} \tag{4.15}$$

where $\bar{\mu}$ is the mean with some noise over grid cells. $z_{i,j}$ and $\eta_{i,j}$ are spatial related stationary Gaussian processes and independent errors, respectively.

For the spatial dependence of $Z$ processes, instead of using Matérn covariance, we switch to the exponential model with nugget

$$C_0(t) = \theta_1 I(t = 0) + \theta_2 \exp^{-\frac{t}{\theta_3}} I(t > 0) \tag{4.16}$$

where $t$ is the distance to the origin. Because we are actually more focusing on those extreme rainfalls, the spatial dependence structure is not the most important issue here. In fact, in practice, it is very hard to distinguish the two models. Even in standard geostatistical applications (when we observe the Gaussian process directly) it tends to be tough to distinguish the different covariance models(Stein, 1999). And the reason we switch from Matérn class of covariance structures to exponential class is because in the exact approaching cases, we need to find every entry of the covariance matrix and the complex forms of Matérn function will let us spend too much time on calculating these entries themselves during MCMC iterations.

Given $\{\theta_1, \theta_2, \theta_3\}$,

$$Z = \{Z_{i,j} : i, j \in \{0, 1, \cdots, 29\}\} \tag{4.17}$$

is an array of zero means Gaussian variables with covariance given by

$$E\left\{Z_{i,j} Z_{i',j'}\right\} = C_0(\sqrt{(i - i')^2 + (j - j')^2}) \tag{4.18}$$

We would like first to re-order the indices to build a new 1-dimensional Gaussian Random field $\hat{Z}$ which is equivalent to $Z$. The 2-dimensional indices $\{(i, j) : i, j \in \{0, 1, 2, \cdots, 29\}\}$ are related to the new indices $\{r : r \in \{0, 1, \cdots, 899\}\}$ by

$$r = i + 30j. \tag{4.19}$$

We assume $\hat{Z}$ process has covariance matrix $\hat{\Sigma}(\theta_1, \theta_2, \theta_3)$.

This lead us to the model

$$\begin{aligned}
\mu|Z &\sim N\left[\bar{\mu}+\hat{K}\hat{Z},\frac{1}{\kappa_\eta}\right],\\
\hat{Z} &\sim N\left[0,\hat{\Sigma}(\theta_1,\theta_2,\theta_3)\right],\\
\bar{\mu} &\sim N\left[\bar{\mu}_0\mathbf{1},\frac{1}{\kappa_\mu}\right],\\
\kappa_\eta &\sim Gamma[q_\eta,r_\eta],\\
\theta_1 &\sim \pi_{\theta_1},\\
\theta_2 &\sim \pi_{\theta_2},\\
\theta_3 &\sim \pi_{\theta_3}.
\end{aligned}$$

where $\hat{K}$ is a $q\times 900$ location matrix($q$ is the number of observational stations) with its column indices $\{v=r+1 : r\in\{0,1,\cdots,899\}\}$ corresponding to the two-dimensional indices $\{(i,j) : i,j\in\{0,1,\cdots,29\}\}$ under the relationship $r=i+30j$.

Under this model, the joint density of $\{\mu,\kappa_\eta,\bar{\mu},Z,\theta_1,\theta_2,\theta_3,Y\}$ is proportional to

$$\begin{aligned}
&\exp\left\{-\frac{\kappa_\mu}{2}(\bar{\mu}-\bar{\mu}_0\mathbf{1})^2\right\}\cdot\kappa_\eta^{q_\eta-1}e^{-r_\eta\kappa_\eta}\cdot\pi_{\theta_1}\cdot\pi_{\theta_2}\cdot\pi_{\theta_3}\cdot|\hat{\Sigma}(\theta_1,\theta_2,\theta_3)|^{-1/2}\exp\left\{-\frac{1}{2}\hat{Z}^T\hat{\Sigma}(\theta_1,\theta_2,\theta_3)^{-1}\hat{Z}\right\}\\
&\cdot\kappa_\eta^{m/2}\exp\left\{-\frac{\kappa_\eta}{2}(\mu-\bar{\mu}-\hat{K}\hat{Z})^T(\mu-\bar{\mu}\mathbf{1}-\hat{K}\hat{Z})\right\}\cdot f(y,\mu). \qquad (4.20)
\end{aligned}$$

The conditional distribution of $\hat{Z}$ turns to be

$$\hat{Z}|rest\sim N\left[\left(\hat{\Sigma}(\theta_1,\theta_2,\theta_3)^{-1}+\kappa_\eta\hat{K}^T\hat{K}\right)^{-1}\kappa_\eta\hat{K}^T(\mu-\bar{\mu}\mathbf{1}),\left(\hat{\Sigma}(\theta_1,\theta_2,\theta_3)^{-1}+\kappa_\eta\hat{K}^T\hat{K}\right)^{-1}\right]. \qquad (4.21)$$

It is not possible to employ Wood-Chan's method directly because they need the value of each entry of covariance matrix to calculate eigenvalues and then lead to simulation of Gaussian process. Here we have to invert $\hat{\Sigma}(\theta_1,\theta_2,\theta_3)$, and it will take too much time if we don't have anything, for example, spectral representation like Wikle did, done preliminarily.

But alternatively, it is possible to write the prior of $\hat{Z}$ in terms of $V=(V_0,V_1,\cdots,V_{\bar{m}-1})^T$ which is a vector of independent $N[0,1]$ random variables, where $\bar{m}=2^g$ is the least number

such that the embedded circulant covariance matrix $C$ is positive definite. That is, write $\hat{Z} = GV$, where $G = (g_{ij}(\theta_1, \theta_2, \theta_3))$. Then we have

$$\begin{aligned}
\mu|Z &\sim N\left[\bar{\mu} + \hat{K}GV, \frac{1}{\kappa_\eta}\right], \\
V &\sim N\left[0, I\right], \\
\bar{\mu} &\sim N\left[\bar{\mu}_0\mathbf{1}, \frac{1}{\kappa_\mu}\right], \\
\kappa_\eta &\sim Gamma[q_\eta, r_\eta], \\
\theta_1 &\sim \pi_{\theta_1}, \\
\theta_2 &\sim \pi_{\theta_2}, \\
\theta_3 &\sim \pi_{\theta_3}.
\end{aligned}$$

The joint density of $\{\mu, \kappa_\eta, \bar{\mu}, V, \theta_1, \theta_2, \theta_3\}$ then is proportional to

$$\begin{aligned}
&\exp\left\{-\frac{\kappa_\mu}{2}(\bar{\mu} - \bar{\mu}_0\mathbf{1})^2\right\} \cdot \kappa_\eta^{q_\eta - 1} e^{-r_\eta\kappa_\eta} \cdot \pi_{\theta_1} \cdot \pi_{\theta_2} \cdot \pi_{\theta_3} \cdot \exp\left\{-\frac{1}{2}V^TV\right\} \\
&\cdot\kappa_\eta^{m/2} \exp\left\{-\frac{\kappa_\eta}{2}(\mu - \bar{\mu} - \hat{K}GV)^T(\mu - \bar{\mu} - \hat{K}GV)\right\} \cdot f(y, \mu) \qquad (4.22)
\end{aligned}$$

These imply the following conditional distributions

$$\begin{aligned}
\mu|\text{rest} &\propto \exp\left\{-\frac{\kappa_\eta}{2}(\mu - \bar{\mu} - \hat{K}GV)^T(\mu - \bar{\mu}\mathbf{1} - \hat{K}GV)\right\} \cdot f(y, \mu), \qquad (4.23) \\
V|\text{rest} &\sim N\left[\left(I + \kappa_\eta G^T\hat{K}^T\hat{K}G\right)^{-1}\kappa_\eta G^T\hat{K}^T(\mu - \bar{\mu}\mathbf{1}), \left(I + \kappa_\eta G^T\hat{K}^T\hat{K}G\right)^{-1}\right], \qquad (4.24) \\
\bar{\mu}|\text{rest} &\sim N\left[\frac{\kappa_\mu\bar{\mu}_0 + \kappa_\eta\mathbf{1}^T(\mu - \hat{K}GV)}{\kappa_\mu + 900\kappa_\eta}, \frac{1}{\kappa_\mu + 900\kappa_\eta}\right], \qquad (4.25) \\
\theta_1 &\sim \pi_{\theta_1} \cdot \exp\left\{-\frac{\kappa_\eta}{2}(\mu - \bar{\mu} - \hat{K}GV)^T(\mu - \bar{\mu} - \hat{K}GV)\right\}, \qquad (4.26) \\
\theta_2 &\sim \pi_{\theta_2} \cdot \exp\left\{-\frac{\kappa_\eta}{2}(\mu - \bar{\mu} - \hat{K}GV)^T(\mu - \bar{\mu} - \hat{K}GV)\right\}, \qquad (4.27) \\
\theta_3 &\sim \pi_{\theta_3} \cdot \exp\left\{-\frac{\kappa_\eta}{2}(\mu - \bar{\mu} - \hat{K}GV)^T(\mu - \bar{\mu} - \hat{K}GV)\right\}. \qquad (4.28)
\end{aligned}$$

Now we define the parameter matrix $G$. Assume $\bar{m}$ has been chosen and discuss this

part lately. Define the following notations for convenience:

$$\begin{aligned} I(m) &= \{u=(i,j): 0 \le i \le m[1]-1, 0 \le j \le m[2]-1\} \\ I^*(m) &= \{u=(i,j): 0 \le |i| \le m[1]-1, 0 \le |j| \le m[2]-1\} \\ C &= \{c_{u_1 u_2}, u_1, u_2 \in I(m)\} \\ \frac{u}{m} &= \left(\frac{i}{m[1]}, \frac{j}{m[2]}\right)^T \end{aligned}$$

where $\bar{m} = m[1] \times m[2]$ and $m[1] = m[2]$ in this case. $c_{u_1 u_2} = c(u_1 - u_2)$,

$$c(h) = C_0(\frac{\tilde{h}}{m}), \qquad h, \tilde{h} \in I^*(m), \tag{4.29}$$

$\tilde{h} = \tilde{h}(h)$ is given by

$$\begin{aligned} \tilde{h}(h)[l] &= h[l] && \text{if} \quad 0 \le |h[l]| \le \frac{m[l]}{2} \\ &= h[l] - m[l] && \text{if} \quad \frac{m[l]}{2} < h[l] \le m[l]-1 \\ &= h[l] + m[l] && \text{if} \quad \frac{m[l]}{2} < -h[l] \le m[l]-1, \end{aligned} \tag{4.30}$$

where $l = 1, 2$ and

$$C_0(t) = \theta_1 I(t=0) + \theta_2 \exp^{-\frac{t}{\theta_3}} I(t>0)$$

as we defined previously. So $C$ is symmetric with covariance structure $\Sigma$ of $Z$ embedded and $C$ has block circulant structure.

The eigenvalues of $C$ can be calculated by two-dimensional discrete Fourier transform of $c$

$$\begin{aligned} \lambda(u') &= \sum_{u \in I(m)} c(u) \exp\left\{-2\pi\sqrt{-1} u^T (\frac{u'}{m})\right\}, \quad u' \in I(m) \\ &= \sum_{u \in I(m)} (\theta_1 I(u=0) + \theta_2 \exp\left\{-\frac{\sqrt{i^2+j^2}}{\theta_3} I(u \ne 0)\right\}) \exp\left\{-2\pi\sqrt{-1}(\frac{ii'}{m[1]} + \frac{jj'}{m[2]})\right\} \end{aligned} \tag{4.31}$$

(4.32)

For each $u \in I(m)$, define a vector $q_j$ of length $\bar{m}$ with components

$$q_u(u') = \bar{m}^{-1/2} \exp\left\{-2\pi\sqrt{-1}(\frac{ii'}{m[1]} + \frac{jj'}{m[2]})\right\}, \quad u' \in I(m). \tag{4.33}$$

This defines a $\bar{m} \times \bar{m}$ diagonal matrix $\Lambda$ and $\bar{m} \times \bar{m}$ matrix $Q$ and according to Wood-Chan's method, we have

$$C = Q\Lambda Q^* \quad \text{and} \quad C^{1/2} = Q\Lambda^{1/2}Q^* \tag{4.34}$$

(4.34) implies that

$$W = C^{1/2}V = Q\Lambda^{1/2}Q^*V \sim N[0, C] \tag{4.35}$$

Finally, we have

$$Z(u) = W(u), \quad u \in I(n) \tag{4.36}$$

where $n = (30, 30)$ is a subset of $W$ and corresponds to the actual grid cells we need to consider and therefore, $G$ is the corresponding sub-matrix of $C^{1/2}$.

