DISSERTATION

UNDERSTANDING EXTREME BEHAVIOR BY OPTIMIZING TAIL DEPENDENCE WITH APPLICATION TO GROUND LEVEL OZONE VIA DATA MINING AND SPATIAL MODELING

Submitted by

Brook T. Russell

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2015

Doctoral Committee:

Advisor: Daniel S. Cooley

Jennifer Hoeting
Haonan Wang
Russ Schumacher

ABSTRACT

UNDERSTANDING EXTREME BEHAVIOR BY OPTIMIZING TAIL DEPENDENCE WITH APPLICATION TO GROUND LEVEL OZONE VIA DATA MINING AND SPATIAL MODELING

This dissertation presents novel work in statistical methods for extremes. Our underlying modeling procedure identifies the linear combination of covariates that is associated with extreme values of a response variable, and is based on the framework of bivariate regular variation. We propose a data mining strategy that is suitable for an analysis of ground level ozone, and spatially model the primary drivers of extreme ozone over a large study region.

In this dissertation, we first review statistical methods for univariate and multivariate extremes. We then discuss tail dependence parameters and their estimators and introduce γ, a tail dependence metric which is better suited for optimization than other existing metrics. We also introduce the idea of tail dependence estimators that utilize a smooth threshold rather than the ‘hard’ threshold common to extremes. A smooth threshold is necessary to perform optimization, which has not previously been considered in extremes studies. We also show consistency of estimators with smooth thresholds.

Subsequently, we outline our procedure for optimizing tail dependence and discuss parameter estimation. We also propose a model selection procedure that is based on cross-validation. Then we give a simulation study where we demonstrate our method’s ability to detect complicated conditions which lead to extreme behavior and compare our approach to competing methods.

Next, we propose a data mining procedure that can be used to find the set of covariates that produces the linear combination that has the highest degree of tail dependence with a response variable. Our data mining procedure is a model selection exercise where the model space is too large to be searched exhaustively. We use an automated model search procedure based on simulated annealing. We also give an analysis of ground level ozone, applying our data mining procedure to data from Atlanta, Georgia and Charlotte, North Carolina. We discuss how our method can be modified to deal with non-continuous covariates such as precipitation.

Lastly, we seek to model how a set of primary drivers varies spatially over a study region. We utilize data from 160 EPA stations in 13 US states plus the District of Columbia. We model the parameters in our extreme value procedure spatially using a hierarchical modeling technique. For inference, we utilize a two-step procedure.

ACKNOWLEDGEMENTS

I am extraordinarily thankful to have supportive and encouraging people in my life, without whom this dissertation would not have been possible. First and foremost, I want to thank my wife, Maya, for standing by me every step of the way. She always knows exactly what to say to motivate me, encourage me, or simply to make me laugh. I also wish to thank my children, Colin and Zara, for showing a great deal of understanding over the past couple of years when dad hasn't been available as much as they would have liked. Through all of the ups and downs that go along with a dissertation, I could always count on their smiles and hugs to put things in the proper perspective.

I am very fortunate to have Dan Cooley as a mentor. I would like to express my gratitude to him for his patience and guidance throughout this entire process. I appreciate the way that he has pushed me out of my comfort zone and given me opportunities that I wouldn't have been able to access otherwise. In particular, I would like to thank him for the extra level of flexibility that he has shown to me as a 'nontradition' graduate student.

I would also like to thank the other members of my graduate committee: Jennifer Hoeting, Haonan Wang, and Russ Schumacher. The feedback that they have provided has been invaluable and has certainly helped to improve the quality of this work.

I also wish to acknowledge the collaborators on the EPA air quality project: Colette Heald, Dan Cooley, Brian Reich, Alma Hodzic, Eric Gilleland, and Will Porter. Having the opportunity to work on this project has been an incredible learning experience.

I acknowledge financial support from EPA STAR Grant RD-83522861-0 and high performance computing support from Yellowstone (ark:/85065/d7wd3xhc) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. I would also like to thank Chihoon Lee for helpful discussion.

Lastly, I would like to thank my parents, Todd Russell and Sue Bonner, for helping to show me the value of education and hard work.

DEDICATION

to Maya, Zara, and Colin

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction and Background

Although extreme events occur infrequently, they typically carry a high cost to society. Analysts in a variety of applied fields recognize the need to better understand such extreme events. Examples of disciplines that utilize extremes analyses include hydrology (modeling floods), the atmospheric sciences (understanding extreme precipitation events), finance (modeling extreme losses), insurance (modeling extreme claims), and engineering (designing structures to withstand extreme events).

Extreme Value Theory (EVT) is the branch of probability that attempts to model these types of extreme events. It is being utilized in an expanding number of applied fields. Extremes is also an active area of statistical research, and many of the recent advances have been made in multivariate and spatial extremes.

Most traditional extremes analyses tend to be descriptive in nature. In univariate extremes, analysts are often interested in estimating the upper tail of a univariate probability distribution function. Return level and value-at-risk analyses are univariate extremes methods that rely on properly estimating the upper tail of a distribution. Multivariate extremes has also seen these types of descriptive analyses. For example, it is common for an analyst to estimate the probability associated with a predetermined risk region. A risk region is a set in the multivariate space that has a small probability of occurance but a high degree of consequence.

The work in this dissertation advances the field of extremes, but these newly developed methods can be described as exploratory and explanatory as opposed to descriptive in nature. Instead of attempting to describe some attribute of a distribution, we seek to find the function of covariates that has the highest possible degree of *tail dependence* with a continuous response variable. In this work we quantify tail dependence using a tail dependence metric that is well-suited for optimization. In addition, we propose a data mining method to decide which subset of a large number of covariates best describes extreme behavior of the response. We then perform a spatial analysis using our extreme value method in order to understand how the tail dependence relationship with covariates changes over a study region.

Others have developed methods to link covariates with extreme behavior, but our methods are new and are better suited for our data and objectives. Extreme value regression methods allow the parameters of an extreme value distribution to be linear combinations of explanatory variables (Beirlant et al., 2004, Ch. 7). Other methods allow for the threshold to change based on the value of covariates (Davison and Smith, 1990). Preprocessing methods attempt to create a stationary series whose residuals can be modeled using an extreme value distribution (Eastoe and Tawn, 2009). Our tail dependence method is fundamentally different than these aforementioned approaches, and is based on the framework of regular variation which is central to multivariate extremes. We feel our approach is better suited to situations where the covariates vary on the same time scale as the response (e.g., when the covariates and the response are recorded daily). We also feel that our approach is better suited for a data mining application where a large number of possible covariates exist and the analyst wishes to find those which are important to extreme values of the response.

1.1.1 Ozone Application

Although the method described in Chapter 3 of this dissertation is applicable to any continuous response variable, this dissertation is motivated by an application to understanding ground level ozone. Ozone is a secondary pollutant, meaning that it is not emitted directly into the atmosphere. Rather, it is created through a series of chemical reactions, when nitrogen oxides (NO_x) and volatile organic compounds (VOCs) in the atmosphere are exposed to ultraviolet radiation from sunlight. Consequently, the meteorological drivers which are associated with high ozone levels (high temperature, low wind speed, high solar radiation) are well known (Jacob and Winner, 2009). However, it is less well known what meteorological conditions distinguish an extreme ozone day from one with merely high ozone levels. The novel methods presented in this dissertation have two main goals: to attempt to find the linear combination of meteorological covariates which maximizes tail dependence with ozone, and to model the primary drivers of ozone over a spatial study area.

Figure 1.1 partially illustrates this idea. Ground level ozone is plotted versus air temperature for Atlanta, Georgia from 1992 to 2010 (April-October). This scatterplot shows that extreme levels of ozone occur when the air temperature is high; however, days with the highest ozone readings do not correspond to the days with the highest temperatures. This study aims to better understand the meteorological drivers of extreme ozone, and is part of an interdisiplinary project that brings together statisticians and atmospheric scientists. This project, entitled *Using advanced statistical techniques to identify the drivers and occurrence of historical and future extreme air quality events in the United States from observations and models*, is funded by EPA STAR Grant RD-83522861-0. In addition to Brook Russell and Dan Cooley of the Colorado State University Statistics Department, project members include Brian Reich (North Carolina State University Department of Statistics), William Porter (Massachusetts Institute of Technology Department of Civil and Environmental Engineering), Colette Heald (Massachusetts Institute of Technology Department of Civil and Environmental Engineering and Department of Earth, Atmospheric and Planetary Sciences), Alma Hodzic (National Center for Atmospheric Research), Eric Gilleland (National Center for Atmospheric Research), and Barbara Brown (National Center for Atmospheric Research).

We aim to find functions of meteorological covariates that have a high degree of tail dependence with ground level ozone. That is, we want to find functions of covariates which tend to be very large when ground level ozone is extreme. As is typical for an extreme value analysis, we only analyze data which are considered to be extreme and disregard that

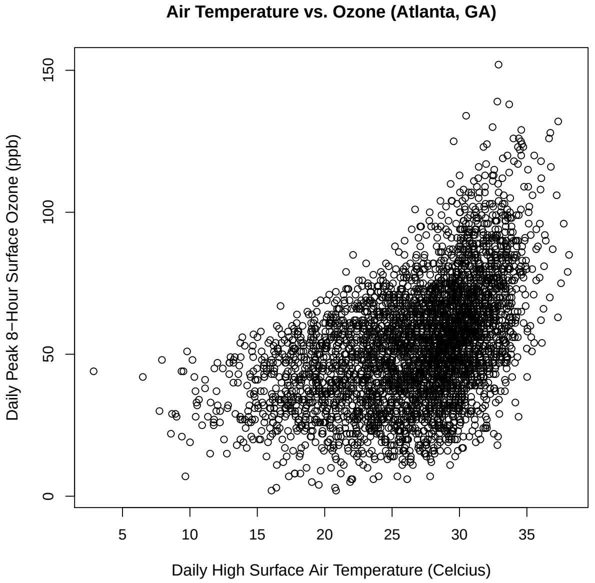

Figure 1.1: A scatterplot of daily high surface air temperature versus peak daily maximum eight-hour surface ozone in Atlanta, Georgia from 1992 to 2010 (April-October days only).

which is non-extreme. We restrict ourselves to linear combinations of (possibly transformed) meteorological covariates. We think of these functions as helping to describe the 'perfect storm' of meteorological covariates which lead to extreme ozone conditions.

We develop several novel methods which are used to describe and model the conditions of this perfect storm. The first is an optimization problem: given a set of covariates, we want to find the coefficients in the linear combination of meteorological covariates that optimize tail dependence with ozone. The second is a data mining problem: we aim to find which of the possible meteorological covariates are associated extreme ozone conditions. These two methods are used in conjunction to search a large number of possible meteorological covariates to determine which are the drivers of extreme ozone levels in Atlanta and Charlotte. We then extend the study to a large number of stations, and we develop methodology to spatially model the parameters which describe the tail dependence between the response and the covariate functional.

This dissertation is organized in the following manner. In the remainder of Chapter 1 we provide the necessary background on EVT. We begin by reviewing classical univariate Extreme Value Theory. We then introduce multivariate extremes and develop the concepts of multivariate regular variation and tail dependence.

In Chapter 2 we discuss tail dependence parameters and their estimators. We introduce γ, a tail dependence metric which is better suited for optimization than other existing metrics. Additionally, we introduce the idea of tail dependence estimators that utilize a smooth threshold rather than the 'hard' threshold common to extremes. A smooth threshold is necessary to perform optimization, which has not previously been considered in extremes studies. We show consistency of estimators with smooth thresholds.

In Chapter 3 we outline our procedure for optimizing tail dependence and discuss parameter estimation. We also propose a model selection procedure that is based on cross-validation. We close this chapter with a simulation study where we demonstrate our method's ability to detect complicated conditions which lead to extreme behavior and compare our approach to competing methods.

Chapter 4 proposes a data mining procedure that can be used to find the set of covariates that produces the linear combination that has the highest degree of tail dependence with a response variable. Our data mining procedure is a model selection exercise where the model space is too large to be searched exhaustively. We use an automated model search procedure based on simulated annealing. We also give an analysis of ground level ozone, applying our data mining procedure to data from Atlanta, Georgia and Charlotte, North Carolina. We discuss how our method can be modified to deal with non-continuous covariates such as precipitation.

Rather than data mining to tease out all the contributors to extreme behavior, Chapter 5 seeks to model how a set of primary drivers varies spatially over our study region, which is made up of EPA Regions 3 and 4. We utilize data from 160 EPA stations in 13 US states plus the District of Columbia. Our extreme value procedure does not use a likelihood, so we model the parameters in our extreme value procedure spatially using a hierarchical modeling technique. For inference, we utilize a two-step procedure. In Chapter 6, we summarize our results and discuss the possible consequences of our work. We also discuss possible extensions of this method and ozone analysis.

The work presented in Chapters 2, 3, and 4 essentially make up the article "Data Mining for Extreme Behavior with Application to Ground Level Ozone", which is currently submitted for publication. Chapter 5 largely corresponds to a manuscript entitled "Investigating the Spatial Effects of Meteorological Covariates on Extreme Ground Level Ozone", which is in preparation.

1.2 Statistical Methods for Extremes

As the methods in this dissertation rely on extremes, we first review classical statistical methods for extremes. The two most common approaches to modeling univariate extremes are the block maxima approach and the threshhold excess approach. We begin by giving a brief review of these two approaches. There are a number of books that cover univariate extremes (see Beirlant et al. (2004), Coles (2001), and De Haan and Ferreira (2007)).

1.2.1 Block Maxima Approach

Assume that we have a sequence of independent random variables $X_1, X_2, X_3, \ldots$ with common distribution function F. In the block maxima approach, define

$$M_n = \max\{X_1, X_2, \ldots, X_n\}.$$

If there exist sequences $a_n > 0$ and b_n such that

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \to G\left(z\right) \text{ as } n \to \infty, \tag{1.1}$$

for some nondegenerate G, then G belongs to one of the following three families:

$$\begin{aligned}
I : G(z) &= \exp\Big\{-\exp\Big[-(z)\Big]\Big\}, -\infty < z < \infty \\
II : G(z) &= \begin{cases} 0 & \text{if } z \leq 0 \\ \exp\Big\{-(z)^{-1/\xi}\Big\} & \text{if } z > 0 \end{cases} \\
III : G(z) &= \begin{cases} \exp\Big\{-\Big[-(z)^{1/\xi}\Big]\Big\} & \text{if } z < 0 \\ 1 & \text{if } z \geq 0. \end{cases}
\end{aligned}$$

Families II and III have an additional parameter, the shape parameter $\xi > 0$. Families I, II, and III are commonly known as the Gumbel, Fréchet, and Weibull families (respectively). This theorem is due to Fisher and Tippett (1928) and Gnedenko (1943).

The Fréchet is a family of max stable probability distributions that have the distribution function $P(Z \leq z) = \exp\Big\{-\big(\frac{z-b}{a}\big)^{-1/\xi}\Big\}\mathbb{I}_{\{z>b\}}$. The Fréchet is a location scale family with location parameter, b, and scale parameter, a. Its probability density function is right skewed, and the scale parameter $\xi > 0$ determines how 'heavy' the tail will be. The m^{th} moment exists if and only if $m < 1/\xi$.

The Gumbel is a location scale family that is defined by its distribution function $P(Z \leq z) = \exp\Big\{-\exp\Big[-\big(\frac{z-b}{a}\big)\Big]\Big\}$ and has support over all of $\mathbb{R}$. The Weibull family described above is sometimes known as the 'reverse Weibull' or 'negative Weibull' and is characterized by its distribution function

$$P(Z \leq z) = \exp\left\{-\Big[-\left(\frac{z-b}{a}\right)^{1/\xi}\Big]\right\}\mathbb{I}_{\{z<b\}} + \mathbb{I}_{\{z\geq b\}}.$$

It is also a location scale family and should not be confused with the Weibull described in Casella and Berger (2002), although negating this type of Weibull random variable will result in a 'reverse Weibull' random variable. If Equation (1.1) holds, then G will have the Generalized Extreme Value distribution (GEV) with parameters $\mu \in \mathbb{R}$, $\sigma > 0$, and $\xi \in \mathbb{R}$.

In this parameterization μ is the location parameter, σ is the scale parameter, and ξ is the shape parameter. The GEV, defined as

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{1.2}$$

for $\{z : 1 + \xi(z-\mu)/\sigma > 0)\}$, can be interpreted as a generalization of the three families: the Gumbel, Fréchet, and the Weibull. In extreme value analyses, the shape parameter ξ, is critical as it characterizes the distribution to which the process will converge. If $\xi < 0$ then the tail is bounded and G is Weibull. If $\xi = 0$ then G will be Gumbel, and for $\xi > 0$ the tail is considered heavy and G is Fréchet.

In reality, the sequences a_n and b_n in Equation (1.1) will not be known. In practice n is fixed and this issue can be resolved by reparameterizing to absorb a_n and b_n into the μ and σ parameters, labeling this GEV G^* (Coles, 2001).

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \approx G\left(z\right) \Longrightarrow P\left(M_n \leq z\right) \approx G\left(\frac{z-b_n}{a_n}\right) = G^*(z) \tag{1.3}$$

In practice, the analyst divides his or her data into non-overlapping "blocks" of large n and uses the maximum of each block of data as an independent realization of a GEV random variable. Analysts often use likelihood based methods to make inference on the GEV parameters. Other estimation methods exist, such as probability weighted moments (Hosking et al., 1985). As the annual maximum is often a meaningful quantity, in many analyses a useful block size is one year.

The main disadvantage of the block maxima approach is that it may discard many observations that could help in modeling the tail behavior. For example, if a researcher is analyzing daily data using blocks of one year, he or she would end up discarding 'extreme' observations whenever two or more such observations occur in the same year. The threshold exceedance approach attempts to resolve that issue.

1.2.2 Threshold Exceedance Approach

In the threshold excess approach, all of the observations above a particular threshold are retained. If the block maxima can be well approximated by a GEV, then the excesses above a high threshold can be modeled by a Generalized Pareto Distribution (GPD) with parameters ξ and $\tilde{\sigma}$ (Coles, 2001). It is important to note that the GPD ξ parameter is that same as the GEV ξ parameter.

The GPD is defined in the following way. If M_n converges to G^* as in (1.3), then for a large enough threshold u the distribution function of the excesses is approximately given by

$$P(X - u < y | X > u) \approx 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}. \tag{1.4}$$

Note that $\tilde{\sigma} = \sigma + \xi(u - \mu)$ and that the function is defined on the set

$$\{y : y > 0 \text{ and } (1 + \xi y/\tilde{\sigma}) > 0\}.$$

One challenge of the threshold excess approach is that a suitable threshold must be chosen. In practice there are a number of diagnostic techniques designed to assist in selecting an appropriate threshold. Scarrott and MacDonald (2012) discuss the challenge of threshold selection and summarize several commonly used methods.

1.3 Multivariate Extremes Statistical Methods

Modeling univariate extremes is relatively straightforward as it essentially involves estimating the parameters in Equation (1.2) or Equation (1.4). Modeling multivariate extremes is more challenging. In this section we begin by introducing the componentwise block maxima approach and conclude with the ideas of multivariate threshold excess and multivariate regular variation.

1.3.1 Componentwise Block Maxima

Consider the d dimensional random vector $\boldsymbol{X}$ with distribution function F. Assume we have n iid realizations $\boldsymbol{X}_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,d})^T$ and define the vector of componentwise maxima

$$\boldsymbol{M}_n = (\bigvee_{i=1}^{n} X_{i,1}, \bigvee_{i=1}^{n} X_{i,2}, \ldots, \bigvee_{i=1}^{n} X_{i,d})^T.$$

Then $\boldsymbol{M}_n$ is a d dimensional vector that includes the maximum of each component in the iid sample from $\boldsymbol{X}$. Note that $\boldsymbol{M}_n$ will not necessarily be an observed data point. Now, assume that there exist vectors of renormalizing sequences $\boldsymbol{a}_n \in \mathbb{R}^d$ and $\boldsymbol{b}_n \geq \boldsymbol{0}$ such that

$$P\left(\frac{\boldsymbol{M}_n - \boldsymbol{a}_n}{\boldsymbol{b}_n} \leq \boldsymbol{x}\right) = F^n(\boldsymbol{b}_n\boldsymbol{x} + \boldsymbol{a}_n) \overset{d}{\to} G(\boldsymbol{x}),$$

where all operations here are componentwise. $G(\boldsymbol{x})$ is defined to be a multivariate extreme value distribution (MVEVD). Since a sequence of random vectors can converge only if all marginals converge,

$$F_j^n(\boldsymbol{b}_n\boldsymbol{x} + \boldsymbol{a}_n) \overset{d}{\to} G_j(\boldsymbol{x}),$$

for $j = 1, \ldots, d$ and where F_j and G_j are the j^{th} univariate marginals of F and G (respectively). This implies that G_j is univariate EVD. However, it should be noted that not every multivariate distribution with univariate EVD marginals will be MVEVD.

It can be shown that every MVEVD has the form $G(\boldsymbol{x}) = \exp\left(-\mu[-\boldsymbol{\infty}, \boldsymbol{x}]^C\right)$ (Resnick, 1987). The measure μ is known as the exponent measure and contains all of the information regarding marginal transformations.

The marginal distributions for a MVEVD must be univariate EVD, so it makes sense to assume common marginals to better characterize μ. For convenience, we choose to transform marginals to unit Fréchet using probability integral transformations. Resnick (1987) shows that the domain of attraction is preserved under such transformations. MVEVDs with unit Fréchet marginals are called ‘simple’ and $G_j^*(z_j) = \exp(-z_j^{-1})$. Then G_* is a MVEVD with unit Fréchet marginals such that

$$G_*(\boldsymbol{z}) = G(G_1^{-1}(\exp(-z_1^{-1})), G_2^{-1}(\exp(-z_2^{-1})), \ldots, G_d^{-1}(\exp(-z_d^{-1})))$$

for $\boldsymbol{z} \in (\boldsymbol{0}, \boldsymbol{\infty})$. Let μ_* denote the measure of the simple MVEVD. It can be shown that for any $s > 0$, μ_* has the scaling property $\mu_*(sA) = s^{-1}\mu_*(A)$. For any norm, let S_{d-1} be the unit sphere, i.e. $S_{d-1} = \{\boldsymbol{z} \in \mathbb{R}_+^d : ||\boldsymbol{z}|| = 1\}$. Now, consider sets in the form of

$$A(r, B) = \{\boldsymbol{z} \in \mathbb{R}_+^d : ||\boldsymbol{z}|| > r, \boldsymbol{z}/||\boldsymbol{z}|| \in B\}$$

for $B \subset S_{d-1}$. For fixed $B \subset S_{d-1}$ define the angular (spectral) measure

$$H(B) := d^{-1}\mu_*(A(1, B)).$$

Due to the scaling property of μ_*,

$$\mu_*(A(r, B)) = r^{-1}\mu_*(A(1, B)) = dr^{-1}H(B). \tag{1.5}$$

This suggests the transformation $T : \mathbb{R}_+^d \to \mathbb{R}_+ \times [0, 1]^{d-1}$ such that

$$T(\boldsymbol{Z}) = (||\boldsymbol{Z}||, \boldsymbol{Z}/||\boldsymbol{Z}||) =: (R, \boldsymbol{W}).$$

R is called the radial component and $\boldsymbol{W}$ is called the angular component. The right hand side of (1.5) is a product measure, revealing that the radial and angular component become independent in the limit. The exponent measure function, V_*, is defined

$$V_*(\boldsymbol{z}) := -\log(G_*(\boldsymbol{z})) = \mu_*([\boldsymbol{0}, \boldsymbol{z}]^C).$$

Thus, G_* can be expressed as

$$G_*(\boldsymbol{z}) = \exp(-\mu_*([\mathbf{0}, \boldsymbol{z}]^C)) = \exp(-V_*(\boldsymbol{z})).$$

When utilizing the L_1 norm: $\|\boldsymbol{Z}\|_1 = Z_1 + \ldots + Z_d$, V_* can be expressed as

$$\begin{aligned}
V_*(\boldsymbol{z}) &= d \iint\limits_{(r,\boldsymbol{w})\in T([\mathbf{0},\boldsymbol{z}]^C)} r^{-2} dr H(d\boldsymbol{w}) \\
&= d \int_{S_{d-1}} \int_{r=\min\limits_{i=1,\ldots,d}\left(\frac{w_i}{z_i}\right)}^{\infty} r^{-2} dr H(d\boldsymbol{w}) \\
&= d \int_{S_{d-1}} [-r^{-1}]^{\infty}_{\min\limits_{i=1,\ldots,d}\left(\frac{w_i}{z_i}\right)} H(d\boldsymbol{w}) \\
&= d \int_{S_{d-1}} \left(\min_{i=1,\ldots,d}\left(\frac{w_i}{z_i}\right)\right)^{-1} H(d\boldsymbol{w}) \\
&= d \int_{S_{d-1}} \max_{i=1,\ldots,d}\left(\frac{w_i}{z_i}\right) H(d\boldsymbol{w}).
\end{aligned}$$

This definition shows the difficulty involved in expressing G_* (and V_*) in Cartesian coordinates and indicates that the polar coordinate transformation, T, may be convenient.

Building on the characterization of the MVEVDs, we next describe the threshold excess approach provided by the regular variation framework where we will further describe H.

1.3.2 Threshold Excesses

One challenge of multivariate threshold exceedance approaches is that one must define what is meant by a threshold exceedance. The multivariate GPD (Rootzen and Tajvidi, 2006) is a model which can be used when thresholds are defined in terms of each univariate marginal. Our approach is to utilize the multivariate regular variation framework which can be used when a threshold is defined in terms of a vector norm.

