DISSERTATION

ADVANCES IN STATISTICAL ANALYSIS AND MODELING OF EXTREME VALUES MOTIVATED BY ATMOSPHERIC MODELS AND DATA PRODUCTS

Submitted by

Miranda J. Fix

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2018

Doctoral Committee:

Advisor: Daniel Cooley

Jennifer Hoeting
Ander Wilson
Elizabeth Barnes

ABSTRACT

ADVANCES IN STATISTICAL ANALYSIS AND MODELING OF EXTREME VALUES MOTIVATED BY ATMOSPHERIC MODELS AND DATA PRODUCTS

This dissertation presents applied and methodological advances in the statistical analysis and modeling of extreme values. We detail three studies motivated by the types of data found in the atmospheric sciences, such as deterministic model output and observational products. The first two investigations represent novel applications and extensions of extremes methodology to climate and atmospheric studies. The third investigation proposes a new model for areal extremes and develops methods for estimation and inference from the proposed model.

We first detail a study which leverages two initial condition ensembles of a global climate model to compare future precipitation extremes under two climate change scenarios. We fit non-stationary generalized extreme value (GEV) models to annual maximum daily precipitation output and compare impacts under the RCP8.5 and RCP4.5 scenarios. A methodological contribution of this work is to demonstrate the potential of a "pattern scaling" approach for extremes, in which we produce predictive GEV distributions of annual precipitation maxima under RCP4.5 given only global mean temperatures for this scenario. We compare results from this less computationally intensive method to those obtained from our GEV model fitted directly to the RCP4.5 output and find that pattern scaling produces reasonable projections.

The second study examines, for the first time, the capability of an atmospheric chemistry model to reproduce observed meteorological sensitivities of high and extreme surface ozone (O_3). This work develops a novel framework in which we make three types of comparisons between simulated and observational data, comparing (1) tails of the O_3 response variable, (2) distributions of meteorological predictor variables, and (3) sensitivities of high and extreme O_3 to meteorological predictors. This last comparison is made using quantile regression and a recent tail dependence optimization approach. Across all three study locations, we find substantial

differences between simulations and observational data in both meteorology and meteorological sensitivities of high and extreme O_3.

The final study is motivated by the prevalence of large gridded data products in the atmospheric sciences, and presents methodological advances in the (finite-dimensional) spatial setting. Existing models for spatial extremes, such as max-stable process models, tend to be geostatistical in nature as well as very computationally intensive. Instead, we propose a new model for extremes of areal data, with a common-scale extension, that is inspired by the simultaneous autoregressive (SAR) model in classical spatial statistics. The proposed model extends recent work on transformed-linear operations applied to regularly varying random vectors, and is unique among extremes models in being directly analogous to a classical linear model. We specify a sufficient condition on the spatial dependence parameter such that our extreme SAR model has desirable properties. We also describe the limiting angular measure, which is discrete, and corresponding tail pairwise dependence matrix (TPDM) for the model.

After examining model properties, we then investigate two approaches to estimation and inference for the common-scale extreme SAR model. First, we consider a censored likelihood approach, implemented using Bayesian MCMC with a data augmentation step, but find that this approach is not robust to model misspecification. As an alternative, we develop a novel estimation method that minimizes the discrepancy between the TPDM for the fitted model and the estimated TPDM, and find that it is able to produce reasonable estimates of extremal dependence even in the case of model misspecification.

ACKNOWLEDGEMENTS

I would not be here without the guidance and steadfast support of my advisor, Dan Cooley. Dan has been a great role model, both as a patient teacher and mentor, and as a researcher with vision and perspective. Our meetings often gave me a renewed sense of purpose and direction. I am grateful to Dan for exposing me to the broader extremes community by funding and encouraging me to attend international conferences and workshops. I feel very lucky to have had the opportunity to learn from and share ideas with many experts in the field.

A cornerstone of my graduate school experience was collaborating with world-class scientists at the National Center for Atmospheric Research (NCAR). Many thanks are due to Steve Sain, Claudia Tebaldi, and Brian O'Neill for including me in the BRACE project that kickstarted my research into climate extremes. I also learned a great deal from analyzing ozone extremes in collaboration with Alma Hodzic and Eric Gilleland. In addition, I am grateful to Doug Nychka and Dorit Hammerling for treating me like family.

I owe many thanks to Emeric Thibaud for his assistance on the extreme SAR project. Emeric is not only a brilliant researcher, but also a wonderfully kind mentor, and I appreciated our brainstorming sessions. I am also grateful to Ben Shaby for taking the time to chat about MCMC, and to Raphaël Huser for a memorable conversation in the Swiss Alps. Thanks also to Brook Russell and Will Porter for paving the way on the ozone project.

Colorado State University has been an incredibly supportive community. I would like to thank my committee members Jennifer Hoeting, Ander Wilson, and Libby Barnes, for their helpful feedback. Though she may not know it, Jennifer was one of the reasons I decided to come to CSU, and I am so grateful to her for believing in me. I would also like to thank Phil Turk for being another advocate as well as my mentor in the realm of statistical consulting. I send heartfelt thanks to Kristina Quynn and the rest of CSU Writes for helping me rediscover the joy and pleasure of writing. Additionally, none of this could have happened without excellent administrative assistance from Kristin Stephens and Katy Jackson.

I was fortunate to receive funding from many sources during my time at CSU, including CSU's Center for Interdisciplinary Mathematics and Statistics fellowship, Environmental Protection Agency grant EPA-STAR RD-83522801-0, and National Science Foundation grant DMS-1243102. Additional travel funding was provided by the Research Network for Statistical Methods for Atmospheric and Oceanic Sciences (STATMOS).

I have had the great fortune to get to know an amazing group of graduate students and post-docs who have inspired and supported me throughout the past five years. Special thanks to Josh Hewitt, Clint Leach, Erin Gorsich, Magda Garbowski, and Henry Scharf for all the times they showed up to write, chat, or toss a disc. Special thanks to the Tomato Club for helping me persevere. Thanks also to my roommate Morgan Hawker and my pseudo-roommate Jessica Mao for their warmth and generosity.

The Fort Collins ultimate community was instrumental in keeping me active and sane. I must also thank the wonderful physical therapists and fitness professionals, including Katie Hall, Bryan Duran, Sue Rutherford, Eric Maxwell, and Kelly Buell-Schoenman, who have helped me move more easily through the world.