Now instead of considering $Z$ or $\hat{Z}$, we consider the whole $W$ process. This won't increase the computational complexity by extending the location matrix $\hat{K}$ to $K$ with dimension $d \times \bar{m}$,where $d$ is still the number of observational stations in certain grid cell. The expanded entries are all set to be 0. The updated model then is

$$\begin{aligned}
\mu|Z &\sim N\left[\bar{\mu} + KC^{1/2}V, \frac{1}{\kappa_\eta}\right], \\
V &\sim N\left[0, I\right], \\
\bar{\mu} &\sim N\left[\bar{\mu}_0\mathbf{1}, \frac{1}{\kappa_\mu}\right], \\
\kappa_\eta &\sim Gamma[q_\eta, r_\eta],
\end{aligned}$$

$$\begin{aligned}\theta_1 &\sim \pi_{\theta_1},\\ \theta_2 &\sim \pi_{\theta_2},\\ \theta_3 &\sim \pi_{\theta_3}.\end{aligned}$$

and the joint density is proportional to

$$\begin{aligned}&\exp\left\{-\frac{\kappa_\mu}{2}(\bar\mu-\bar\mu_0\mathbf{1})^2\right\}\cdot\kappa_\eta^{q_\eta-1}e^{-r_\eta\kappa_\eta}\cdot\pi_{\theta_1}\cdot\pi_{\theta_2}\cdot\pi_{\theta_3}\cdot\exp\left\{-\frac{1}{2}V^TV\right\}\\ &\cdot\kappa_\eta^{m/2}\exp\left\{-\frac{\kappa_\eta}{2}(\mu-\bar\mu-\hat KC^{1/2}V)^T(\mu-\bar\mu-KC^{1/2}V)\right\}\cdot f(y|\mu)\end{aligned}\tag{4.37}$$

These imply that the conditional distributions are

$$\mu|\text{rest} \propto \exp\left\{-\frac{\kappa_\eta}{2}(\mu-\bar\mu-KC^{1/2}V)^T(\mu-\bar\mu\mathbf{1}-KC^{1/2}V)\right\}\cdot f(y|\mu),\tag{4.38}$$

$$V|\text{rest} \sim N\left[\left(I+\kappa_\eta C^{1/2}K^TKC^{1/2}\right)^{-1}\kappa_\eta G^T\hat K^T(\mu-\bar\mu\mathbf{1}),\left(I+\kappa_\eta C^{1/2}K^TKC^{1/2}\right)^{-1}\right],\tag{4.39}$$

$$\bar\mu|\text{rest} \sim N\left[\frac{\kappa_\mu\bar\mu_0+\kappa_\eta\mathbf{1}^T(\mu-KC^{1/2}V)}{\kappa_\mu+900\kappa_\eta},\frac{1}{\kappa_\mu+900\kappa_\eta}\right],\tag{4.40}$$

$$\theta_1|\text{rest} \sim \pi_{\theta_1}\cdot\exp\left\{-\frac{\kappa_\eta}{2}(\mu-\bar\mu-KC^{1/2}V)^T(\mu-\bar\mu-KC^{1/2}V)\right\},\tag{4.41}$$

$$\theta_2|\text{rest} \sim \pi_{\theta_2}\cdot\exp\left\{-\frac{\kappa_\eta}{2}(\mu-\bar\mu-KC^{1/2}V)^T(\mu-\bar\mu-KC^{1/2}V)\right\},\tag{4.42}$$

$$\theta_3|\text{rest} \sim \pi_{\theta_3}\cdot\exp\left\{-\frac{\kappa_\eta}{2}(\mu-\bar\mu-KC^{1/2}V)^T(\mu-\bar\mu-KC^{1/2}V)\right\}.\tag{4.43}$$

Here $C^{1/2}=Q\Lambda^{1/2}Q^*$, and now the computation issue becomes how to handle $\left(I+\kappa_\eta C^{1/2}K^TKC^{1/2}\right)^{-1}$. We have Woodbury Formula as follows:

$$(A+UV^T)^{-1}=A^{-1}-\left[A^{-1}U(I+V^TA^{-1}U)^{-1}V^TA^{-1}\right],\tag{4.44}$$

so

$$\begin{aligned}&\left(I+\kappa_\eta C^{1/2}K^TKC^{1/2}\right)^{-1}\\ =\;& I-\left[\kappa_\eta C^{1/2}K^T(I+\kappa_\eta KCK^T)^{-1}KC^{1/2}\right]\\ =\;& I-\left[\kappa_\eta Q\Lambda^{1/2}Q^*K^T(I+\kappa_\eta KQ\Lambda Q^*K^T)^{-1}KQ\Lambda^{1/2}Q^*\right]\end{aligned}\tag{4.45}$$

Because $(I + \kappa_\eta KCK^T)^{-1}$ is depending on the number of observational stations which is much lower than the number of sub-grid cells and then is much lower than $\bar{m}$(for example, in CCSM grid cell around RDU, there are only 31 stations, or in NCEP grid cell around RDU, there are 66 stations).

In principle, $\bar{m}$ is chosen to be the least number greater than $2r - 1$ which let $C$ be positive definite. Currently, the existence of $\bar{m}$ is proved by Wood and Chan(1994) but there are no theoretical criteria to pick such $\bar{m}$ in different scenarios. In our case, because parameters are updated after each MCMC iteration, we have to check if $C$ is positive definite every time. The way to do it is very straightforward. We will have to calculate eigenvalues of $C$ every time by using DFFT, these eigenvalues will provide direct information about the positive definite property of $C$. If in some step, $C$ is not positive definite, then we will need to increase $\bar{m}$ for the following iterations.

## 4.4 Computational Results

### 4.4.1 Computational Results Based on Spectral Generalized Linear Mixed Model

We apply the model to the data collected in Aprils of 1978-1982 within the NCEP grid cell around RDU. For these data, eight Markov chains were obtained after 20000 iterations. The last 100 samples of the $\mu$ process are retained to compute the posterior summaries. Other history of global parameters' samples are also retained. The Raftery and Lewis(1992b)'s convergence diagnostic is used for test the convergence of all the samples. The diagnostic suggests that all parameters have converged with a small number of iterations as "burn-in"(after discarding 5000 iterations) after estimating 2.5% quantile of each parameter to within $\pm 0.005$ with 95% probability. Convergence was also assessed by using Gelman-Rubin diagnostic(Gelman and Rubin 1992).

Posterior means and standard deviations for model parameters are given in the following table(4.1). The MCMC history for parameter $\kappa_\eta$ is shown in Figure(4.2) which, although indicating high autocorrelation, does appear stable and well-mixing. The two parameters $\mu_{15,34}$ and $\mu_{136,20}$ which refer to the $34th$ grid cell of day 15 and the $136th$ grid cell of day

| $Variable$ | $PosteriorMean$ | $StandardDeviation$ |
|---|---|---|
| $\mu_{15,34}$ | 0.326 | 0.0123 |
| $\mu_{136,20}$ | 0.333 | 0.0120 |
| $\bar{\mu}$ | 0.328 | $5.85 \times 10^{-4}$ |
| $\theta_1$ | 24.59 | 0.641 |
| $\kappa_\eta$ | 5670.50 | 513.47 |
| $\kappa_\epsilon$ | $2.856 \times 10^{-2}$ | $8.157 \times 10^{-4}$ |

Table 4.1: *Summary of MCMC Output(Spectral GLMM Approach)*

136, respectively are randomly picked for illustration.

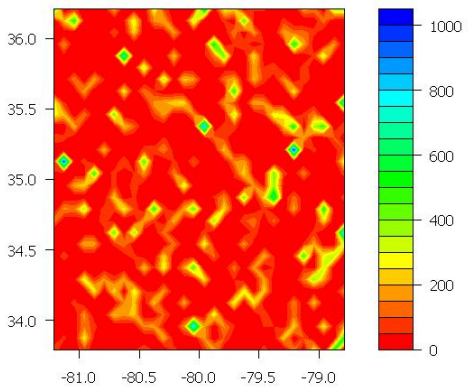
Now we are able to create a simulated rainfall on each grid-point within the NCEP grid cell around RDU, as illustrations, at day 15, day 100 and day 147(Figure (4.1)). The simulated grid cell based averages comparing with NCEP data indicates on most of the days, the simulation is quite close to the re-analyses except for day 80-90(Figure (4.3)). The scatter plot indicates both the NCEP data and simulated data are in same scale and the correlation(0.7416) provides an evaluation on how well NCEP is doing.

### 4.4.2 Computational Results Based on Embedded Block Circulant Approach

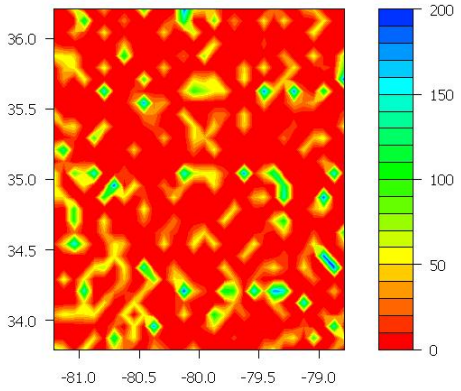
We still apply the model to the data collected in Aprils of 1978-1982 within the NCEP grid cell around RDU. Six Markov chains were obtained after 20000 iterations. The last 100 samples of the $\mu$ process are retained to compute the posterior summaries. Other history of global parameters' samples are also retained. The Raftery and Lewis(1992b)'s convergence diagnostic is used for test the convergence of all the samples. The diagnostic suggests that all parameters have converged with a small number of iterations as "burn-in"(after discarding 5000 iterations) after estimating 2.5% quantile of each parameter to within $\pm 0.005$ with 95% probability. Convergence was also assessed by using Gelman-Rubin diagnostic(Gelman and Rubin 1992). The $\bar{m} = 2^{11}$ all over the MCMC iterations.

Posterior means and standard deviations for model parameters are given in the table 4.2. The MCMC history for parameter $\kappa_\eta$ is shown in Figure 4.5 which, although still indicating high autocorrelation, does appear stable and well-mixing.

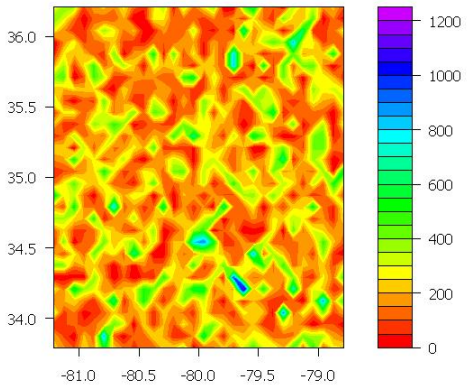
The two parameters $\mu_{15,34}$ and $\mu_{136,20}$ which refer to the $34th$ grid cell of day 15 and the




(a) Spatial rainfall plot on April 15, 1978




(b) Spatial rainfall plot on April 10, 1981




(c) Spatial rainfall plot on April 28, 1982

Figure 4.1: *Simulated Spatial Plots on NCEP Grid Cell around RDU*

| $Variable$ | $PosteriorMean$ | $StandardDeviation$ |
|---|---|---|
| $\mu_{15,34}$ | 0.315 | 0.0201 |
| $\mu_{136,20}$ | 0.381 | 0.0150 |
| $\bar{\mu}$ | 0.346 | $4.83 \times 10^{-4}$ |
| $\kappa_\eta$ | 5836.17 | 468.82 |
| $\theta_1$ | 1.021 | 0.020 |
| $\theta_2$ | 0.903 | 0.012 |
| $\theta_3$ | 1.874 | 0.067 |

Table 4.2: *Summary of MCMC output(Embedded Block Circulant Approach)*

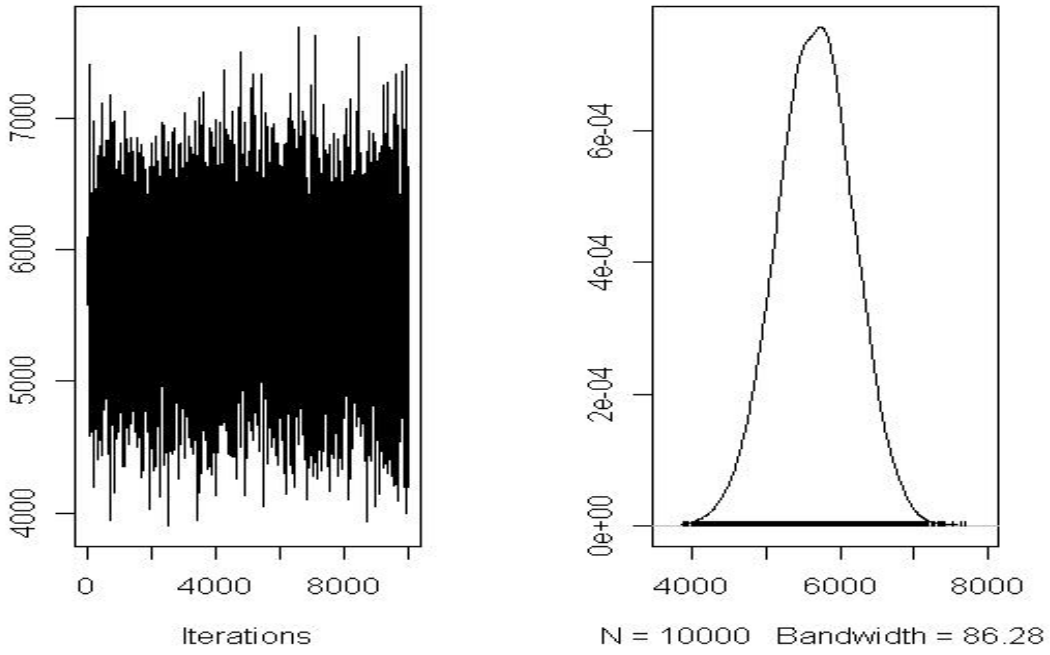



Figure 4.2: *MCMC history for $\kappa_\eta$ and its estimated posterior distribution*

$136th$ grid cell of day 136, respectively are picked the same as spectral GLMM approach for comparison.

The simulated rainfalls on each grid-point within the NCEP grid cell around RDU can be obtained then, and as illustrations, at day 10, day 48 day 125 and day 150(Figure 4.4). The simulated grid cell based averages comparing with NCEP data indicates similar result comparing with using spectral method(Figure 4.6) and also provides an evaluation on re-analyses by NCEP. The scales of NCEP and simulated results are similar with each other and the correlation(0.7398) implies how well NCEP is doing.

### 4.4.3 Precipitation Predictions based on CCSM grid cell

The Community Climate System Model(CCSM), as one of the global climate models, provides future rainfall data that could lead us to detailed spatial rainfall predictions. To be more specific, after getting spatial dependence information from past data and given the future area averages, our goal is to predict future rainfalls on the sub-grid cells of CCSM grid cells within certain areas.