1.3.3 Regular Variation and Multivariate Exceedences

Multivariate regular variation (MVRV) is a probabilistic framework for modeling multivariate extremes. MVRV implies that the joint tail decays like a power function and is defined only in terms of the joint tail. There are a number of equivalent definitions of MVRV. A random vector $\boldsymbol{Z} \in [0, \infty)^d$ is regularly varying if there exists a sequence $b(n)$ such that

$$nP\left(\frac{\boldsymbol{Z}}{b(n)} \in \cdot\right) \xrightarrow{v} \nu(\cdot), \tag{1.6}$$

where v denotes vague convergence on $\mathbb{E} = [0, \infty]^d \backslash \{\mathbf{0}\}$ and $\|\cdot\|$ is any norm (Resnick, 2007). For $\alpha = 1$, the measure ν described in (1.6) is equivalent to the exponent measure μ_* in the previous section. If we choose a set $A \in \mathbb{E}$ and $s > 0$, the limit measure ν also has the scaling property, $\nu(sA) = s^{-\alpha}\nu(A)$. This scaling property shows the heavy-tailed nature of the MVRV framework. The tail index α determines the power-law behavior of the tail.

The scaling property leads to an equivalent and useful definition in terms of psuedo-polar coordinates, where we transform $\boldsymbol{Z}$ from Cartesian to psuedo-polar coordinates using T. Then there exists a finite measure H on S_{d-1} such that for any $H-$continuity Borel subset B of S_{d-1}

$$nP\left(b(n)^{-1}R > r, \boldsymbol{W} \in B\right) \xrightarrow{v} r^{-\alpha}H(B) \text{ as } n \to \infty. \tag{1.7}$$

The measure H described in (1.7) is equivalent to the measure H in the previous section. Because the right hand side of (1.7) is a product measure, we are again reminded that the radial and angular components become independent in the limit. We can characterize the tail behavior through knowing α and the angular measure H, which completely describes dependence. $b(n)$ can be chosen so that H is a probability measure.

Although the regular variation framework implies that the univariate marginal distributions are each heavy-tailed with tail index α, the framework can be used to model multivariate data with differing tail behavior, which may or may not be heavy-tailed. Resnick (1987) shows that monotone transformations of the univariate marginal distributions do not change the fundamental nature of the tail dependence, in the sense that the domain of attraction is preserved. If the data come from $\boldsymbol{Y}$ which is not regularly varying, we assume that there exist probability integral transformations T_i, such that $T_i(Y_i) = Z_i$, and $\boldsymbol{Z} = (Z_1, \ldots Z_d)^T$ is regular varying.

Statistical practice for multivariate extremes typically requires transforming marginals to a common, convenient distribution under which the dependence structure is more easily described. When dealing with real data, the appropriate transformation T_i will be unknown. An analyst will typically transform each margin using $\hat{T}_i$, where $\hat{T}_i$ is a probability integral transformation based on an estimated distribution function. Copula approaches (Nelsen, 2006), which model dependence after marginal transformation to Uniform[0,1] are similar in spirit. However, unlike copula approaches, our approach aims only to describe the tail.

We choose to transform our marginals to the unit Fréchet distribution. The unit Fréchet is regularly varying with $\alpha = 1$, therefore our transformed data will be very heavy-tailed. Since $\alpha = 1$, we find it convenient to use the L_1 norm. Although the Fréchet is max-stable, this is unimportant for our purposes as we model threshold exceedances rather than block maxima.

After marginal transformation, α is assumed known, and one needs to model H in (1.7) to describe the dependence in the tail. In this dissertation, we only need to consider tail dependence for bivariate vectors. Let $\boldsymbol{Z} = (Z_1, Z_2)^T$ be a bivariate regular-varying $\alpha = 1$ random vector with common marginal distributions. To begin to understand H, consider the limit

$$\chi := \lim_{z \to \infty} P(Z_2 > z | Z_1 > z). \tag{1.8}$$

If $\chi = 0$, then the variables are termed asymptotically independent, whereas if $\chi > 0$, the variables exhibit asymptotic (tail) dependence. The two extreme cases for χ, asymptotic independence and perfect dependence, yield specific forms for H. In the asymptotic independence case, H puts all of its mass on the axes, placing mass of .5 at (0,1) and (1,0). In the case of perfect asymptotic dependence, H places all of its mass at (.5,.5). In general, as H puts more mass towards the center of the distribution the amount of tail dependence increases.

When exploring data, a scatterplot after transformation to a heavy tailed distribution can help one to understand tail dependence. Figure 1.2 gives a scatterplot of air temperature versus ozone in Atlanta, where both variables are transformed to have unit Fréchet marginals via rank transformations. Notice that the majority of the points have congregated near the origin, and one is left to view the behavior of the large points. As there are points in the interior of the plot, this suggests that this data does not exhibit asymptotic independence. However, as many of the large points occur near the axes, it also seems to indicate that the level of tail dependence is relatively weak; that is, that temperature alone does not describe extreme ozone conditions.

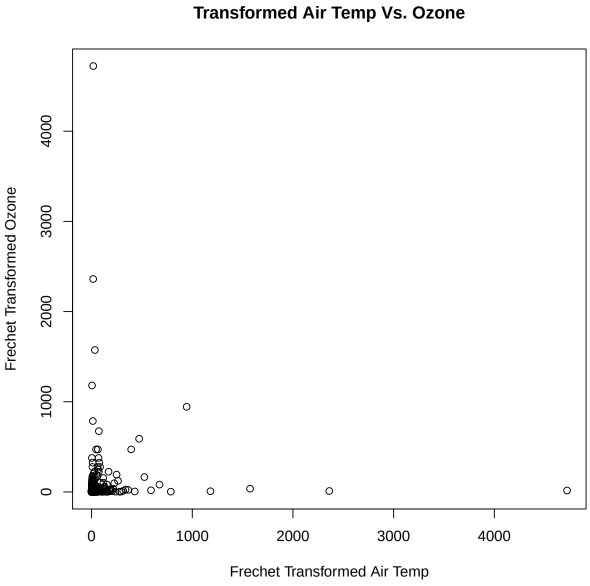

Figure 1.2: A scatterplot of daily high surface air temperature versus peak daily maximum eight-hour surface ozone in Atlanta, Georgia from 1992 to 2010 (April-October) after transforming to the unit Fréchet scale via rank transformations.

1.4 Utilizing Regular Variation

Our method relies on the MVRV framework, but does so differently than typical MVRV methods. A typical problem in multivariate extremes is to estimate the probability of seeing observations in a risk region. The risk region has a small probability of occurance yet often has great consequences. Figure 1.3 illustrates this problem. As there are no observations in the risk region, it is difficult to esimate the probability associated with it. A parametric framework such as MVRV can be useful, as it allows for extrapolation beyond the observed data.

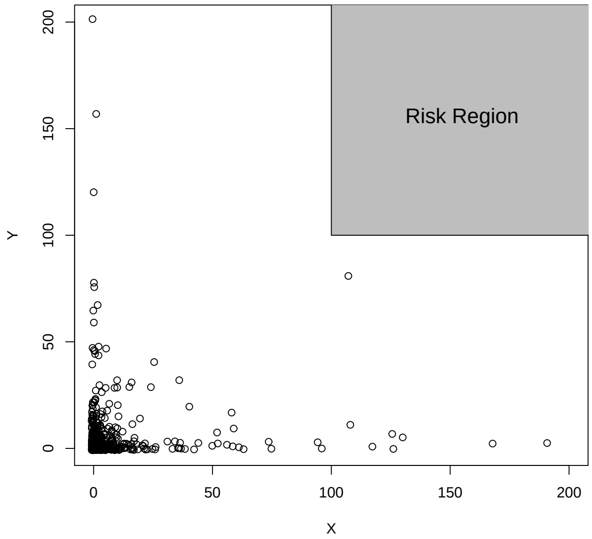

Figure 1.3: Estimating the probability of getting realizations in a predetermined risk region is a typical utilization of the MVRV framework. MVRV can be especially useful in this case as it allows the analyst to extrapolate beyond the observed data set.

The methods in this dissertation utilize MVRV, but do so in a very different way. One primary goal of this dissertation is to develop a procedure to estimate the coefficients in the linear combination of continuous covariates that has the highest degree of tail dependence with a continuous response variable. Changing the coefficients in the linear combination has no effect on the response variable. However, such an adjustment causes the points in the Unit Fréchet transformed scatterplot to shift horizontally, as seen in Figure 1.4. We seek to find the parameter estimates that put as many points as possible with large radial component in the interior, away from the axes. Informally, the goal is to get the points with large radial component as close as possible to the 45° line, thus maximizing tail dependence. In the next two chapters we introduce tail dependence metrics and then discuss optimization of tail dependence.

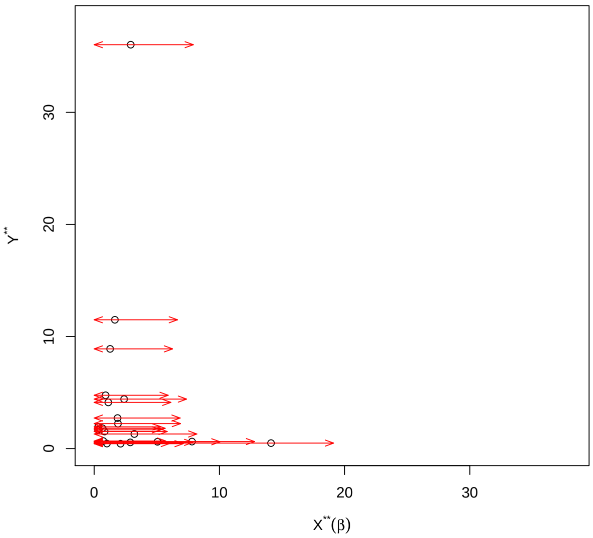

Figure 1.4: An illustration of a possible scatterplot of the linear combination of covariates, $(X^{**}(\boldsymbol{\beta})$, versus the response variable, Y^{**}. Changing the parameter vector, $\boldsymbol{\beta}$, causes the points to shift horizontally. As the red arrows indicate, the points will shift different distances to the left or right. Our method seeks to find the $\boldsymbol{\beta}$ that produces the linear combination that has the highest degree of tail dependence with the response variable.

CHAPTER 2

TAIL DEPENDENCE

2.1 Tail Dependence Metrics and Estimators

In Chapter 1 we introduced statistical methods for extremes and the idea of MVRV which can be used to model multivariate threshold exceedances. We also made the claim that our objective is unique for a MVRV analysis. Instead of describing dependence within the MVRV framework, we wish to find the linear combination of covariates that has the highest degree of tail dependence with a continuous response variable. This requires us to outline a new tail dependence metric and to consider alternatives to traditional extremes thresholding techniques.

In this chapter we introduce γ, the tail dependence metric we utilize in this work. We also propose, $\hat{\gamma}_n^{(S)}$, its corresponding estimator that utilizes a 'smooth' threshold. We conclude this chapter by showing the consistency of $\hat{\gamma}_n^{(S)}$ under certain conditions.

In 'traditional' statistics, it is common for analysts to utilize correlation metrics such as Pearson's, Kendall's, and Spearman's to describe association. While these metrics do well at describing association in the bulk of the data, they are poorly suited for describing tail dependence. This is because dependence in the tail can differ from dependence in the bulk of the distribution. For example, any bivariate Gaussian random vector with correlation parameter less than one can be shown to be asymptotically independent (Sibuya, 1959).

As its definition is strictly in terms of tail behavior, the regular variation framework provides a method for describing tail dependence. Within this framework, the angular measure H introduced in Chapter 1 completely describes tail dependence, but is not easily summarized. In order to perform a numerical optimization later, we need to be able to summarize tail dependence with a single value. Several tail dependence metrics have been suggested.

Coles et al. (1999) describe χ, as given in Equation (1.8). Schlather and Tawn (2003) propose the extremal coefficient and Davis and Mikosch (2009) propose the extremogram, which can be viewed as a generalization of χ.

We introduce a new tail dependence metric γ, which is a particular instance of an extremal dependence measure (Resnick (2004) and Larsson and Resnick (2012)). Important for our purposes, γ has a natural estimator $\hat{\gamma}$ which is well suited for optimization. Furthermore, we will propose employing a smooth threshold for estimation. These tail dependence parameter estimation issues have not been previously encountered in extremes work as optimizing tail dependence has not previously been attempted. At the end of this chapter, we show consistency of $\hat{\gamma}_n^{(smooth)}$, our estimator with smooth threshold.

2.1.1 Metrics with Counting-type Estimators

We begin by describing some tail dependence metrics and their estimators. Coles et al. (1999) describe the parameter χ, given in Equation (1.8), which could serve as a tail dependence summary measure. Note since we assume $\boldsymbol{Z}$ is bivariate regularly varying, χ can also be defined in terms the limit measure ν:

$$\chi = \frac{\nu([u,\infty] \times [u,\infty])}{\nu([u,\infty] \times [0,\infty])}.$$

Clearly χ is the ratio of the measures of two sets. Figure 2.1 depicts these two sets graphically.

Let $\boldsymbol{Z}_t = (Z_{t,1}, Z_{t,2})$ for $t = 1, \ldots, n$ be iid copies of $\boldsymbol{Z}$. An estimator of χ can be obtained by replacing ν with the observed counts:

$$\hat{\chi}(u) = \frac{\sum_{t=1}^{n} (\mathbb{I}\{Z_{t,1} > u\})(\mathbb{I}\{Z_{t,2} > u\})}{\sum_{t=1}^{n} \mathbb{I}\{Z_{t,1} > u\}}. \tag{2.1}$$

This estimator is typically calculated for an increasing sequence of values of u, and then $\hat{\chi}(u)$ is plotted versus u or $\hat{F}_n(u)$. If $\hat{\chi}(u)$ approaches zero as u gets larger the bivariate vector shows asymptotic independence, whereas if $\hat{\chi}(u)$ does not approach zero the bivariate vector

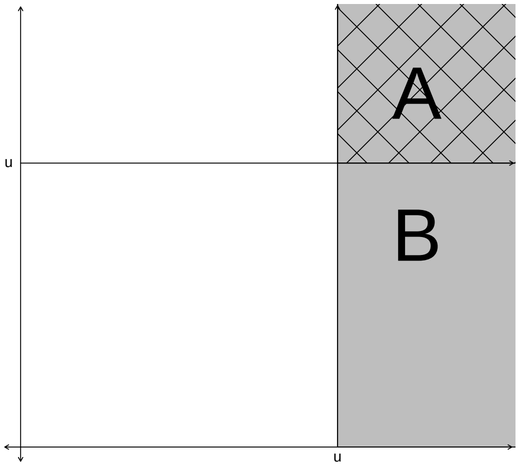

Figure 2.1: In the regular variation framework, the tail dependence parameter $\chi = \nu([u,\infty]\times[u,\infty])/\nu([u,\infty]\times[0,\infty])$ can be described as the ratio of the measure of two sets. Here, A (the gray 'criss-crossed' region) has measure $\nu(A) = \nu([u,\infty]\times[u,\infty])$ and B (the entire gray region) has measure $\nu(B) = \nu([u,\infty]\times[0,\infty])$.

shows asymptotic dependence. Figure 2.2 gives this plot for the Atlanta ozone and temperature data. The plot indicates that air temperature and ozone appear to be asymptotically dependent as $\hat{\chi}(u)$ does not appear to become 0 as u increases, but the degree of dependence is rather low.

Coles et al. (1999) give an estimator for χ which is symmetric in the sense that it also considers Z_1 conditioned on Z_2 exceeding u. The extremogram (Davis and Mikosch, 2009), can be viewed as a generalization of $\hat{\chi}$. Like the estimator in Equation (2.1), the estimators proposed by Coles et al. (1999) and Davis and Mikosch (2009) are ratios of sums of indicators. We call such estimators 'counting estimators' for obvious reasons.

For a fixed u, $\hat{\chi}(u)$ gives the number of points that exceed u in both components divided by the number of points that exceed u in the first component. Counting estimators have proven useful for describing tail dependence, but they cannot serve as the objective function in numerical optimization. This is because many unique arrangements of points can result in

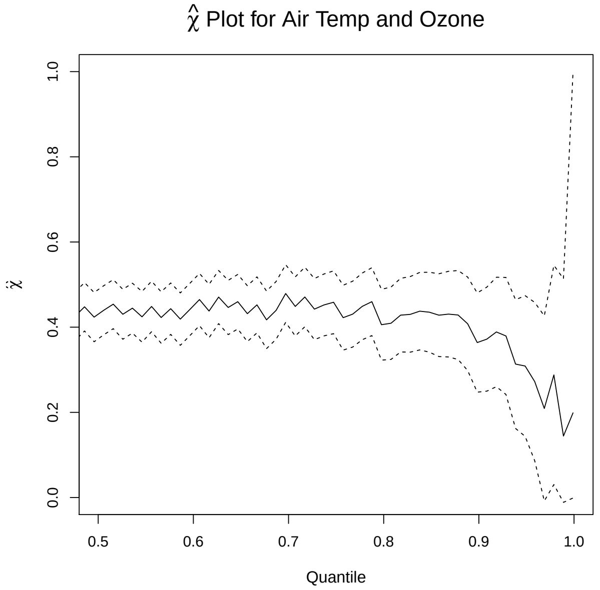

Figure 2.2: Daily high surface air temperature and peak daily maximum eight-hour surface ozone in Atlanta, Georgia from 1992 to 2010 (April-October) are transformed to the unit Fréchet scale via rank transformations. We plot of $\hat{\chi}$ using a sequence of quantiles for the threshold (with 95% confidence bands).

identical values of $\hat{\chi}$. This idea is illustrated in Figure 2.3. As the number of points in each respective region is the same, both scatterplots will produce identical values of $\hat{\chi}$ despite being distinct arrangements of points.

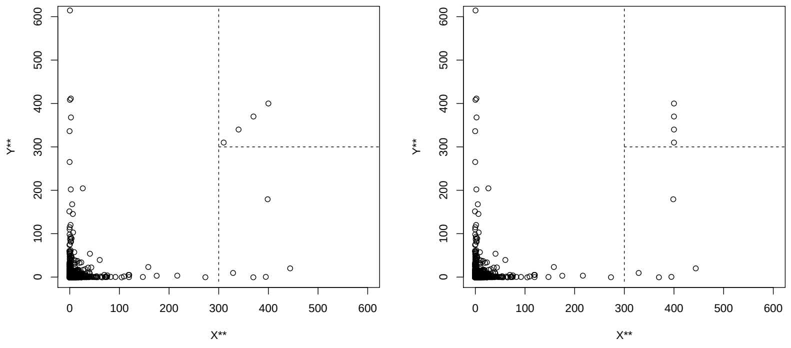

Figure 2.3: Although the data sets used to create the scatterplots in the (L) and (R) panels are not identical, they produce identical values of $\hat{\chi}$. This illustrates the fact that $\hat{\chi}$ is poor for use as an objective function.

2.1.2 Tail Dependence Metrics from H

If $\boldsymbol{Z}$ is bivariate regularly varying with angular measure H, tail dependence metrics can be created by integrating with respect to H. Larsson and Resnick (2012, subsection 2) show that any tail dependence metric based on ν in the form $\int_{\mathbb{E}} f(x)d\nu(x)$ can be written as an integral with respect to the angular measure H. Larsson and Resnick (2012) describe these dependence summary parameters based on H in the form

$$\rho_\kappa = \int_{S_1} \kappa(w)dH(w) \tag{2.2}$$

for $\kappa : S_1 \to [0, \infty)$. In earlier work, Resnick (2004) proposed the extremal dependence measure (EDM), a specific case of (2.2) where $\kappa(\theta) = (4/\pi)^2\theta(\pi/2 - \theta)$ and H is defined on the unit ball given by the L_2 norm.

Others have utilized the EDM in analyses intending to describe tail dependence in specific applications. Hernandez-Campos et al. (2005) use this EDM in an analysis of internet file sizes and rate of transfer and D'Auria and Resnick (2008) make use of the EDM in modeling inputs to data networks.

Resnick (2004) and Larsson and Resnick (2012) propose estimators and show consistency by relying on the intermediate asymptotics common to EVT. Let $k := k(n)$ be a sequence such that $k \to \infty$ and $k/n \to 0$. Begin with the empirical estimator for ν:

$$\hat{\nu}(\cdot) := \frac{1}{k}\sum_{t=1}^{n} \epsilon_{\frac{\boldsymbol{Z}_t}{b(n/k)}}(\cdot),$$

where $\epsilon_{\boldsymbol{x}}(A) = 1$ if $\boldsymbol{x} \in A$ and 0 otherwise, and b is defined as in (1.7). It can be shown that $\hat{\nu} \Rightarrow \nu$ (Resnick, 2007, Theorem 4.1). Then, for sets in S_1, define

$$\hat{H}^{(hard)}(\cdot) := \frac{\hat{\nu}\{\boldsymbol{z} \mid \|\boldsymbol{z}\| > 1, \boldsymbol{z}\|\boldsymbol{z}\|^{-1} \in (\cdot)\}}{\hat{\nu}\{\boldsymbol{z} \mid \|\boldsymbol{z}\| > 1\}} = \frac{\sum_{t=1}^{n} \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)\mathbb{I}\left\{\frac{\boldsymbol{Z}_t}{\|\boldsymbol{Z}_t\|} \in \cdot\right\}}{\sum_{t=1}^{n} \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)},$$

where $\delta^{(hard)} : \mathbb{R}_+ \mapsto [0,1]$ is defined as

$$\delta^{(hard)}(z) = \begin{cases} 0 & \text{for } z < 1; \\ 1 & \text{for } z \geq 1. \end{cases}$$

The superscript $(hard)$ denotes the standard 'hard' threshold at 1 which is standard for extremes. Letting $\boldsymbol{W}_t = \boldsymbol{Z}_t / \|\boldsymbol{Z}_t\|$, we obtain the estimator for the tail dependence measure

$$\hat{\rho}_\kappa^{(hard)} = \int_{S_1} \kappa(w) \hat{H}^{(hard)}(dw) = \frac{\sum_{t=1}^{n} \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \kappa(\boldsymbol{W}_t)}{\sum_{t=1}^{n} \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)}. \tag{2.3}$$

As they are functions of $\hat{\nu}$, $\hat{H}^{(hard)} \Rightarrow H$ and $\hat{\rho}_\kappa \xrightarrow{p} \rho$ follow from the continuous mapping theorem. Resnick 2007, p. 301 further states that $b(n/k)$ can be replaced with an estimator $\hat{b}(n/k)$, and so long as $\hat{b}(n/k)/b(n/k) \xrightarrow{p} 1$, the above convergences will hold.

As mentioned in Chapter 1, we choose to work with the L_1 norm, thus we use a dependence summary parameter based on the distance $|Z_1 - Z_2|$. When assuming common marginal distributions, $|Z_1 - Z_2|$ informally represents the distance to the 45° line. This makes sense heuristically as the 45° line is where all points with large radial component would reside under perfect dependence. The madogram (Cooley et al., 2006) is a tail dependence metric also based on an L_1 distance, but in the context of block maxima rather than threshold exceedances.

After converting to psuedo-polar coordinates, we get $|Z_1 - Z_2| = |RW - R(1-W)| = R|2W-1|$. We require a function that is not dependent upon R, so we choose $\kappa(w) = |2w-1|$, noting

$$\frac{|Z_1 - Z_2|}{Z_1 + Z_2} = \frac{R|2W-1|}{R} = |2W-1|.$$

Define the parameter

$$\gamma = \int_{S_1} |2w-1| dH(w). \tag{2.4}$$

Note that $0 \leq \gamma \leq 1$, and that a smaller value of γ implies a higher degree of tail dependence.

For illustrative purposes, we now calculate γ for three specific forms for H. In the case of asymptotic dependence $\gamma = 1$, whereas perfect tail dependence yields $\gamma = 0$. If the angular component is uniformly distributed over S_1,

$$\gamma = \int_0^1 |2w - 1| dw = \int_0^{.5} (1 - 2w) dw + \int_{.5}^1 (2w - 1) dw = .25 + .25 = .5.$$

The tail dependence between air temperature and ground level ozone is fairly weak, evidenced by Figure 2.4 where $\hat{\gamma}$ is approximately .65 when using a threshold at the .95 quantile.

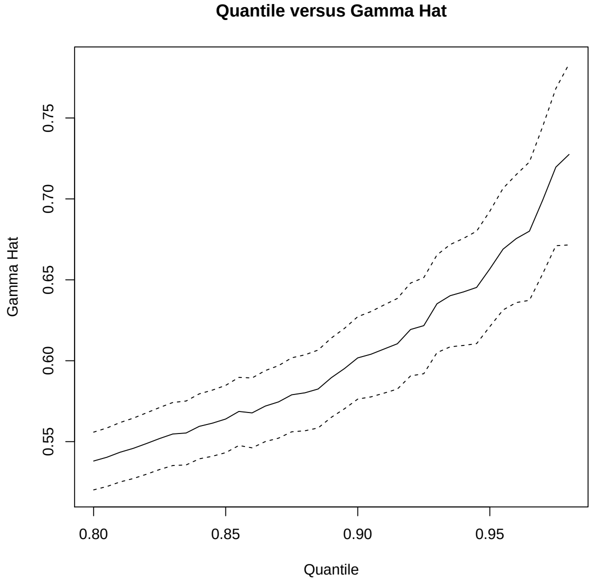

Figure 2.4: Daily high surface air temperature and peak daily maximum eight-hour surface ozone in Atlanta, Georgia from 1992 to 2010 (April-October) are transformed to the unit Fréchet scale via rank transformations. We plot of $\hat{\gamma}$ using a sequence of quantiles for the threshold (with 95% confidence bands).