No acknowledgements section would be complete without recognizing the fundamental support from my family. I would like to thank my parents, Charlotte Lo and Douglas Fix, for motivating me to pursue a PhD in the first place, and for their unwavering confidence in my ability to do whatever I set my mind to. Finally, I am immeasurably grateful to my partner, David Betz, for being my cheerleader and my rock. You have the biggest heart and inspire me to be better every day. Thanks for waiting.

DEDICATION

For David

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Motivation

In September 2013, Colorado experienced its second most costly natural disaster (Lukas et al., 2014). Severe flooding inundated the Front Range during one of the region's most extreme rainfall events on record. In total, there were eight flood-related fatalities and damages exceeded $2 billion (Gochis et al., 2015). This flooding event, and others like it, are extreme events. Such events occur infrequently but can impose a high cost on society. Extreme value theory (EVT) was developed in the early part of the 20th century to understand the asymptotic distribution of the largest order statistic. Development of statistical methodology for extreme events was spurred by the devastating North Sea flood of 1953 that killed more than 1800 people in the Netherlands, and statistical analyses contributed to decision making about the height of the Dutch dykes (de Haan, 1990). While the earliest applications of EVT focused primarily on hydrology and civil engineering, these days the need to characterize the likelihood and severity of extreme events applies to a wide variety of fields, including atmospheric science, finance, telecommunications, forest fire science, and public health.

This dissertation presents applied and methodological advances in the analysis of extreme values. The main goal of an extreme value analysis is to describe the upper (or lower) tail of a probability distribution. In the univariate case, the aim is to characterize the tail of the distribution of a single variable, such as daily precipitation at a weather station. A multivariate analysis may consider the joint tail (i.e., the tail dependence structure) of several variables, such as daily ozone levels and air temperature at a certain location, or, in the spatial setting, a single variable measured at several neighboring locations. Traditional statistical methods that aim to describe the bulk of a distribution, and may employ such summary measures as means, variances or correlations, are not useful for describing the tail. Because the focus is on the tail, extreme value analyses typically use only the extreme observations, discarding the bulk of the data. By

definition, extremes are rare, so in the study of extremes we are always data poor. Moreover, extreme value analyses often require estimating the probability of events that are more extreme than any that have previously been observed. EVT provides a theoretical framework for such extrapolations, with extreme value models generated by asymptotic arguments.

The work in this dissertation is largely motivated by the types of data found in the atmospheric sciences. We consider large ensembles of global climate model output (Chapter 2), compare high resolution atmospheric chemistry model output to station observations and reanalysis (Chapter 3), and propose a new model and develop new inference methods for areal extremes motivated by the prevalence of atmospheric datasets that are indexed by regular grids (Chapter 4 and Chapter 5).

1.2 Outline

The remainder of this chapter (Chapter 1) provides an overview of EVT concepts essential to this dissertation. We first briefly review classical univariate EVT, including block maxima and threshold exceedance approaches. We then introduce the framework of multivariate regular variation and tie it back to multivariate extreme value distributions. We end with a few notes on extremes in the (spatial) process setting. Our review is far from exhaustive, thus we point the reader to several excellent books on the probability theory underlying the study of extremes, including Resnick (1987) and de Haan and Ferreira (2006), as well as the statistical analysis of extremes, such as Coles (2001) and Beirlant et al. (2006).

Chapter 2 leverages two initial condition ensembles of a global climate model to compare future precipitation extremes under two climate change scenarios. A methodological contribution of this work is to demonstrate the potential of a “pattern scaling" approach for extremes. This chapter is based on an article published in *Climatic Change*.[1]

[1] Fix, M. J., Cooley, D., Sain, S. R., & Tebaldi, C. (2018). A comparison of US precipitation extremes under RCP8.5 and RCP4.5 with an application of pattern scaling. *Climatic Change*, 146(3):335–347.

Chapter 3 considers the complex question of how to evaluate the ability of a high resolution atmospheric chemistry model to reproduce observed relationships between meteorology and high or extreme surface ozone. A contribution of this work is to develop a novel framework for comparing simulations and observational data products. We uniquely investigate meteorological sensitivities, extending a recent approach relating extreme responses to a set of atmospheric drivers. This chapter is based on an article published in *Atmospheric Environment*.[2]

Chapter 4 is motivated by the preponderance of large gridded products in the atmospheric sciences. Existing models for spatial extremes, such as max-stable process models, tend to be geostatistical in nature as well as very computationally intensive. The goal of this work is to develop a simple spatial extremes model that is computationally feasible for high-dimensional areal data. To this end, Chapter 4 proposes extremal versions of the classical simultaneous autoregressive (SAR) model. Linear models are ubiquitous in traditional (non-extreme) statistics, but to our knowledge this is the first extremes model that is directly analogous to a classical linear model. One property of our extreme SAR model is that its limiting angular measure, which characterizes the tail dependence structure, is discrete in nature. Similar extremal models have been found to be challenging for inference. Chapter 4 also describes the tail pairwise dependence matrix (TPDM) which summarizes this dependence.

Building upon Chapter 4, Chapter 5 presents several approaches to estimation and inference for the extreme SAR model, and discusses their respective challenges. We investigate a Bayesian approach, which relies on a likelihood specification. Likelihood approaches for extremes models with discrete dependence structure have not been previously investigated. We also develop a novel estimation method that minimizes the discrepancy between the TPDM for the fitted model and the estimated TPDM, and find that it is able to produce reasonable dependence estimates even in the case of model misspecification.

[2]Fix, M. J., Cooley, D., Hodzic, A., Gilleland, E., Russell, B. T., Porter, W. C., & Pfister, G. G. (2018). Observed and predicted sensitivities of extreme surface ozone to meteorological drivers in three US cities. *Atmospheric Environment*, 176:292–300.

Finally, Chapter 6 concludes the dissertation with a brief summary of the work completed and remarks on future extensions of each of the projects.