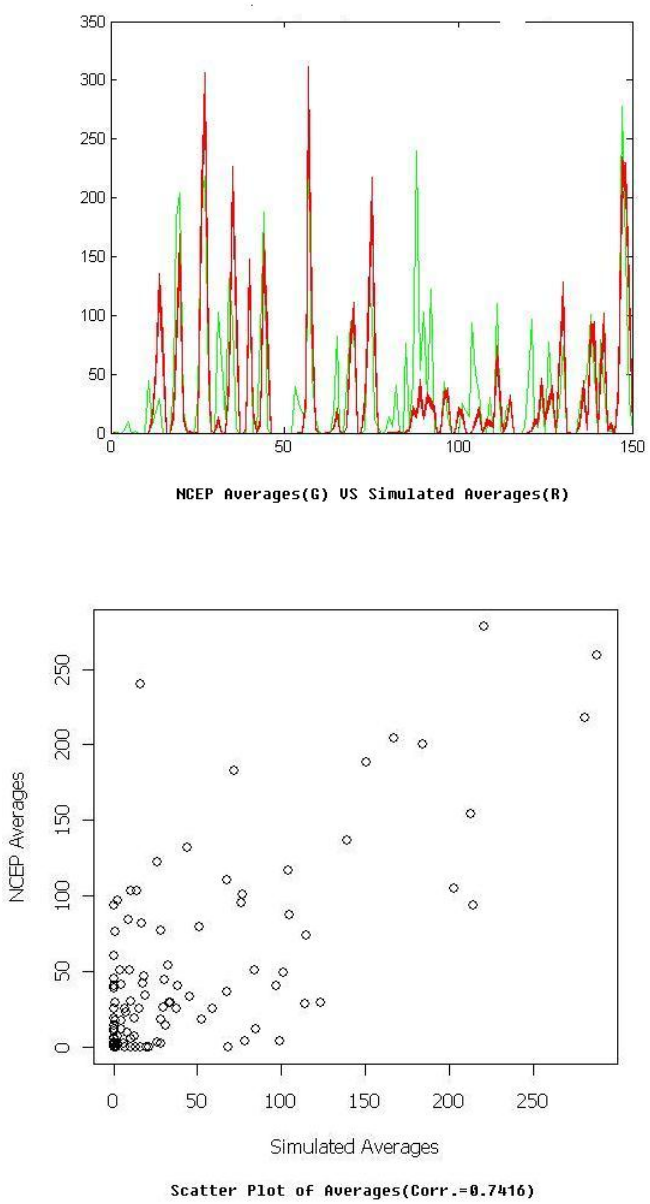



Figure 4.3: *Simulated Grid Cell Averages vs NCEP Averages*

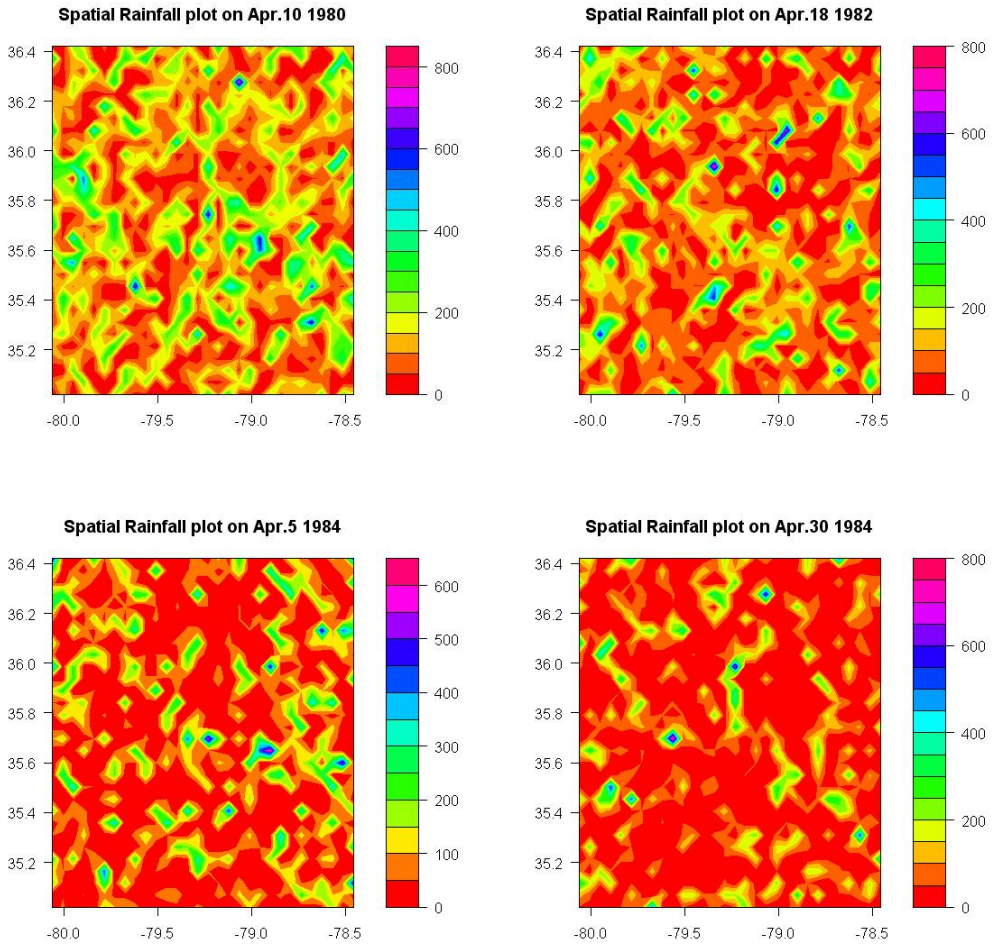



Figure 4.4: *Simulated Spatial Plots on NCEP Grid Cell around RDU*

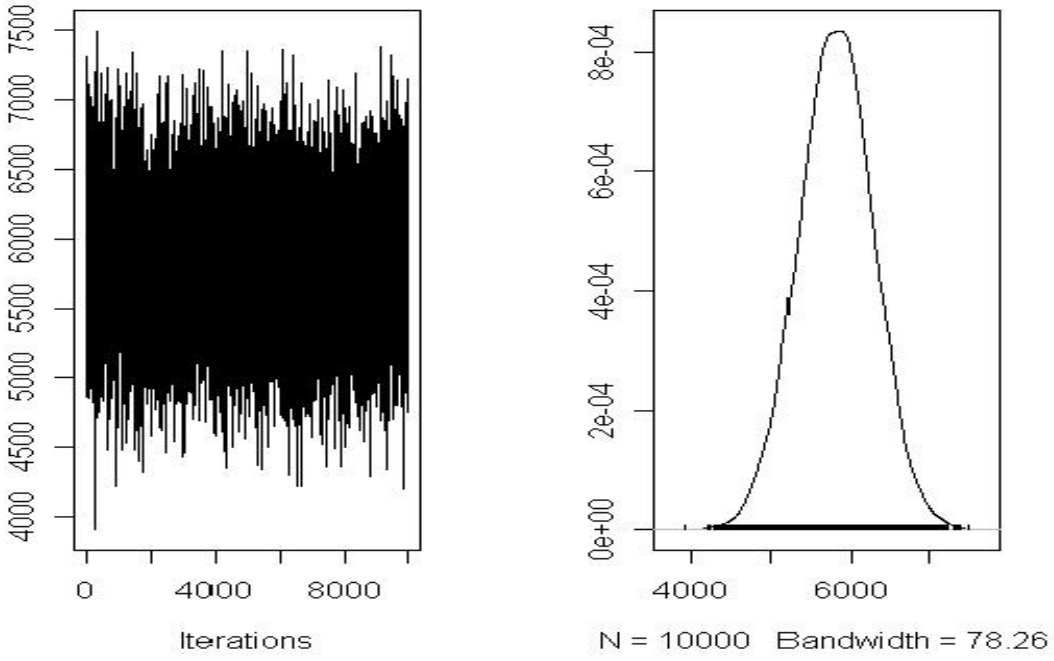



Figure 4.5: *MCMC history for $\kappa_\eta$ and its estimated posterior distribution*

**Computational Results from Spectral GLMM**

First of all, we need to get the spatial dependence structure of rainfalls. We use the CCSM grid cell around RDU as an illustration to do our simulation. A CCSM grid cell is of dimensions $1.4^o$ latitude $\times$ $1.4^o$ longitude. We still split this area into $30 \times 30$ sub-grid cells to have good spatial coverage. We consider the 150 days' data from Aprils of year 1980 $-$ year 1984 then make pointwise precipitation predictions for Aprils of year 2080 $-$ year 2084 based on CCSM data.

Figure.4.7 shows the density of CCSM data and the density of simulated mean(Aprils of 1980-1984) within the grid cell around RDU. Table 4.3 shows the mean and variance from two sources. Comparing the distributions for 1980-1984 is intended to validate the CCSM model for its use in future prediction, because table 4.3 provides the information how far the simulated data are away from CCSM data.

And the simulation also provides the detailed spatial rainfall information as showing on figure.4.8

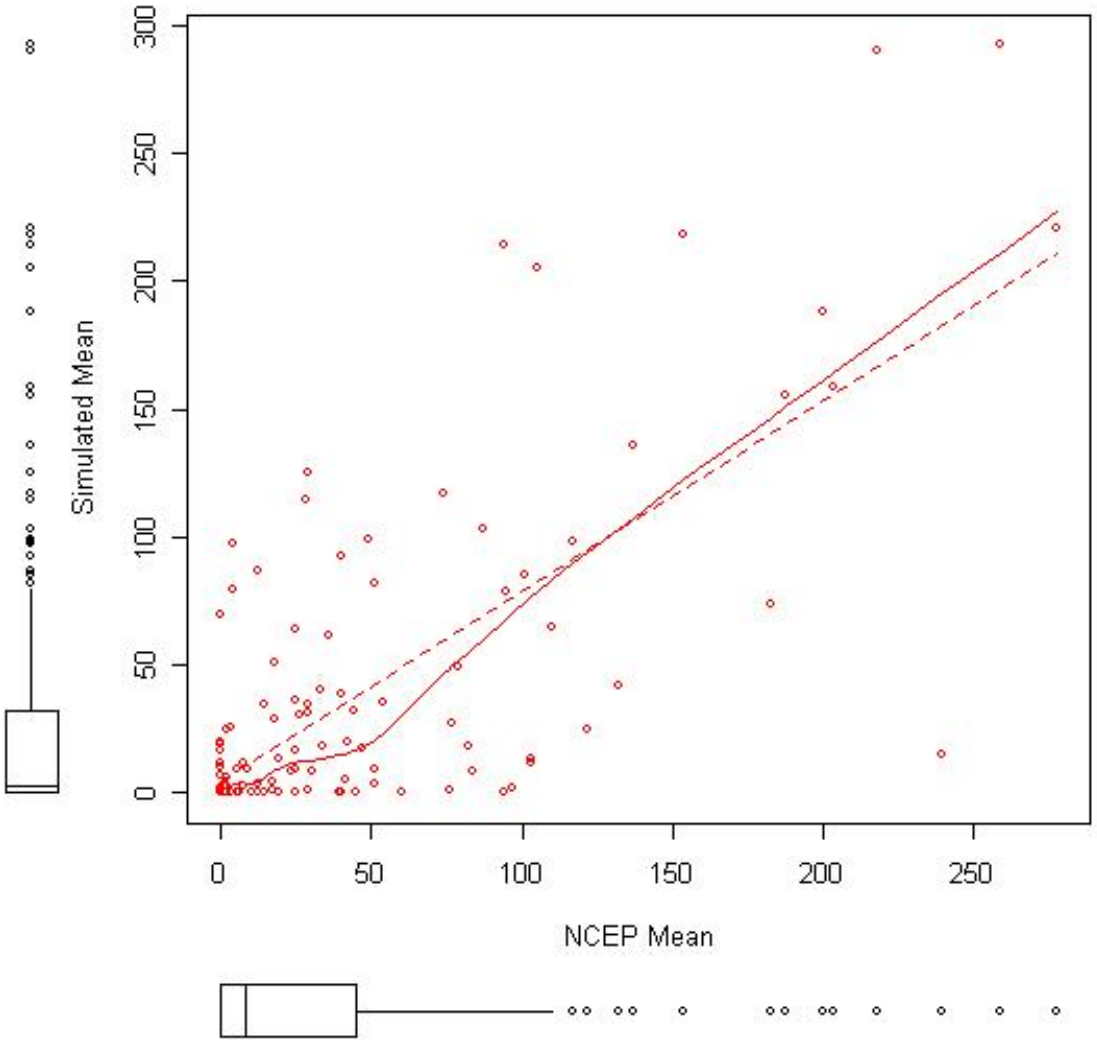



Figure 4.6: *Simulated Grid Cell Averages vs NCEP Averages, Corr=0.7398*

| | Mean | Variance |
|---|---|---|
| CCSM Data | 30.91 | 2873.79 |
| Simulated Data | 29.03 | 2349.61 |