We could estimate γ with an estimator of the form (2.3), however, the hard threshold does not lend itself well to performing optimization, and we explore using a smooth threshold in the next section.

2.1.3 Tail Dependence Estimation with a Smooth Threshold

As we wish to optimize tail dependence in terms of our tail dependence estimator, the hard threshold typically used in EVT is problematic. This has to do with the nature of our procedure. During optimization, points will likely move back and forth across the threshold, and ultimately cause the optimizer not to converge. This is illustrated in Figure 2.5. When using the 'hard' threshold typical of extremes analyses, points in region A are given a weight of 1 and all other points are given a weight of 0. As the parameter vector changes during the course of the optimization, points are shifted horizontally and the subset of points over which we are optimizing can change from iteration to iteration.

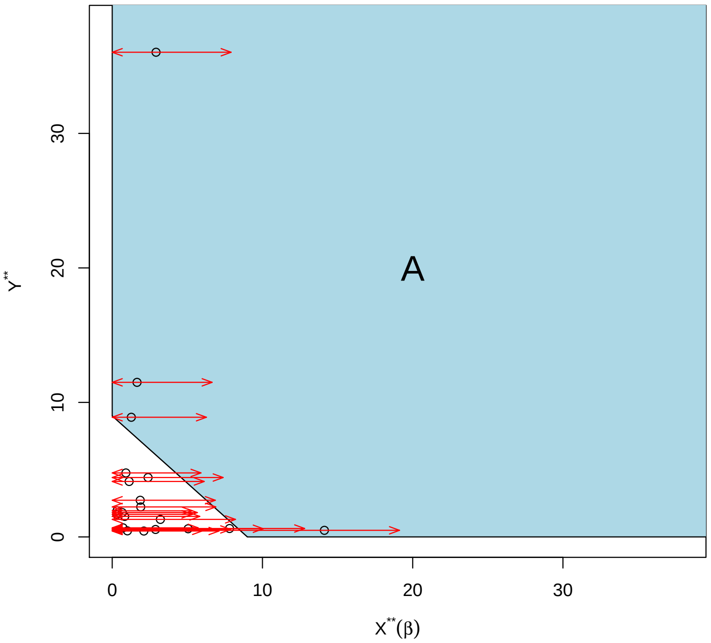

Figure 2.5: When using the 'hard' threshold typical of extremes analyses, points in region A are given a weight of 1 and all other points are given a weight of 0. As the parameter vector changes during the course of the optimization, points are shifted horizontally. This underscores the need for a smooth threshold that gradually increases the weight from 0 to 1 as the distance from the origin increases.

Chaudhuri and Solar-Lezama (2010 and 2011) propose using 'smooth interpretations' of discontinuous functions in numeric optimization. We use these techniques for extremes by replacing the hard threshold with a smooth one, which gradually increases the weights from 0 to 1 as the radial component increases.

Let

$$\hat{H}_n^{(smooth)} := \frac{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)\kappa(\boldsymbol{W}_t)}{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)}, \tag{2.5}$$

where $\delta_n^{(smooth)}$ is a function which converges pointwise to $\delta_n^{(hard)}$ on $(0,1)\cup(1,\infty)$. At the end of this chapter, we show that sufficient conditions on $\delta_n^{(smooth)}$ for $H_n^{(smooth)} \Rightarrow H$ are:

1. $\delta_n^{(smooth)} : \mathbb{R} \mapsto [0,1]$ is non-decreasing,
2. there exists an $a \in (0,1)$ such that $n\ \delta_n^{(smooth)}(a) \to 0$ as $n \to \infty$,
3. $dP_{\frac{|Z_1|}{b(n/k)}}(z) \sim \frac{k}{n} z^{-2}$ for $z > a$,
4. $k\int_a^1 \delta_n^{(smooth)}(z) z^{-2} dz \to 0$ and $k\int_1^\infty (1-\delta_n^{(smooth)}(z)) z^{-2} dz \to 0$.

Consistency of

$$\begin{aligned}
\hat{\gamma}_n^{(smooth)} &:= \int_{S_1} |2w-1| \hat{H}^{(smooth)}(dw) \\
&= \frac{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) |2\boldsymbol{W}_t - 1|}{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} \\
&= \left(\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)\right)^{-1} \sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \frac{|\boldsymbol{Z}_{t,1} - \boldsymbol{Z}_{t,2}|}{|\boldsymbol{Z}_{t,1} + \boldsymbol{Z}_{t,2}|} \quad (2.6)
\end{aligned}$$

follows by the continuous mapping theorem from the weak convergence of $H_n^{(smooth)}$ to H.

Before going through the consistency proof, we first develop some intuition for conditions 1-4. The sufficient conditions for $\delta_n^{(smooth)}$ can be understood by considering the sequence of processes given by $\sum_{t=1}^n \epsilon_{\boldsymbol{Z}_t/b(n/k)}$. Informally, as $n \to \infty$, points with radial components near 0 pile up with rate n, and points pile up with rate k for regions bounded away from 0. $\hat{H}^{(hard)}$ behaves nicely since all points with radial components less than 1 get weights of zero because of the hard threshold. When considering the smooth threshold, one needs $\delta_n^{(smooth)}$ to go to zero fast enough to keep mass from accumulating near zero (thus requiring condition 2), and one needs $\delta_n^{(smooth)}$ to go to zero or one less rapidly for regions bounded away from 0. As condition 2 is more stringent than 4, it is more critical. For example, the function $\delta_n^{(smooth)}(z) = \pi^{-1}\arctan(a_n(z-1)) + 1/2$ will meet condition 2 if $n^{-1}a_n \to \infty$ as $n \to \infty$.

In practice, since n is fixed, the convergence rate of $\delta_n^{(smooth)}$ is irrelevant. In the simulation study in Chapter 3, we test the optimization procedure's sensitivity to different degrees of smoothness. Also, as is typical of extremes procedures, in Equations (2.5) and (2.6) $b(n/k)$ is replaced by a suitably chosen threshold r.

2.2 Consistency of $\hat{H}_n^{(smooth)}$

Lemma 1. *Let $\hat{H}_n^{(smooth)}$ be defined as in (2.5). If*

$$\frac{\sum_{t=1}^{n}\left|\delta_n^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) - \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)\right|}{\sum_{t=1}^{n}\delta_n^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} \xrightarrow{p} 0, \tag{2.7}$$

then $\hat{H}_n^{(smooth)} \Rightarrow H$.

Although this condition is not easily interpretable, it does have meaning. Since the denominator grows at rate k, (2.7) will hold if the sum of the absolute differences in the numerator grows slower than k.

Proof: Our approach is to first show that $|\hat{H}_n^{(hard)} - \hat{H}_n^{(smooth)}| \xrightarrow{p} 0$.

$$
\begin{aligned}
& |\hat{H}_n^{(hard)}(\cdot) - \hat{H}_n^{(smooth)}(\cdot)| \\
= & \left| \frac{\sum_{t=1}^n \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \mathbb{I}\left\{\frac{\boldsymbol{Z}_t}{\|Z_t\|} \in \cdot\right\}}{\sum_{t=1}^n \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} - \frac{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \mathbb{I}\left\{\frac{\boldsymbol{Z}_t}{\|Z_t\|} \in \cdot\right\}}{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} \right| \\
= & \left| \left(\frac{\sum_{t=1}^n \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \mathbb{I}\left\{\frac{\boldsymbol{Z}_t}{\|Z_t\|} \in \cdot\right\}}{\sum_{t=1}^n \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} - \frac{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \mathbb{I}\left\{\frac{\boldsymbol{Z}_t}{\|Z_t\|} \in \cdot\right\}}{\sum_{t=1}^n \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} \right) \right. \\
& \left. - \left(\frac{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \mathbb{I}\left\{\frac{\boldsymbol{Z}_t}{\|Z_t\|} \in \cdot\right\}}{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} - \frac{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \mathbb{I}\left\{\frac{\boldsymbol{Z}_t}{\|Z_t\|} \in \cdot\right\}}{\sum_{t=1}^n \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} \right) \right| \\
= & \left| \frac{\sum_{t=1}^n \left(\delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) - \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \right) \mathbb{I}\left\{\frac{\boldsymbol{Z}_t}{\|Z_t\|} \in \cdot\right\}}{\sum_{t=1}^n \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} \right. \\
& \left. - \frac{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \mathbb{I}\left\{\frac{\boldsymbol{Z}_t}{\|Z_t\|} \in \cdot\right\} \left(\sum_{t=1}^n \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) - \sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \right)}{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \sum_{t=1}^n \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} \right| \\
\leq & \frac{\sum_{t=1}^n \left| \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) - \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \right| \mathbb{I}\left\{\frac{\boldsymbol{Z}_t}{\|Z_t\|} \in \cdot\right\}}{\sum_{t=1}^n \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} \\
& + \frac{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \mathbb{I}\left\{\frac{\boldsymbol{Z}_t}{\|Z_t\|} \in \cdot\right\}}{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} \frac{\sum_{t=1}^n \left| \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) - \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \right|}{\sum_{t=1}^n \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} \\
\leq & \; 2 \, \frac{\sum_{t=1}^n \left| \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) - \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \right|}{\sum_{t=1}^n \delta^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)}.
\end{aligned}
$$

Thus, by assumption $|\hat{H}_n^{(hard)} - \hat{H}_n^{(smooth)}| \xrightarrow{p} 0$. Knowing that $\hat{H}_n^{(hard)} \Rightarrow H$, we then employ Slutsky's theorem (Resnick, 2007, Thm. 3.4) giving $\hat{H}_n^{(smooth)} \Rightarrow H$.

The Lemma's condition for consistency is not very satisfying as it tells us little about the behavior of $\delta_n^{(smooth)}$. The theorem below provides sufficient conditions on $\delta_n^{(smooth)}$ such that (2.7) holds.

Theorem 1. *If*

1. $\delta_n^{(smooth)} : \mathbb{R} \mapsto [0,1]$ *is non-decreasing,*
2. *there exists an* $a \in (0,1)$ *such that* $n\ \delta_n^{(smooth)}(a) \to 0$ *as* $n \to \infty$,
3. $dP_{\frac{|Z_1|}{b(n/k)}}(z) \sim \frac{k}{n} z^{-2}$ *for* $z > a$,
4. $k \int_a^1 \delta_n^{(smooth)}(z) z^{-2} dz \to 0$ *and* $k \int_1^\infty (1 - \delta_n^{(smooth)}(z)) z^{-2} dz \to 0.$

then (2.7) holds.

Proof: Notice

$$\begin{aligned}\frac{\sum_{t=1}^n \left| \delta_n^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) - \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)\right|}{\sum_{t=1}^n \delta_n^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} &= \frac{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \mathbb{I}\left\{\frac{\|\boldsymbol{Z}_t\|}{b(n/k)} < a\right\}}{\sum_{t=1}^n \delta_n^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} \\ &+ \frac{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \mathbb{I}\left\{\frac{\|\boldsymbol{Z}_t\|}{b(n/k)} \in [a,1)\right\}}{\sum_{t=1}^n \delta_n^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} \\ &+ \frac{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \mathbb{I}\left\{\frac{\|\boldsymbol{Z}_t\|}{b(n/k)} \geq 1\right\}}{\sum_{t=1}^n \delta_n^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)}.\end{aligned}$$

We consider the three terms individually and use the law of large numbers for triangular arrays. The law of large numbers for triangular arrays states that for $\mu_n' = E[S_n']$ and $\sigma_n^{2'} = Var[S_n']$, if $(a_n')^{-2}\sigma_n^{2'} \to 0$ then $(a_n')^{-1}(S_n' - \mu_n') \xrightarrow{p} 0$ (Durrett, 2010, Theorem 5.4).

Consider the first term

$$\frac{\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right) \mathbb{I}\left\{\frac{\|\boldsymbol{Z}_t\|}{b(n/k)} < a\right\}}{\sum_{t=1}^n \delta_n^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} \leq \frac{\delta_n^{(smooth)}(a) \sum_{t=1}^n \mathbb{I}\left\{\frac{\|\boldsymbol{Z}_t\|}{b(n/k)} < a\right\}}{\sum_{t=1}^n \delta_n^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)}.$$

Define

$$a_n' := \sum_{t=1}^n \delta_n^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)$$

and

$$S_n' := \delta_n^{(smooth)}(a) \sum_{t=1}^n \mathbb{I}\left\{\frac{\|\boldsymbol{Z}_t\|}{b(n/k)} < a\right\}.$$

Also, note that $\lim_{n\to\infty} = k$. Notice

$$\begin{aligned}
\mu_n' &= E\left[S_n'\right] \\
&= E\left[\delta_n^{(smooth)}(a)\sum_{t=1}^{n}\mathbb{I}\left\{\frac{\|\boldsymbol{Z}_t\|}{b(n/k)} < a\right\}\right] \\
&= \delta_n^{(smooth)}(a)nP\left(\frac{\|\boldsymbol{Z}_1\|}{b(n/k)} < a\right) \\
&\approx \delta_n^{(smooth)}(a)(n - ka^{-1}) \\
&\to 0
\end{aligned}$$

by assumption 2. Also notice

$$\begin{aligned}
\sigma_n^{2'} &= Var\left[S_n'\right] \\
&= Var\left[\delta_n^{(smooth)}(a)\sum_{t=1}^{n}\mathbb{I}\left\{\frac{\|\boldsymbol{Z}_t\|}{b(n/k)} < a\right\}\right] \\
&\sim \left(\delta_n^{(smooth)}(a)\right)^2(ka^{-1} - k^2n^{-1}a^{-2}).
\end{aligned}$$

Consider

$$\begin{aligned}
(a_n')^{-2}\sigma_n^{2'} &= \left(\sum_{t=1}^{n}\delta_n^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)\right)^{-2} Var\left[\delta_n^{(smooth)}(a)\sum_{t=1}^{n}\mathbb{I}\left\{\frac{\|\boldsymbol{Z}_t\|}{b(n/k)} < a\right\}\right] \\
&\sim \left(\delta_n^{(smooth)}(a)\right)^2\left((ka)^{-1} - (na^2)^{-1}\right) \\
&\to 0.
\end{aligned}$$

This implies that $(a_n')^{-1}(S_n' - \mu_n') \xrightarrow{p} 0$. Now, consider

$$\begin{aligned}
P\left(\left|\frac{S_n'}{a_n'}\right| > \varepsilon\right) &= P\left(\left|\frac{S_n' - \mu_n'}{a_n'} + \frac{\mu_n'}{a_n'}\right| > \varepsilon\right) \\
&\leq P\left(\left|\frac{S_n' - \mu_n'}{a_n'}\right| + \left|\frac{\mu_n'}{a_n'}\right| > \varepsilon\right) \\
&\leq P\left(\left|\frac{S_n' - \mu_n'}{a_n'}\right| > \frac{\varepsilon}{2}\right) + P\left(\left|\frac{\mu_n'}{a_n'}\right| > \frac{\varepsilon}{2}\right) \\
&\to 0,
\end{aligned}$$

thus

$$\frac{\delta_n^{(smooth)}(a)\sum_{t=1}^n \mathbb{I}\left\{\frac{\|\boldsymbol{Z}_t\|}{b(n/k)} < a\right\}}{\sum_{t=1}^n \delta_n^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)} \xrightarrow{p} 0.$$

Note that

$$P\left(\left|\frac{S_n' - \mu_n'}{a_n'}\right| > \frac{\varepsilon}{2}\right) \to 0$$

because $(a_n')^{-1}(S_n' - \mu_n') \xrightarrow{p} 0$, and

$$P\left(\left|\frac{\mu_n'}{a_n'}\right| > \frac{\varepsilon}{2}\right) \le P\left(|\mu_n'| > \frac{\varepsilon}{2}\right) \to 0.$$

For the second term, define

$$S_n'' := \sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|Z_t\|}{b(n/k)}\right)\mathbb{I}\left\{\frac{\|Z_t\|}{b(n/k)} \in [a,1)\right\}.$$

We apply law of large numbers for triangular arrays for $\mu_n'' = E[S_n'']$ and $\sigma_n^{2''} = Var[S_n'']$, where if $(a_n')^{-2}\sigma_n^{2''} \to 0$ then $(a_n')^{-1}(S_n'' - \mu_n'') \xrightarrow{p} 0$ (Durrett, 2010, Theorem 5.4). Note that

$$\begin{aligned}
\mu_n'' &= E\left[S_n''\right] \\
&= E\left[\sum_{t=1}^n \delta_n^{(smooth)}\left(\frac{\|Z_t\|}{b(n/k)}\right)\mathbb{I}\left\{\frac{\|Z_t\|}{b(n/k)} \in [a,1)\right\}\right] \\
&= n\int_a^1 \delta_n^{(smooth)}(z)dP_{\frac{\|Z_1\|}{b(n/k)}}(z) \\
&\sim k\int_a^1 \delta_n^{(smooth)}(z)z^{-2}dz \\
&\to 0
\end{aligned}$$

by assumption 4. Also note that

$$\begin{aligned}
(a_n')^{-2}\sigma_n^{2''} &= (a_n')^{-2}Var\left[\sum_{t=1}^{n}\delta_n^{(smooth)}\left(\frac{\|Z_t\|}{b(n/k)}\right)\mathbb{I}\left\{\frac{\|Z_t\|}{b(n/k)}\in[a,1)\right\}\right] \\
&= n(a_n')^{-2}\left(\int_a^1\left(\delta_n^{(smooth)}(z)\right)^2 dP_{\frac{\|Z_1\|}{b(n/k)}}(z)-\left(\int_a^1\delta_n^{(smooth)}(z)dP_{\frac{\|Z_1\|}{b(n/k)}}(z)\right)^2\right) \\
&\sim k^{-1}\left(\int_a^1\left(\delta_n^{(smooth)}(z)\right)^2 z^{-2}dz-\left(\int_a^1\delta_n^{(smooth)}(z)z^{-2}dz\right)^2\right) \\
&\leq k^{-1}\int_a^1\left(\delta_n^{(smooth)}(z)\right)^2 z^{-2}dz \\
&\leq k\int_a^1\delta_n^{(smooth)}(z)z^{-2}dz \\
&\to 0
\end{aligned}$$

by assumption 4. Hence, $(a_n')^{-1}(S_n''-\mu_n'')\xrightarrow{p}0$. Now, consider

$$\begin{aligned}
P\left(\left|\frac{S_n''}{a_n'}\right|>\varepsilon\right) &= P\left(\left|\frac{S_n''-\mu_n''}{a_n'}+\frac{\mu_n''}{a_n'}\right|>\varepsilon\right) \\
&\leq P\left(\left|\frac{S_n''-\mu_n''}{a_n'}\right|+\left|\frac{\mu_n''}{a_n'}\right|>\varepsilon\right) \\
&\leq P\left(\left|\frac{S_n''-\mu_n''}{a_n'}\right|>\frac{\varepsilon}{2}\right)+P\left(\left|\frac{\mu_n''}{a_n'}\right|>\frac{\varepsilon}{2}\right) \\
&\to 0.
\end{aligned}$$

Therefore

$$\frac{\sum_{t=1}^{n}\delta_n^{(smooth)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)\mathbb{I}\left\{\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\in[a,1)\right\}}{\sum_{t=1}^{n}\delta_n^{(hard)}\left(\frac{\|\boldsymbol{Z}_t\|}{b(n/k)}\right)}\xrightarrow{p}0.$$

The third term follows a similar argument to the second.

CHAPTER 3

OPTIMIZING TAIL DEPENDENCE

3.1 Introduction

In this chapter, we develop a method to estimate the coefficients in the linear combination of covariates that optimizes tail dependence with a continuous response variable. When modeling a single response variable using a linear combination of covariates, it is common for researchers to consider a type of regression analysis such as standard linear regression, logistic regression, or quantile regression. Standard linear regression, which models the expected response, tends to be a poor method for describing extremes since it describes the behavior in the center of the distribution. Approaches such as quantile regression or logistic regression can be tailored to focus on large values of the response.

Logistic regression and quantile regression can do well in terms of describing the covariates that are associated with high values of the response; however, our approach of optimizing tail dependence is fundamentally different from both logistic regression and quantile regression. Although our approach is different, one can make an analogy between our approach and standard regression. Standard regression aims to find the linear combination of covariates which optimizes correlation with the response (in terms of R-squared), and our approach aims to find the coefficients which optimize our measure of tail dependence, γ.

Rather than using a regression approach, our method relies on the framework of bivariate regular variation, explained in Chapter 1. To model the relationship between a linear combination of covariates and ozone the framework allows us to model bivariate threshold exceedences, thus focusing only on extreme behavior. We use the framework of regular variation to define a measure of bivariate tail dependence which is used to measure the dependence between the function of covariates and the response. Tail dependence is then optimized subject to a constraint which imposes a required marginal condition. Our method also differs from conditional or regression approaches for extremes (Beirlant et al., 2004, Ch. 7) which model the parameters of a univariate extremes model (e.g., generalized extreme value (GEV) or generalized Pareto distribution (GPD)) as functions of covariates.

Conditional models have often been applied when the covariate is measured on a longer time scale (e.g. annual) than the response, thus allowing the researcher to extract data (e.g. annual maxima or threshold exceedances) which are considered extreme for the given the covariate. Conditional models for extremes have been used in atmospheric science to study trends by conditioning on year or to study the relationship with slowly-evolving climatological regimes (Sillmann et al., 2011; Maraun et al., 2011). However, threshold exceedance approaches have successfully been applied when the covariates vary on the same time scale as the response (e.g., Reich et al. (2013)), and we will compare the conditional approach to our approach in Section 3.4.5. Our approach is specifically designed for covariates which vary on the same time scale as the response, and would not seem well suited to modeling 'non-random' covariates such as a time trend. Importantly, there is a subtle difference between the questions answered by a conditional extremes approach and our proposed approach. Because it models the tail conditional on the covariates, the conditional approach answers the question "Given certain conditions, what is the extreme behavior?" By optimizing tail dependence, our approach answers a slightly different question of "What conditions are most strongly associated with the extreme observations?"

3.2 Maximizing Tail Dependence

Our first goal is to be able to estimate the parameter vector in the linear combination of continuous covariates that has the highest degree of tail dependence with a continuous response variable. In our particular application the covariates are meteorological variables and our response variable is ground level ozone.

3.2.1 Method

We now formalize the method to find the linear combination of covariates that optimizes tail dependence with the response variable in terms of our tail dependence measure γ as in Equation 2.4. Let the response at time $t \in \mathbb{N}$ be given by the continuous random variable Y_t (on its original scale). Also, assume (for now) that its distribution function $F_Y(y)$ is known. Furthermore, assume that there exists a $k-$dimensional random vector of continuous covariates (on their original scales) at time t given by

$$\boldsymbol{X}_t = (X_{t,1}, X_{t,2}, X_{t,3}, \dots, X_{t,k})^T.$$

Assume (for now) that the distribution function for the i^{th} covariate, $F_{X_i}(x)$, is known.

To make use of the bivariate regular variation framework, we would like the response variable and the linear combination of covariates to have unit Fréchet marginal distributions. We can easily transform Y_t to the unit Fréchet scale. We define

$$Y_t^{**} = G^{-1}[F_Y(Y_t)],$$

where G is the unit Fréchet distribution function.

To ensure that our linear combination of covariates has a unit Fréchet marginal, we do a two-step transformation procedure. We first transform each covariate to the $N(0,1)$ scale using the probability integral transformation,

$$X_{t,i}^* = \Phi^{-1}[F_{X_i}(X_{t,i})],$$

where Φ represents the Gaussian distribution function with mean 0 and variance 1. Define the vector

$$\boldsymbol{X}_t^* = (X_{t,1}^*, X_{t,2}^*, \dots, X_{t,k}^*)^T,$$

letting Σ^* denote its covariance matrix.

For any $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)^T \in \mathbb{R}^k$, construct the linear combination of covariates

$$\boldsymbol{X}_t^{*T}\boldsymbol{\beta} = \beta_1 X_{t,1}^* + \ldots + \beta_k X_{t,k}^*. \tag{3.1}$$

Note that $E[\boldsymbol{X}_t^{*T}\boldsymbol{\beta}] = 0$ and $Var[\boldsymbol{X}_t^{*T}\boldsymbol{\beta}] = \boldsymbol{\beta}^T\Sigma^*\boldsymbol{\beta}$. Furthermore, we make the modeling assumption that $\boldsymbol{X}_t^{*T}\boldsymbol{\beta}$ is approximately Gaussian. If $\boldsymbol{X}_t^*$ were multivariate Gaussian then $\boldsymbol{X}_t^{*T}\boldsymbol{\beta}$ would be exactly $N(0, \boldsymbol{\beta}^T\Sigma^*\boldsymbol{\beta})$.

We would like to transform $\boldsymbol{X}_t^{*T}\boldsymbol{\beta}$ to be unit Fréchet using a probability integral transformation, but its variance is a function of $\boldsymbol{\beta}$. By imposing the constraint $\boldsymbol{\beta}^T\Sigma^*\boldsymbol{\beta} = 1$, we ensure that $\boldsymbol{X}_t^{*T}\boldsymbol{\beta} \sim N(0,1)$. Thus the probability integral transformation

$$X_t^{**}(\boldsymbol{\beta}) = G^{-1}[\Phi(\boldsymbol{X}_t^{*T}\boldsymbol{\beta})]$$

will be distributed as unit Fréchet.