1.3 Univariate Extremes

1.3.1 Block maxima approach

Asymptotic theory

Classical extreme value models arise from arguments regarding the limiting distribution of suitably renormalized block maxima. Let $X_1, \ldots, X_n$ represent independent copies of a random variable X with common ("parent") distribution function F. Consider $M_n = \max(X_1, \ldots, X_n)$, and note that $\mathbb{P}(M_n \leq x) = F^n(x)$. For $F(x) < 1$, $F^n(x) \to 0$ as $n \to \infty$, so $F^n(x)$ converges to a degenerate distribution function with a point mass at the upper endpoint $\sup\{x : F(x) < 1\}$. Instead, we consider limiting distributions of renormalized maxima $\frac{M_n - b_n}{a_n}$, for appropriate choices of $a_n > 0$ and $b_n \in \mathbb{R}$. The Extremal Types Theorem, first introduced by Fisher and Tippett (1928) and later rigorously proven by Gnedeko (1943), provides a fundamental result. If there exist sequences of constants $a_n > 0$ and b_n such that, as $n \to \infty$,

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \to G(x) \tag{1.1}$$

for some non-degenerate distribution function G, then G must belong to one of the following types of extreme value distributions:

$$\text{Type I (Gumbel):} \qquad G(x) = \exp\{-\exp(-x)\}\,,\ x \in \mathbb{R} \tag{1.2}$$

$$\text{Type II (Fréchet):} \qquad G(x) = \begin{cases} 0, & x \leq 0, \\ \exp(-x^{-\alpha}), & x > 0 \end{cases} \tag{1.3}$$

$$\text{Type III (reverse Weibull):} \qquad G(x) = \begin{cases} \exp(-(-x)^{\alpha}), & x < 0, \\ 1, & x \geq 0 \end{cases} \tag{1.4}$$

for some $\alpha > 0$. This result is powerful because when the limiting distribution G exists, it follows one of the above distributions irrespective of the parent distribution F. We say F is in the max-domain of attraction (MDA) of G. The extreme value distributions are equivalent to the max-stable distributions, defined as the distributions for which there exist sequences $a_n > 0$ and $b_n \in \mathbb{R}$ such that $G^n(a_n x + b_n) = G(x)$ for all positive integers n and all $x \in \mathbb{R}$.

The three extremal types above can be combined into a single parametric family known as the generalized extreme value (GEV) distribution with the form

$$G(x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_{+}^{-1/\xi}\right\}, \tag{1.5}$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter, $\xi \in \mathbb{R}$ is the shape parameter, and $y_+ := \max(y, 0)$. The support of G is $\{x \in \mathbb{R} : 1+\xi(x-\mu)/\sigma > 0\}$. The shape parameter ξ determines the tail behavior (see Figure 1.1). If $\xi = 0$ (we take the limit of (1.5) as $\xi \to 0$), then the tail is light and G is Gumbel. The Gamma and Gaussian distributions are examples of distributions that are in the MDA of G with $\xi = 0$. If $\xi > 0$, then the tail is heavy and G is Fréchet. The Student t distribution is in the MDA of G with $\xi = \frac{1}{d.f.} > 0$. If $\xi < 0$, then the upper tail is bounded and G is reverse Weibull. The Beta distribution is in the MDA of G with $\xi < 0$.

Statistical inference

In practice, the limit in (1.1) is interpreted as an approximation for large n, and the GEV family can be used for modeling the distribution of block (e.g., annual or seasonal) maxima. We do not need to estimate the renormalizing sequences a_n and b_n because they can be absorbed into the μ and σ parameters of the GEV (Coles, 2001). Suppose one begins with iid data $x_{b,i}$ indexed by block $b = 1, \ldots, B$ and time-within-block $i = 1, \ldots, n$. Letting $m_{n,b} = \max_{i=1,\ldots,n} x_{b,i}$, one obtains B block maxima from which the parameters of the GEV distribution can be estimated

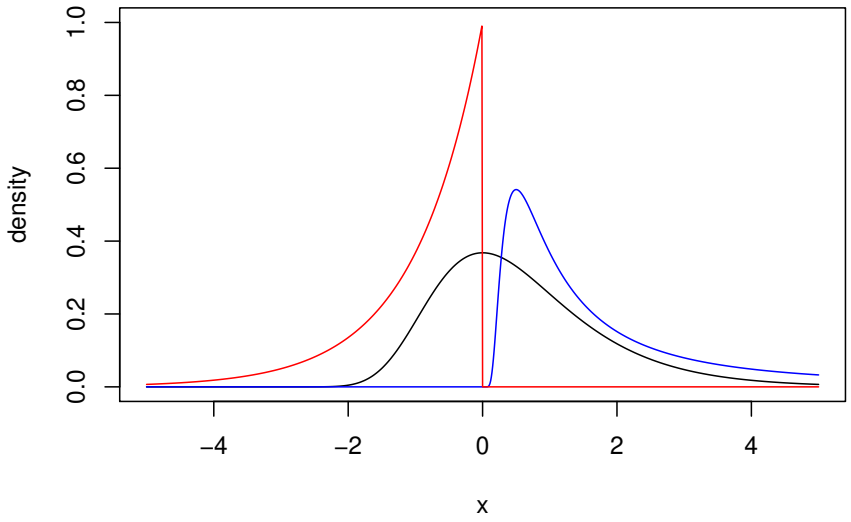

Figure 1.1: Illustration of the three extremal types: standard Gumbel (black), unit Fréchet (blue), and unit reverse Weibull (red) density functions.

via maximum likelihood (Prescott and Walden, 1980; Smith, 1985) or moment-based methods like probability-weighted moments (Hosking et al., 1985). From the fitted model, we can obtain estimates of high quantiles (so-called “return levels" in environmental science) along with the associated uncertainties.

The extreme value distributions can be thought of as “ultimate" approximations that hold in the limit, i.e., as block size n goes to infinity. In practice, accuracy may be increased by using a penultimate approximation to the distribution of M_n for finite n (Embrechts et al., 2012; Katz, 2013). This was noted as early as Fisher and Tippett (1928), who showed that taking $\xi < 0$ (reverse Weibull type) provides a better approximation for maxima of finite samples from the Gaussian distribution than does the limiting Gumbel distribution. This means that there is no advantage to constraining the estimation procedure to $\xi = 0$ even if the correct limiting distribution is known to be Gumbel (Furrer and Katz, 2008).

1.3.2 Threshold exceedance approach

Asymptotic theory

A disadvantage of the block maxima approach is that it may be wasteful of data. Although many extreme events could have occurred in the same block, only one event per block is retained and the rest discarded. An alternative is to fix a high threshold u and retain all observations above this threshold. The classical asymptotically motivated model for exceedances over a high threshold is the generalized Pareto distribution (GPD). Suppose F is in the MDA of a GEV distribution G. Then for large enough u, the distribution function of threshold excesses is approximately given by the GPD:

$$\mathbb{P}(X - u \leq x \mid X > u) \approx 1 - \left(1 + \frac{\xi x}{\sigma_u}\right)_+^{-1/\xi}, \tag{1.6}$$

where ξ is equivalent to the shape parameter of the GEV distribution and $\sigma_u = \sigma + \xi(u - \mu)$ (Balkema and de Haan, 1974; Pickands, 1975). In the case $\xi \to 0$, the right-hand side of (1.6) becomes $1 - \exp(-x/\sigma_u)$.