Table 4.3: *Mean and Variance(CCSM vs. Simulated Data)*

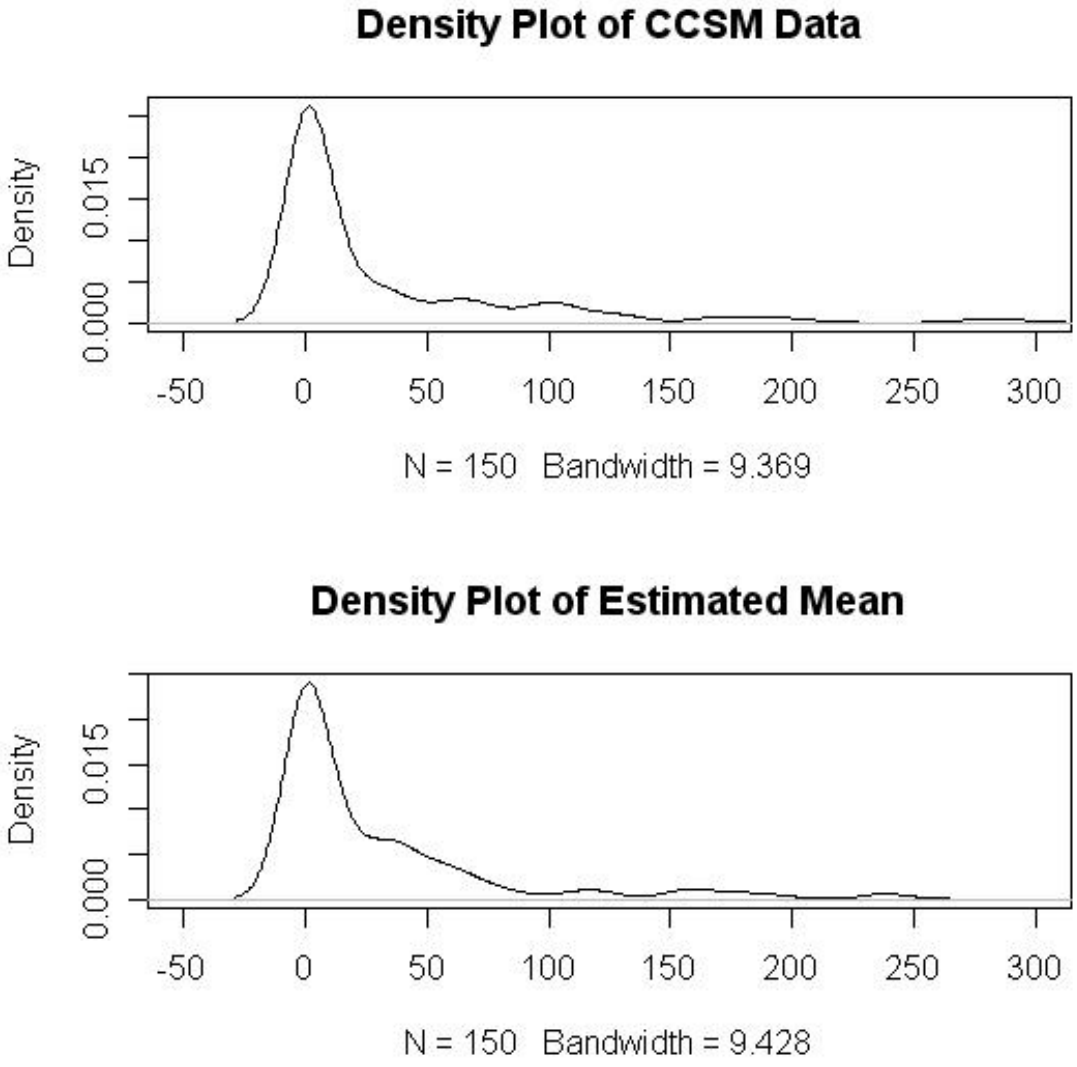



Figure 4.7: *Comparison of two densities*

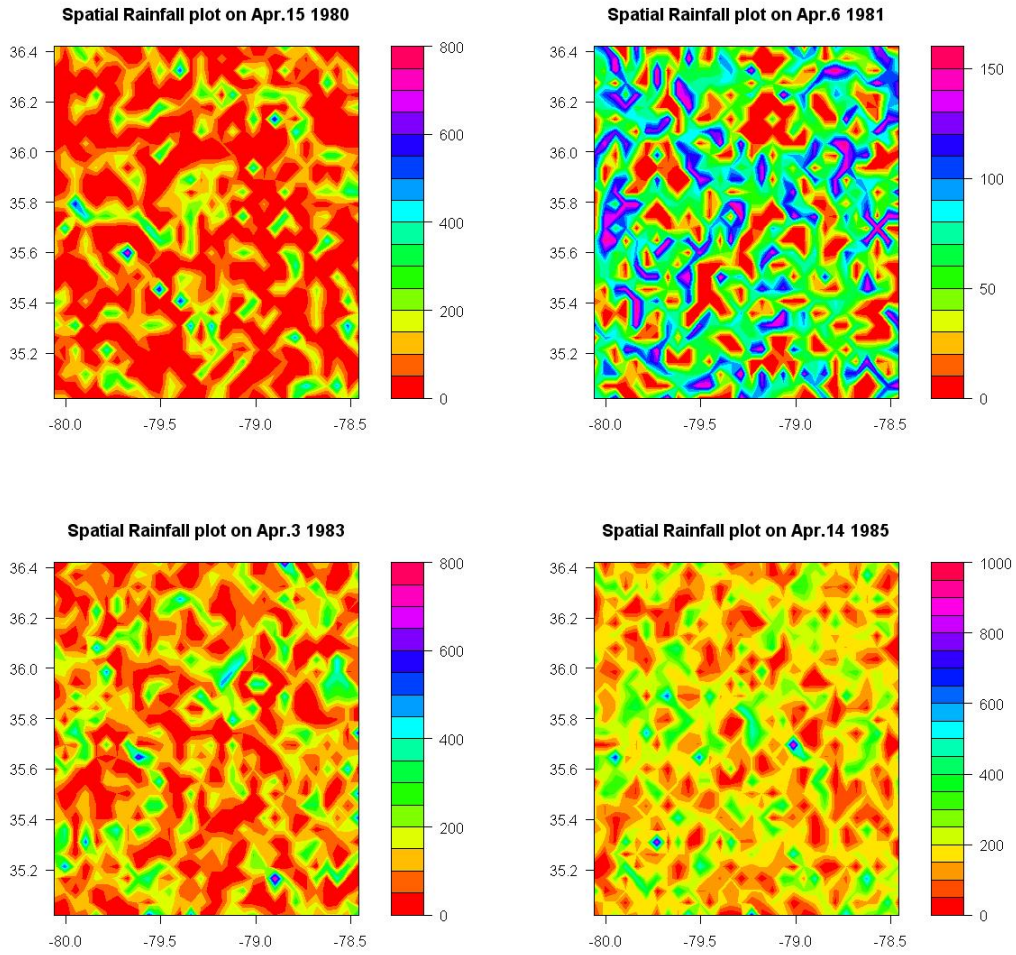



Figure 4.8: *Simulated Spatial Rainfall Plots*

**Precipitation Predictions**

Now we try to make predictions on future precipitations. There are two beneficial things we want to do here. The first is try to generate pointwise precipitation based on CCSM mean for grid cells. This will give us the detailed precipitation predictions in the future. Another thing is to find the $N$-year return values and try to identify the trends of extreme precipitations which can't be derived directly from mean values. To make such predictions, we still assume the future rainfalls follow the piecewise distribution similar to the past days. There are two ways to adjust the previous piecewise distribution. One is to perturb the parameters but keep the probabilities of the rainfall falling into three categories$\{0, (0, threshold], (threshold, +\infty)\}$ the same. The other is to change the probabilities but keep parameters the same. I am choosing the second way because it is much easier to keep the spatial dependence fixed.

Because we have assumed daily independence, we link any future day with a past day with closest area mean to get simulated averages. As mentioned before, the mean of CCSM data and simulated data give us a measurement of how close could be considered as reasonable. In our prediction, the difference between simulated prediction mean and CCSM mean is limited to less than 1% of CCSM mean. And apparently, the predictions are drawn from the posterior piecewise distribution with perturbation on probabilities falling into three different categories.

Now we have the initial rainfall predictions on each future day and we can use sampling technique to refine the predictions. Before doing this, we would like to change the piecewise distribution a little as follows:

- For the data exceed our threshold $u$, we assume $y$ follows POT approach, i.e.

$$f_{n,k}(y) \sim \frac{1}{\phi_{n,k}}\left(1+\xi_{n,k}\frac{y_{n,k}-\mu_{n,k}}{\phi_{n,k}}\right)_{+}^{-1/\xi_{n,k}-1}\exp\left\{-\left(1+\xi_{n,k}\frac{u-\mu_{n,k}}{\phi_{n,k}}\right)^{-1/\xi_{n,k}}\right\}$$

  where $\mu_{n,k}$ is a location parameter, $\xi_{n,k}$ is a shape parameter and $\phi_{n,k}$ is a scale parameter.

- For $0 \le y \le \epsilon$, $F_{n,k}(y)$ is linear.

- For $\epsilon < y \leq u$(threshold), divide the range into $T$ equiprobable intervals(depending on the number of data we have in this interval) and assume $F_{n,k}(y)$ is piecewise linear within each interval.

All of the parameters used here are refer to posterior parameters of the corresponding day in the past. The reason to make such modification is because when we do the sampling, we also want 0 to be perturbed.

According to the modified piecewise distribution, we can write down the corresponding likelihood components. Suppose $l(y_{n,k}, \mu_{n,k})$ is the likelihood component at day $n$, location $k$, then perturb $y_{n,k}$ with the constraint that the mean of $y_{n,k}$ at day $n$ is not far away from CCSM data(the difference should not be greater than 1% of CCSM mean) and the proposal densities of generating $\hat{y_{n,k}}$ here are as follows:

$$Q(\hat{y_{n,k}}; y_{n,k}) \sim N(y_{n,k}, \sigma^2_{n,k}),$$

where $\sigma_{n,k}$ is a unique value for day $n$ and location $k$ and is decided by the initial $y_{n,k}$'s and the mean values from CCSM data. Now let

$$\alpha = min\{1, \frac{\prod_{k=1}^{150} l(\hat{y}_{n,k}, \mu_{n,k})}{\prod_{k=1}^{150} l(y_{n,k}, \mu_{n,k})}\} \tag{4.46}$$

We then perform Metropolis sampling and accept $\hat{y}_{n,k}$ with probability $\alpha$ at day $n$.

After running 20000 iterations, we have the posterior rainfall predictions. The following figure.4.9 shows some snapshots of detailed spatial rainfall predictions. These are generated from the distributions we updated for future days and are just illustrations. And the spatial dependence structures are embedded in the distributions.

We also have the difference plot between simulated means and CCSM means as 4.10 where the red line refers to the CCSM means and the points represent the simulated means. This shows how the actual sequence of means used in the MCMC generation compares with the CCSM mean that we are conditioning on.

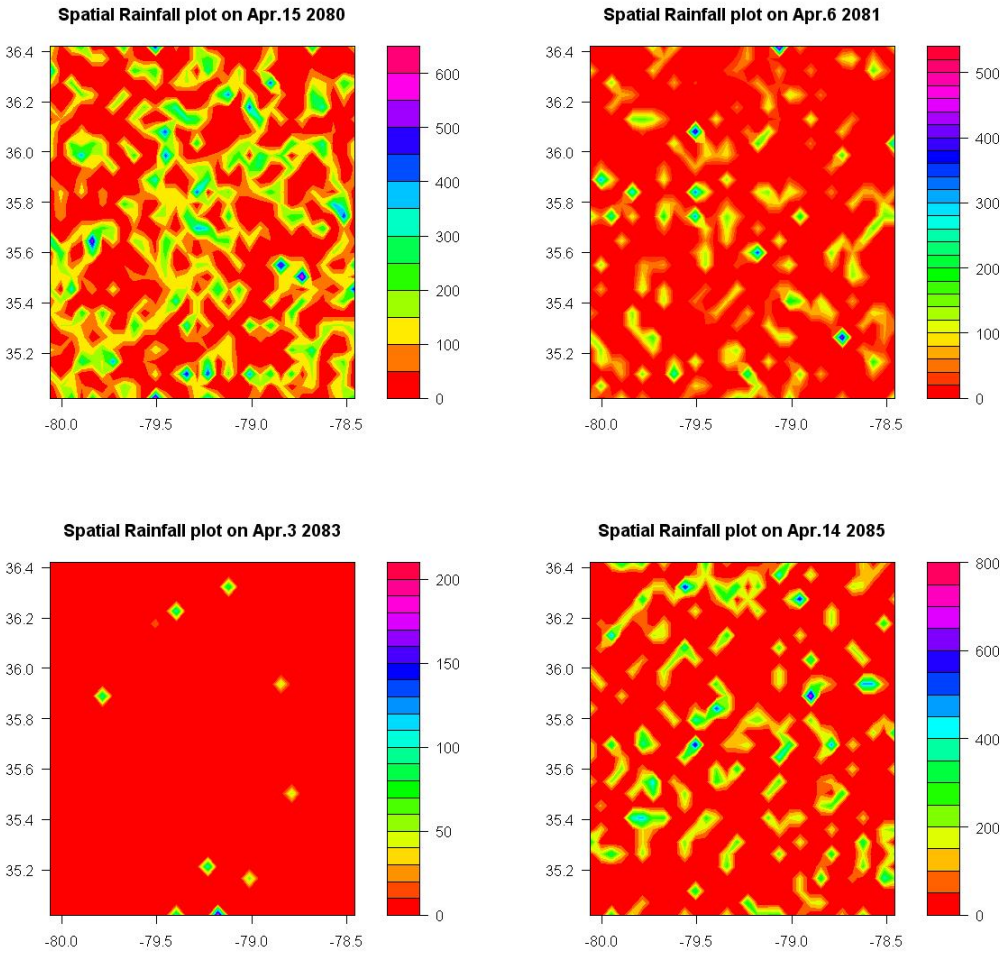



Figure 4.9: *Simulated Spatial Rainfall Predictions*

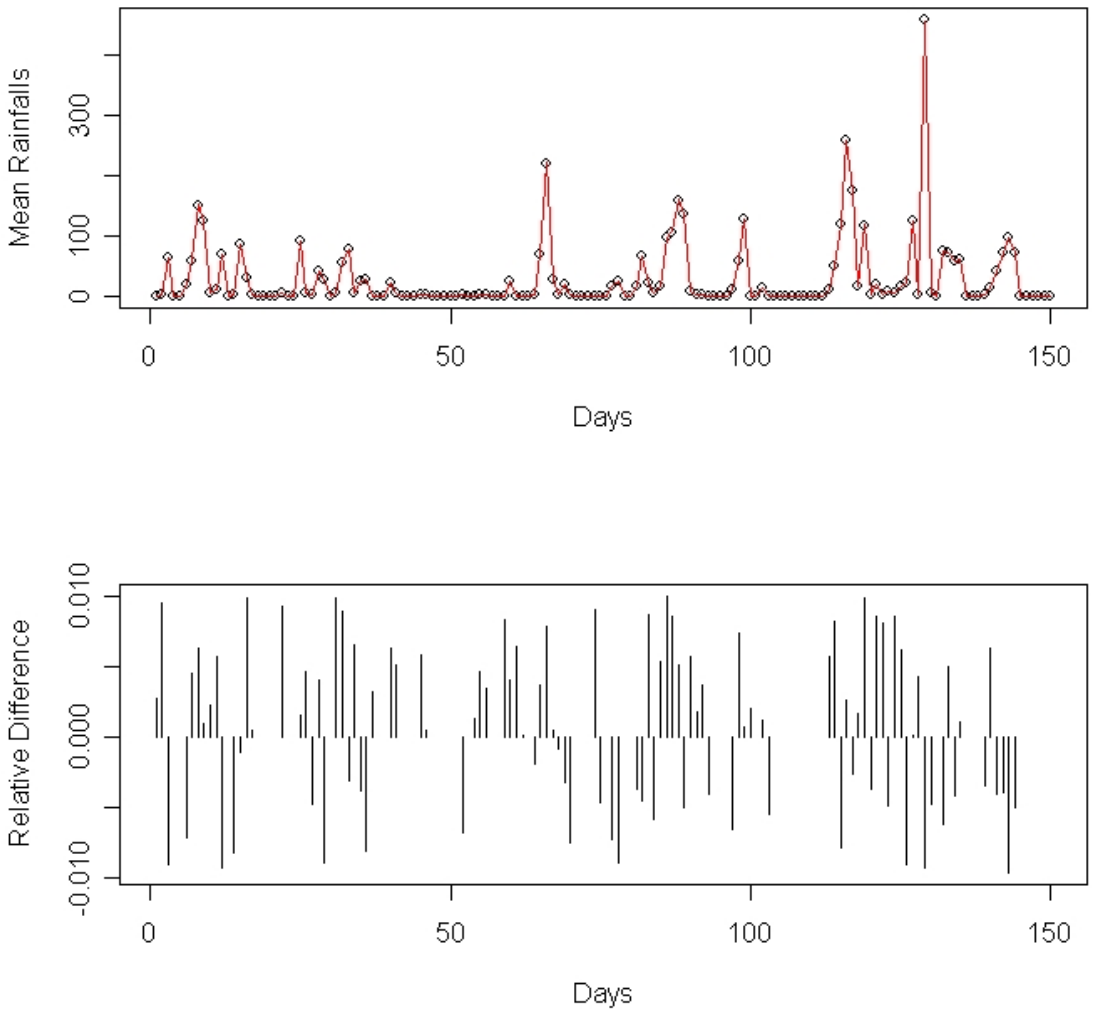



Figure 4.10: *Comparison of Means from CCSM and Simulation*

### Return Years and $N$-year Return Values

As we described in Chapter 2, the increase in intense precipitation, rather than the change in mean precipitation, is consistent with changes expected in a warmer atmosphere due to an acceleration of the hydrological cycle(e.g. Trenberth 1999) and reflects the climate changes forced by scenarios of greenhouse gas and sulfate aerosol emissions. We first take a look at the difference between present means and future means generated by CCSM. Figure 4.11 shows there is a trend that we will have more heavy rainfall events in the future than present days.