We now have the bivariate sequence $\{(X_t^{**}(\boldsymbol{\beta}), Y_t^{**})\}_{t\in\mathbb{N}}$ which we assume is a bivariate regularly varying process with unit Fréchet marginals for all $\boldsymbol{\beta} \in \mathbb{R}^k$ such that $\boldsymbol{\beta}^T\Sigma^*\boldsymbol{\beta} = 1$. Next, we wish to find the $\boldsymbol{\beta}$ that optimizes the tail dependence in terms of γ. The angular measure H is dependent upon the choice of $\boldsymbol{\beta}$, and as a reminder, we denote the angular measure as $H_{\boldsymbol{\beta}}$. Define the constrained optimum

$$\boldsymbol{\beta}^* = \underset{\boldsymbol{\beta}\in\mathbb{R}^k}{\operatorname{argmin}} \int_{S_1} |2w-1| dH_{\boldsymbol{\beta}}(w) \text{ subject to } \boldsymbol{\beta}^T\Sigma^*\boldsymbol{\beta} = 1,$$

recalling that the linear combination $\boldsymbol{X}_t^{*T}\boldsymbol{\beta}^*$ will have the highest degree of tail dependence with the response in terms of γ.

In practice, we observe $(\boldsymbol{x}_t, y_t)$ for $t = 1, 2, \ldots, n$ without any knowledge of the underlying distribution functions. We proceed by using the procedure described in Section 3 using estimated marginal distributions. To transform the response variable we use a rank transformation $y_t^{**} = G^{-1}[\hat{F}_Y(y_t)]$ where $\hat{F}_Y(a) = n^{-1}\sum_{t=1}^n \mathbb{I}\{y_t \leq a\}$ is the empirical distribution function for Y. To transform each covariate to the $N(0,1)$ scale, we perform a rank transformation using each corresponding empirical distribution function: $x^*_{t,i} = \Phi^{-1}[\hat{F}_{X_i}(x_{t,i})]$ for $i = 1, 2, \ldots, k$. To transform the linear combination to the unit Fréchet scale, we use the probability integral transformation $x^{**}_t(\boldsymbol{\beta}) = G^{-1}[\Phi(\boldsymbol{\beta}^T \boldsymbol{x}^*_t)]$, subject to the constraint $\boldsymbol{\beta}^T \hat{\Sigma}^* \boldsymbol{\beta} = 1$ and where $\hat{\Sigma}^*$ is the observed covariance matrix of the transformed covariates. Then, letting $\boldsymbol{z}^{**}_t = (x^{**}_t(\boldsymbol{\beta}), y^{**}_t)$,

$$\hat{\boldsymbol{\beta}}^* = \underset{\{\boldsymbol{\beta} \in \mathbb{R}^k : \boldsymbol{\beta}^T \hat{\Sigma}^* \boldsymbol{\beta} = 1\}}{\text{argmin}} \left(\sum_{t=1}^{n} \delta^{(smooth)}(\|\boldsymbol{z}^{**}_t\|/r_0) \right)^{-1} \sum_{t=1}^{n} \delta^{(smooth)}(\|\boldsymbol{z}^{**}_t\|/r_0) \frac{|z^{**}_{t,1} - z^{**}_{t,2}|}{z^{**}_{t,1} + z^{**}_{t,2}}$$

We use

$$\delta^{(smooth)}(\|\boldsymbol{z}\|/r_0) := \Phi\left(\frac{\frac{\|\boldsymbol{z}\|}{r_0} - 1}{r_0 \sigma} \right) = \Phi\left(\frac{\|\boldsymbol{z}\| - r_0}{\sigma} \right) \tag{3.2}$$

as our weight function, where σ determines the amount of smoothness and r_0 is the hard threshold. As $\sigma \to 0$, $\delta^{(smooth)}$ becomes a degenerate Gaussian distribution function and $\delta^{(smooth)} \to \delta^{(hard)}$. Chaudhuri and Solar-Lezama (2011) comment that a large standard deviation will help with convergence, but can cause the optimizer to converge to a local optimum that is far from the global optimum. We explore the sensitivity to the choice of σ in the simulation study.

3.2.2 Numerical Optimization

Even with the implementation of the smoothed threshold, the optimization is non-trivial. We use the statistical computing environment R (R Development Core Team, 2011) for numeric optimization. We find the constrained optimizer in the 'alabama' package (Varadhan, 2011) to work well when given good starting values. However, we suggest using more robust optimization algorithms.

We find that the differential evolution algorithm in the 'DEoptim' package (Mullen et al., 2011) and the generalized simulated annealing algorithm in the 'GenSA' (Yang Xiang et al., 2013) package to be effective and to work equally well in most cases. Unfortunately, neither of these optimizers offers the user a way to implement constraints except simple box constraints. To use DEoptim and GenSA we impose the constraint via a transformation to spherical coordinates.

The motivation for a spherical transformation is that the constraint $\boldsymbol{\beta}^T\Sigma^*\boldsymbol{\beta} = 1$ is equivilant to the statement that the optimum $\boldsymbol{\beta}^* \in \mathbb{R}^k$ will occur on the $k-$dimensional ellipse $\boldsymbol{\beta}^T\Sigma^*\boldsymbol{\beta} = 1$. We consider $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_{k-1})^T$ and $r > 0$ where $\theta_j \in (0, \pi)$ for $j = 1, 2, \ldots, k-2$ and $\theta_{k-1} \in (0, 2\pi)$. We note that the constraints on the $\boldsymbol{\theta}$ are simple box constraints which nearly all optimizers are able to handle. To ensure that $(\boldsymbol{\theta}, r)$ is on the ellipse, we constrain $r = (\boldsymbol{\beta}^T\Sigma^*\boldsymbol{\beta})^{-1/2}$. Once $\boldsymbol{\theta}$ is chosen and r is determined, we can transform back to rectangular coordinates using

$$\begin{aligned}
\beta_1 &= r\cos\theta_1 \\
\beta_2 &= r\sin\theta_1\cos\theta_2 \\
\beta_3 &= r\sin\theta_1\sin\theta_2\cos\theta_3 \\
&\vdots \\
\beta_{k-1} &= r\left(\prod_{i=1}^{k-2}\sin\theta_i\right)\cos\theta_{k-1} \\
\beta_k &= r\left(\prod_{i=1}^{k-1}\sin\theta_i\right).
\end{aligned}$$

Doing the optimization in spherical coordinates reduces the problem to a $k-1$ dimensional optimization problem with upper and lower bounds on the parameters as opposed to a k dimensional optimization with a quadratic constraint.

3.2.3 Understanding the Optimization Surface

In order to better understand how this method works, we use a relatively simple model to examine the optimization surface and how the choice of $\boldsymbol{\beta}$ affects the scatterplot of $X^{**}(\boldsymbol{\beta})$

versus Y^{**}. The model we choose has two covariates: X_1 and X_4. Since this model has just two covariates, we can attempt to visualize the optimization surface. Figure 3.1 plots contours of $\hat{\gamma}$ given β_{X_1} and β_{X_4}. The constraint we impose means that we are restricted to combinations of β_{X_1} and β_{X_4} on the ellipse. This restriction is to ensure that $X^{**}(\boldsymbol{\beta})$ will be unit Fréchet.

Next, we examine three points on the ellipse more closely. In Figure 3.1 Point A is the constrained optimum, point B is a somewhat poor choice for $\boldsymbol{\beta}$, and point C is one of the worst choices for $\boldsymbol{\beta}$. Figure 3.2 plots $X^{**}(\boldsymbol{\beta})$ versus Y^{**} for the three choices of $\boldsymbol{\beta}$ implied by points A, B, and C in Figure 3.1. Figure 3.2 also gives the corresponding histograms for the angular component for each of the three choices of $\boldsymbol{\beta}$.

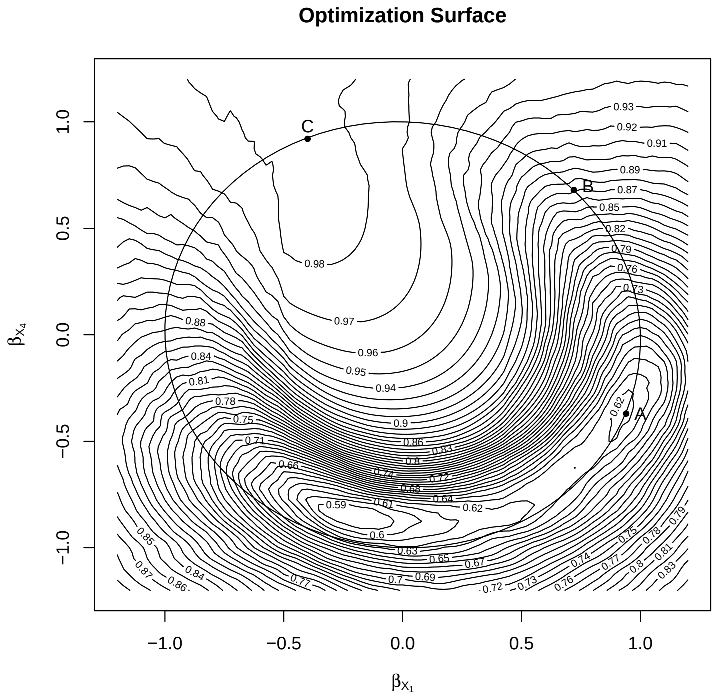

Figure 3.1: For a model with two covariates, we give the value of $\hat{\gamma}$ (the contours) for values of $\hat{\beta}_{X_1}$ and $\hat{\beta}_{X_1}$ between -1 and 1. The points on the ellipse give the region satisfying the constraint $\boldsymbol{\beta}^T\hat{\Sigma}^*\boldsymbol{\beta} = 1$. We consider three points on the ellipse: point A is the constrained optimum, point B yields a moderate degree of tail dependence, and point C yields an extremely low degree of tail dependence. We give the scatterplots and histograms estimating H generated using these three $\boldsymbol{\beta}$s in Figure 3.2.

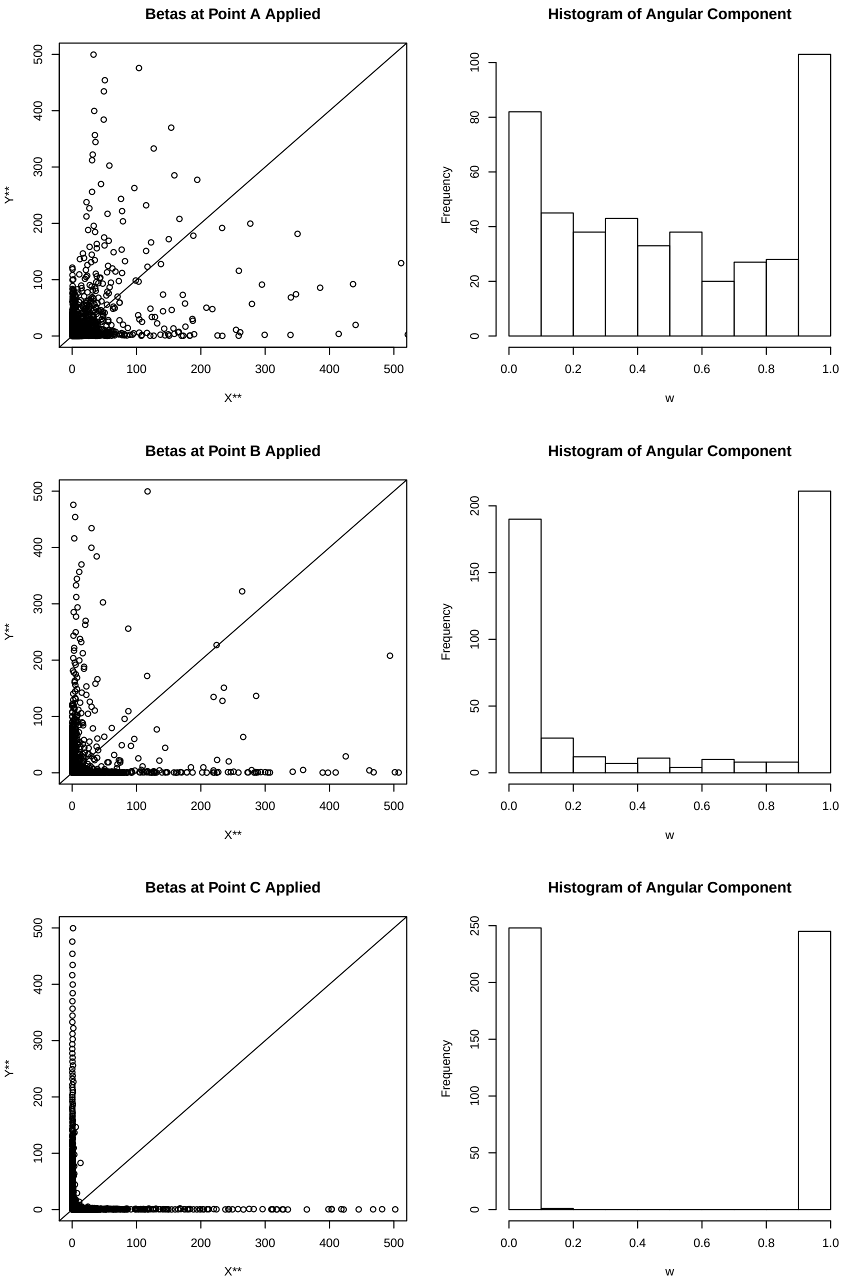

Figure 3.2: In Figure 3.1 we choose three $\boldsymbol{\beta}$s in the constrained region. We then apply the $\boldsymbol{\beta}$s to the simulated data on the Unit Fréchet scale. The top two panels give the scatterplot and histogram of w, the angles of the top 5% of points in terms of radial component, for the constrained optimum (point A in Figure 3.1). The bottom two panels give analogous plots for a choice of $\boldsymbol{\beta}$ that produces a low degree of tail dependence (point B in Figure 3.1). The middle two panels give analogous plots for a choice of $\boldsymbol{\beta}$ that produces a moderate degree of tail dependence (point C in Figure 3.1).

The scatterplot and histogram implied by point C indicate something close to asymptotic independence based on the fact that the points with large radial component hug the axes. Using the $\boldsymbol{\beta}$ at point B shows a higher degree of asymptotic dependence. This is seen by the fact that there are some points with large radial component in the interior. However, many of the large points are still near the axes. The constrained optimum, at point A, shows a higher degree of tail dependence as there are a large number of points in the interior.

In examining the scatterplots, it is important to point out that the values on the vertical axis are fixed. Choosing a different $\boldsymbol{\beta}$ shifts the points in the scatterplot horizontally, while maintaining unit Fréchet marginal distributions for $X^{**}(\boldsymbol{\beta})$ and Y^{**}. To see a higher degree of tail dependence than we see at the constrained optimum, we may need to consider additional covariates. We pursue this in the model fitting section.

3.3 Model Comparison Using Cross Validation

For model comparison, we need a criterion to compare two subsets of covariates in terms of their ability to optimize tail dependence with ozone. Using $\hat{\gamma}$ as a criterion could lead to overfitting as models would not be penalized for including unnecessary covariates. We employ the following 10-fold cross-validation procedure.

Given a particular set of covariates, we first obtain y_t^{**} and $x_{t,i}^{*}$ as described in Section 3.1. We then randomly partition these transformed observations into 10 equally sized subsets. Let the observation numbers corresponding to the i^{th} partition be given by $\Gamma_i \subset \{1, 2, \ldots, n\}$. Similarly, let the indices corresponding to all observations except the i^{th} partition be given by $\Gamma_{-i} = \{1, 2, \ldots, n\} \backslash \Gamma_i$. At the i^{th} iteration (for $i = 1, \ldots, 10$), we obtain the parameter estimates using all observations except those in the i^{th} partition:

$$\hat{\boldsymbol{\beta}}^{(-i)} = \underset{\boldsymbol{\beta} \in \mathbb{R}^k}{\operatorname{argmin}} \left(\sum_{t \in \Gamma_{-i}} \delta_{t;n}^{(smooth)} \right)^{-1} \sum_{t \in \Gamma_{-i}} \delta_{t;n}^{(smooth)} \frac{|x_t^{**}(\boldsymbol{\beta}) - y_t^{**}|}{x_t^{**}(\boldsymbol{\beta}) + y_t^{**}} \text{ subject to } \boldsymbol{\beta}^T \hat{\Sigma}^* \boldsymbol{\beta} = 1.$$

After obtaining $\hat{\boldsymbol{\beta}}^{(-i)}$, we calculate $\hat{\gamma}_{(-i)}$ by applying $\hat{\boldsymbol{\beta}}^{(-i)}$ to the held out data:

$$\hat{\gamma}_{(-i)} = \left(\sum_{t \in \Gamma_i} \delta_{t;n}^{(smooth)} \right)^{-1} \sum_{t \in \Gamma_i} \delta_{t;n}^{(smooth)} \frac{|x_t^{**}(\hat{\boldsymbol{\beta}}^{(-i)}) - y_t^{**}|}{x_t^{**}(\hat{\boldsymbol{\beta}}^{(-i)}) + y_t^{**}}.$$

After iterating over the 10 partitions, we define

$$CV = 10^{-1} \sum_{i=1}^{10} \hat{\gamma}_{(-i)}.$$

We also note that $CV \in [0, 1]$ and that a smaller CV value implies a higher degree of tail dependence.

3.4 Simulation Study

In order to illustrate and test our method we undertake a simulation study.

3.4.1 Description of Simulated Data

We randomly generate 5,000 iid realizations of five covariates, X_1, X_2, X_3, X_4, X_5. The first four covariates are four-dimensional Gaussian with non-identity covariance matrix. The fifth covariate is drawn independently of the first four covariates uniformly on the unit interval, i.e. $X_5 \sim U(0, 1)$. The response, Y, is a linear combination of functions of the five covariates plus noise:

$$Y_t = -.3X_{t,1} + X_{t,2} - .75X_{t,4} - (X_{t,2})^2 + 6\Phi[(X_{t,1} - X_{1;.95})/.35]X_{t,5} + \varepsilon_t, \qquad (3.3)$$

where $X_{1;.95}$ represents the .95 quantile of X_1 and $\varepsilon_t \sim$ iid N(0, .00125). The idea is to create a response which has a nonlinear relationship with the covariates, and which only appears when conditions are extreme. Specifically, the term $6\Phi[(X_{t,1} - X_{1;.95})/.35]X_{t,5}$ only contributes to Y_t when $X_{t,1}$ is extremely large.

Figure 3.3 shows scatterplots of the response variable versus each of the five covariates; we focus on the relationships driving the large response values.. Large values of Y_t are clearly associated with large values of $X_{t,1}$. The quadratic term causes large values of the response to occur when $X_{t,2}$ is between 0 and 1.5. $X_{t,3}$ does not seem to be related to large response values and the relationship between $X_{t,4}$ and the large values of Y_t seems slight. Finally, the evidence of $X_{t,5}$'s influence on the extreme behavior is slight, and its interaction with $X_{t,1}$ is not apparent from these plots.

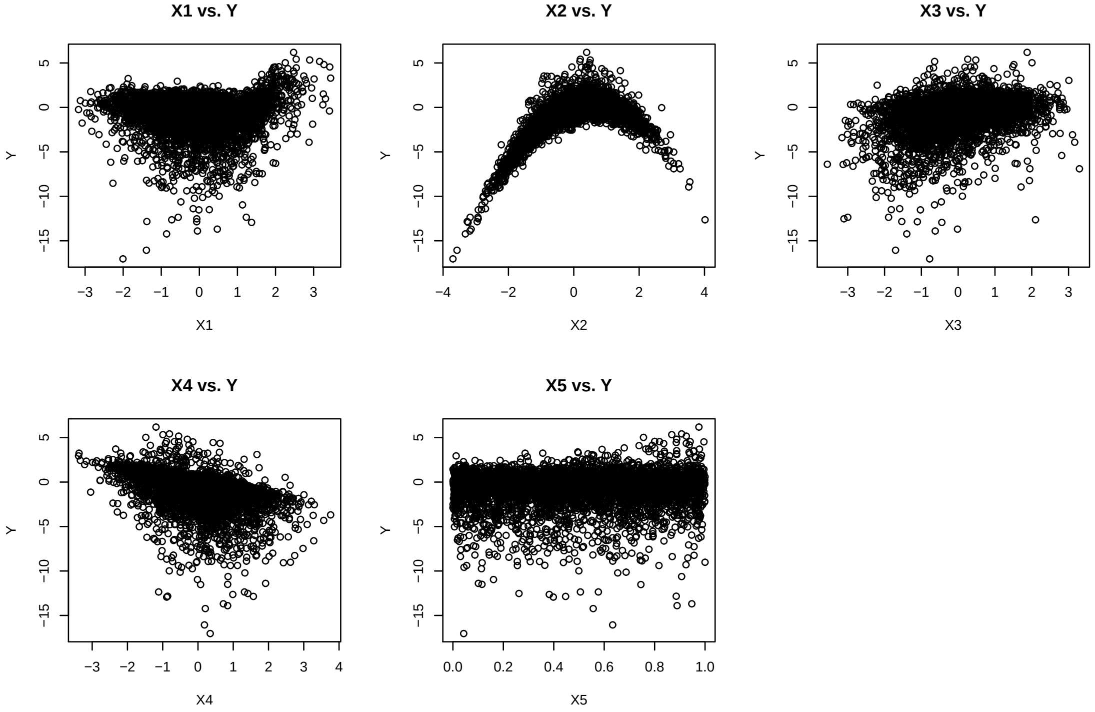

Figure 3.3: Scatterplots for each of the five covariates versus the response in the simulation study. All variables are on their original scales.

3.4.2 Model Selection

To analyze the simulated data, we consider the models M1-M6 listed below. The first model includes all five covariates. The models M2-M4 each leave out a single covariate: X_3, X_2, X_1 respectively. Model five leaves out X_3 but adds an interaction between X_1 and

X_5. Model six adds $(X_2)^2$ to model five. Note that none of these models corresponds exactly to Equation (3.3), but that M6 is closest that it includes some type of interaction between X_1 and X_5 and the quadratic behavior of X_2.

M1: $x_t^{**} = [X_{t,1}^*, X_{t,2}^*, X_{t,3}^*, X_{t,4}^*, X_{t,5}^*]\boldsymbol{\beta}_1$

M2: $x_t^{**} = [X_{t,1}^*, X_{t,2}^*, X_{t,4}^*, X_{t,5}^*]\boldsymbol{\beta}_2$

M3: $x_t^{**} = [X_{t,1}^*, X_{t,3}^*, X_{t,4}^*, X_{t,5}^*]\boldsymbol{\beta}_3$

M4: $x_t^{**} = [X_{t,2}^*, X_{t,3}^*, X_{t,4}^*, X_{t,5}^*]\boldsymbol{\beta}_4$

M5: $x_t^{**} = [X_{t,1}^*, X_{t,2}^*, X_{t,4}^*, X_{t,5}^*, X_{t,1}^* \times X_{t,5}]\boldsymbol{\beta}_5$

M6: $x_t^{**} = [X_{t,1}^*, X_{t,2}^*, X_{t,4}^*, X_{t,5}^*, X_{t,1}^* \times X_{t,5}^*, (X_{t,2}^*)^2]\boldsymbol{\beta}_6$

We optimize the tail dependence with a smooth threshold where $r_0 = 40$, the .95 quantile of the radial components $\|\boldsymbol{z}_t\| = z_{t,1} + z_{t,2}$ $t = 1, \dots, n$ and $\sigma = 1.25$. Marginal transformations from the original scale were based on the rank transform.

Table 3.1 gives $\hat{\gamma}$ and the CV value for each of the six models. Comparing models M1 and M2 shows that the CV method is effective in selecting covariates which influence extreme responses. The overall score $\hat{\gamma}$ is higher (worse) for M2 as it is a submodel of M1; however, the CV score for M2 is better indicating that $X_{t,3}$ is not useful for describing extreme response values. The CV score for M4 shows that leaving out X_1 is clearly not a good idea as the CV value is by far the highest of the six models. M5, which adds the interaction of X_1 and X_5, gives a noticeably improved CV value, and M6, which is closest to the generating model has the best CV score.

Table 3.1: The score (the optimized value of $\hat{\gamma}$) and the 10-fold cross-validation value for each of the six models considered in the simulation study.

	M1	M2	M3	M4	M5	M6
$\hat{\gamma}$	0.4930	0.4950	0.5125	0.6053	0.4603	0.4060
CV	0.5052	0.5017	0.5240	0.6196	0.4703	0.4120

The plots in Figure 3.4 give a visual approach to assess how well each model is capturing the tail dependence. For each model, $x_t^{**}(\hat{\boldsymbol{\beta}}^*)$ is plotted versus y_t^{**}. Note that the large points for the models with lower $\hat{\gamma}$ scores and CV values, like M5 and M6, occur in the interior of the positive orthant, whereas models with higher values of $\hat{\gamma}$ (M4) have more points near the axes.