Below we outline a justification for the GPD from Coles (2001). Assume $X, X_1, X_2, \ldots, \overset{iid}{\sim} F$ and $F \in \text{MDA}(G)$. Then for large enough n, following from (1.1) we have the approximation

$$n \log F(x) \approx -\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]_+^{-1/\xi}. \tag{1.7}$$

For large values of x, a Taylor expansion gives $\log F(x) \approx -[1 - F(x)]$. Hence for large u,

$$1 - F(u) \approx n^{-1}\left[1 + \xi\left(\frac{u - \mu}{\sigma}\right)\right]_+^{-1/\xi}, \tag{1.8}$$

and similarly for $x > 0$,

$$1 - F(u + x) \approx n^{-1}\left[1 + \xi\left(\frac{u + x - \mu}{\sigma}\right)\right]_+^{-1/\xi}. \tag{1.9}$$

Thus for sufficiently large u, we obtain

$$\begin{aligned}\mathbb{P}(X > u + x | X > u) &= \frac{1 - F(u+x)}{1 - F(u)} \\ &= \frac{n^{-1}\left[1 + \xi\left(\frac{u+x-\mu}{\sigma}\right)\right]_+^{-1/\xi}}{n^{-1}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}} \\ &= \left(1 + \frac{\xi x}{\sigma_u}\right)_+^{-1/\xi}, \end{aligned} \tag{1.10}$$

where $\sigma_u = \sigma + \xi(u - \mu)$ and (1.10) is the survival function of the GPD.

Statistical Inference

The result (1.6) suggests a framework for statistical modeling of threshold exceedances. Starting with the original data $x_1, \ldots, x_N$, extreme events can be identified by selecting a high threshold u above which the distribution's tail is well approximated by a GPD. The observations that exceed u, i.e., $\{x_i : x_i > u\}$, can be labeled $x_{(1)}, \ldots, x_{(N_u)}$, where N_u is the number of exceedances. Then the threshold excesses $y_j = x_{(j)} - u, j = 1, \ldots, N_u$, can be used to estimate the parameters σ_u and ξ by either numerical maximum likelihood or moment-based approaches.

Statistical application of the GPD requires the choice of a suitable threshold, which involves a bias-variance tradeoff. The threshold must be sufficiently high for the GPD to be an appropriate model for the tail. On the other hand, raising the threshold reduces the sample size and thus increases the variance of the parameter estimates. In practice there are several diagnostic aids for threshold selection. Scarrott and MacDonald (2012) provide a review of some of the existing approaches. The penultimate approximation described in Section 1.3.1 for the GEV distribution applies equally well to the corresponding GPD (Furrer and Katz, 2008).

1.4 Multivariate Extremes and Regular Variation

There are several related approaches for modeling multivariate extremes. The classical approach is based on the multivariate extreme value distributions (MVEVDs; see Section 1.4.2). These arise as the limiting distributions of suitably normalized componentwise maxima and can be developed in a multivariate analogue to the argument leading to the Extremal Types

Theorem and the GEV distribution (de Haan and Ferreira, 2006, Sections 6.1 and 6.2). As such, MVEVDs provide sensible models for componentwise sample maxima. Note that componentwise maxima are not an entirely intuitive concept, as maxima in different components can occur at different times.

Akin to univariate extremes, threshold exceedance approaches can also be taken for modeling multivariate extremes. Such approaches require defining what is meant by a multivariate threshold exceedance. The multivariate GPD of Rootzén and Tajvidi (1997) can be used when thresholds are defined in terms of each univariate marginal, i.e., extreme events are described in terms of Cartesian coordinates. The probabilistic framework of multivariate regular variation (MVRV; see Section 1.4.1) is well-suited for defining a threshold in terms of a vector norm, and is easily described in terms of pseudo-polar coordinates.

The above approaches to multivariate extreme value modeling share several commonalities. In all cases, there is no finite parameterization for the class of dependence structures (but parametric models may be specified). For each approach, characterization of the dependence structure is made easier by imposing assumptions on the marginal behavior, and consequently estimation of the dependence structure from data often requires transformation of the marginals. As in the univariate case, models for multivariate extremes are fit only to observations deemed extreme. The approaches are also theoretically linked, as each can be tied to the so-called angular measure, which we will describe in more detail in Section 1.4.1. This angular measure fully characterizes dependence in the limit.

Within this dissertation, we will focus on the MVRV framework for modeling multivariate threshold exceedances. MVRV provides a probabilistic characterization of the joint (upper) tail of a random vector, and is defined entirely in terms of the joint tail. The MVRV framework assumes the heavy-tailed case, i.e., it implies that the joint tail decays like a power function. MVRV is most easily understood via a pseudo-polar decomposition. The limiting angular measure mentioned above arises from this decomposition. Below we provide definitions and back-

ground on MVRV (Section 1.4.1), explain its connection to MVEVDs for componentwise maxima (Section 1.4.2), and discuss methods of inference (Section 1.4.3).

Although MVRV can be defined on $\mathbb{R}^p$ (Resnick, 2007, Section 6.5.5), we model in the nonnegative orthant to focus attention on the upper tail. In applications, there is often a natural direction in which one wants to assess risk. Consider wildfire risk, which is related to many factors including fuel and weather conditions. For example, high windspeed and low humidity (high dryness) are associated with increased fire risk. The left panel of Figure 1.2 shows daily summary measures for windspeed and dryness at a weather station in southern California. More details about these data can be found in Cooley et al. (2018). Wildfires, such as those indicated by the colored points, tend to occur in the upper right, when both variables are high. Thus this is the direction of interest. Note that the raw data are not regularly varying, so for a MVRV analysis it is necessary to perform a marginal transformation as discussed earlier. As an example, the right panel of Figure 1.2 shows the same data after transformation to Fréchet($\alpha = 2$) marginals. With the transformed data, one could use the MVRV on the nonnegative orthant to model the tail dependence between the daily windspeed and dryness variables, and ultimately estimate the probabilities associated with risk regions where both variables are high.