According to the simulated results, we can construct the detailed $N$-year return values pointwisely in the grid-cell around RDU. The return year and $N$-year return value are defined by the exceedance probability as follows:

If we have

$$P\{Annual\ maximum\ value > y_N\} = \frac{1}{N}, \tag{4.47}$$

The following two tables 4.4 and 4.5 provide the return years and $N$-year return values for the whole area, that is, the CCSM grid-cell around RDU. Without changing much in the total mean values between present days (29.03) and future days(28.68), we can certainly find the trend of increasing in intensive precipitations in the future.

| Return Year | Return Value(Present) | Return Value(Future) | Ratio(Future/Present) |
|---|---|---|---|
| 10 | 111.1643 | 104.3863 | 0.939 |
| 20 | 159.5820 | 220.0126 | 1.379 |
| 50 | 266.5093 | 342.9191 | 1.287 |
| 100 | 349.2094 | 430.1455 | 1.232 |

Table 4.4: *Return Values' Comparison (Present days vs Future days)*

| Return Value | Return Year(Present) | Return Year(Future) | Ratio(Future/Present) |
|---|---|---|---|
| 111.1643 | 10 | 10.1964 | 1.0196 |
| 159.5820 | 20 | 12.9051 | 0.6453 |
| 266.5093 | 50 | 28.1074 | 0.5615 |
| 349.2094 | 100 | 51.8433 | 0.5184 |
| 430.1455 | 223.8806 | 100 | 0.4467 |

Table 4.5: *Return Years' Comparison (Present days vs Future days)*

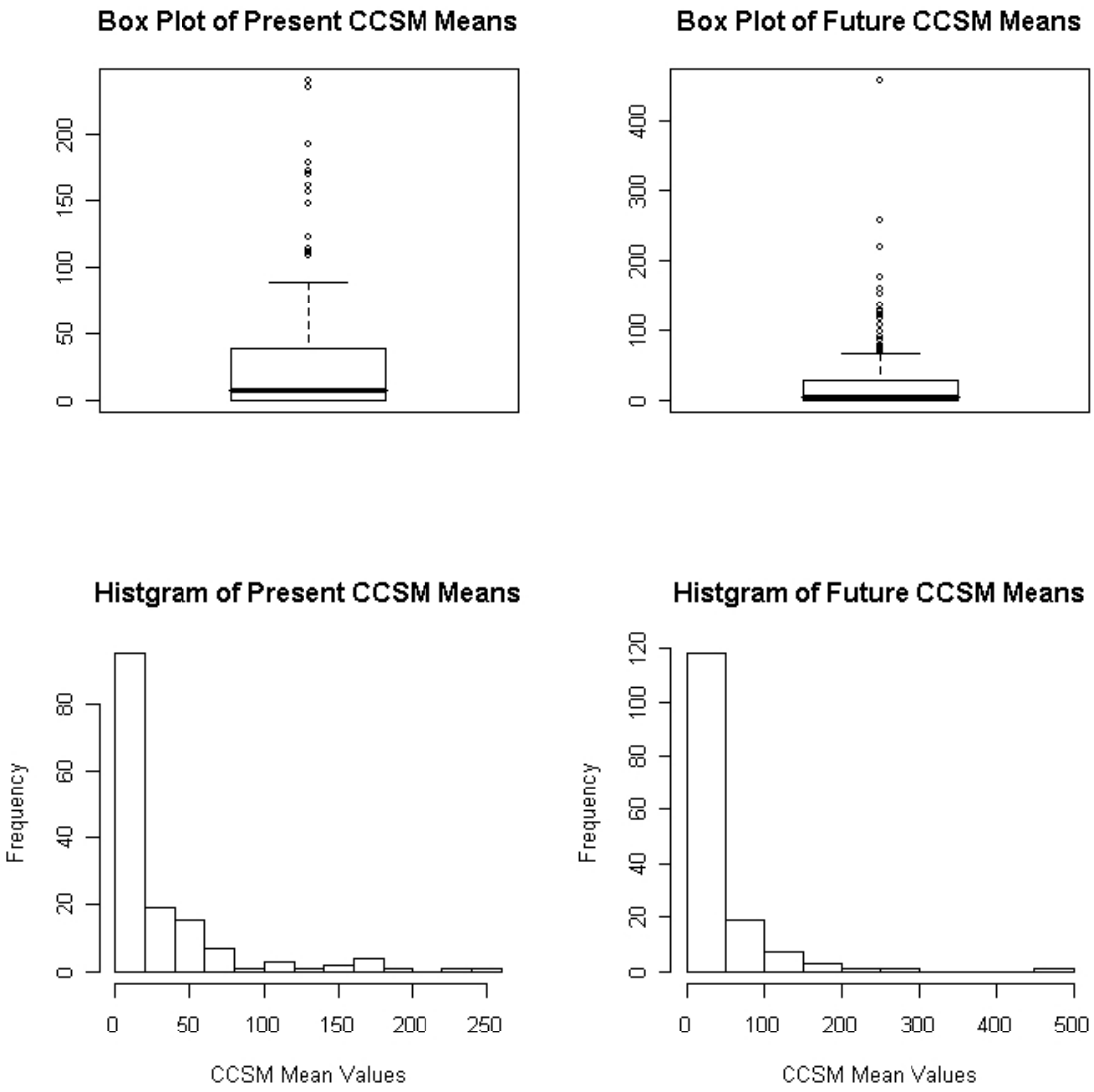



Figure 4.11: *Comparison of Current CCSM Means and Future CCSM Means*

### 4.4.4 Conclusions

The Generalized Linear Mixed Model provides a quite useful way to estimate the spatial dependence structure over a certain area. Further more, this model be applied not only to those extreme value distributions, but also to any other parameterized distributions.

Spectral representation on high dimensional spatial covariance matrices provides an efficient way to do the sampling implementations(MCMC). The approximate decomposition method can also be extended to quite general scenarios without limitations on distributions. Embedding the spatial covariance matrices in matrices with block circulant structure will lose some of the computational efficiency comparing with the spectral method. But on the other hand, it is still quite efficient and worth doing because it is an exact approach. According to the computational results provided in the previous section, the two methods produce very similar results. There is a potential defect on the embedded block circulant approach, which is that it is possible when parameters change, the least number $\bar{m}$ to let the block circulant matrix $C$ be positive definite could increase significantly. In this case, we probably won't have computational efficiency any more.

The simulated results also show the intensive rainfalls, rather than the mean of the rainfalls, will be increasing much in the future. This trend is consistent with the research done by meteorologists before(Karl and Trenberth 2003, etc.). The warming environment will tend to produce more extreme events.

CHAPTER 5

# Models for Rainfall Extremes Based on Mixture of Gaussian Processes

The Generalized Linear Mixed Model works well when the data sets are not too large, but it would require an enormous amount of computation to reconstruct the data for the whole US for a long time period, using this method. As an alternative approach, Smith (2006) has suggested using a transformed, thresholded Gaussian process model. Some details of this approach are outlined in Section 5.1. Nevertheless, it appears that the transformed Gaussian model may not fit the data. The new contribution of the present chapter is to propose an extension of Smith's model that is based on a mixture of Gaussian processes. Some initial results about this are given in Section 5.2 and elaborated in Appendix B.

## 5.1 Single Transformed, Thresholded Gaussian Approach for Rainfall Extremes

Another statistical approach to the rainfall extremes is to think of a transformed, thresholded Gaussian model. The objectives of this approach is quite similar with the previous model:

1. Use spatial statistics to interpolate daily data and estimate a grid-cell average for each day

2. Estimate long-term(50-year) return values based on estimated daily grid-cell averages

3. Compare with NCEP re-analyses and give an evaluation of how good NCEP data are.

### 5.1.1 Statistical Model

For Rainfall data, we still would like to assume they follow a piecewise distribution.

- For the data exceed a threshold $u$, Fit GEV to tails of distribution which is equivalent to

$$1 - F(y) = Pr(Y > y) \approx \frac{1}{T}\left(1 + \xi\frac{y-\mu}{\psi}\right)_+^{-1/\xi} \tag{5.1}$$

  where T is number of relevant days per year

- Estimate $F(0)$ by sample portion of zeros.

- For $0 < y \leq u$(threshold), divide range into some equiprobable intervals(depending on the number of data we have in this interval) and assume $F(y)$ is piecewise linear within each interval.

Now define $Z = \Phi^{-1}(F(Y))$ so that $Z$ has marginal $N[0,1]$ distribution. $u^* = \Phi^{-1}(F(0))$ then becomes the natural threshold of $Z$ values, i.e. $Z$ values are censored at $u^*$. Assume that the $Z$ process is a Gaussian spatial process.

There are several related topics in the literature. Coles and Tawn(1996) proposed a similar approach with assuming $Z$ is max-stable. Their paper is limited to a particular representation of max-stable processes and the estimation methods are very intensive in computation for our problem. Sansó and Guenni(2000,2004) developed an embedded Gaussian model which is similar to the one proposed here, but due to much less data they used, MCMC can proceed directly in their case which is not implementable in the high dimensional cases.

### 5.1.2 Parameter Estimation of a Thresholded Gaussian Process

Suppose $Z = (Z_1, ..., Z_n)^T$ has a multivariate normal distribution $N[0, \Sigma(\theta)]$ where $\Sigma(\theta)$ is a known function of finite-dimensional parameter $\theta$. Also assume $Z_i$'s are censored at $u$, which means we only observe those values $Z_i$ for which $Z_i > u$.

After considering various alternatives, Smith proposed a method for fitting this model based on a pairwise approximation to the likelihood. First choose a fixed ordering of the

indices. Then use a pairwise approximation

$$L(Z_1) \cdot L(Z_2|Z_1) \cdot L(Z_3|Z_2) \cdots \tag{5.2}$$

Here $L(Z_{i+1}|Z_i)$ is the conditional likelihood of $Z_{i+1}$ given $Z_i$ which allows for censoring,i.e.

1. $L = f(Z_{i+1}|Z_i)$, if $Z_i > u, Z_{i+1} > u$,
2. $L = Pr\{Z_{i+1} \leq u|Z_i\}$, if $Z_i > u, Z_{i+1} \leq u$,
3. $L = \frac{Pr\{Z_i \leq u|Z_{i+1}\}f(Z_{i+1})}{Pr\{Z_i \leq u\}}$, if $Z_i \leq u, Z_{i+1} > u$,
4. $L = \frac{Pr\{Z_i \leq u, Z_{i+1} \leq u\}}{Pr\{Z_i \leq u\}}$, if $Z_i \leq u, Z_{i+1} \leq u$.

The estimating equations are unbiased. The method is motivated by similar methods in the spatial statistics literature using approximations to the likelihood function (Vecchia (1988), Stein, Chi and Welty (2004)).

### 5.1.3 Application to Rainfall Data

#### Some Computational Results by Smith

There is a paper in preparation by Smith(2006) which have some results using this method applied to the North American Rainfall Data(by NCDC). For instance, within the grid cell in Alabama, using natural threshold($\Phi^{-1}(F(0))$) $u^* = 0.5306$, and other two alternative thresholds $u^* = 1$ and $u^* = 2$, the results(parameter estimations and SE's) are as follows:

| Threshold | $\hat{\theta}_1$ | $\hat{\theta}_3$ | $\hat{\theta}_3$ |
|---|---|---|---|
| 0.5306 | 1.037(0.015) | 0.665(0.015) | 1.37(0.04) |
| 1 | 1.007(0.13) | 0.630(0.17) | 0.84(0.04) |
| 2 | 1.008(0.18) | 0.724(0.62) | 0.15(0.37) |

The estimation is based on exponential covariance function with nugget via pairwise likelihood with ordering of stations that minimizes total distance.

The standard errors for $\theta_3$ are much smaller than the differences between the different estimates of $\theta_3$ associated with different thresholds. Therefore, this points to a weakness

in the model itself, rather than just meaning that we get different estimates because of random variation and we need to derive an alternative model that is consistent with different thresholds.