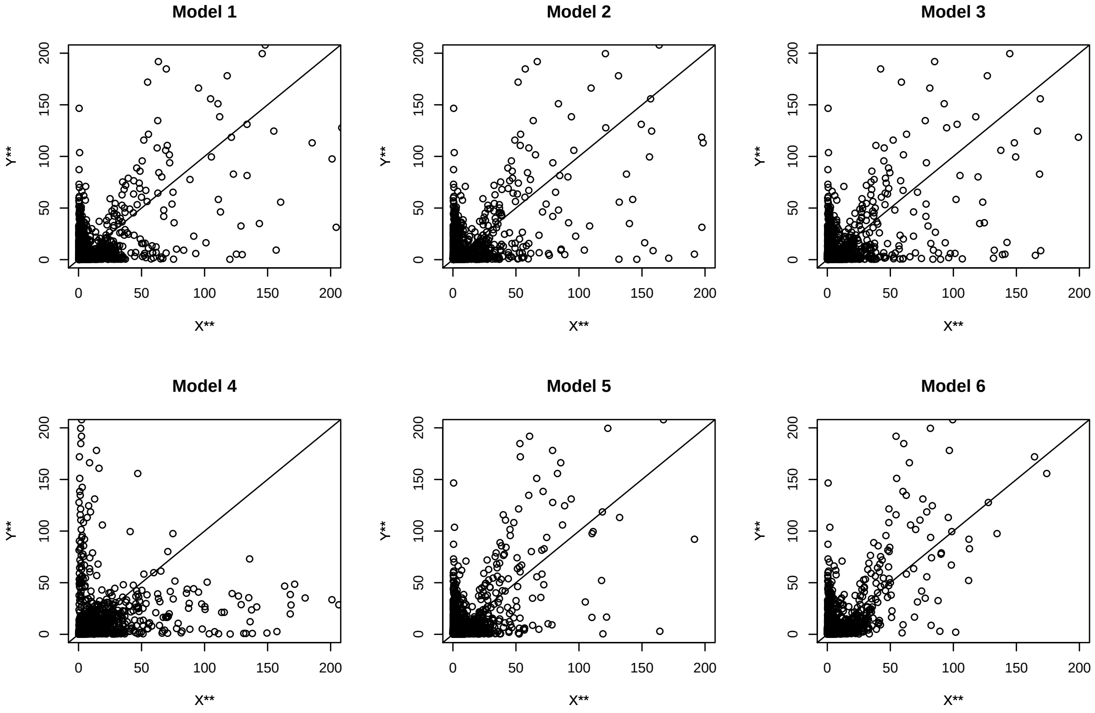

Figure 3.4: For each of the six models we consider in the simulation study, the scatterplot of $x_t^{**}(\hat{\boldsymbol{\beta}}^*)$ versus y_t^{**} is given. Models that yield a linear combination with a higher degree of tail dependence with the response will result in a scatterplot with large points closer to the identity line.

3.4.3 Parameter Estimates: Interpretation and Uncertainty

For a given model, the parameter estimates that we obtain can be useful, however care needs to be taken with respect to their interpretation. Due to the constraint and the marginal transformations, only the relative magnitude and sign of the parameter estimates are interpretable. Table 3.2 reports the parameter estimates for M6, with standard errors in parentheses. Standard errors are calculated using the nonparametric paired bootstrap (Givens and Hoeting, 2005, p. 257), where $(\boldsymbol{x}_i, y_i)$ are resampled, and utilizing 640 bootstrap replications. The estimate with largest magnitude corresponds to X_1, indicating that it plays an important role in describing extreme levels of the response. Standard errors indicate that the $\hat{\beta}_{X_1}$, $\hat{\beta}_{X_4}$, $\hat{\beta}_{X_1X_5}$ and $\hat{\beta}_{(X_2)^2}$ terms are significant, and the signs of these agree with the behavior shown in Figure 3.3.

Table 3.2: The parameter estimates for M6 in the simulation study with bootstrap standard errors in parentheses. We utilize the nonparametric paired bootstrap (Givens and Hoeting, 2005, p. 257), where $(\boldsymbol{x}_i, y_i)$ are resampled, and 640 bootstrap replications.

Coefficient	$\hat{\beta}_{X_1}$	$\hat{\beta}_{X_2}$	$\hat{\beta}_{X_4}$	$\hat{\beta}_{X_5}$	$\hat{\beta}_{X_1X_5}$	$\hat{\beta}_{(X_2)^2}$
Estimate	.61 (.12)	.10 (.10)	-.22 (.07)	-.19 (.14)	.47 (.15)	-.17 (.08)

3.4.4 Smooth Threshold Sensitivity Analysis

To explore sensitivity to the smoothing parameter, we fit M6 with $\sigma = .3125, .625, 1.25, 2.5$, and 5.0. Table 3.3 gives the parameter estimates for M6 for the various smoothing levels, and the optimization seems to be robust to the choice of standard deviation for values close to $\sigma = 1.25$.

Table 3.3: The parameter estimates for M6 using .3125, .625, 1.25, 2.5, and 5.0 as the standard deviation of the Gaussian cdf in the smoothed threshold.

	$\hat{\beta}_{X_1}$	$\hat{\beta}_{X_2}$	$\hat{\beta}_{X_4}$	$\hat{\beta}_{X_5}$	$\hat{\beta}_{X_1X_5}$	$\hat{\beta}_{(X_2)^2}$
$\sigma = .3125$	0.6489	0.1070	-0.1978	-0.1322	0.4138	-0.1595
$\sigma = .625$	0.6356	0.1006	-0.2035	-0.1497	0.4322	-0.1621
$\sigma = 1.25$	0.6319	0.0965	-0.2079	-0.1497	0.4349	-0.1673
$\sigma = 2.5$	0.5922	0.1005	-0.2141	-0.1990	0.4864	-0.1751
$\sigma = 5.0$	0.5480	0.0921	-0.2245	-0.2394	0.5446	-0.1780

3.4.5 Comparison to Other Methods

We compare our tail dependence method to several other approaches: (standard) multiple regression, logistic regression, quantile regression, a conditional extremes approach, and the preprocessing approach of Eastoe and Tawn (2009). For each of the three regression procedures, we use M6 and obtain parameter estimates ($\hat{\boldsymbol{\beta}}_{\text{mr}}$, $\hat{\boldsymbol{\beta}}_{\text{lr}}$, and $\hat{\boldsymbol{\beta}}_{\text{qr}}$ respectively). For logistic regression we transform the response such that it takes on the value 1 for the top 5% of Y values and 0 otherwise. For quantile regression, we model the .95 quantile.

The conditional extremes approach employs a covariate-varying threshold with a conditional GPD model for threshold exceedances. The threshold model is a quantile regression fit to the .90 quantile (chosen to have an adequate sample of exceedances), and the conditional GPD models the log-scale parameter as a function of covariates. To follow the preprocessing approach of Eastoe and Tawn (2009), we perform a Box-Cox transformation of the response using $\lambda = 3.5$, which is suggested by the plot of the profile log-likelihood versus λ. We define $y_t^* := (y_t^\lambda - 1)/\lambda$ and use Gaussian maximum likelihood to obtain parameter estimates for the functions $\mu(\boldsymbol{x}_t) = (1, \boldsymbol{x}_t^T)\boldsymbol{\mu}_{\text{pr}}$ and $\sigma(\boldsymbol{x}_t) = \exp((1, \boldsymbol{x}_t^T)\boldsymbol{\sigma}_{\text{pr}})$ where $\boldsymbol{\mu}_{\text{pr}}$ and $\boldsymbol{\sigma}_{\text{pr}}$ are the corresponding parameter vectors. The preprocessed response is obtained by defining $z_{t_{\text{pr}}} := (y_t^* - \mu(\boldsymbol{x}_t))/\sigma(\boldsymbol{x}_t)$. A conditional GPD model is then fit to the preprocessed responses, a threshold of 2 (.97 empirical quantile) was suggested by diagnostic plots. For both the conditional extremes and preprocessing approaches, AIC results suggested using M6 for the initial fit (quantile regression or preprocessing) and also M6 for the conditional modeling of the GPD's log-scale parameter.

Because it is invariant to marginal transformations, we employ Kendall's τ to assess dependence between the various methods' predicted values and the response. For the regression models, predicted values are obtained in the usual way: $\hat{y}_{t_{\text{mr}}} = (1, \boldsymbol{x}_t^T)\hat{\boldsymbol{\beta}}_{\text{mr}}$, $\hat{y}_{t_{\text{lr}}} = \text{logit}^{-1}\left((1, \boldsymbol{x}_t^T)\hat{\boldsymbol{\beta}}_{\text{lr}}\right)$, and $\hat{y}_{t_{\text{qr}}} = (1, \boldsymbol{x}_t^T)\hat{\boldsymbol{\beta}}_{\text{qr}}$. There is not an obvious way to obtain predicted values for the conditional extremes and preprocessing approaches. However, as only order matters for Kendall's τ, we obtain predicted values from the .5 quantile of the fitted GPD, with the final predicted values $\hat{y}_{t_{\text{cond}}}$ and $\hat{y}_{t_{\text{pr}}}$ obtained by adding the threshold or undoing the preprocessing step. As we wish to focus only on the tail behavior, we only use points which exceed the .97 empirical quantile in either marginal in our calculation. Our tail dependence method exhibits positive concordance among these large values with $\tau = .35$. Standard regression, logistic regression, quantile regression, the conditional extremes approach and the preprocessing methods yield values of τ = -.21, -.14, -.17, -.10, and -.21 respectively.

Figure 3.5 shows scatterplots for the response versus predicted values for logistic regression, the conditional extremes approach, the preprocessing method, and our tail dependence method. Points which exceed the .97 quantile in either marginal are marked with black circles, while nonexceedances are marked with gray crosses. A method which does a good job of describing tail dependence will produce large predicted values that correspond with large values of the response. Logistic regression only does a fair job modeling the most extreme events, as the 0/1 coding yields no information about the actual response values above the .95 quantile. The scatterplots for standard multiple regression and quantile regression (not shown) are similar. The conditional extremes approach offers an improvement over logistic regression as presumably the conditional GPD helps to further describe the behavior of the response given the covariate conditions. The preprocessing approach shows an increased number of 'false alarms': points whose predicted values are large when the actual response is not. Contrary to the previous methods which all do a fair job of modeling the bulk of the distribution even though the methods themselves are extremes-focused, our tail dependence method only models the relationship between the covariates and the response at extreme levels.

Obviously, the generating function (3.3) was designed to have interesting extreme behavior, which would tend to favor our approach over methods which model the entire distribution. Although inference for the tail dependence approach is involved, it is a one-step process, and two step procedures such as the conditional extremes or preprocessing methods could be challenging for a data-mining application as we describe in the next section.

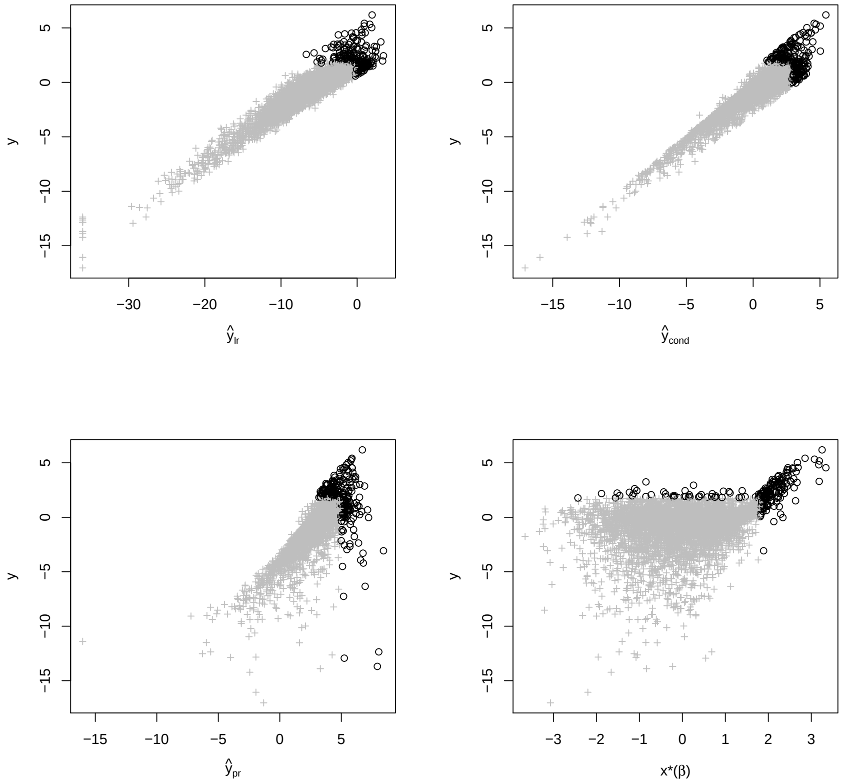

Figure 3.5: Scatterplots of y vs. the predicted values for logistic regression (top left), the conditional extremes approach (top right), the preprocessing method (bottom left), and our tail dependence method (bottom right). The points which exceed the .97 quantile in either marginal are marked with black circles, while nonexceedances are marked with gray crosses.

CHAPTER 4

DATA MINING FOR EXTREME BEHAVIOR

4.1 Application to Ground Level Ozone Pollution

We now employ our method in a data mining capacity in order to better understand the meteorological drivers of extreme ground level ozone. We analyze ozone data for Atlanta, Georgia and Charlotte, North Carolina because of their geographical proximity and consistent data records for ground level ozone.

Previous studies have used EVT to model ground level ozone. Smith (1989) describes a method to model extremes for non-stationary sequences of ozone data, allowing for the parameters of the extreme value distribution to depend on covariates. Tobias and Scotto (2005) analyze extreme ozone in Barcelona, Spain using a method similar to Smith (1989). Heffernan and Tawn (2004) utilize a conditional approach to model the multivariate distribution of several air pollutants, including ozone. Essentially, this allows the analyst to define what is extreme based on temporal or meteorological covariates, using the variables on their original scales. Eastoe and Tawn (2009) introduce a preprocessing method to account for non-stationarity time series and apply this to ground level ozone. They describe this method as analogous to the common approach of preprocessing a non-stationary time series. Ali et al. (2012) describe a variation of the preprocessing method proposed by Eastoe and Tawn (2009) and give an analysis of extreme ozone in Malaysia.

4.1.1 Data

An EPA website[1] provides ground level ozone data as well as data on other pollutants. We select station 13-121-0055 in Atlanta and station 37-119-1005 in Charlotte because of their

[1] See http://www.epa.gov/airdata/ad_maps.html.

long data records and relative lack of missing values. Although the ozone level is measured hourly at each station, the response variable we use is the maximum eight hour average ozone as it is a value on which United States National Ambient Air Quality Standards are based. The EPA-defined ozone season for North Carolina is April through October while the ozone season for Georgia is March through October. In our analysis, we use daily responses from April through October for both locations. Our analysis is based on data from the years 1992 through 2010, providing a total of 4,037 observations for Atlanta and 4,055 observations for Charlotte.

Ground level ozone readings have been decreasing over most of the United States in recent years. On its website, the EPA reports that "nationally, average ozone levels declined in the 1980s, leveled off in the 1990s, and showed a notable decline after 2002."[2] This is a trend that we notice in our exploratory data analysis. To account for non-stationarity, we transform the response variable by partitioning the data into nonoverlapping four-year blocks. In each block, we fit a gamma distribution to the observations below the .95 quantile and a generalized Pareto distribution for observations above the .95 quantile. We then use these estimated distribution functions to transform the response to unit Fréchet via a probability integral transformation. We employ a parametric model for marginal transformation as a rank transform resulted in common values in the tail if blocks have the same number of observations.

Not surprisingly, a seasonal effect is also apparent in the ozone data. Because our aim is to link extreme ozone levels to meteorological conditions, we do not deseasonalize the ozone data. Rather, we assume that by conditioning on the relevant meteorological variables by including them in the model, we are able to account for the seasonal behavior in the ozone response. Meteorological covariates are obtained from the North American Regional Reanalysis (NARR)[3] (Mesinger et al., 2006). The data are in gridded cells, approximately

[2] See http://www.epa.gov/airtrends/ozone.html.

[3] NARR data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/

30km by 30km in size, differing from our ozone data which correspond to point locations. Understanding the relationship between ozone and meteorological variables of large spatial scales is motivated by the larger project's goal of investigating the simulation of air quality extremes in atmospheric chemistry models. The NARR provides a large number of meteorological variables. Based on guidance from the atmospheric chemists collaborators, we initially use 18 NARR covariates in our models; however, we also consider these variables on different spatial scales and transformations of these variables leading to a longer overall list of possible covariates. We expect that many of these variables will be irrelevant in terms of explaining extreme ozone behavior. Furthermore, we aim to explore whether any interactions between covariates are useful to explain extreme ozone behavior. Hence, covariate selection is a primary goal of this study.

We do not include information about the ozone precursors $\mathrm{NO_x}$ and VOCs among our covariates. The EPA monitors NO and $\mathrm{NO_2}$, and although the $\mathrm{NO_x}$ data record is not as extensive as it is for ozone, many studies (e.g., Eastoe (2009)) have found $\mathrm{NO_x}$ measurements to be helpful when modeling ozone. The larger aim of our project is to provide information to improve atmospheric chemistry models which require information about $\mathrm{NO_x}$ *emissions*, rather than measurements. $\mathrm{NO_x}$ emissions are not known at the daily level, and are presumed by modelers to be relatively constant. Our aim is to link extreme ozone to meteorological conditions, and we believe that the daily variability found in $\mathrm{NO_x}$ measurements is largely attributable to meteorology rather than fluctuations in emissions.

4.1.2 Handling a Non-continuous Covariate: Precipitation

The method described in Section 3 requires continuous covariates to perform the two-step marginal transformation leading to $X_t^{**}(\boldsymbol{\beta})$. We wish to investigate precipitation's effect on extreme ozone, but precipitation has a positive probability of being exactly zero. Exploratory analysis indicates that the presence of precipitation likely affects extreme ozone, but the amount of precipitation may not be important. Thus, we extend the method to account for variables like precipitation by including a precipitation indicator and interactions with this indicator.

In Chapter 2, it was critical that the distribution of $\boldsymbol{X}_t^{*T}\boldsymbol{\beta}$ be known for any $\boldsymbol{\beta}$, which we achieved with the constraint. Let $X_{t,P}$ be the amount of precipitation at time t, where $P(X_{t,P} = 0) > 0$. Assume there are $k + l - m$ total covariates included: k 'main effects' as before, and l effects to be included in the precipitation interaction term, m of which were already included as main effects. Including the precipitation covariate and interactions and letting $X_{t,1}, \ldots, X_{t,m}$ be the overlapping covariates changes Equation (3.1) to

$$\begin{aligned} \boldsymbol{X}_t^{*T}\boldsymbol{\beta} &= \beta_1 X_{t,1}^* + \ldots + \beta_m X_{t,m}^* + \beta_{m+1} X_{t,m+1}^* + \ldots + \beta_k X_{t,k}^* + \qquad (4.1) \\ & \mathbb{I}_{\{X_{t,P}>c\}}(\beta_0^{(P)} + \beta_1^{(P)} X_{t,1}^* + \ldots + \beta_m^{(P)} X_{t,m}^* + \beta_{m+1}^{(P)} X_{t,k+1}^* + \ldots + \beta_l^{(P)} X_{t,k+l-m}^*), \end{aligned}$$

for $\boldsymbol{\beta} = (\boldsymbol{\beta}_{(P^C)}^T, \boldsymbol{\beta}_{(P)}^T)^T = ((\beta_1, \ldots, \beta_k), (\beta_0^{(P)}, \beta_1^{(P)}, \ldots, \beta_l^{(P)}))^T$. As before, if $\boldsymbol{X}_t^{*T}\boldsymbol{\beta}|(X_{t,P} \leq c) \sim N(0,1)$ under the constraint $\boldsymbol{\beta}_{(P^C)}^T \Sigma^* \boldsymbol{\beta}_{(P^C)} = 1$. However, given $X_{t,P} > c$, (4.1) becomes

$$\begin{aligned} \boldsymbol{X}_t^{*T}\boldsymbol{\beta}|(X_{t,P} > c) &= (\beta_1 + \beta_1^{(P)}) X_{t,1}^* + \ldots + (\beta_m + \beta_m^{(P)}) X_{t,m}^* + \\ & \beta_{m+1} X_{t,m+1}^* + \ldots + \beta_k X_{t,k}^* + \beta_0^{(P)} + \\ & \beta_{m+1}^{(P)} X_{t,k+1}^* + \ldots + \beta_l^{(P)} X_{t,k+l-m}^*, \qquad (4.2) \end{aligned}$$

which is distributed $N(\beta_0^{(P)}, \boldsymbol{\lambda}^T \Psi^* \boldsymbol{\lambda})$ where Ψ^* is the covariance matrix of the continuous covariates and $\boldsymbol{\lambda} = ((\beta_1 + \beta_1^{(P)}), \ldots, (\beta_m + \beta_m^{(P)}), \beta_{m+1}, \ldots, \beta_k, \beta_{m+1}^{(P)}, \ldots, \beta_l^{(P)})^T$. $\boldsymbol{X}_t^{*T}\boldsymbol{\beta}$ has a distribution which is a known mixture of Gaussians for any $\boldsymbol{\beta}$ and we can proceed with the second transformation as before, employing sample covariance matrices and the observed mixture proportion.

4.2 Data Mining Procedure and Results

To describe the model space, we consider binary strings $\boldsymbol{\omega}$ in the space $\{0,1\}^{k'}$, where k' is the total number of covariates (including interactions and transformations). In these strings, a 1 (or 0) in the j^{th} position indicates the presence (or absence) of the j^{th} covariate. Thus, $\boldsymbol{\omega}$ corresponds to the unique representation of one particular model. We denote the CV value for the model given by $\boldsymbol{\omega}$ by $CV(\boldsymbol{\omega})$ and want to find models for which $CV(\boldsymbol{\omega})$ is small. We believe that ozone readings are conditionally independent given the value of meteorological covariates. However, general cross validation methods require independence. We address this issue by considering a block cross validation method. Analysis reveals that the block cross validation method produces results that are very similar to traditional cross validation. For this reason we use traditional cross validation methods in this work.

As we are not able to fit all possible models, we approach the task of data mining in a multiple-step procedure, and we will discuss results at each step. We begin by obtaining CV scores for all possible models with up to four covariates. From this we obtain some understanding of the main effects which influence extreme ozone. We then explore all possible models that are generated by adding up to four additional variables to our best four covariate model. We then perform an automated model search procedure over $\{0,1\}^{k'}$ using a discrete optimization technique, using the preliminary results to give us starting values. We use the Yellowstone computing system (Computational and Information Systems Laboratory, 2012) to perform computations. Also, based on exploratory analysis, we employ a smooth threshold with mean equal to the .95 quantile of the radial components and $\sigma = 1.25$.

We note that others have proposed variable selection methods for high dimensional problems. For example, Fan and Lv (2008) propose Sure Independence Screening (SIS) as a method to filter out poorly performing explanatory variables. To the best of our knowledge, SIS has not been utilized in an extremes setting. To use SIS, the analyst filters out unnecessary variables based on bivariate correlations between each predictor variable and the response. He or she then utilizes a method that is more appropriate for lower dimensional problems, such as the LASSO (Tibshirani, 1996). Fan and Lv (2008) indicate that SIS is especially well suited for problems where the number of predictor variables is larger than the sample size. Our data mining strategy is somewhat similar in spirit, in the sense that we initially use an exhaustive approach to gain understanding of potential covariates. Our model search proceeds in three stages which increase in complexity and in the extent of the model space which we searched. We feel that our data mining approach is an effective strategy given our problem and goals.

Fan and Lv (2008) also propose Iterative Sure Independence Screening (ISIS) to "overcome some weak points" of SIS. This iterative adaptation of SIS models the residuals at each step. Fan and Lv (2008) claim that ISIS is able to utilize the joint information contained in the explanatory variables, whereas SIS only takes advantage of marginal information. Fan et al. (2009) extend ISIS to be used in a general psuedo-likelihood framework. However, it is not clear how ISIS could be applied to our modelling procedure.

4.2.1 Model Exploration: Four Variable Models

As a first step, we fit all possible models with up to four covariates because an exhaustive search of such models is possible. There are nearly 35,000 models of this type, and we fit all of the approximately 10,000 models that do not include highly correlated covariates. We allow the models to include the precipitation indicator and interactions between continuous covariates and the precipitation indicator, but do not consider interactions between continuous covariates. We define the precipitation indicator to be $\mathbb{I}\{X_{t,P} > .01\text{in.}\}$.

Table 4.1 reports the covariates in the five models with the lowest CV scores for Atlanta and Charlotte. In Atlanta, variables such as temperature, wind speed, downward shortwave radiative flux (dswrf, a measure of sunshine), the precipitation indicator, height of the planetary boundary layer (hpbl) at different times of day, and relative humidity seem to be in the best fitting models. In Charlotte, similar variables appear along with northwest or west (shown as a negative coefficient on east) wind directions. Most of these variables are not surprising, as one would suspect high ozone to be associated with hot sunny days with low wind speeds. We view the results from this first step mostly as confirmatory: the approach is choosing sensible covariates when limited to only four variables at a time. That the precipitation indicator appears is somewhat interesting as Jacob and Winner (2009) noted that precipitation has little effect on ground level ozone pollution. While precipitation may have little effect on mean levels of ground level ozone, it seems reasonable that the most extreme ozone levels do not occur on days where there is precipitation.

We also note that the four-covariate models with the best CV scores are identical for Atlanta and Charlotte, and this allows us to compare the parameter estimates for these common models. Table 4.1 also gives the parameter estimates and bootstrap standard errors for the top model, using the nonparametric paired bootstrap (Givens and Hoeting, 2005, p. 257), where $(\boldsymbol{x}_i, y_i)$ are resampled, and utilizing 640 bootstrap replications. We note that the corresponding parameter estimates are similar and the signs of the parameter estimates are sensible. Both locations show that air temperature and downward shortwave radiation flux have positive relationship with extreme ozone while wind speed and precipitation have negative relationships. Standard errors show there is large uncertainty associated with the parameter estimates, but our CV-based model selection procedure should protect us from identifying irrelevant covariates.