1.4.1 Asymptotic theory of multivariate regular variation

Resnick (2007) presents several equivalent definitions of MVRV, including the following. Let $\mathbf{Z}$ be a random vector taking values in the nonnegative orthant $\mathbb{R}^d_+ = [0,\infty)^d$. Let $M_+(\mathfrak{C})$ denote the space of nonnegative Radon measures on $\mathfrak{C} = [0,\infty]^d \setminus \{\mathbf{0}\}$. We say $\mathbf{Z}$ is *regularly varying* if there exists a sequence $c_n \to \infty$ and a limit measure $\nu(\cdot)$ on the Borel subsets of $\mathbb{R}^d_+$ such that

$$n\mathbb{P}\left(\frac{\mathbf{Z}}{c_n} \in \cdot\right) \xrightarrow{v} \nu(\cdot) \tag{1.11}$$

in $M_+(\mathfrak{C})$ as $n \to \infty$, where $\xrightarrow{v}$ denotes vague convergence of measures. One recognizes from (1.11) the Poisson convergence, and in fact Resnick and other authors give equivalent defini-

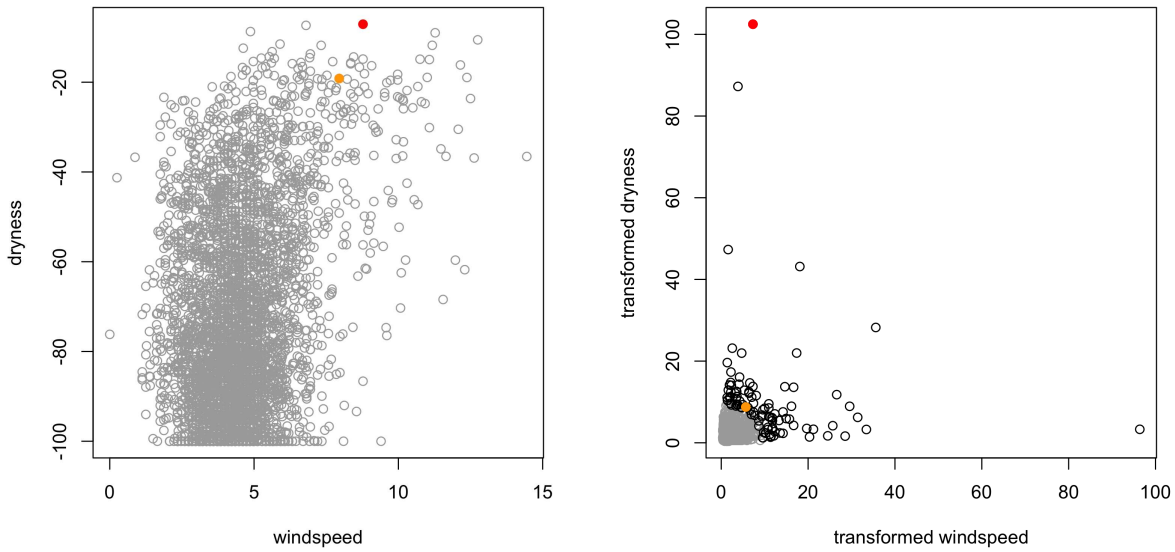

Figure 1.2: Illustration of the directionality of risk: wildfire example. Left: daily summary measures of windspeed and dryness at a weather station in southern California during the years 1973-2015. Right: Transformed data with Fréchet($\alpha = 2$) marginals. Black circles indicate observations exceeding the 0.975 quantile of the radial component (L_2 norm). In both panels, orange and red points correspond to the initial days of two of the most destructive wildfires during the observation period.

tions of MVRV in terms of counting measures converging to a Poisson random measure with mean measure ν (Resnick, 2007, Section 6.2).

The limit measure ν can be shown to have the scaling property

$$\nu(pA) = p^{-\alpha}\nu(A) \tag{1.12}$$

for any set $A \subset \mathfrak{C}$ and any $p > 0$, where $\alpha > 0$ is termed the *tail index* and determines the power-law behavior of the tail. We denote a d-dimensional regularly varying random vector $\mathbf{Z}$ with tail index α by $\mathbf{Z} \in RV_+^d(\alpha)$.

The scaling property (1.12) suggests a transformation to pseudo-polar coordinates, and leads to an equivalent definition of MVRV. Given any norm $||\cdot||$, denote the nonnegative unit sphere $\mathbb{S}_+^{d-1} = \{\mathbf{z} \in \mathbb{R}_+^d : ||\mathbf{z}|| = 1\}$. We define "radial" and "angular" components by $R = ||\mathbf{Z}||$ and $\boldsymbol{\Theta} = ||\mathbf{Z}||^{-1}\mathbf{Z}$, respectively. Then the random vector $\mathbf{Z}$ is regular varying if there exists a sequence $c_n \to \infty$ and a finite measure H on $\mathbb{S}_+^{d-1}$ such that for any H-continuity Borel set $B \subset \mathbb{S}_+^{d-1}$, and for $r > 0$,

$$n\mathbb{P}\left(\frac{R}{c_n} > r, \boldsymbol{\Theta} \in B\right) \xrightarrow{v} r^{-\alpha} H(B) \tag{1.13}$$

as $n \to \infty$. The right-hand side of (1.13) is a product measure, indicating that the radial and angular components become independent in the limit. H is termed the *angular measure* (sometimes also referred to as the *spectral measure*), and completely characterizes the limiting tail dependence structure of $\mathbf{Z}$. The normalizing sequence $\{c_n\}$ can be chosen such that H is a probability measure, although in many modeling situations it may make sense for H to have total mass other than one. It is often assumed that $\mathbf{Z}$ has common marginal distributions, which imposes the following balance condition on H (Resnick, 1987):

$$\int_{\mathbb{S}_+^{d-1}} \theta_1 H(d\boldsymbol{\theta}) = \int_{\mathbb{S}_+^{d-1}} \theta_j H(d\boldsymbol{\theta}), \qquad j = 2, \ldots, d. \tag{1.14}$$

To provide some intuition on the angular measure, consider the bivariate case $\mathbf{Z} = (Z_1, Z_2)^\top \in RV_+^2(\alpha)$ with common marginal distributions. There are two limiting cases. If $\mathbf{Z}$ exhibits perfect dependence, i.e., if Z_1 determines Z_2 exactly, then H consists of a single point mass on the interior of $\mathbb{S}_+^1$. In the asymptotic independence case, i.e., if $\lim_{z\to\infty} \mathbb{P}(Z_2 > z | Z_1 > z) = 0$, then H consists of two point masses, one at each end of the one-dimensional unit sphere. In general, dependence increases as the mass of H concentrates toward the center of $\mathbb{S}_+^{d-1}$ and decreases as it moves towards the edges and vertices of $\mathbb{S}_+^{d-1}$.