## 5.2 Alternative approach based on mixtures of Gaussian processes

In fact, in the previous model, $Z$ is treated as a single Gaussian process which is defined by a set of parameters $\{\theta_1, \theta_2, \theta_3\}$. But as we stated before, the estimates of $\theta_3$ are not consistent across different thresholds, so it is possible, $\theta_3$, along with $\theta_1$ and $\theta_2$ are not just coming from one Gaussian Process. That's why we are trying to think about mixtures of Gaussian process as an alternative approach.

The first thing we did is that suppose our previous assumption was correct, i.e., $Z$ is a single Gaussian process, then what the maximum likelihood estimations of parameters are. The parameters should satisfy the equation:

$$E\left\{\frac{\partial \ell_i}{\partial \theta_j}\right\} = 0, \tag{5.3}$$

where $\ell_i$ is the approximate likelihood function of $f(Z_i|Z_{i-1})$ and $j \in \{1, 2, 3\}$. The calculation is not trivial due to the fact that censored data are used here. Appendix B includes the detailed check-up calculation.

But, what we are really interested is to get a formula for $E\left\{\frac{\partial \ell_i}{\partial \theta_j}\right\}$ under wrong model. More clearly, we assume the previous thresholded Gaussian process model is not correct, and the correct $Z$ process is a mixture of Gaussian processes. Then the parameters are coming from different Gaussian processes with certain priors. After getting the formula of $E\left\{\frac{\partial \ell_i}{\partial \theta_j}\right\}$, we are able to estimate those real parameters. We cannot do this directly because apparently, the estimations of parameters we get from data is under a wrong model. This is practically solvable and we would like to introduce a theorem by Huber(1964,1972) first.

**Theorem 5.2.1** (Maximum likelihood type estimators(M-estimators)). *Let $\rho$ be a real valued function of a real parameter, with derivative $\psi = \rho'$. Define a statistics $T_n =$*

*$T_n(X_1, ..., X_n)$ either by*

$$\sum_{i=1}^{n} \rho(X_i - T_n) = \inf_t \sum_{i=1}^{n} \rho(X_i - t), \tag{5.4}$$

*or by*

$$\sum_{i=1}^{n} \psi(X_i - T_n) = 0 \tag{5.5}$$

*Under quite general conditions, $T_n$ converges to $T(F)$, defined by*

$$\int \psi(x - T(F))F(dx) = 0, \tag{5.6}$$

*and $\sqrt{n}(T_n - T(F))$ is asymptotically normal with asymptotic mean 0 and asymptotic variance*

$$\sigma_M^2(F) = \int \Omega_F(x)^2 F(dx), \tag{5.7}$$

*where*

$$\Omega_F(x) = \frac{\psi(x - T(F))}{\int \psi'(x - T(F))F(dx)}. \tag{5.8}$$

If we choose

$$\psi_0(x) = -f_0'(x)/f_0(x), \tag{5.9}$$

for $\psi(x)$, then $T_n$ is the maximum likelihood estimator of $\theta$ for distribution function $F_0$ and will be asymptotically efficient for $F = F_0$.(Huber(1964)).

In our case, because we are using pairwise approximation of likelihood components, i.e.:

$$L(Z_1) \cdot L(Z_2|Z_1) \cdot L(Z_3|Z_2) \cdots ,$$

we are able to apply the theorem stated above and regulate the problem as follows:

We define a mixture model in the following way. Assume there are $M$ Gaussian processes with parameters $\{\theta^{(1)}, ..., \theta^{(M)}\}$. Assume there are also mixture probabilities $\{\pi_1, ..., \pi_M\}$. On any day, we first choose index $m \in \{1, ..., M\}$ with probability $\pi_m$, and then, conditional on that choice, generate a Gaussian process with parameter $\theta^{(m)}$. Suppose we choose $R$ thresholds $u_1, ..., u_R$ and, for each threshold, calculate the estimator $\hat{\theta}^{(r)},\ r = 1, ..., R$ under

the method of Section 4.1. If we can calculate

$$E_{\theta_j^{(m)},\pi}\left\{\frac{\partial \ell_i}{\partial \theta_j^{(m)}}\left(\hat{\theta}^{(r)}\right)\right\},$$

where $m \in 1, 2, ..., M$ and $r \in 1, 2, ..., R$, then by solving the set of equations:

$$\sum_{i=1}^{n} E_{\theta_j^{(m)},\pi}\left\{\frac{\partial \ell_i}{\partial \theta_j^{(m)}}\left(\hat{\theta}^{(r)}\right)\right\} = 0, \tag{5.10}$$

we can finally find the approximations for $\{\theta^{(1)}, \theta^{(2)}, ..., \theta^{(M)}\}$ when $n$ is sufficiently large. This will lead us to the correct mixture of Gaussian processes model.

CHAPTER 6

# Future Works

## 6.1 Computation Based on Mixture of Transformed, Thresholded Gaussian Approach

The transformed, thresholded Gaussian approach will provide us a more reasonable way to handle the whole rainfall data set over the United States if the mixture assumption is correct. After setting certain thresholds and making calculations under "wrong" model, we can obtain the "correct" parameters by using the relationship between the "wrong" model and the "correct" model(Single Gaussian process v.s. Mixture of Gaussian Processes).

## 6.2 Further Consideration on Approximation Method of Embedded Block Circulant Approach

The spectral representation is a good approximation method as we stated before, but it has its own disadvantage. Though each of the approximated eigenvalues is very close to the real one, it is still hard to know how well the global performance in high dimensional cases. So it is probably promising to consider some other matrix approximation methods under more meaningful matrix norms based on the embedded block circulant approach.

### 6.2.1 Refined Embedded Block Circulant Approach to Spatial Dependence Structures

In chapter 3, an exact statistical approach on estimating spatial dependence structures has been developed(embedded block circulant approach). But the efficiency of the approach mostly depends on the number $\bar{m}$, which is the least number letting the embedded matrix $C$

be positive definite. In some cases, especially when we update the corresponding parameters of $C$, the number $\bar{m}$ will become very hard to control. So it is necessary to find an alternative way to refine the approach.

Now let us still consider the embedded block circulant matrix $C$ as follows:

$$C = \{c_{u_1u_2}\}, \qquad u_1, u_2 \in I^*(m) \tag{6.1}$$

where

$$I^*(m) = \{u = (i,j) : 0 \leq |i| \leq m[1]-1, 0 \leq |j| \leq m[2]-1\} \tag{6.2}$$

and $c_{u_1u_2}$ is defined by $C_0$ which is the exponential covariance function with nugget as we indicated before.

Now suppose $C$ is not a positive definite matrix. In the previous chapter, the way to handle it is to increase the number $\bar{m}$ until $C$ is positive definite and let it be a covariance matrix of a real random process, but it is possible to think it in different ways.

One possibility is to consider an approximation to $C$. Because $C$ is not positive definite any more, so some of the eigenvalues, though still real and can be efficiently calculated by two-dimensional discrete Fast Fourier Transform, are negative. Now we still assume by eigendecompsition

$$C = Q\Lambda Q^* \tag{6.3}$$

and write

$$\Lambda = \Lambda_+ - \Lambda_- \tag{6.4}$$

where the non-zero entries of $\Lambda_+$ and $\Lambda_-$ corresponding to positive and negative eigenvalues of $C$.

Suppose the approximation of $C$ has the form

$$C^* = \rho^2 Q \Lambda_+ Q^* \tag{6.5}$$

and set the trace of $C*$ and $C$ be the same, which brings us

$$\rho^2 = \frac{tr(\Lambda)}{tr(\Lambda_+)} \tag{6.6}$$

The approximation of $\hat{Z}$ has the same marginal distribution with $\hat{Z}$(Wood and Chan, 1994). Alternatively, letting $\rho^2$ be $\left(\frac{tr(\Lambda)}{tr(\Lambda_+)}\right)^2$ will also lead us to a good approximation.

### 6.2.2 Other Thoughts

We've considered trace norm, but it is definitely not a good matrix norm at any point. There are two essential matrix norms in the literature – the operator norm(strong norm) and the Hilbert-Schmidt norm(weak norm or Frobenius norm). We still would like to break the eigenvalue matrix $\Lambda$ into $\Lambda_+$ and $\Lambda_-$, and try to find approximations only using positive numbers, which could be from both positive and negative eigenvalues, under these two norms. Then after investigating issues like marginal distributions, we may be able to find some more accurate approximation comparing with the spectral method. The computational efficiency is still retained because the main computational issue is still the discrete Fourier transform.

# APPENDIX A: SAMPLING TECHNIQUES

Sampling techniques are always used for simulating samples from the posterior distribution of the parameters of interest which does not have an analytic closed form in Bayesian inference.

There are several indirect sampling techniques available in the literature, for instance, importance sampling, rejection sampling, weighted bootstrap and MCMC.

## .1 Importance Sampling

This approach was stated by Hammersley and Handscomb(1964) first and discussed for Bayesian analysis by Geweke(1989).

Suppose we want to approximate a posterior expectation:

$$E[f\,(\theta|x)] = \frac{\int f(\theta)L(\theta|x)\pi(\theta)d\theta}{\int L(\theta|x)\pi(\theta)d\theta} \tag{7}$$

where $L(\theta|x) \propto p(x|\theta)$ is likelihood function and $\pi(\theta)$ is the prior distribution of parameters. If we are able to approximate the normalized likelihood times the prior, $\frac{L(\theta|x)\pi(\theta)}{c}$, by some density $g(\theta)$, then we can define the weight function

$$\omega(\theta) = \frac{L(\theta|x)\pi(\theta)}{g(\theta)} \tag{8}$$

We have

$$\begin{aligned} E[f(\theta|x)] &= \frac{\int f(\theta)L(\theta|x)\pi(\theta)d\theta}{\int L(\theta|x)\pi(\theta)d\theta} \\ &\approx \frac{\frac{1}{N}\sum_{j=1}^{N} f(\theta_j)\omega(\theta_j)}{\frac{1}{N}\sum_{j=1}^{N}\omega(\theta_j)} \end{aligned} \tag{9}$$

where $\theta_j$ are i.i.d samples from *importance function* $g(\theta)$.

## .2 Rejection Sampling

Rejection sampling is widely used in random variate generation. The detailed discussion are given in the books by Ripley(1987) and Devroye(1986).

Suppose there exists a constant $M > 0$ and a smooth density $g(\theta)$, called the *envelope function*, such that

$$L(\theta|x)\pi(\theta) < Mg(\theta) \tag{10}$$

for all $\theta$.

The algorithm of *rejection sampling* is as following:

1. Generate $\theta_j \sim g(\theta)$
2. Generate $U \sim Uniform(0, 1)$
3. If $MUg(\theta_j) < L(\theta_j)\pi(\theta_j)$, then accept $\theta_j$, otherwise reject $\theta_j$
4. Return to the first step and repeat until the desired sample $\theta_j, j = 1, ..., N$ is obtained.

## .3 Weighted Bootstrap

This method was proposed by Smith and Gelfand(1992). Suppose we are not able to find an appropriate $M$ for the rejection sampling, while we have a sample $\theta_1, ..., \theta_N$ from some approximating density $g(\theta)$. Define

$$\omega_i = \frac{L(\theta_i|x)\pi(\theta_i)}{g(\theta_i)},$$
$$q_i = \frac{\omega_i}{\sum_{i=1}^{N} \omega_i}.$$

Draw $\theta^*$ from the discrete distribution over $\{\theta_1, ..., \theta_N\}$, which places mass $q_i$ at $\theta_i$. Then $\theta^*$ is a sample from

$$p(\theta|x) = \frac{L(\theta|x)\pi(\theta)}{\int L(\theta|x)\pi(\theta)d\theta}, \tag{11}$$

with the approximation improving as $N \to \infty$. This is a *weighted bootstrap*.

## .4 Markov Chain Monte Carlo Methods

For many problems, especially *high dimensional* ones, it is quite difficult to apply the above sampling techniques because they are all noniterative methods. They draw a sample of size $N$ and stop. Furthermore, it is very hard or even impossible to find an importance sampling density or envelop function which is reasonably accurate for the log posterior. Markov Chain Monte Carlo(MCMC) methods are standard ways to solve these problems.

### .4.1 Gibbs Sampling

Suppose we have a collection of $k$ random variables $U = (U_1, ..., U_k)$. We assume that the full conditional distributions

$$p(u_i|u_j), j \neq i, i = 1, ..., k \tag{12}$$

can be calculated up to a normalized constant, and therefore are available for sampling.

Besag(1974) proves under mild conditions, the one-dimensional distributions uniquely determine the full joint distribution $[U_1, ..., U_k]$, and hence all marginal distributions $[U_i], i = 1, ..., k$. The *Gibbs Sampling* algorithm proceeds as following:

0) Suppose we have a set of arbitrary starting values $U_1^{(0)}, ..., U_k^{(0)}$.

1) Draw $U_1^{(1)}$ from $[U_1|U_2^{(0)}, ..., U_k^{(0)}]$

2) Draw $U_2^{(1)}$ from $[U_2|U_1^{(1)}, U_3^{(0)}, ..., U_k^{(0)}]$

$\vdots$

k) Draw $U_k^{(1)}$ from $[U_k|U_1^{(1)}, ..., U_{k-1}^{(1)}]$

This is one iteration of the Gibbs sampler and after $t$ such iterations, we are able to obtain $(U_1^{(t)}, ..., U_k^{(t)})$

The convergence of Gibbs sampler has been proved by Geman and Geman(1984) and Schervish & Carlin(1992) and we have the following result: For the Gibbs sampling algorithm stated above, $(U_1^{(t)}, ..., U_k^{(t)}) \xrightarrow{d} [U_1, ..., U_k]$ as $t \to \infty$ and the convergence is exponential in $t$.

### .4.2 Metropolis-Hastings Algorithm

This method is first discussed by Metropolis *et al.*(1953). Hastings(1970) generalized the algorithm and introduced it for statistical problems.

Suppose we want to sample from the joint distribution $[U_1, ..., U_k][U]$ with density $p(u)$. Let $q(v, u)$ be a density in $u$ such that $q(u, v) = q(v, u)$. The function $q$ is called a candidate or proposal density. The Metropolis algorithm can be written as following:

1) Draw $v \sim q(\cdot, u)$, where $u = U^{(t-1)}$, the current state of the Metropolis algorithm.