We also note that our simple (non-block) bootstrap could underestimate variability due to temporal dependence. However, we believe that the magnitude of the underestimation is likely small since we assume that $(Y_{t_0}|\boldsymbol{X}_{t_0})$ is independent of $Y_t, \boldsymbol{X}_t$ for $t < t_0$, the $\boldsymbol{\beta}$s estimate the conditional relationship of $Y_t|\boldsymbol{X}_t$, and the temporal dependence is short-lived. It is clear that since we are able to include only four variables at a time, we have limited ability to explore the model space. However, results from this first step suggest a method for further exploration of the model space.

Table 4.1: Covariates for the top five models containing four covariates for Atlanta and Charlotte are given along with their respective CV scores. Parameter estimates with bootstrap standard errors are reported for the top model at each location. We use the nonparametric paired bootstrap (Givens and Hoeting, 2005, p. 257), where $(\boldsymbol{x}_i, y_i)$ are resampled, and utilize 640 bootstrap replications.

Model	CV				
Atlanta 1	.5398	temp	wnd spd	dswrf	precip
		.45 (.25)	-.44 (.36)	.53 (.64)	-5.82 (.20)
Atlanta 2	.5508	temp	wnd spd	rel hum	precip
		.59	-.46	-.28	-5.46
Atlanta 3	.5512	temp	wnd spd	dswrf	cape
		.80	-.39	.34	-.35
Atlanta 4	.5513	temp	wnd spd	hpbl 7am	cape
		.87	-.36	-.24	-.28
Atlanta 5	.5519	temp	hpbl 7am	rel hum	precip
		.53	-.39	-.48	-5.10
Charlotte 1	.5452	temp	wnd spd	dswrf	precip
		.44 (.33)	-.50 (.42)	.53 (.61)	-5.86 (.43)
Charlotte 2	.5770	temp	wnd spd	NW wind	dswrf
		.54	-.50	.12	.38
Charlotte 3	.5772	temp	wnd spd	dswrf	rel hum
		.52	-.44	.34	-.19
Charlotte 4	.5774	temp	wnd spd	E wind	dswrf
		.54	-.46	-.10	.45
Charlotte 5	.5781	temp	wnd spd	dswrf	tcdc
		.55	-.51	.46	.15

4.2.2 'Core Plus Four' Model Search

The main result from the first model search stage is that necessary conditions for extreme ozone are high temperature, high sunshine, low wind speed, and lack of precipitation. These conditions are largely explained by the four 'core' variables which appear in the best fitting models for Atlanta and Charlotte: temperature, wind speed, dswrf, and the precipitation indicator. We continue our model search by fitting all possible models which include these four variables, plus up to four additional main effects. A total of 534 models were fit at this stage.

Results from the first stage also suggested slightly altering our list of covariates. One change that was made was to include only the minimum and maximum hpbl values rather than the 8 values recorded throughout each day. We also introduce a new cloud covariate that is a linear combination of cloud variables from the NARR. Looking at a longer list of good-fitting four variable models showed that many of these models would swap out dswrf and replace it with one of several cloud variables. The various cloud variables and dswrf were mutually strongly correlated. As we will include dswrf as one of our core variables, we seek to define a variable which captures information in the cloud variables and which is residual to the information in dswrf. We define the new variable as a linear combination of five of the cloud variables in the NARR: $\boldsymbol{x} = [x_{\text{cdcon}}, x_{\text{cdlyr}}, x_{\text{lcdc}}, x_{\text{mcdc}}, x_{\text{hcdc}}]^T$. Specifically, we find the parameter vector of unit length, $\boldsymbol{a}$ such that $Var(\boldsymbol{a}^T\boldsymbol{x}) = \boldsymbol{a}^T\Sigma\boldsymbol{a}$ is maximized and $Cov(x_{\text{dswrf}}, \boldsymbol{a}^T\boldsymbol{x}) = 0$. We estimate $\boldsymbol{a}$ via constrained optimization at several locations in the East and Southeast United States (including Atlanta and Charlotte) and find $\boldsymbol{a}$ to be similar at all locations. Thus, we define the new cloud variable:

$$x_{\text{new.cloud}} = .47x_{\text{cdcon}} - .45x_{\text{cdlyr}} - .37x_{\text{lcdc}} + .46x_{\text{mcdc}} + .48x_{\text{hcdc}}. \tag{4.3}$$

Table 4.2 compares the top five 'core-plus-four' models at Atlanta and Charlotte to the core-only model. In Atlanta, we see a convincing drop in the CV scores of the best fitting core-plus-four models compared to the core-only model. The top models tend to include minimum planetary boundary layer height (in 5 of the top 5 and 9 of the top 10), relative humidity (5/5 and 9/10), tropospheric height (3/5 and 5/10) and NE wind direction (3/5 and 4/10). The negative coefficients indicate that lower levels of planetary boundary layer height and relative humidity tend to be associated with extreme ozone, and the negative coefficient of the NE wind direction would indicate that extreme ozone tends to occur in Atlanta when wind is from the southwest. In Charlotte, we see a less convincing drop in the CV scores when we compare the best fitting core-plus-four models to the core only model. However, there are some variables which are associated with most of the best fitting models. The top models in Charlotte tend to include the new cloud variable (in 3 of the top 5 and 7 of the top 10), tropospheric height (3/5 and 6/10), and E wind direction (3/5 and 6/10).

The differences in the two cities' variables may illustrate that different factors lead to extreme ozone in the two cities, and the difference in the predominant wind direction may illustrate local differences in emissions sources. That the tropospheric height variable has a negative coefficient in Charlotte and a positive coefficient in Atlanta illustrates some of the difficulty in interpreting the parameter estimates. The difference in sign between the two cities may be due to the fact that in Atlanta, tropospheric height appears in models which also include planetary boundary layer height, whereas this was not the case in Charlotte.

Table 4.2: CV scores for the core only model and the top five 'core-plus-four' models for both Atlanta and Charlotte. We also include the parameter estimates for the four non-core covariates.

Rank	CV	Covariates				
Atl Core	0.5398	Core				
Atl 1	0.5046	Core	hpbl min	rel hum	pres	ht tropo
			-0.34	-0.19	-0.12	0.22
Atl 2	0.5075	Core	hpbl min	rel hum	NE wnd	ht tropo
			-0.30	-0.17	-0.16	0.15
Atl 3	0.5083	Core	hpbl max	hpbl min	rel hum	ht tropo
			-0.10	-0.32	-0.35	0.16
Atl 4	0.5090	Core	hpbl min	rel hum	NE wnd	pres
			-0.39	-0.10	-0.19	-0.06
Atl 5	0.5097	Core	hpbl min	rel hum	NE wnd	lwrf
			-0.36	-0.17	-0.17	0.08
Char Core	0.5452	Core				
Char 1	0.5412	Core	E wnd	pres	lwrf	ht tropo
			-0.11	0.08	0.13	-0.12
Char 2	0.5415	Core	cloud	E wnd	pres chng	pres
			-0.11	-0.09	-0.04	0.14
Char 3	0.5415	Core	cloud	E wnd	ht tropo	
			-0.14	-0.10	-0.11	
Char 4	0.5420	Core	hpbl max	NW wnd	lwrf	ht tropo
			-0.10	0.06	0.04	-0.05
Char 5	0.5421	Core	cloud	rel hum	N wnd	pres chng
			-0.06	-0.21	0.04	-0.07

Table 4.3 gives bootstrap standard errors, using the nonparametric paired bootstrap (Givens and Hoeting, 2005, p. 257) where $(\boldsymbol{x}_i, y_i)$ are resampled, and utilizing 640 bootstrap replications, for the best fitting models in Atlanta and Charlotte. Because we are using a

small subset of extreme data, and because these models include a large number of covariates which are likely dependent, it is not surprising that the standard errors are quite large. Because our aim is to uncover possible covariates for further exploration rather than to give a definitive model, we are not overly concerned with the large standard errors.

Table 4.3: Parameter estimates of the best ‘core plus four’ models for Atlanta and Charlotte with bootstrap standard errors in parentheses. We use the nonparametric paired bootstrap (Givens and Hoeting, 2005, p. 257), where $(\boldsymbol{x}_i, y_i)$ are resampled, and utilizing 640 bootstrap replications.

Atlanta	temp	wnd spd	dswrf	precip
	.40 (.27)	-.31 (.29)	.29 (.50)	-2.12 (.96)
	hpbl min	rel hum	pres	ht ropo
	-.34 (.47)	-.19 (.28)	-.12 (.26)	.22 (.28)
Charlotte	temp	wnd spd	dswrf	precip
	.46 (.31)	-.41 (.34)	.55 (.48)	-4.35 (.89)
	E wnd	pres	lwrf	ht tropo
	-.11 (.39)	.08 (.17)	.13 (.24)	-.12 (.27)

4.2.3 Automated Model Search Procedure

Our model search procedure thus far has been limited to at most eight main effects. We would like to further explore the model space to investigate whether interactions or a larger number of covariates would show even stronger tail dependence. Because a systematic model search becomes infeasible, we perform an automated model search.

We implement a slightly modified version of the simulated annealing procedure utilized in the R optim function (using the SANN method). Simulated annealing is a well known stochastic global optimization method. Kirkpatrick et al. (1983) give an interesting introduction to simulated annealing and Goffe et al. (1994) discuss simulated annealing in the context of parameter estimation for statistical models.

The simulated annealing procedure we employ works in the following manner. We begin with an initial value, $\boldsymbol{\omega}_0 \in \{0,1\}^{k'}$ and rely on a function $f : \{0,1\}^{k'} \to \{0,1\}^{k'}$ to choose a new string. We can construct f to exclude certain undesired models, such as models with highly correlated covariates. At the i^{th} step we calculate $CV(f(\boldsymbol{\omega}_{i-1}))$. If $CV(\boldsymbol{\omega}_{i-1}) > CV(f(\boldsymbol{\omega}_{i-1}))$ then we define $\boldsymbol{\omega}_i := f(\boldsymbol{\omega}_{i-1})$ and proceed to the next iteration. If $CV(\boldsymbol{\omega}_{i-1}) \leq CV(f(\boldsymbol{\omega}_{i-1}))$ then

$$\boldsymbol{\omega}_i := \begin{cases} f(\boldsymbol{\omega}_{i-1}) & \text{with probability} \quad \exp\{-\Delta CV/\text{Temp}_i\} \\ \boldsymbol{\omega}_{i-1} & \text{with probability} \quad 1 - \exp\{-\Delta CV/\text{Temp}_i\} \end{cases}$$

where Temp_i is the current global temperature in the simulated annealing process and $\Delta CV = CV(f(\boldsymbol{\omega}_{i-1})) - CV(\boldsymbol{\omega}_{i-1})$. The global temperature is a parameter that is lowered throughout the optimization according to a cooling schedule. As in the R function optim using the SANN method, we use the logarithmic cooling schedule outlined in Bélisle (1992). When the global temperature is high the process is more likely to move to $\boldsymbol{\omega}$s with higher CV values, reducing the chances of finding a local optimum. When the global temperature is low, the process is unlikely to move to $\boldsymbol{\omega}$s with higher CV values.

Possible covariates include all the main effects considered in the previous 'core plus four' exploration, 77 interactions between continuous covariates, and 15 interactions between continuous covariates and the precipitation indicator. We include the four core variables in all considered models, as this reduces the search to a region of the model space where extremes are known to occur. Starting values are chosen by using the best core plus four models at each location. We perform 640 runs at each location.

The covariates in the top five models at each location are given in Table 4.4. In Atlanta, and to a greater extent in Charlotte, we see a convincing drop between the CV scores of the best fitting models found during the model search and the best core-plus-four model from the previous section. In both cities, we see relative humidity appears in many of the top models, although in Charlotte it tends to appear in an interaction. We further notice that many of the interactions in these top models include include a core variable such as wind speed or downward short wave radiative flux, and a planetary boundary layer height or pressure variable. Interestingly, few of these interactions include air temperature. We also notice that just one of these top models contains an interaction with the precipitation indicator, which may suggest that the presence of precipitation, regardless of other variables, is enough to discourage the most extreme ozone events.

Table 4.4: The covariates and interactions in the best five models in the automated model search applied to the Atlanta (top) and Charlotte data (bottom). Interactions between continuous covariates are indicated with a '×'. The four core main effects were also included in all models, but do not appear in the covariate list.

Rank	CV			
Atl C+4	0.5046	best core-plus-four model		
Atl 1	0.4812	rel hum	wnd spd×pres chg	wnd spd×ht tropo
		dswrf×NE wnd	dswrf×pres	dswrf×hpbl.max
Atl 2	0.4823	hpbl min	rel hum	ht tropo
		dswrf×NE wnd	wnd spd×hpbl min	N wnd ×precip
Atl 3	0.4836	pres	wnd spd×NE wnd	wnd spd×pres
		wnd spd×ht tropo	dswrf×pres	rel hum×NW wnd
Atl 4	0.4837	wnd spd×ht tropo	dswrf×ht tropo	hpbl min×NE wnd
		wnd spd×hpbl min	dswrf×hpbl min	hpbl min×rel hum
Atl 5	0.4868	rel hum	wnd spd×pres chg	wnd spd×ht tropo
		new.cloud×pres	temp×hpbl max	dswrf×hpbl max
Char C+4	0.5412	best core-plus-four model		
Char 1	0.5085	hpbl.max	pres chg	rel hum×lwrf
		wnd spd×hpbl min	dswrf×hpbl.max	hpbl.max×rel hum
Char 2	0.5172	hpbl min	rel hum	hpbl.max×NW wnd
		wnd spd×hpbl min	dswrf×hpbl.max	dswrf×hpbl min
Char 3	0.5175	dswrf×NE wnd	rel hum×lwrf	temp×hpbl min
		wnd spd×hpbl min	dswrf×hpbl min	dswrf×rel hum
Char 4	0.5177	dswrf×pres	hpbl.max×pres	hpbl.max×ht tropo
		rel hum×lwrf	wnd spd×hpbl min	dswrf×new.cloud
Char 5	0.5181	dswrf×NE wnd	dswrf×pres	hpbl min×pres
		hpbl min×ht tropo	new.cloud×E wnd	wnd spd×hpbl min

CHAPTER 5

MODELING THE SPATIAL BEHAVIOR OF THE METEOROLOGICAL DRIVERS OF EXTREME OZONE

5.1 Introduction

In Chapter 4 we learned a great deal about the primary meteorological drivers of extreme ground level ozone in two locations: Atlanta, Georgia and Charlotte, North Carolina. We were able to identify a core set of four covariates that begin to explain extreme ozone. We refer to this set of variables (air temperature, wind speed, downward shortwave radiative flux, and the presence/absence of precipitation) as the 'core four' covariates. We also learned that other variables, such as the minimum height of the planetary boundary layer and relative humidity, play a role in extreme ozone at these two locations. In this chapter, we have a different goal. Here we wish to understand how the primary drivers of extreme ground level ozone change over a large study area.

We take the following approach. We begin by performing a model selection exercise at a number of locations throughout the study area. Although extensive, this model selection exercise will not be as large in scope as the data mining procedure done for Atlanta and Charlotte in Chapter 4. As we build on our knowledge from Chapter 4, our aim is to use the results of this exercise to construct a 'common' model for our study area. This common model will contain covariates that include the primary drivers of ozone throughout the region. After determining this common model, we will spatially model the parameter estimates for these covariates.

The primary products of the spatial analysis will be maps of the study region which relate the driving meteorological variables to extreme ozone. For each parameter estimate, we will produce two maps: one that models the point estimate and one that gives the amount of variability. Additionally, we aim to reduce the amount of uncertainty associated with the parameter estimates. For example, exploratory analysis suggests that air temperature should be at least somewhat important in terms of describing extreme ground level ozone. The parameter estimates for air temperature in the models that contain only the four core covariates in Atlanta and Charlotte are .45 and .44 respectively. However, the standard errors are .25 and .33 respectively. These standard errors make it difficult to gain a clear understanding of the primary meteorological drivers of extreme ground level ozone. Although estimators with a high degree of variability are not necessarily unusual in extremes analyses, our spatial method is able to borrow strength across multiple stations which will hopefully reduce variability and increase understanding of the primary drivers of extreme ozone.

Our model is hierarchical; however as our inference method in Chapter 3 is not likelihood-based, we do not employ Bayesian inference. Instead, we use a two-step inference approach which accounts for uncertainty in the initial estimates. Our approach models the parameters in the common model as a function of geographical covariates plus spatially correlated Gaussian random effects. To build our multivariate spatial model, we use the method of coregionalization.

As our method aims to explain how the drivers of extreme behavior vary over the region, it differs fundamentally from other spatial extremes studies. Previous spatial extremes studies can roughly be divided into two categories. The first type aims to understand how the marginal extreme behavior varies over a region. A popular approach for this type of study is to build a hierarchical model in which the parameters of the GEV or GPD vary spatially (Cooley et al. (2007), Sang and Gelfand (2009), Dyrrdal et al. (2015)). This approach is useful when one is interested in quantities that can be expressed as a function of these GEV or GPD parameters, such as return levels. A second type of spatial extremes study seeks to model the spatial extent of extreme events, such as storms. The leading approach is to fit a max-stable process to block maximum data (Kabluchko et al., 2009) or a Pareto process to threshold exceedances (Ferreira et al., 2014). When observed at a finite number of locations, the dependence structure of either a max-stable process with unit Fréchet marginals or a Pareto process with regular-varying $\alpha = 1$ tails can be described with the angular measure we introduced in Section 1.3.3. Although our approach is hierarchical, and since we aim to spatially model the meteorological drivers of extreme ozone rather than the behavior of ozone itself, it is different in both aim and procedure from these other methods.

In this chapter, we begin by defining the study region we will use and the data that we use for estimation. We also describe the process that we use to select the common model. Next, we formalize the spatial hierarchical model that we use for analysis. The following section describes the method we use for estimation and gives results of inference on our model. We conclude by using our parameter estimates to interpolate over the study area.

5.2 Study Area and Common Model

Before describing the details of our spatial model, we need to formalize our study area and decide which meteorological variables need to be included in the common model. We begin by describing the study area.

5.2.1 Study Area and Data

Since our project involves air quality in general and ground level ozone in particular, we feel that it is natural to consider regions designated by the EPA. The EPA has 10 regional offices, each serving one or more US states and/or territories. In addition to Georgia and North Carolina, EPA Region 4 includes Florida, Alabama, Mississippi, Tennessee, Kentucky, and South Carolina. This makes it a natural choice to be included in our study area. In order to make the study area slightly larger and more geographically diverse, we also include EPA Region 3. EPA Region 3 includes Virginia, West Virginia, Maryland, Pennsylvania, Delaware, and the District of Columbia.

We use data from 160 EPA air monitoring stations spread throughout the study area described above. Figure 5.1 shows the locations of these stations which are selected for their spatial coverage and temporal records. Ozone data is obtained via the EPA website (http://www.epa.gov/airdata/ad_maps.html). Because different stations will have different lengths of data records, we alter the method we use to remove the downward trend in ozone. Rather than transforming the data in non-overlapping blocks as described in Chapter 4, we transform in the same way but utilize five-year moving windows.

As in Chapter 4, the meteorological covariates we use are downloaded from the NARR[4] (Mesinger et al., 2006). We expect that the primary meteorological drivers of extreme ozone will vary over the study region. To make comparisons, we wish to spatially model the parameter estimates from the modeling procedure in Chapter 3. It is important to recall that these types of comparisons only make sense when exactly the same model is fit. Thus, we need to select a single common model that includes all of the primary drivers throughout the study region.

5.2.2 Detemining the Common Model

In order to make sure that we include the primary drivers in our common model, we first utilize data from 18 stations spread throughout the study area. The analysis of Atlanta and Charlotte indicates that the four core covariates (air temperature, downward shortwave radiative flux, wind speed, and precipitation as an indicator variable) should be included in any model. To decide which other covariates need to be included, we fit all possible 'core plus four' models at each of the 18 locations. Recall from Chapter 4 that core plus four models always include the four core covariates, and also include up to four additional covariates. We then consider the top 50 core plus four models at each location and compare the variables in the best models. Figure 5.2 gives bar charts showing the number of times that variables added to the core four are included in the best 50 core plus four models for four locations.

The top left panel of Figure 5.2 shows that relative humidity is present in nearly all of the top 50 models in Washington, D.C. Exploratory analysis finds that relative humidity appears

[4] NARR data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/

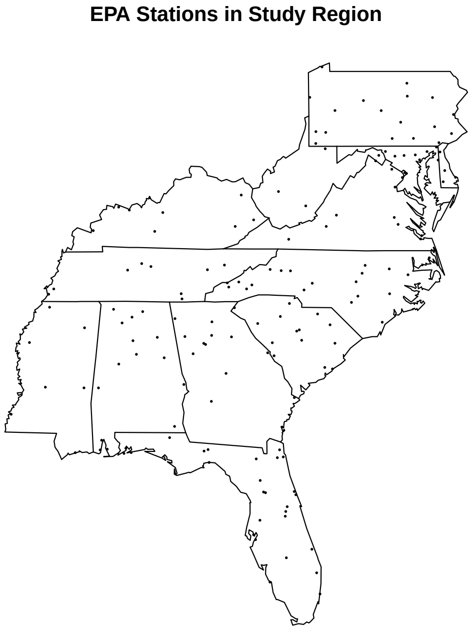

Figure 5.1: Our study area consists of EPA Regions 3 and 4. Here, we plot the locations of the 160 stations we use in our spatial analysis.

in many of the top models in many of the locations in the northern portion of the study area. We find that the minimum height of the planetary boundary layer is an important covariate in large portions of the study area, but especially in the Western portion. The top right and bottom left panels of Figure 5.2 show that the minimum height of the planetary boundary layer is present in most of the best 50 models in Memphis, Tennessee and Knoxville, Tennessee.

The bottom right panel of Figure 5.2 gives the covariates in the top 50 core plus four models in Jacksonville, Florida. This figure reveals that our cloud variable (described in Section 4.2.2) and downward long wave radiative flux are most important. We also notice this type of pattern in other Southeastern locations such as Daytona, Florida, Tallahassee, Florida, and Jackson, Mississippi. Wind direction variables show up in a few of the top models in some locations, exploratory analysis suggests that wind direction effects are quite local, and may be related to the direction to emissions sources. We choose not to include wind direction in our spatial analysis.

We wish to limit the number of covariates in our common model to a manageable number. Including too many variables in the common model could make convergence increasingly difficult. The exploratory analysis described here suggests that we can meet our objective by including three additional covariates in our common model: the daily minimum height of the planetary boundary layer, relative humidity, and our cloud variable described in Equation (4.3). We do not include downward longwave radiative flux in our common model, as find it to be correlated with the other two cloud/sunlight variables, and find adding it to a model with downward shortwave radiative flux and the cloud variable adds little. We believe that this common model includes the primary drivers throughout the region and we aim to understand how these drivers vary spatially.

We next fit this seven covariate model to each of the 160 locations in our study area. Figures 5.3 and 5.4 show the parameter estimates at each location. Some interesting patterns seem to be evident in the plots. As evidenced by the magnitude of the parameter estimates, air temperature seems to be more important in the Carolinas and the Mid-Atlantic regions. Air temperature looks to be less important in Southern Georgia and Northern Florida. Although its parameter estimate is not significantly different from 0, we note that one outlier in Southern Georgia has a negative parameter estimate for air temperature.

Wind speed looks to be most important in North Georgia, North Alabama, and Western North Carolina. Unexpectedly, wind speed appears to be much less important in the far Western portion of our study area and in Virginia and West Virginia. The minimum value of the height of the planetary boundary layer looks to be an important covariate, but moreso in the Western portion of the study area. Relative humidity seems to be an important covariate throughout the region, but the sign of its parameter estimate seems to change in different portions of the study area. Higher values of relative humidity are associated with extreme ozone in North Georgia, the Carolinas, and Pennsylvania whereas lower values of

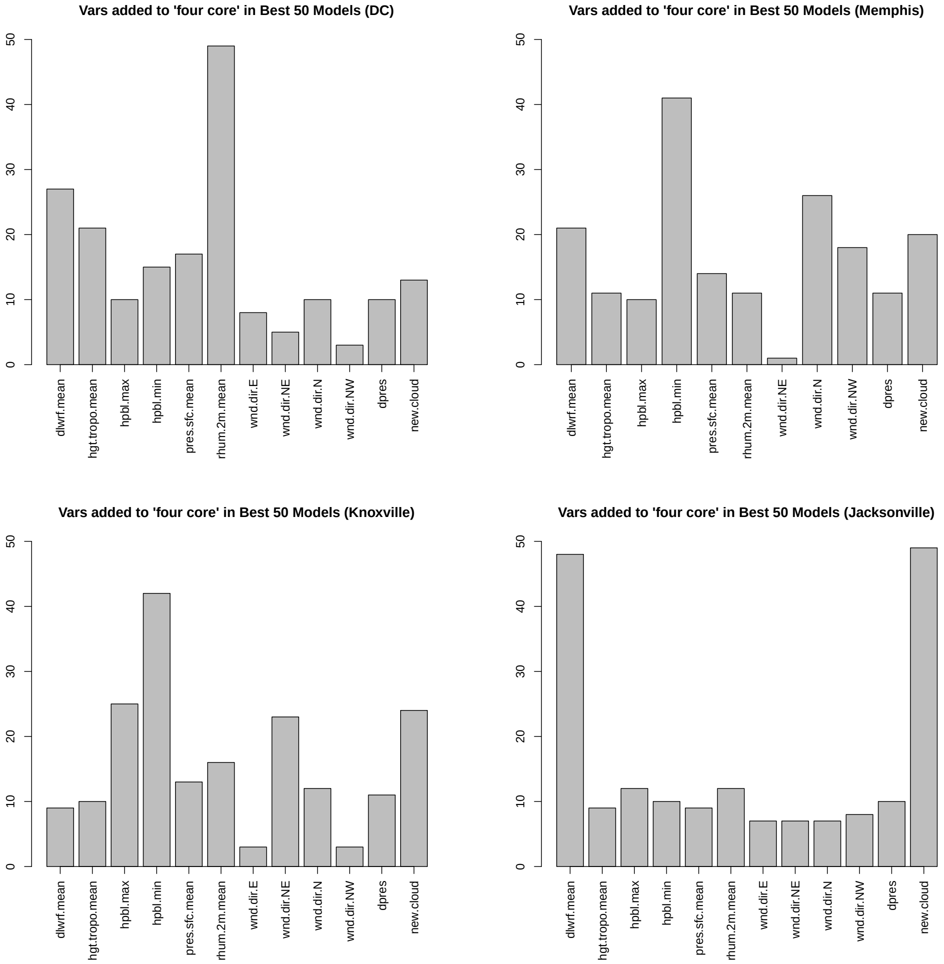

Figure 5.2: We fit all possible core plus four models at 18 locations in the study area. Here, we give bar charts showing the number of times each covariate is added to the core four is included in the best 50 core plus four models for Washington, D.C., Memphis, TN, Knoxville, TN, and Jacksonville, FL.

relative humidity are associated with extreme ozone in the rest of the study area. We also see this type of interesting behavior in the cloud variable. However, there is large uncertainty associated with these point estimates and the aforementioned spatial patterns are noisy.