Although the polar-coordinate geometry is natural for describing regular variation, it can be hard to reconcile with the Cartesian geometry required by cumulative distribution functions and other familiar notions. Geometric arguments can give needed expressions. For a given norm $||\cdot||$, tail index α, and angular measure H, consider the set $A = [\mathbf{0}, \mathbf{z}]^\mathsf{c}$ for $\mathbf{z} = (z_1, \ldots, z_d)^\top \in \mathfrak{C}$. Then

$$\begin{aligned} \nu(A) &= \int_A \alpha r^{-\alpha-1} dr dH(\boldsymbol{\theta}) \\ &= \int_{\mathbb{S}_+^{d-1}} \int_{r=\bigwedge_{i=1}^d \frac{z_i}{\theta_i}}^{\infty} \alpha r^{-\alpha-1} dr dH(\boldsymbol{\theta}) \end{aligned}$$

$$= \int_{\mathbb{S}_+^{d-1}} \bigvee_{i=1}^{d} \left(\frac{z_i}{\theta_i}\right)^{-\alpha} H(d\boldsymbol{\theta}). \tag{1.15}$$

1.4.2 Link to multivariate extreme value distributions

Analogous to the univariate block maxima approach described in Section 1.3.1, classical multivariate EVT considers the limiting distribution of the vector of appropriately renormalized componentwise maxima. Let $\mathbf{X} = (X_1, \ldots, X_d)^\top$ be a d-dimensional random vector with marginals $F_1, \ldots, F_d$ and joint distribution F. Suppose we have n iid replicate vectors $\{\mathbf{X}_i\}_{i=1}^n$ and define the vector of componentwise maxima $\mathbf{M}_n = \left(\bigvee_{i=1}^n X_{i,1}, \ldots, \bigvee_{i=1}^n X_{i,d}\right)^\top$. Note that $\mathbf{M}_n$ does not necessarily correspond to an observed data point, as maxima may not occur at the same time in each margin. If there exist renormalizing sequences of vectors $\mathbf{a}_n \geq \mathbf{0}$ and $\mathbf{b}_n \in \mathbb{R}^d$ such that

$$\mathbb{P}\left(\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \leq \mathbf{z}\right) = F^n(\mathbf{a}_n \mathbf{z} + \mathbf{b}_n) \to G(\mathbf{z}) \tag{1.16}$$

(non-degenerate) as $n \to \infty$, where all operations are componentwise, then G is a multivariate extreme value distribution (MVEVD). A sequence of random vectors can only converge if all the marginals converge, so $F_j^n(\mathbf{a}_n \mathbf{z} + \mathbf{b}_n) \to G_j(\mathbf{z})$ for $j = 1, \ldots, d$. Thus G has univariate GEV marginals. Although the marginals are fully parameterized, no finite parameterization exists for the dependence structure of the d components.

The MVRV framework can be tied back to classical multivariate EVT, in that MVRV is a condition implying that a distribution is in the MDA of a MVEVD (Beirlant et al., 2006; Resnick, 1987). Let $\mathbf{Z} \in RV_+^d(\alpha)$ with limit measure ν as in (1.11), and let $\mathbf{M}_n$ be the vector of componentwise maxima of n iid realizations of $\mathbf{Z}$. Then there exist renormalizing sequences of vectors $\mathbf{a}_n \geq \mathbf{0}$ and $\mathbf{b}_n \in \mathbb{R}^d$ such that (1.16) holds with $G(\mathbf{z}) = \exp\{-\nu[\mathbf{0}, \mathbf{z}]^c\}$, and the marginals of G are GEV with $\xi = 1/\alpha > 0$. In other words, $\mathbf{Z}$ is in the MDA of a MVEVD with Fréchet(α) margins and dependence structure characterized by the measure ν from (1.15).

1.4.3 Statistical inference for multivariate extremes

The classical approach to statistical inference for multivariate extremes consists of partitioning a sample of multivariate observations into blocks and fitting a MVEVD to the sample of componentwise block maxima. Beirlant et al. (2006) describes both parametric and nonparametric techniques to estimate the MVEVD dependence structure. More efficient inference can be performed using threshold methods, and in this dissertation we focus on statistical inference for multivariate threshold exceedances within the MVRV framework. There are several possible approaches to define a threshold exceedance in the multivariate setting. Here we consider two definitions, one in terms of the norm of a random vector, and one in terms of the marginals, as illustrated in Figure 1.3. In the first case, an exceedance is defined as an observation whose norm (radial component) exceeds a suitably high threshold. In the second case, a multivariate observation is considered an exceedance if at least one of its components exceeds its marginal threshold.

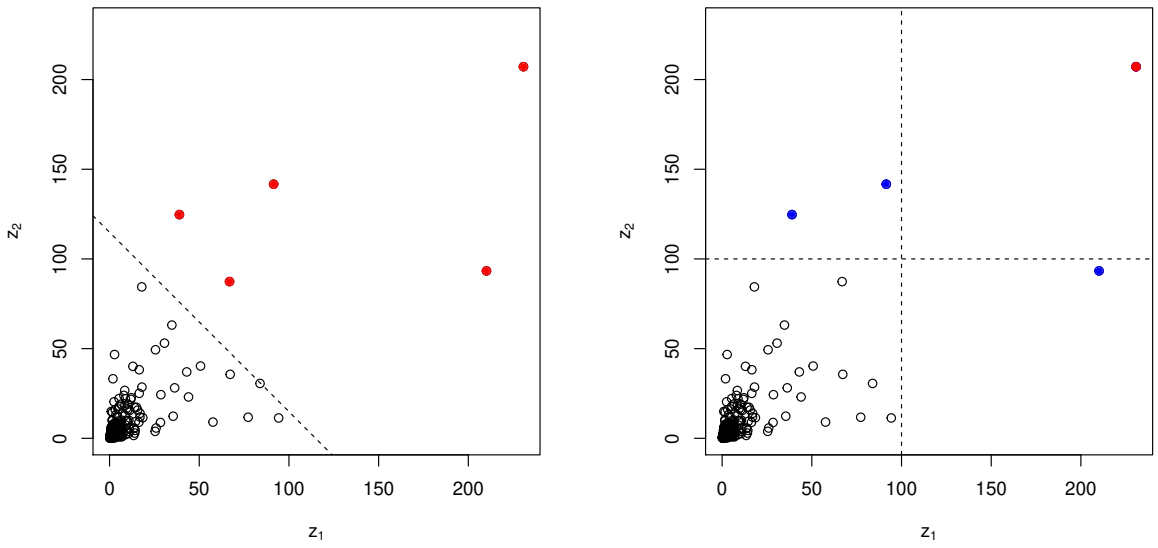

Figure 1.3: Illustration of two definitions for multivariate threshold exceedances, for $d = 2$ and with unit Fréchet margins. Colored points are exceedances, and dotted lines indicate the norm threshold (left) or marginal thresholds (right). On the right, blue points are those which exceed the marginal threshold in only one component; these observations would be censored under a censored likelihood approach.