2) Compute the odds ratio

$$r = \frac{p(v)}{p(u)} = \frac{L(v)\pi(v)}{L(u)\pi(u)}$$

3) If $r \geq 1$, accept $v$ and let $U^{(t)} = v$

4) If $r < 1$, set

$$U^{(t)} = \begin{cases} v & \textit{with probability } r \\ u & \textit{with probability } 1 - r \end{cases} \tag{13}$$

**Theorem .4.1.** *For the Metropolis algorithm, under mild conditions,* $[U^{(t)} \rightarrow [U]]$ *as* $t \rightarrow \infty$.

Hastings(1970) drops the requirement that $q(\cdot, \cdot)$ is symmetric and redefined the odds ratio as

$$r = \frac{p(v)q(v, u)}{p(u)q(u, v)}. \tag{14}$$

The modification generalized the sampling technique and the algorithm still converges to the required distribution for any candidate density $q$. The detailed proof due to Hastings (1970).

# APPENDIX B: MAXIMUM LIKELIHOOD CALCULATIONS FOR THRESHOLDED GAUSSIAN PROCESS

Suppose $Z$ is a thresholded Gaussian process. "Thresholded" means if we have some observations $\{Z_1, ..., Z_n\}$, we are only interested in those values above a fixed threshold $u$ and the rest are censored. Furthermore, we assume $\{Z_1, ..., Z_n\}$ follow a multivariate normal distribution $MVN[0, \Sigma(\Theta)]$.

The way to estimate the parameters $\Theta$ here is to apply pairwise principle to the whole likelihood under some fixed ordering of indices. That is, concentrate on the pairwise approximation of likelihood function:

$$L(Z_1) \cdot L(Z_2|Z_1) \cdot L(Z_3|Z_2) \cdots \tag{15}$$

instead of the exact likelihood:

$$L(Z_1) \cdot L(Z_2|Z_1) \cdot L(Z_3|Z_2, Z_1) \cdots . \tag{16}$$

So we have

- $L_i = f(Z_{i+1}|Z_i)$, if $Z_i > u, Z_{i+1} > u$,
- $L_i = Pr\{Z_{i+1} \leq u|Z_i\}$, if $Z_i > u, Z_{i+1} \leq u$,
- $L_i = \frac{Pr\{Z_i \leq u|Z_{i+1}\}f(Z_{i+1})}{Pr\{Z_i \leq u\}}$, if $Z_i \leq u, Z_{i+1} > u$,
- $L_i = \frac{Pr\{Z_i \leq u, Z_{i+1} \leq u\}}{Pr\{Z_i \leq u\}}$, if $Z_i \leq u, Z_{i+1} \leq u$.

and the corresponding negative log-likelihood's are as follows:

- $l_i = \frac{1}{2}\log(1 - \rho_i^2) + \frac{1}{2}\frac{(Z_i - \rho_i Z_{i-1})^2}{1-\rho_i^2}$, if $Z_{i-1} > u$, $Z_i > u$

- $l_i = -\frac{1}{2}\log\left\{\Phi\left(\frac{u-\rho_i Z_{i-1}}{\sqrt{1-\rho_i^2}}\right)\right\}$, if $Z_{i-1} > u$, $Z_i \leq u$
- $l_i = -\frac{1}{2}\log\left\{\Phi\left(\frac{u-\rho_i Z_i}{\sqrt{1-\rho_i^2}}\right)\right\}$, if $Z_{i-1} \leq u$, $Z_i > u$
- $l_i = -\log\{\Phi_2(u,u,\rho_i)\}$, if $Z_{i-1} \leq u$, $Z_i \leq u$

where $\rho_i = \rho_i(\theta) = Corr(Z_{i-1}, Z_i)$ is the correlation function and $\Phi$ & $\Phi_2$ are the normal distribution function and bivariate normal distribution function, respectively. Consider $\begin{pmatrix} Z_{i-1} \\ Z_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix}\right]$, i.e., we will be evaluating the expectation assuming the model is correct first. The estimating equations are unbiased for this method:

$$E\left\{\frac{\partial l_i}{\partial \theta_j}\right\} = 0. \tag{17}$$

(17) gives out the relationship between correlation parameters $\theta_j$ and threshold $u$.

For the case $Z_{i-1} > u$ & $Z_i > u$, we have:

$$\frac{\partial l_i}{\partial \theta_j} = -\frac{\rho_i}{1-\rho_i^2}\frac{\partial \rho_i}{\partial \theta_j} + \frac{(\rho_i Z_i - Z_{i-1})(Z_i - \rho_i Z_{i-1})}{(1-\rho_i^2)^2}\frac{\partial \rho_i}{\partial \theta_j}, \tag{18}$$

For the case $Z_{i-1} \leq u$ & $Z_i \leq u$, we have:

$$\frac{\partial l_i}{\partial \theta_j} = \frac{1}{\Phi_2(u,u,\rho_i)}\int_{-\infty}^{u}\int_{-\infty}^{u}\left(-\frac{\rho_i}{1-\rho_i^2} + \frac{(\rho_i y - x)(y - \rho_i x)}{(1-\rho_i^2)^2}\right)\phi(x,y,\rho_i)\frac{\partial \rho_i}{\partial \theta_j}dxdy \tag{19}$$

For the case $Z_{i-1} > u$ & $Z_i \leq u$, we have:

$$\frac{\partial l_i}{\partial \theta_j} = -\frac{1}{\sqrt{2\pi}}\frac{1}{2\Phi(\frac{u-\rho_i Z_{i-1}}{\sqrt{1-\rho_i^2}})}\exp^{-\frac{(u-\rho_i Z_{i-1})^2}{2(1-\rho_i^2)}}\frac{u\rho_i - Z_{i-1}}{(1-\rho_i^2)^{3/2}}\frac{\partial \rho_i}{\partial \theta_j} \tag{20}$$

For the case $Z_{i-1} \leq u$ & $Z_i > u$, we have:

$$\frac{\partial l_i}{\partial \theta_j} = -\frac{1}{\sqrt{2\pi}}\frac{1}{2\Phi(\frac{u-\rho_i Z_i}{\sqrt{1-\rho_i^2}})}\exp^{-\frac{(u-\rho_i Z_i)^2}{2(1-\rho_i^2)}}\frac{u\rho_i - Z_i}{(1-\rho_i^2)^{3/2}}\frac{\partial \rho_i}{\partial \theta_j} \tag{21}$$

As preparation, we would like to evaluate the following expressions first:

1. $E\{Z_i^2 I(Z_{i-1} > u, Z_i > u)\}$,

2. $E\{Z_{i-1}Z_iI(Z_{i-1} > u, Z_i > u)\}$,

3. $E\{Z_i^2 I(Z_{i-1} \le u, Z_i \le u)\}$,

4. $E\{Z_{i-1}Z_iI(Z_{i-1} \le u, Z_i \le u)\}$.

In fact, suppose $Z_{i-1} = \rho_i Z_i + \sqrt{1-\rho_i^2}\hat{Z}_i$, where $\hat{Z}_i \sim N[0,1]$ is a Gaussian process independent with $Z_{i-1}$, and do the calculation term by term, we have:

$$
\begin{aligned}
& E\{Z_i^2 I(Z_{i-1} > u, Z_i > u)\} \\
= \; & E\{Z_i^2(1-\Phi(\frac{u-\rho_i Z_i}{\sqrt{1-\rho_i^2}}))I(Z_i > u)\} \\
= \; & u\phi(u) + 1 - \Phi(u) - \int_u^\infty x^2 \Phi(\frac{u-\rho_i x}{\sqrt{1-\rho_i^2}})\phi(x)dx \\
= \; & u\phi(u) + 1 - \Phi(u) + \int_u^\infty x^2 \Phi(\frac{u-\rho_i x}{\sqrt{1-\rho_i^2}})d\Phi(x) \\
= \; & u\phi(u) + 1 - \Phi(u) - \left[x\Phi(\frac{u-\rho_i x}{\sqrt{1-\rho_i^2}})\phi(x)\right]_u^\infty \\
& - \int_u^\infty \phi(x)\Phi(\frac{u-\rho_i x}{\sqrt{1-\rho_i^2}})dx + \frac{\rho_i}{\sqrt{1-\rho_i^2}}\int_u^\infty x\phi(x)\phi(\frac{u-\rho_i x}{\sqrt{1-\rho_i^2}})dx \\
= \; & u\phi(u) + 1 - \Phi(u) - u\Phi(u\sqrt{\frac{1-\rho_i}{1+\rho_i}})\phi(u) - Pr\{Z_i > u, Z_{i-1} \le u\} \\
& + \frac{\rho_i}{2\pi\sqrt{1-\rho_i^2}}\int_u^\infty x \exp^{-\frac{x^2}{2}} \exp^{-\frac{(u-\rho_i x)^2}{2(1-\rho_i^2)}} dx \\
= \; & u\phi(u) + 1 - \Phi(u) - u\Phi(u\sqrt{\frac{1-\rho_i}{1+\rho_i}})\phi(u) - \Phi(u) + \Phi_2(u,u,\rho_i) \\
& + \frac{\rho_i}{2\pi\sqrt{1-\rho_i^2}}\exp^{-\frac{u^2}{2}}\int_u^\infty (x - u\rho_i + u\rho_i)\exp^{-\frac{(x-u\rho_i)^2}{2(1-\rho_i^2)}} dx \\
= \; & u\phi(u) + \Phi_2(u,u,\rho_i) + 1 - 2\Phi(u) - u\Phi(u\sqrt{\frac{1-\rho_i}{1+\rho_i}})\phi(u) \\
& - \left[\frac{\rho_i\sqrt{1-\rho_i^2}}{2\pi}\exp^{-\frac{u^2}{2}}\exp^{-\frac{(x-u\rho_i)^2}{2(1-\rho_i^2)}}\right]_u^\infty + u\rho_i^2\phi(u)\bar{\Phi}(u\sqrt{\frac{1-\rho_i}{1+\rho_i}}) \\
= \; & \Phi_2(u,u,\rho_i) + 1 - 2\Phi(u) + u(1+\rho_i^2)\bar{\Phi}(u\sqrt{\frac{1-\rho_i}{1+\rho_i}})\phi(u) + \frac{\rho_i\sqrt{1-\rho_i^2}}{2\pi}\exp^{-\frac{u^2}{1+\rho_i}} \quad (22)
\end{aligned}
$$

and

$$
\begin{aligned}
& E\{Z_i^2 I(Z_{i-1} \le u, Z_i \le u)\} \\
= \; & E\{Z_i^2 \Phi(\frac{u-\rho_i Z_i}{\sqrt{1-\rho_i^2}}) I(Z_i \le u)\} \\
= \; & \int_{-\infty}^{u} x^2 \Phi(\frac{u-\rho_i x}{\sqrt{1-\rho_i^2}}) \phi(x) dx \\
= \; & -\int_{-\infty}^{u} x \Phi(\frac{u-\rho_i x}{\sqrt{1-\rho_i^2}}) d\phi(x) \\
= \; & -\left[ x\Phi(\frac{u-\rho_i x}{\sqrt{1-\rho_i^2}})\phi(x) \right]_{-\infty}^{u} + \int_{-\infty}^{u} \phi(x)\Phi(\frac{u-\rho_i x}{\sqrt{1-\rho_i^2}}) dx \\
& -\frac{\rho_i}{\sqrt{1-\rho_i^2}} \int_{-\infty}^{u} x\phi(x)\phi(\frac{u-\rho_i x}{\sqrt{1-\rho_i^2}}) dx \\
= \; & -u\Phi(u\sqrt{\frac{1-\rho_i}{1+\rho_i}})\phi(u) + \Phi_2(u,u,\rho_i) + \frac{\rho_i\sqrt{1-\rho_i^2}}{2\pi} \exp^{-\frac{u^2}{1+\rho_i}} - u\rho_i^2 \Phi(u\sqrt{\frac{1-\rho_i}{1+\rho_i}})\phi(u) \\
= \; & -u(1+\rho_i^2)\Phi(u\sqrt{\frac{1-\rho_i}{1+\rho_i}})\phi(u) + \Phi_2(u,u,\rho_i) + \frac{\rho_i\sqrt{1-\rho_i^2}}{2\pi} \exp^{-\frac{u^2}{1+\rho_i}} \qquad (23)
\end{aligned}
$$

and

$$
\begin{aligned}
& E\{Z_i Z_{i-1} I(Z_{i-1} > u, Z_i > u)\} \\
= \; & E\{Z_i(\rho_i Z_i + \sqrt{1-\rho_i^2}\hat{Z}_i) I(Z_i > u, \hat{Z}_i > \frac{u-\rho_i Z_i}{\sqrt{1-\rho_i^2}})\} \\
= \; & \rho_i E\{Z_i^2 \bar{\Phi}(\frac{u-\rho_i Z_i}{\sqrt{1-\rho_i^2}}) I(Z_i > u)\} + \sqrt{1-\rho_i^2} E\{Z_i \phi(\frac{u-\rho_i Z_i}{\sqrt{1-\rho_i^2}}) I(Z_i > u)\} \\
= \; & \rho_i \left( \Phi_2(u,u,\rho_i) + 1 - 2\Phi(u) + u(1+\rho_i^2)\bar{\Phi}(u\sqrt{\frac{1-\rho_i}{1+\rho_i}})\phi(u) + \frac{\rho_i\sqrt{1-\rho_i^2}}{2\pi} \exp^{-\frac{u^2}{1+\rho_i}} \right) \\
& \frac{1-\rho_i^2}{\rho_i} \left( u\rho_i^2 \bar{\Phi}(u\sqrt{\frac{1-\rho_i}{1+\rho_i}})\phi(u) + \frac{\rho_i\sqrt{1-\rho_i^2}}{2\pi} \exp^{-\frac{u^2}{1+\rho_i}} \right) \\
= \; & \rho_i(\Phi_2(u,u,\rho_i) + 1 - 2\Phi(u)) + 2u\rho_i\bar{\Phi}(u\sqrt{\frac{1-\rho_i}{1+\rho_i}})\phi(u) + \frac{\sqrt{1-\rho_i^2}}{2\pi} \exp^{-\frac{u^2}{1+\rho_i}} \qquad (24)
\end{aligned}
$$

and

$$
\begin{aligned}
& E\{Z_i Z_{i-1} I(Z_{i-1} \le u, Z_i \le u)\} \\
= \; & E\{Z_i(\rho_i Z_i + \sqrt{1-\rho_i^2}\hat{Z}_i) I(Z_i \le u, \hat{Z}_i \le \frac{u-\rho_i Z_i}{\sqrt{1-\rho_i^2}})\} \\
= \; & \rho_i E\{Z_i^2 \Phi(\frac{u-\rho_i Z_i}{\sqrt{1-\rho_i^2}}) I(Z_i \le u)\} - \sqrt{1-\rho_i^2} E\{Z_i \phi(\frac{u-\rho_i Z_i}{\sqrt{1-\rho_i^2}}) I(Z_i \le u)\} \\
= \; & \rho_i \left( -u(1+\rho_i^2)\Phi(u\sqrt{\frac{1-\rho_i}{1+\rho_i}})\phi(u) + \Phi_2(u,u,\rho_i) + \frac{\rho_i\sqrt{1-\rho_i^2}}{2\pi} \exp^{-\frac{u^2}{1+\rho_i}} \right) \\
& + \frac{1-\rho_i^2}{\rho_i} \left( -u\rho_i^2 \Phi(u\sqrt{\frac{1-\rho_i}{1+\rho_i}})\phi(u) + \frac{\rho_i\sqrt{1-\rho_i^2}}{2\pi} \exp^{-\frac{u^2}{1+\rho_i}} \right) \\
= \; & \rho_i \Phi_2(u,u,\rho_i) - 2u\rho_i \Phi(u\sqrt{\frac{1-\rho_i}{1+\rho_i}})\phi(u) + \frac{\sqrt{1-\rho_i^2}}{2\pi} \exp^{-\frac{u^2}{1+\rho_i}} . \qquad (25)
\end{aligned}
$$