The parameter estimate for the precipitation indicator seems to be nearly the same throughout the study region, indicating that precipitation plays a similar role everywhere in the region. Recall that including the precipitation indicator in the model (as described in Chapter 4) allows for the linear combination of transformed covariates to have a distribution described by a mixture of Gaussians. If precipitation is important, the parameter estimate for the precipitation indicator variable will be large enough for the two Gaussians overlap minimally. Practically speaking, this means that the most extreme ground level ozone events tend to occur on days without precipitation. As the precipitation indicator's parameter estimate varies little spatially, we exclude it from subsequent analysis.

5.3 Statistical Model

For each of the remaining six parameters in the common model and location $\boldsymbol{s}$ in the study area $\mathcal{D} \subset \mathbb{R}^2$, we assume that

$$\begin{aligned} \beta_1(\boldsymbol{s}) &= \boldsymbol{X}_1^T(\boldsymbol{s})\boldsymbol{\alpha}_1 + \eta_1(\boldsymbol{s}), \\ \beta_2(\boldsymbol{s}) &= \boldsymbol{X}_2^T(\boldsymbol{s})\boldsymbol{\alpha}_2 + \eta_2(\boldsymbol{s}), \\ &\vdots \\ \beta_6(\boldsymbol{s}) &= \boldsymbol{X}_6^T(\boldsymbol{s})\boldsymbol{\alpha}_6 + \eta_6(\boldsymbol{s}), \end{aligned} \tag{5.1}$$

where $\boldsymbol{X}_i^T(\boldsymbol{s})$ $(i = 1, \ldots, 6)$ are vectors of covariates. We use the coregionalization model for the random spatial effects (Wackernagel, 2003),

$$\begin{aligned}\begin{pmatrix} \eta_1(\boldsymbol{s}) \\ \eta_2(\boldsymbol{s}) \\ \vdots \\ \eta_6(\boldsymbol{s}) \end{pmatrix} &= A \begin{pmatrix} \delta_1(\boldsymbol{s}) \\ \delta_2(\boldsymbol{s}) \\ \vdots \\ \delta_6(\boldsymbol{s}) \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & 0 & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} & 0 & 0 \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & 0 \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{pmatrix} \begin{pmatrix} \delta_1(\boldsymbol{s}) \\ \delta_2(\boldsymbol{s}) \\ \vdots \\ \delta_6(\boldsymbol{s}) \end{pmatrix},\end{aligned}$$

where $\delta_i(\boldsymbol{s})$ $(i = 1, \ldots, 6)$ are independent second-order stationary Gaussian processes with mean 0 and variance 1. We assume isotropy, and use the exponential covariance function $Cov(\delta_j(\boldsymbol{s}), \delta_j(\boldsymbol{s}')) = \exp(-||\boldsymbol{s} - \boldsymbol{s}'||/\rho_j)$ for $\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{D}$. The lower triangular formulation of A is suggested by Finley et al. (2008).

For fixed location $\boldsymbol{s} \in \mathcal{D}$ we calculate the covariance matrix of $(\beta_1(\boldsymbol{s}), \ldots, \beta_6(\boldsymbol{s}))^T$,

$$\begin{aligned}
Cov\begin{pmatrix} \beta_1(\boldsymbol{s}) \\ \vdots \\ \beta_6(\boldsymbol{s}) \end{pmatrix} &= Cov\begin{pmatrix} \eta_1(\boldsymbol{s}) \\ \vdots \\ \eta_6(\boldsymbol{s}) \end{pmatrix} \\
&= Cov\left(A \begin{pmatrix} \delta_1(\boldsymbol{s}) \\ \vdots \\ \delta_6(\boldsymbol{s}) \end{pmatrix} \right) \\
&= ACov\begin{pmatrix} \delta_1(\boldsymbol{s}) \\ \vdots \\ \delta_6(\boldsymbol{s}) \end{pmatrix} A^T \\
&= AIA^T \\
&= AA^T. \qquad (5.2)
\end{aligned}$$

For fixed $i, j \in \{1, \ldots, 6\}$ and $\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{D}$ the cross covariances of the random effects is

$$\begin{aligned}
Cov(\eta_i(\boldsymbol{s}), \eta_j(\boldsymbol{s}')) &= Cov\left(\sum_{k=1}^{i} a_{ik}\delta_k(\boldsymbol{s}), \sum_{l=1}^{j} a_{jl}\delta_l(\boldsymbol{s}') \right) \\
&= \sum_{k=1}^{i}\sum_{l=1}^{j} a_{ik}a_{jl}Cov(\delta_k(\boldsymbol{s}), \delta_l(\boldsymbol{s}')) \\
&= \sum_{k=1}^{\min\{i,j\}} a_{ik}a_{jk}Cov(\delta_k(\boldsymbol{s}), \delta_k(\boldsymbol{s}')) \\
&= \sum_{k=1}^{\min\{i,j\}} a_{ik}a_{jk} \exp\left(-||\boldsymbol{s} - \boldsymbol{s}'||/\rho_k\right). \qquad (5.3)
\end{aligned}$$

If $i = j$ and $\boldsymbol{s} = \boldsymbol{s}'$ Equation (5.3) reduces to $\sum_{k=1}^{i} a_{ik}^2$ and corresponds to the diagonal elements in AA^T from (5.2). When $i \neq j$ and $\boldsymbol{s} = \boldsymbol{s}'$ Equation (5.3) reduces to $\sum_{k=1}^{\min\{i,j\}} a_{ik}a_{jk}$ and corresponds to the off diagonal elements in AA^T from (5.2). For $i = j$ and $\boldsymbol{s} \neq \boldsymbol{s}'$ Equation (5.3) simplifies to $\sum_{k=1}^{i} a_{ik}^2 \exp\left(-||\boldsymbol{s} - \boldsymbol{s}'||/\rho_k\right)$.

5.4 Estimation and Inference

5.4.1 Two Stage Inference for Spatial Model Parameters

Of course, we do not observe $\beta_i(\boldsymbol{s})$ $(i = 1, \ldots, 6)$ in Equation (5.1). Instead, we have estimates of $\beta_i(\boldsymbol{s})$ $(i = 1, \ldots, 6)$ and we wish to account for the uncertainty in these estimates. At the first stage of inference we obtain parameter estimates, $\tilde{\beta}_i$ $(i = 1, \ldots, 6)$, at each of the 160 stations individually, using the optimization method described in Chapter 3. We assume

$$\begin{aligned}
\tilde{\beta}_1(\boldsymbol{s}_l) &= \beta_1(\boldsymbol{s}_l) + \epsilon_1(\boldsymbol{s}_l), \\
\tilde{\beta}_2(\boldsymbol{s}_l) &= \beta_2(\boldsymbol{s}_l) + \epsilon_2(\boldsymbol{s}_l), \\
&\vdots \\
\tilde{\beta}_6(\boldsymbol{s}_l) &= \beta_6(\boldsymbol{s}_l) + \epsilon_6(\boldsymbol{s}_l),
\end{aligned}$$

where $(\epsilon_1(\boldsymbol{s}_l), \epsilon_2(\boldsymbol{s}_l), \ldots, \epsilon_6(\boldsymbol{s}_l))^T$ represents estimation error. We assume that

$$(\epsilon_1(\boldsymbol{s}_l), \epsilon_2(\boldsymbol{s}_l), \ldots, \epsilon_6(\boldsymbol{s}_l))^T \sim N(\mathbf{0}, \Sigma(\boldsymbol{s}_l)),$$

and that $(\epsilon_1(\boldsymbol{s}_l), \epsilon_2(\boldsymbol{s}_l), \ldots, \epsilon_6(\boldsymbol{s}_l))^T$ is independent of $(\epsilon_1(\boldsymbol{s}_{l'}), \epsilon_2(\boldsymbol{s}_{l'}), \ldots, \epsilon_6(\boldsymbol{s}_{l'}))^T$ for all $\boldsymbol{s}_l \neq \boldsymbol{s}_{l'}$. We also assume that $(\epsilon_1(\boldsymbol{s}_l), \epsilon_2(\boldsymbol{s}_l), \ldots, \epsilon_6(\boldsymbol{s}_l))^T$ is independent of $\eta_i(\boldsymbol{s}_l)$. At location l we estimate $\Sigma(\boldsymbol{s}_l)$ via the nonparametric paired bootstrap (Givens and Hoeting, 2005, p. 257), where $(\boldsymbol{x}_i, y_i)$ are resampled and utilizing 640 bootstrap replications, obtaining $\tilde{\Sigma}(\boldsymbol{s}_l)$. Define

$$\tilde{\boldsymbol{\beta}}_i = (\tilde{\beta}_i(\boldsymbol{s}_1), \tilde{\beta}_i(\boldsymbol{s}_2), \ldots, \tilde{\beta}_i(\boldsymbol{s}_n))^T$$

for $i = 1, \ldots, 6$.

In stage two, we use $\tilde{\boldsymbol{\beta}}_i$ to estimate $\boldsymbol{\alpha}_i$, A, and ρ_i (for $i = 1, \ldots, 6$). Denote

$$\boldsymbol{\beta}_i = (\beta_i(\boldsymbol{s}_1), \beta_i(\boldsymbol{s}_2), \ldots, \beta_i(\boldsymbol{s}_n))^T$$

for $i = 1, \ldots, 6$. Similarly, let

$$X_i = (\boldsymbol{X}_i(\boldsymbol{s}_1)^T, \boldsymbol{X}_i(\boldsymbol{s}_2)^T, \ldots, \boldsymbol{X}_i(\boldsymbol{s}_n)^T)^T$$

be the matrix of covariates for the i^{th} parameter. We also let K be the covariance matrix of $(\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \ldots, \boldsymbol{\beta}_6^T)^T$. Finally, we define

$$X = \begin{pmatrix} X_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & X_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & X_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & X_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & X_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & X_6 \end{pmatrix} \text{ and } \boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \boldsymbol{\alpha}_3 \\ \boldsymbol{\alpha}_4 \\ \boldsymbol{\alpha}_5 \\ \boldsymbol{\alpha}_6 \end{pmatrix}.$$

This implies that

$$(\tilde{\boldsymbol{\beta}}_1^T, \tilde{\boldsymbol{\beta}}_2^T, \ldots, \tilde{\boldsymbol{\beta}}_6^T)^T \sim N(X\boldsymbol{\alpha}, K + \Sigma).$$

In order to estimate $\boldsymbol{\alpha}_i$ (for $i = 1, \ldots, 6$), A, and ρ_i (for $i = 1, \ldots, 6$), we use sequential maximum likelihood. At each iteration we first maximize $\boldsymbol{\alpha_i}$ (for $i = 1, \ldots, 6$), then A, then ρ_i (for $i = 1, \ldots, 6$).

5.4.2 Interpolation of $\beta_i(\boldsymbol{s})$

Consider $\boldsymbol{s}_0 \in \mathcal{D}$, that may or may not be an observed location. Our main interest is in estimating $\beta_1(\boldsymbol{s}_0), \beta_2(\boldsymbol{s}_0), \ldots,$ and $\beta_6(\boldsymbol{s}_0)$ along with the corresponding uncertainty estimates. These estimates are obtained via universal co-kriging based on the fitted process and $\tilde{\boldsymbol{\beta}}_1, \ldots, \tilde{\boldsymbol{\beta}}_6$. Let K_{12} be $Cov((\tilde{\boldsymbol{\beta}}_1^T, \ldots, \tilde{\boldsymbol{\beta}}_6^T)^T, (\beta_1(\boldsymbol{s}_0), \ldots, \beta_6(\boldsymbol{s}_0))^T)$. The estimate of $(\hat{\beta}_1(\boldsymbol{s}_0), \ldots, \hat{\beta}_6(\boldsymbol{s}_0))^T = \boldsymbol{w}^T(\tilde{\boldsymbol{\beta}}_1^T, \ldots, \tilde{\boldsymbol{\beta}}_6^T)^T$ where

$$\boldsymbol{w} = (K+\Sigma)^{-1}K_{12} + (K+\Sigma)^{-1}X(X^T(K+\Sigma)^{-1}X)^{-1}(X_0^T - X_0^T(K+\Sigma)^{-1}K_{12})$$

and where

$$X_0 = \begin{pmatrix} \boldsymbol{X}_1(\boldsymbol{s}_0)^T & 0 & 0 & 0 & 0 & 0 \\ 0 & \boldsymbol{X}_2(\boldsymbol{s}_0)^T & 0 & 0 & 0 & 0 \\ 0 & 0 & \boldsymbol{X}_3(\boldsymbol{s}_0)^T & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{X}_4(\boldsymbol{s}_0)^T & 0 & 0 \\ 0 & 0 & 0 & 0 & \boldsymbol{X}_5(\boldsymbol{s}_0)^T & 0 \\ 0 & 0 & 0 & 0 & 0 & \boldsymbol{X}_6(\boldsymbol{s}_0)^T \end{pmatrix}.$$

We obtain the universal kriging mean squared prediction error by

$$K_{22} - K_{12}^T(K+\Sigma)^{-1}K_{12} + (X_0^T - X^T(K+\Sigma)^{-1}K_{12})^T(X^T(K+\Sigma)^{-1}X)(X_0^T - X^T(K+\Sigma)^{-1}K_{12})$$

where $K_{22} = Cov((\beta_1(\boldsymbol{s}_0), \ldots, \beta_6(\boldsymbol{s}_0))^T, (\beta_1(\boldsymbol{s}_0), \ldots, \beta_6(\boldsymbol{s}_0))^T)$.

5.5 Analysis of EPA Regions 3 and 4

5.5.1 Interpreting Parameter Estimates

Using AIC as the model selection criterion, we choose to use the model where $X_i = \mathbf{1}$ for $i = 1, \ldots, 6$. Using sequential maximum likelihood, we estimate

$$\hat{\boldsymbol{\alpha}} = (0.46, 0.40, -0.39, -0.14, -0.13, -0.06)^T,$$

$$\hat{\boldsymbol{\rho}} = (129.54, 87.97, 113.50, 21.00, 9.45, 43.81)^T,$$

and

$$\hat{A} = \begin{pmatrix} 0.166 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.032 & 0.054 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.030 & 0.079 & 0.101 & 0.000 & 0.000 & 0.000 \\ 0.104 & 0.100 & -0.040 & 0.048 & 0.000 & 0.000 \\ 0.024 & -0.045 & 0.012 & -0.044 & 0.066 & 0.000 \\ 0.011 & -0.037 & -0.014 & 0.037 & 0.062 & 0.078 \end{pmatrix}.$$

Although these estimates do not have a great deal of meaning by themselves, they can be used to estimate various attributes of the spatial process we are modeling. For example, we can use the parameter estimates to estimate the covariance matrix of $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \ldots, \beta_6(\cdot))^T$

$$\widehat{Cov}(\boldsymbol{\beta}(\boldsymbol{s})) = \hat{A}\hat{A}^T = \begin{pmatrix} 0.028 & 0.005 & 0.005 & 0.017 & 0.004 & 0.002 \\ 0.005 & 0.004 & 0.005 & 0.009 & -0.002 & -0.002 \\ 0.005 & 0.005 & 0.017 & 0.007 & -0.002 & -0.004 \\ 0.017 & 0.009 & 0.007 & 0.025 & -0.005 & -0.000 \\ 0.004 & -0.002 & -0.002 & -0.005 & 0.009 & 0.004 \\ 0.002 & -0.002 & -0.004 & -0.000 & 0.004 & 0.013 \end{pmatrix}.$$

We can then use this estimated covariance matrix to calculate the estimated correlation matrix,

$$\widehat{Cor}(\boldsymbol{\beta}(\boldsymbol{s})) = \begin{pmatrix} 1.000 & 0.510 & 0.228 & 0.661 & 0.256 & 0.099 \\ 0.510 & 1.000 & 0.631 & 0.884 & -0.277 & -0.229 \\ 0.228 & 0.631 & 1.000 & 0.335 & -0.131 & -0.270 \\ 0.661 & 0.884 & 0.335 & 1.000 & -0.304 & -0.011 \\ 0.256 & -0.277 & -0.131 & -0.304 & 1.000 & 0.390 \\ 0.099 & -0.229 & -0.270 & -0.011 & 0.390 & 1.000 \end{pmatrix}. \tag{5.4}$$

This sample covariance matrix is interesting to interpret. It suggests that air temperature is positively correlated with each of the other six covariates, which seems to contradict our constraint on the parameter estimates. Closer inspection reveals that this does not violate the constraint on the parameter estimates. Since the constraint effectively deals with the magnitude of the parameter estimates, the sign of the estimate makes a large difference. The first two covariates (air temperature and downward shortwave radiative flux) tend to have positive parameter estimates while the last four covariates (minimum height of planetary boundary layer, relative humidity, wind speed, and the cloud variable) tend to have negative parameter estimates. As the air temperature parameter estimate increases, it is associated with an increase in the parameter estimate for wind speed (for example). Since wind speed tends to have a negative parameter estimate, an increase would likely result in a parameter estimate that is smaller in magnitude.

5.5.2 Sill and Effective Range

The parameter estimates for $\boldsymbol{\alpha}$, A, and $\boldsymbol{\rho}$ can also be used to estimate the sill and the effective range for our model. Let $d = ||\boldsymbol{s} - \boldsymbol{s}'||$ and define the function $C_i(d) := \sum_{k=1}^{i} a_{ik}^2 \exp\{-d/\rho_k\}$. The sill for β_i will be $C_i(0)$, which we estimate with $\hat{C}_i(d)$ where $\hat{C}_i(d) := \sum_{k=1}^{i} \hat{a}_{ik}^2 \exp\{-d/\hat{\rho}_k\}$. This yields the estimated sill, $\hat{C}_i(0) = \sum_{k=1}^{i} \hat{a}_{ik}^2$. Table 5.1 gives the estimated sills for $\beta_1, \ldots, \beta_6$ based on $\hat{A}$.

Table 5.1: We report the estimated sills for $\beta_1(\cdot), \ldots, \beta_6(\cdot)$ based on $\hat{A}$.

	Estimated Sill
$\beta_1(\cdot)$	0.028
$\beta_2(\cdot)$	0.004
$\beta_3(\cdot)$	0.017
$\beta_4(\cdot)$	0.025
$\beta_5(\cdot)$	0.009
$\beta_6(\cdot)$	0.013

We estimate the effective range for $\beta_i(\cdot)$ by finding smallest d such that $\hat{C}_i(d)/\hat{C}_i(0) \leq .05$. To the nearest tenth of one kilometer, the estimated effective ranges for $\beta_1(\cdot), \ldots, \beta_6(\cdot)$ are displayed in Table 5.2. It is interesting to note that the estimated effective ranges for air temperature, downward shortwave radiative flux, minimum height of the planetary boundary layer, and relative humidity seem to be similar. The estimated effective ranges for wind speed and the cloud variable also seem to be similar, but smaller than the first four.

Table 5.2: We report the estimated effective ranges for $\beta_1(\cdot), \ldots, \beta_6(\cdot)$ (to the nearest tenth of one kilometer).

	Effective Range
$\beta_1(\cdot)$	388.10 km
$\beta_2(\cdot)$	300.20 km
$\beta_3(\cdot)$	317.90 km
$\beta_4(\cdot)$	323.50 km
$\beta_5(\cdot)$	177.20 km
$\beta_6(\cdot)$	138.80 km

5.5.3 Reducing Uncertainty

A secondary goal for this spatial analysis is to reduce the uncertainty associated with individual station parameter estimates. In order to explore the degree to which the spatial model is able to accomplish this, we select four locations spread throughout the study region and report the parameter estimates and standard errors. Tables 5.3, 5.4, 5.5, and 5.6 give the parameter estimates and standard errors based on the extreme value model described in Chapter 3, as well as the parameter estimates and standard errors based on the spatial model outlined in this chapter. As anticipated, the standard errors are reduced significantly at each of these four locations. We also note that in this spatial model, Σ acts as a multivariate nugget effect. This means that at a fixed location $\boldsymbol{s'} \in \mathcal{D}$, $\hat{\beta}_i(\boldsymbol{s'}) \neq \tilde{\beta}_i(\boldsymbol{s'})$.

Table 5.3: We report the parameter estimates and standard errors based on the extreme value model from Chapter 3 and the parameter estimates and standard errors based on the spatial model from this chapter for EPA station number 13-135-0002 in Gwinnett County, GA.

	$\tilde{\beta}_{\text{EV}}$	$SE(\tilde{\beta}_{\text{EV}})$	$\hat{\beta}_{\text{spatial}}$	$SE(\hat{\beta}_{\text{spatial}})$
Air Temp	0.23	0.35	0.23	0.07
DSWRF	0.56	0.15	0.41	0.03
Min HPBL	-0.44	0.25	-0.40	0.07
Rel Hum	-0.24	0.17	-0.18	0.07
Wind Spd	-0.00	0.14	-0.18	0.07
Cloud	-0.59	0.38	-0.17	0.10

Table 5.4: We report the parameter estimates and standard errors based on the extreme value model from Chapter 3 and the parameter estimates and standard errors based on the spatial model from this chapter for EPA station number 21-043-0500 in Carter County, KY.

	$\tilde{\beta}_{\text{EV}}$	$SE(\tilde{\beta}_{\text{EV}})$	$\hat{\beta}_{\text{spatial}}$	$SE(\hat{\beta}_{\text{spatial}})$
Air Temp	0.20	0.39	0.32	0.09
DSWRF	0.15	0.20	0.35	0.04
Min HPBL	-0.61	0.20	-0.51	0.09
Rel Hum	-0.17	0.15	-0.22	0.07
Wind Spd	-0.40	0.18	-0.20	0.07
Cloud	-0.09	0.32	-0.08	0.10

5.5.4 Analysis

Utilizing our method for interpolation, we obtain estimates for each of the six parameters at each location on the NARR grid. Figure 5.5 plots these values over the entire study area. Figure 5.6 gives the square root of the mean squared prediction errors for each of the six variables in the model.

When looking at the maps in Figure 5.5, one of the first things that stands out is the difference between parameter estimates in the Carolinas and the Mid-Atlantic versus the rest of the study area. This is most obvious in the air temperature map (top left panel of Figure 5.5), but seems to be present to a lesser degree for the estimated parameter surfaces. The portions of the region where air temperature is most important roughly corresponds

Table 5.5: We report the parameter estimates and standard errors based on the extreme value model from Chapter 3 and the parameter estimates and standard errors based on the spatial model from this chapter for EPA station number 11-001-0041 in Washington DC.

	$\tilde{\beta}_{\text{EV}}$	$SE(\tilde{\beta}_{\text{EV}})$	$\hat{\beta}_{\text{spatial}}$	$SE(\hat{\beta}_{\text{spatial}})$
Air Temp	0.39	0.22	0.51	0.06
DSWRF	0.17	0.19	0.36	0.03
Min HPBL	-0.49	0.20	-0.41	0.07
Rel Hum	-0.45	0.22	-0.18	0.07
Wind Spd	-0.12	0.16	-0.08	0.06
Cloud	-0.00	0.22	0.05	0.08

Table 5.6: We report the parameter estimates and standard errors based on the extreme value model from Chapter 3 and the parameter estimates and standard errors based on the spatial model from this chapter for EPA station number 12-095-2002 in Orange County, FL.

	$\tilde{\beta}_{\text{EV}}$	$SE(\tilde{\beta}_{\text{EV}})$	$\hat{\beta}_{\text{spatial}}$	$SE(\hat{\beta}_{\text{spatial}})$
Air Temp	0.12	0.19	0.32	0.06
DSWRF	0.40	0.15	0.34	0.03
Min HPBL	-0.60	0.21	-0.68	0.06
Rel Hum	-0.13	0.17	-0.19	0.06
Wind Spd	-0.27	0.14	-0.18	0.07
Cloud	-0.23	0.35	-0.04	0.09

to the areas where emissions are highest. In general, higher local emissions should lead to higher temperature sensitivities. We also observe differences between Florida and other parts of the region. For example, the minimum height of the planetary boundary layer seems to play a large role in extreme ozone events in the central part of the state, even moreso than the rest of the region. Air temperature and downward shortwave radiative flux seem to play less of a role compared to other parts of the study area. As a peninsula, it is not unexpected that the primary drivers are different in Florida.