The MVRV framework implies that the marginal distributions are univariate regularly varying with common tail index α. In statistical practice, a standard approach is to transform the univariate marginals to a convenient distribution with common α. This can be done by first estimating the marginal distributions, either parametrically or nonparametrically, and then applying the probability integral transform. Doing so retains the tail dependence structure, as Proposition 5.10 of Resnick (1987) shows the MDA is preserved under monotone transformations of the marginals. This approach is similar in spirit to a copula, but aims only to describe the tail.

After estimating the marginal effects, it is possible to focus on estimating the tail dependence structure, which is the crux of inference for multivariate extremes. Below we describe two general approaches to likelihood inference, corresponding to two definitions of threshold exceedances. The first approach consists of fitting an angular measure model to norm exceedances. Finite-sample estimation of the angular measure assumes that the limit (1.13) is an equality for $r > r_0$, where r_0 is a sufficiently high threshold. Assume that H is continuously differentiable with Radon-Nikodym derivative $h(\boldsymbol{\theta};\boldsymbol{\eta})$, which is termed the *angular density*. Then, given a parametric model for the angular density (see, e.g., Ballani and Schlather, 2011; Coles and Tawn, 1991; Cooley et al., 2010), it is possible to write down the corresponding likelihood for the points $\mathbf{z}_1,\dots,\mathbf{z}_{N_0}$ for which $||\mathbf{z}_i|| = r_i > r_0$, which is given by a Poisson point process approximation:

$$\begin{aligned} L(\boldsymbol{\eta}; z_1, \dots z_n) &= \exp\left(-\frac{r_0}{c_n}\right)\left\{\prod_{i=1}^{N_0}\left(\alpha r_i^{-\alpha-1}/c_n\right)h(\boldsymbol{\theta}_i;\boldsymbol{\eta})\right\}/N_0! \\ &\propto \prod_{i=1}^{N_0} h(\boldsymbol{\theta}_i;\boldsymbol{\eta}). \end{aligned} \tag{1.17}$$

Parameters can then be estimated via numerical maximum likelihood.

Alternatively, when an observation is defined to be a threshold exceedance if at least one of its components exceeds its marginal threshold, a censored likelihood approach is often used, where the non-extreme components are censored at their marginal thresholds (Ledford and

Tawn, 1996; Smith et al., 1997). Censored likelihood approaches reduce the bias that may arise from non-applicability of the limiting model in regions where not all components are extreme (Huser et al., 2016). More details on inference via censored likelihood will be given in Chapter 5.

Likelihood-based inference for multivariate extremes can be challenging in high dimensions. Moreover, some models may fail to yield densities. Non-likelihood approaches are particularly useful for models whose angular measures consist of discrete point masses, as we will return to in Chapter 5. Recent examples include M-estimators based on the continuous ranked probability score (Yuen and Stoev, 2014) and the stable tail dependence function (Einmahl et al., 2016), both of which are related to the underlying multivariate cumulative distribution function. However, to our knowledge, applications of these methods to spectrally discrete models have been restricted to low-dimensional examples.

1.5 Process Setting

Thus far, we have introduced EVT in a finite dimensional setting, i.e., extremes of random variables or vectors. The asymptotic arguments of Section 1.3 can be extended to the infinite dimensional setting, i.e., extremes of stochastic processes. Extremal processes have primarily been used to model spatial extremes. Such models are typically geostatistical in nature, and are also challenging to fit. In Chapters 4 and 5 we will introduce and fit a very different type of spatial model than the extremal process models discussed here. Our proposed model is finite-dimensional, so a process is not needed. This multivariate model is very simple and is best suited for areal data, e.g., data on a regular grid. Our point in briefly discussing extremal process models here is to serve as contrast to the model we propose. A more comprehensive overview of statistical modeling for spatial extremes can be found in Davison et al. (2012).

1.5.1 Asymptotic theory

The natural extension of the GEV is the class of max-stable processes, which, under mild conditions, are the only possible non-degenerate limits of rescaled pointwise maxima of iid

random processes (de Haan, 1984). Let S be a compact subset of $\mathbb{R}^p$, typically representing the spatial region of interest. Consider a random process $Y = \{Y(\mathbf{s})\}_{\mathbf{s}\in S}$ defined over S, with continuous sample paths, and let $Y_1, Y_2, \ldots$ be independent replicates of Y. If there exist sequences of continuous functions $\{a_n(\mathbf{s})\}_{\mathbf{s}\in S} > 0$ and $\{b_n(\mathbf{s})\}_{\mathbf{s}\in S}$ such that the limiting process $Z = \{Z(\mathbf{s})\}_{\mathbf{s}\in S}$ defined by

$$\frac{\max_{i=1,\ldots,n} Y_i(\mathbf{s}) - b_n(\mathbf{s})}{a_n(\mathbf{s})} \to Z(\mathbf{s}), \qquad \mathbf{s} \in S, \quad n \to \infty, \tag{1.18}$$

is non-degenerate, then Z must be a max-stable process (de Haan, 1984; de Haan and Ferreira, 2006). We say that Y is in the MDA of Z. The marginals of Z are GEV-distributed, and can be transformed to a convenient distribution. For the remainder of this section, we will restrict attention to the commonly used *simple* max-stable processes with unit Fréchet margins. For every set of sites $\mathbf{s}_1, \ldots, \mathbf{s}_d \in S$,

$$\mathbb{P}\{Z(\mathbf{s}_1) \le z_1, \ldots, Z(\mathbf{s}_d \le z_d\} = \exp\{-V(z_1, \ldots, z_d)\}, \qquad z_1, \ldots, z_d > 0, \tag{1.19}$$

where the exponent measure function V satisfies $V(kz_1, \ldots, kz_d) = k^{-1}V(z_1, \ldots, z_d)$, $k > 0$ and $V(\infty, \ldots, \infty, z, \infty, \ldots, \infty) = z^{-1}$.