Furthermore, we have

$$
\begin{aligned}
& E\left\{ \frac{\partial l_i}{\partial \theta_j} I(Z_{i-1} \le u, Z_i > u) \right\} = E\left\{ \frac{\partial l_i}{\partial \theta_j} I(Z_{i-1} > u, Z_i \le u) \right\} \\
= \; & E\left\{ -\frac{1}{\sqrt{2\pi}} \frac{1}{2\Phi(\frac{u-\rho_i Z_i}{\sqrt{1-\rho_i^2}})} \exp^{-\frac{(u-\rho_i Z_i)^2}{2(1-\rho_i^2)}} \frac{u\rho_i - Z_i}{(1-\rho_i^2)^{3/2}} \frac{\partial \rho_i}{\partial \theta_j} \Phi(\frac{u-\rho_i Z_i}{\sqrt{1-\rho_i^2}}) I(Z_i > u) \right\} \\
= \; & -\frac{1}{2(1-\rho_i^2)^{3/2}} \frac{\partial \rho_i}{\partial \theta_j} E\left\{ (u\rho_i - Z_i)\phi(\frac{u-\rho_i Z_i}{\sqrt{1-\rho_i^2}}) I(Z_i > u) \right\} \\
= \; & -\frac{1}{2(1-\rho_i^2)^{3/2}} \frac{\partial \rho_i}{\partial \theta_j} \left( u\rho_i \int_u^\infty \phi(\frac{u-\rho_i x}{\sqrt{1-\rho_i^2}})\phi(x)dx - \int_u^\infty x\phi(\frac{u-\rho_i x}{\sqrt{1-\rho_i^2}})\phi(x)dx \right) \\
= \; & -\frac{u\rho_i}{2(1-\rho_i^2)} \frac{\partial \rho_i}{\partial \theta_j} \bar{\Phi}(u\sqrt{\frac{1-\rho_i}{1+\rho_i}})\phi(u) \\
& + \frac{u\rho_i}{2(1-\rho_i^2)} \frac{\partial \rho_i}{\partial \theta_j} \bar{\Phi}(u\sqrt{\frac{1-\rho_i}{1+\rho_i}})\phi(u) + \frac{1}{4\pi\sqrt{1-\rho_i^2}} \frac{\partial \rho_i}{\partial \theta_j} \exp^{-\frac{u^2}{1+\rho_i}} \\
= \; & \frac{1}{4\pi\sqrt{1-\rho_i^2}} \frac{\partial \rho_i}{\partial \theta_j} \exp^{-\frac{u^2}{1+\rho_i}}, \qquad (26)
\end{aligned}
$$

$$
\begin{aligned}
& E\left\{\frac{\partial l_i}{\partial \theta_j} I(Z_{i-1} > u, Z_i > u)\right\} \\
= & -\frac{\rho_i}{1-\rho_i^2}\frac{\partial \rho_i}{\partial \theta_j} Pr\{Z_{i-1} > u, Z_i > u\} + \frac{2\rho_i}{(1-\rho_i^2)^2}\frac{\partial \rho_i}{\partial \theta_j} E\left\{Z_i^2 I(Z_{i-1} > u, Z_i > u)\right\} \\
& -\frac{1+\rho_i^2}{(1-\rho_i^2)^2}\frac{\partial \rho_i}{\partial \theta_j} E\left\{Z_i Z_{i-1} I(Z_{i-1} > u, Z_i > u)\right\} \\
= & -\frac{1}{4\pi\sqrt{(1-\rho_i^2)}} \exp^{-\frac{u^2}{1+\rho_i}} \frac{\partial \rho_i}{\partial \theta_j}, \qquad (27)
\end{aligned}
$$

and

$$
\begin{aligned}
& E\left\{\frac{\partial l_i}{\partial \theta_j} I(Z_{i-1} \le u, Z_i \le u)\right\} \\
= & -\frac{\rho_i}{1-\rho_i^2}\frac{\partial \rho_i}{\partial \theta_j} Pr\{Z_{i-1} \le u, Z_i \le u\} + \frac{2\rho_i}{(1-\rho_i^2)^2}\frac{\partial \rho_i}{\partial \theta_j} E\left\{Z_i^2 I(Z_{i-1} \le u, Z_i \le u)\right\} \\
& -\frac{1+\rho_i^2}{(1-\rho_i^2)^2}\frac{\partial \rho_i}{\partial \theta_j} E\left\{Z_i Z_{i-1} I(Z_{i-1} \le u, Z_i \le u)\right\} \\
= & -\frac{1}{4\pi\sqrt{(1-\rho_i^2)}} \exp^{-\frac{u^2}{1+\rho_i}} \frac{\partial \rho_i}{\partial \theta_j}. \qquad (28)
\end{aligned}
$$

So, we have

$$
\begin{aligned}
E\left\{\frac{\partial l_i}{\partial \theta_j}\right\} = & E\left\{\frac{\partial l_i}{\partial \theta_j} I(Z_{i-1} \le u, Z_i > u)\right\} + E\left\{\frac{\partial l_i}{\partial \theta_j} I(Z_{i-1} > u, Z_i \le u)\right\} \\
& + E\left\{\frac{\partial l_i}{\partial \theta_j} I(Z_{i-1} \le u, Z_i \le u)\right\} + E\left\{\frac{\partial l_i}{\partial \theta_j} I(Z_{i-1} > u, Z_i > u)\right\} \\
= & \ 0.
\end{aligned}
$$

# BIBLIOGRAPHY


[1] Banerjee, S., Carlin, B.P. and Gelfand, A.E. *Hierarchichal Modeling and Analysis for Spatial Data* CRC Press/Chapman and Hall, Boca Raton, FL, 2004

[2] Chib, Siddhartha and Greenberg E. "Understanding the Metropolis–Hastings Algorithm" *American Statistician*, **49(4)**, 327–335, 1995.

[3] Coles, S.G. and Tawn, J.A. "Modelling Extremes of a Real Rainfall Process" *J.R. Statist. Soc. B* **58**, 329–347, 1996.

[4] Cressie, N.A.C. *Statistics of Spatial Data* Wiley, New York, 2nd edition, 1993.

[5] Devroye, L. *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, 1986.

[6] Diggle, P.J., Tawn, J. and Moyeed, R.A. "Model-based Geostatistics (with discussion)" *Applied Statistics* **47**, 299–350, 1998.

[7] Donnelly, T.G. "Algorithm 462. Bivariate Normal Distribution" *Commun. Assoc. Comput. Mach.* **16**, 638, 1973.

[8] Geman, S. and Geman, D. "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images" *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721-741, 1984.

[9] Groisman, P.Y., Karl, T.R., Easterling, D.R., Knight, R.W., Jamason, P. F., Hennessy, K.J., Suppiah, R., Page, C.M., Wibig, J., Foruniak, K. Razuvaev, V.N., Douglas, A., Forland, E. and Zhai,P "Changes in the Probability of Heavy Precipitation: Important Indicators of Climatic Change" *Climatic Change* **42**, 243–283, 1999.

[10] Groisman, P.Y., Knight, R.W. and Karl, T.R. "Heavy Precipitation and High Streamflow in the Contiguous United States: Trends in the 20th Century" *Bull. Amer. Meteor. Soc.* **82**, 219-246, 2001.

[11] Groisman, P., Knight, R., Easterling, D.R., Karl, T.R. and Hegerl, G. "Trends in Precipitation Intensity in the Climate Record" *J. Climate*, submitted, 2004.

[12] Hammersley, J.M. and Handscomb, D.C. *Monte Carlo Methods* Wiley, New York, 1964.

[13] Hastings W.K. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications" *Biometrika*, **57(1)**, 97-109, 1970.

[14] Huber, P.J. "Robust Estimation of a Location Parameter" *Ann. Math. Statist.* **35**, 73–101, 1964.

[15] Huber, P.J. "Robust Statistics: A Review" *Ann. Math. Statist.* **43**, 1041–1067, 1972.

[16] Leadbetter, M.R., Lindgren G. and Rootzén H. *Extremes and Related Properties of Random Sequences and Series* Springer-Verlag, New York, 1983.

[17] Matheron, G. "Principles of Geostatistics" *Economic Geology*, **58**, 1246–1266, 1963

[18] Metropolis N.,Rosenbluth A.N.,Rosenbluth M.N.,Teller A. H., and Teller E. "Equation of State Calculation by Fast Computing Machines" *Journal of Chemical Physics*, **21(6)**, 1087–1092, 1953.

[19] Nychka, D., C.K. Wikle, and J.A. Royle "Multiresolution Models for Nonstationary Spatial Covariance Functions" *Statistical Modelling: An International Journal* **2**, 315–331, 2002.

[20] Pickands, J. "Statistical inference using extreme order statistics" *Ann. Statist.* **3**, 119–131, 1975.

[21] Resnick, S.I. *Extreme Values, Regular Variation and Point Processes* Springer-Verlag, New York, 1987.

[22] Ripley, B. *Stochastic Simulation* Wiley, New York, 1987.

[23] Royle, J.A., and C.K. Wikle "Efficient Statistical Mapping of Avian Count Data" *Ecological and Environmental Statistics*, to appear, 2005.

[24] Sansó, B. and Guenni, L. "A Bayesian Approach to Compare Observed Rainfall Data to Deterministic Simulations" *Environmetrics* **15**, 597–612, 2004.

[25] Sansó, B. and Guenni, L. "A Non-stationary Multisite Model for Rainfall" *J. Am. Statist. Assoc.* **95**, 1064-1089, 2000.

[26] Smith A.F.M. and Gelfand A.E. "Bayesian Statistics Without Tears: A Sampling-Resampling Perspective" *The American Statistican*, **46**, no. 2, 84–88, 1992.

[27] Smith, R.L. "Spatial Modelling for Rainfall Data" In book *Statistics for the Environment* Edited by V. Barnett & F. Turkman, Wiley, New York, 19–41, 1994.

[28] Smith, R.L. *Environmental Statistics*, Course Notes. Presented as a CBMS lecture series at the University of Washington, June 2001. Under revision for book publication. `http://www.unc.edu/depts/statistics/postscript/rs/envnotes.pdf`

[29] Smith, R.L. "Trends in Rainfall Extremes" Unpublished, `http://www.stat.unc.edu/postscript/rs/rx.pdf`, 1999.

[30] Smith, R.L. "Spatial Statistics in Environmental Science" In book *Nonlinear and Nonstationary Signal Processing* Edited by Fitzgerald, W.J., Smith, R.L., Walden, A.T. and Young, P.C., Cambridge University Press, 152–183, 2000.

[31] Smith, R.L. "Statistics of Extremes, with Applications in Environment, Insurance and Finance" Chapter 1 of *Extreme Values in Finance, Telecommunications and the Environment* Edited by B. Finkenstädt and H. Rootzén, Chapman and Hall/CRC Press, London, 1–78, 2003.

[32] Smith, R.L. "A Transformed, Thresholded Gaussian Model for Precipitation Extremes" In preparation; preliminary version available from `http://www.stat.unc.edu/faculty/rs/talks/rls_033006.pdf`, 2006.

[33] Stein, M.L. *Interpolation of Spatial Data: Some Theory of Kriging* Springer-Verlag, New York, 1999.

[34] Stein, M.L., Chi, Z. and Welty, L.J. "Approximating Likelihoods for Large Spatial Data Sets" *J.R. Statist. Soc. B* **66**, 275–296, 2004.

[35] Trenberth, K.E. "Conceptual Framework for Changes of Extremes of the Hydrological Cycle with Climate Change" *Climatic Change* **42**, 327–339, 1999.

[36] Vecchia, A.V. "Estimation and Identification for Continuous Spatial Processes" *J. Roy. Statist B* **50**, 297–312, 1988.

[37] Wikle, C.K. "Spatial Modeling of Count Data: A Case Study in Modelling Breeding Bird Survey Data on Large Spatial Domains" In *Spatial Cluster Modelling*, A. Lawson and D. Denison, Chapman and Hall, 199–209, 2002.

[38] Wood A.T.A., Chan G. "Simulation of Stationary Gaussian Processes in $[0,1]^d$" *J. Comput. Graph. Statist.* **3**, 409-432, 1994.

[39] Yun, S., Smith, R.L. "Spatial Trends and Spatial Extremes in South Korean Ozone" *Journal of the Korean Statistical Society* **32: 4**, 313-335, 2003.