We also notice that the several of the estimated parameter surfaces have a ‘doughnut hole’ effect South of Charlotte, NC to Columbia, SC. We see this behavior in all of the surfaces except air temperature.

In Chapter 4 we felt strongly enough about the importance of wind speed that we included it in our core set of covariates. The bottom left panel of Figure 5.5 shows that wind speed is most important in central and North Georgia. Although wind speed seems to play a significant role throughout the remainder of the study area, its effect seems to be lessened somewhat.

The minimum height of the planetary boundary layer tends to have large negative parameter estimates throughout the study region, indicating its importance in extreme ozone. The reason for this is not entirely clear, but a low planetary boundary layer height may suggest the presence of an inversion, which could act to trap air pollutants near the ground. The results of this analysis suggest that the minimum height of the planetary boundary layer may be every bit as important to extreme ozone as the four core covariates.

Relative humidity and the cloud variable are interesting in the sense that the sign of their parameter estimates seems to change over different parts of the region. The cloud variable tends to be slightly positive in portions of Alabama, Florida, Maryland, and Pennsylvania as well as the doughnut hole in South Carolina. The reason for this is not clear, but the parameter estimates in these locations are likely not significantly different from 0 based on the estimates of variability in Figure 5.6. Relative humidity tends to have a positive parameter estimate in the Carolinas (except for the doughnut hole) and portions of West Virginia and Pennsylvania. Again, we do not know the reason for this behavior, but we wish to point out that the spatial surface appears remarkably similar to the estimated surface for downward shortwave radiative flux. This is reinforced by the fact that the corresponding estimated correlation in Equation 5.4 is larger than any other correlation estimate.

Not surprisingly, the square root of MSPE tends to be larger in places that are farther away from the 160 station locations. However, we again wish to point out that the standard errors are reduced significanly when compared to the standard errors from the extreme value model. Tables 5.3, 5.4, 5.5, and 5.6 show this for four locations.

5.6 Conclusion

The spatial analysis given here is very different from a typical spatial extremes analysis. Often, spatial extremes analyses have one of two main goals: estimating marginal effects via a hierarchical model, or understanding the spatial extent of extreme events through a max-stable process. Our goal is to understand how the primary drivers of extreme ozone vary spatially. We accomplish this goal by first fitting a common model extreme value model (relying on the model described in Chapter 3) at a large number of locations in our study area. We then utilize the vectors of parameter estimates and estimated covariance matrices at each location to fit a spatial surface for each parameter in the common model. The spatial surfaces are fit using a hierarchical model that uses the coregionalization method for random effects.

We have found out that the minimum height of the planetary boundary layer is even more important in extreme ozone than we initially believed. We also are able to determine that air temperature, a variable that is commonly understood to play a large role in ground level ozone, may be less important in certain parts of the region. We are also able to uncover additional questions to which there are not clear cut answers. For example, we do not completely understand the reason for the doughnut hole over South Carolina.

CHAPTER 6

CONCLUSIONS

This dissertation presented novel work in statistical methods for extremes. Although the application in this dissertation focused on ground level ozone, the methods introduced here could also be applied to any other field where understanding extremes is important.

In Chapter 2 we introduced γ, a tail dependence metric based on the angular measure H. We also introduced its corresponding estimator, which is well suited for use as an objective function in optimization. Additionally, we proposed the idea of tail dependence estimators that utilize a smooth threshold as opposed to the hard threshold typically used in extremes analyses. A smooth threshold is necessary to perform optimization, which has not previously been considered in extremes analyses. We also showed consistency of our estimator with a smooth threshold.

In Chapter 3 we outlined our procedure for finding the linear combination of covariates that optimizes tail dependence with a continuous response variable. We also proposed a model selection procedure that is based on cross-validation. We concluded the chapter by introducing a simulation study, where we demonstrate our method's ability to detect complicated conditions which lead to extreme behavior. We also utilized the simulation study to compare our approach to competing methods.

In Chapter 4, we proposed a data mining procedure that can be used to find the set of covariates the produces the linear combination with the highest degree of tail dependece with a response variable. We gave the results of an analysis of ground level ozone based on our data mining procedure using data from Atlanta, Georgia and Charlotte, North Carolina. We began by comparing cross validation scores for all possible four variable models, and found that the best model in both locations contained air temperature, downward shortwave radiative flux, wind speed, and precipitation. Next, we continued the search procedure by adding up to four more covariates to this core model. Since the model space is too large to be searched exhaustively, as a final step, we used an automated model search procedure based on simulated annealing. We also discussed how our method can be modified to include non-continuous covariates such as precipitation.

Instead of using our data mining procedure to find all of the drivers at fixed locations, in Chapter 5 we modeled the primary drivers of ozone spatially over EPA Regions 3 and 4. We began by performing exploratory analysis to find a common model that included the primary drivers for all locations in the study area. We then modeled the parameters of this common model spatially using a heirarchical modeling technique, and for inference, we utilized a two-step procedure. This spatial analysis yielded interesting results. For example, we found that air temperature was less important than expected in the Gulf Coast region and that the minimum height of the planetary boundary layer was more important than expected over the Western and Southern portions of the study region.

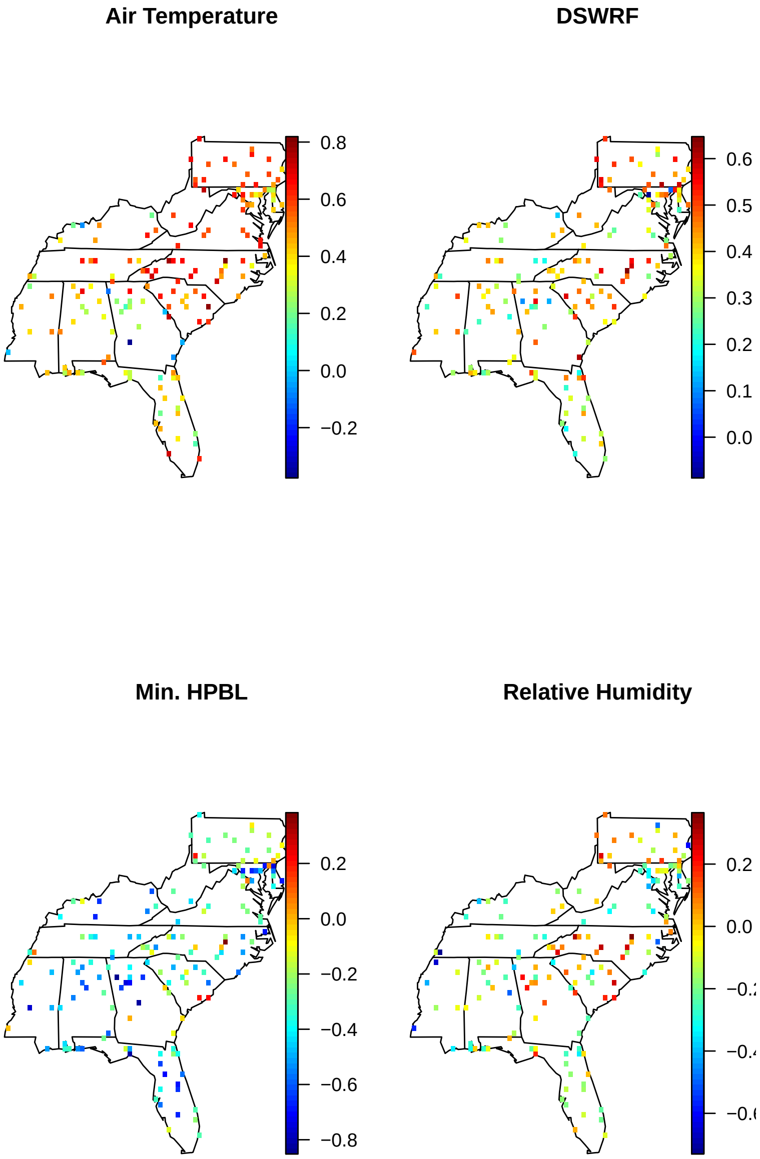

Figure 5.3: We plot the parameter estimates for four of the seven variables in the common model (air temperature, downward shortwave radiative flux, minimum height of the planetary boundardy layer, and relative humidity) at each of the 160 locations. The parameter estimates for the remaining three variables are given in Figure 5.4

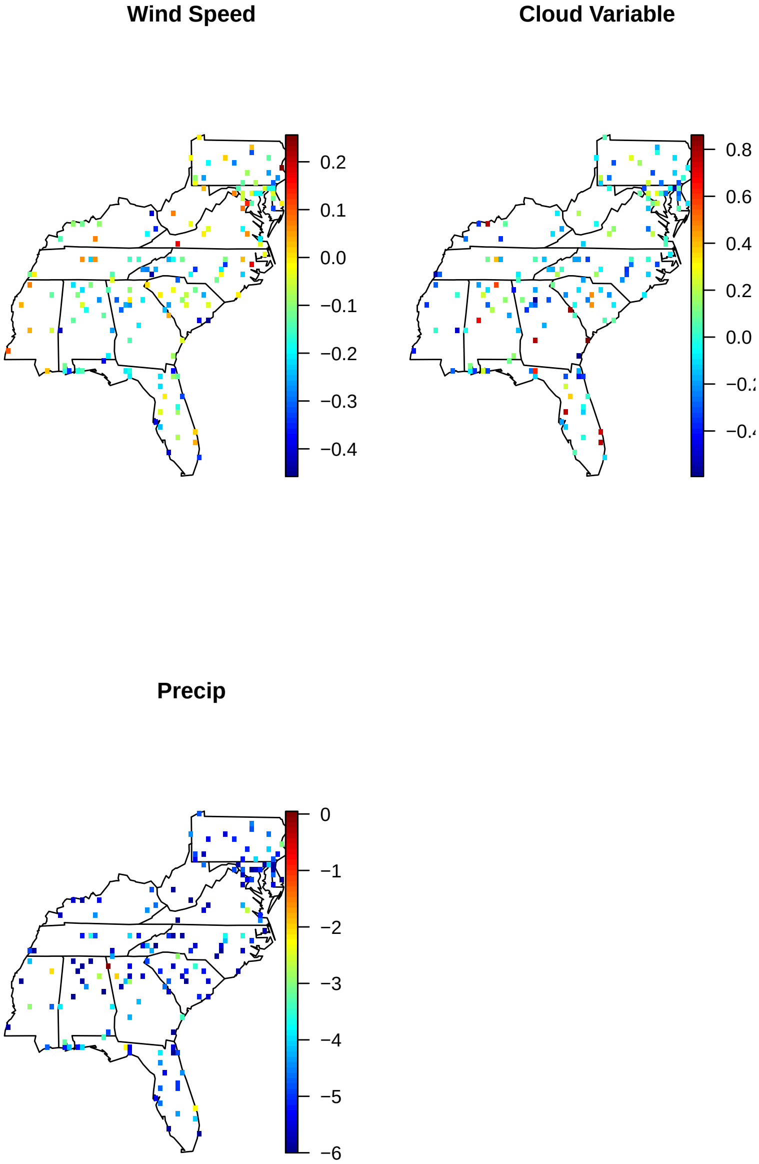

Figure 5.4: We plot the parameter estimates for three of the seven variables in the common model (wind speed, the cloud variable, and the precipitation indicator) at each of the 160 locations.

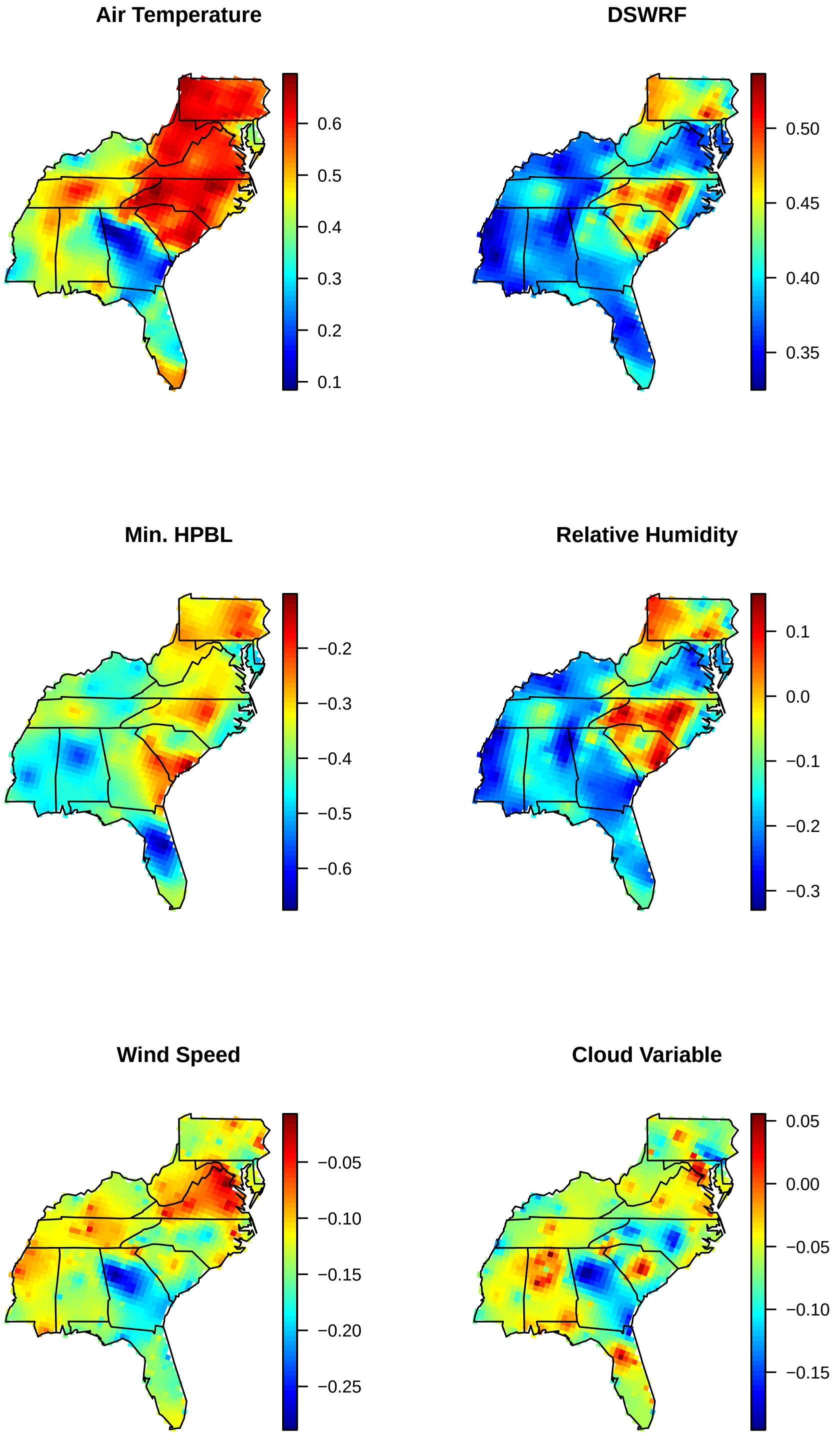

Figure 5.5: Using universal co-kriging, we plot the parameter estimates for each of the six variables in the spatial model. We note that each plot has its own scale.

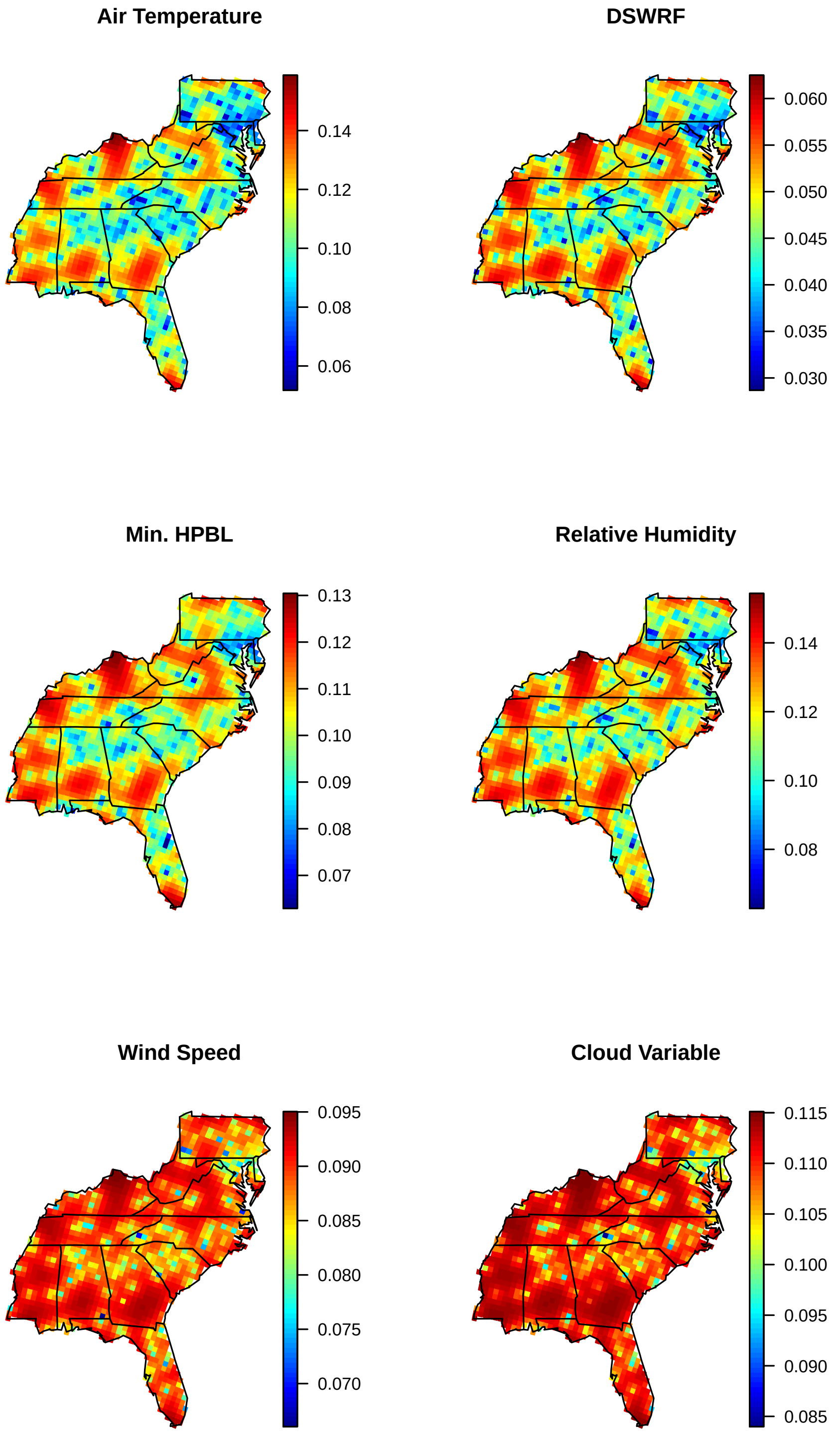

Figure 5.6: Using universal co-kriging, we plot the square root of the mean squared prediction errors for each of the six variables in the spatial model. We note that each plot has its own scale.

REFERENCES

Ali, N. B., bin Adam, M. B., Ibrahim, N. A. B., and Daud, I. B. (2012). Statistical analysis of extreme ozone data. *Journal of Statistical Modeling and Analytics*, 3(1):11–18.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., Waal, D. D., and Ferro, C. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, New York.

Bélisle, C. J. (1992). Convergence theorems for a class of simulated annealing algorithms on $\mathbb{R}^d$. *Journal of Applied Probability*, 29(4):885–895.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury Pacific Grove, CA, 2nd edition.

Chaudhuri, S. and Solar-Lezama, A. (2010). Smooth interpretation. *ACM Sigplan Notices*, 45(6):279–291.

Chaudhuri, S. and Solar-Lezama, A. (2011). Smoothing a program soundly and robustly. In *Computer Aided Verification*, pages 277–292. Springer.

Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analysis. *Extremes*, 2(4):339–365.

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer-Verlag London Ltd., London.

Computational and Information Systems Laboratory (2012). Yellowstone: IBM iDataPlex System (University Community Computing). Boulder, CO: National Center for Atmospheric Research. http://n2t.net/ark:/85065/d7wd3xhc.

Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for spatial max-stable random fields. In *Dependence in Probability and Statistics*, pages 373–390. Springer.

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840.

D'Auria, B. and Resnick, S. I. (2008). The influence of dependence on data network models. *Advances in Applied Probability*, 40(1):60–94.

Davis, R. and Mikosch, T. (2009). The extremogram: A correlogram for extreme events. *Bernoulli*, 15(4):977–1009.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3):393–442.

De Haan, L. and Ferreira, A. (2007). *Extreme Value Theory: An Introduction*. Springer Science & Business Media.

Durrett, R. (2010). *Probability: Theory and Examples*. Cambridge university press, 3rd edition.

Dyrrdal, A. V., Lenkoski, A., Thorarinsdottir, T. L., and Stordal, F. (2015). Bayesian hierarchical modeling of extreme hourly precipitation in Norway. *Environmetrics*, 26(2):89–106.

Eastoe, E. F. (2009). A hierarchical model for non-stationary multivariate extremes: a case study of surface-level ozone and $\mathrm{NO_x}$ data in the UK. *Environmetrics*, 20(4):428–444.

Eastoe, E. F. and Tawn, J. A. (2009). Modelling nonstationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C*, 58(1):25–45.

Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5):849–911.

Fan, J., Samworth, R., and Wu, Y. (2009). Ultrahigh dimensional feature selection: beyond the linear model. *The Journal of Machine Learning Research*, 10:2013–2038.

Ferreira, A., De Haan, L., et al. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20(4):1717–1737.

Finley, A. O., Banerjee, S., Ek, A. R., and McRoberts, R. E. (2008). Bayesian multivariate process modeling for prediction of forest attributes. *Journal of Agricultural, Biological, and Environmental Statistics*, 13(1):60–83.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190.

Givens, G. H. and Hoeting, J. A. (2005). *Computational Statistics*. John Wiley & Sons.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44(3):423–453.

Goffe, W. L., Ferrier, G. D., and Rogers, J. (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics*, 60(1):65–99.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society, Series B*, 66(3):497–546.

Hernandez-Campos, F., Jeffay, K., Park, C., Marron, J., and Resnick, S. I. (2005). Extremal dependence: Internet traffic applications. *Stochastic Models*, 21(1):1–35.

Hosking, J., Wallis, J. R., and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261.

Jacob, D. J. and Winner, D. A. (2009). Effect of climate change on air quality. *Atmospheric Environment*, 43(1):51–63.

Kabluchko, Z., Schlather, M., and De Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, 37(5):2042–2065.

Kirkpatrick, S., Jr., D. G., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598):671–680.

Larsson, M. and Resnick, S. I. (2012). Extremal dependence measure and extremogram: the regularly varying case. *Extremes*, 15(2):231–256.

Maraun, D., Osborn, T., and Rust, H. (2011). The influence of synoptic airflow on UK daily precipitation extremes. Part I: Observed spatio-temporal relationships. *Climate Dynamics*, 36(1):261–275.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jovic, D., Woollen, J., Rogers, E., Berbery, E. H., et al. (2006). North American Regional Reanalysis. *Bulletin of the American Meteorological Society*, 87(3):343–360.

Mullen, K., Ardia, D., Gil, D., Windover, D., and Cline, J. (2011). DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6):1–26.

Nelsen, R. (2006). *An Introduction to Copulas, 2nd Edition*. Lecture Notes in Statistics No. 139. Springer, New York.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Reich, B., Cooley, D., Foley, K., Napelenok, S., and Shaby, B. (2013). Extreme value analysis for evaluating ozone control strategies. *Annals of Applied Statistics*, 7(2):739–762.

Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag, New York.

Resnick, S. (2004). The extremal dependence measure and asymptotic independence. *Stochastic Models*, 20(2):205–227.

Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Series in Operations Research and Financial Engineering. Springer, New York.

Rootzen, H. and Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli*, 12(5):917–930.

Sang, H. and Gelfand, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, 16(3):407–426.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33–60.

Schlather, M. and Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90(1):139–156.

Sibuya, M. (1959). Bivariate extreme statistics, I. *Annals of the Institute of Statistical Mathematics*, 11(2):195–210.

Sillmann, J., Croci-Maspoli, M., Kallache, M., and Katz, R. W. (2011). Extreme cold winter temperatures in Europe under the influence of North Atlantic atmospheric blocking. *Journal of Climate*, 24(22):5899–5913.

Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science*, 4(4):367–393.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.

Tobias, A. and Scotto, M. G. (2005). Prediction of extreme ozone levels in Barcelona, Spain. *Environmental Monitoring and Assessment*, 100(1-3):23–32.

Varadhan, R. (2011). *alabama: Constrained nonlinear optimization.* R package version 2011.9-1.

Wackernagel, H. (2003). *Multivariate Geostatistics*. Springer Science & Business Media.

Yang Xiang, Gubian, S., Suomela, B., and Hoeng, J. (2013). Generalized simulated annealing for global optimization: the GenSA package. *The R Journal*, 5(1):13–29.