Several stationary parametric models for simple max-stable processes have been proposed, including the Smith (1990) process and the Schlather (2002) process. The popular Brown-Resnick process (Brown and Resnick, 1977; Kabluchko et al., 2009) is a flexible model that has been found valuable for a variety of applications. This process is defined by

$$Z(\mathbf{s}) = \max_{i\in\mathbb{N}} P_i^{-1} \exp\left\{\epsilon_i(\mathbf{s}) - \gamma(\mathbf{0}, \mathbf{s})\right\}, \qquad \mathbf{s} \in S, \tag{1.20}$$

where $\{P_i\}_{i\in\mathbb{N}}$ is a unit rate Poisson process on $(0, \infty)$, and the $\{\epsilon_i(\mathbf{s})\}_{\mathbf{s}\in S}$ are independent replicates of a zero-mean Gaussian process with stationary increments and variogram $2\gamma(\mathbf{s}, \mathbf{s}') = \mathbb{E}\left[\{\epsilon(\mathbf{s}) - \epsilon(\mathbf{s}')\}^2\right]$, for $\mathbf{s}, \mathbf{s}' \in S$. The dependence structure depends only on $\gamma(\cdot)$, and a large class of models can be attained via the choice of different variograms. For example, the bivariate ex-

ponent measure of a Brown-Resnick process $Z(\mathbf{s})$ with unit Fréchet margins at the pair of sites $\{\mathbf{s}_1, \mathbf{s}_2\}$ is given by

$$V(z_1, z_2) = \frac{1}{z_1}\Phi\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{z_1}{z_2}\right)\right\} + \frac{1}{z_2}\Phi\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{z_2}{z_1}\right)\right\}, \tag{1.21}$$

where $z_1 = z(\mathbf{s}_1)$, $z_2 = z(\mathbf{s}_2)$, $a = \left\{2\gamma(\mathbf{s}_1, \mathbf{s}_2)\right\}^{1/2}$, and $\Phi(\cdot)$ is the standard normal distribution function (Huser and Davison, 2013).

1.5.2 Statistical inference for spatial extremes

Due to the complicated form of the distribution of a max-stable process, likelihood inference is computationally challenging, and often intractable, for high-dimensional data. For example, although the d-dimensional distribution functions of the Brown-Resnick process are known (Genton et al., 2011; Huser and Davison, 2013), the number of terms in the corresponding d-variate density grows with dimension like the Bell numbers (Ribatet, 2013; Wadsworth and Tawn, 2014). This computational burden has motivated development of less expensive methods such as (pairwise) composite likelihood (Padoan et al., 2010) or the inclusion of partition information (Stephenson and Tawn, 2005; Thibaud et al., 2016). Recently, more efficient inference methods based on full (and possibly censored) likelihoods have been developed for threshold exceedances of processes in the MDA of a Brown-Resnick process (Engelke et al., 2015; Thibaud and Opitz, 2015; Wadsworth and Tawn, 2014), however these methods are still computationally challenging and applications have been restricted to $d \approx 30$ spatial locations.

Chapter 2

A Comparison of US Precipitation Extremes Under RCP8.5 and RCP4.5 with an Application of Pattern Scaling[3]

2.1 Introduction

Extreme weather events have serious environmental and socioeconomic impact. In order to prepare for future impactful events, there has been recent effort to project how extreme weather events will change in an altered climate. A recent IPCC report focusing on extreme events and their impacts states “It is likely that the frequency of heavy precipitation or the proportion of total rainfall from heavy falls will increase in the 21st century over many areas of the globe" (IPCC, 2012, page 11). In an analysis of the CMIP5 multi-model ensemble, Kharin et al. (2013) found that the magnitude of precipitation extremes over land will increase appreciably with global warming, and return periods of late 20th century extreme precipitation events are projected to become shorter.

In this chapter, we use general circulation model (GCM) output to investigate future extreme precipitation associated with Representative Concentration Pathways (RCPs) 8.5 and 4.5 (Van Vuuren et al., 2011) over the contiguous United States. RCP8.5 corresponds to the pathway with the highest greenhouse gas emissions, while RCP4.5 describes a moderate mitigation pathway. This study is part of a larger project on the Benefits of Reducing Anthropogenic Climate changE (BRACE; O’Neill and Gettelman, 2018) which focuses on characterizing the difference in impacts driven by climate outcomes resulting from the forcing associated with these two pathways. Specifically, we utilize a “Large Ensemble” of 30 perturbed initial condition runs under RCP8.5 (CESM-LE; Kay et al., 2015) and a “Medium Ensemble” of 15 perturbed initial condition runs under RCP4.5 (CESM-ME; Sanderson et al., 2018) to statistically model how ex-

[3] Fix, M. J., Cooley, D., Sain, S. R., & Tebaldi, C. (2018). A comparison of US precipitation extremes under RCP8.5 and RCP4.5 with an application of pattern scaling. *Climatic Change*, 146(3):335–347.

treme precipitation is affected by changes in global mean temperature. All ensemble members use a single CMIP5 coupled climate model: the NCAR Community Earth System Model, version 1, with the Community Atmosphere Model, version 5 (CESM1(CAM5); Hurrell et al., 2013) at approximately 1° horizontal resolution in all model components. Each member within an initial condition ensemble has a unique climate trajectory due to small round-off level differences in the initial atmospheric state. The initial condition ensemble provides us with a large data set of extreme precipitation events which allows us to reduce the uncertainty associated with the quantities we estimate. This is in contrast to typical extreme value studies of either observational data or single realizations of model output which often have large uncertainties associated with quantities of interest such as return levels.

It is important to note that GCM precipitation output, particularly extreme precipitation output, should not be interpreted in the same manner as observations recorded at weather stations (the disparate nature of these two data types is illustrated in Figure 2.1 for Boulder, CO). By their nature as point measurements, precipitation extremes obtained from individual station records are not directly comparable to gridded model output, as Chen and Knutson (2008) argue that it is more appropriate to interpret GCM output as an areal average, rather than as an estimate corresponding to a point location. To link GCM output to station observations, it could be necessary to resort to statistical or dynamical downscaling. The main value of GCMs lies in describing how extreme precipitation is likely to *change* under an altered climate, and in particular under various climate change scenarios. In this study, we restrict our focus to studying the change in GCM output.

Since we analyze annual maximum daily precipitation, the generalized extreme value (GEV) distribution forms the foundation of our statistical model. A number of studies have applied the GEV distribution to describe the behavior of extreme precipitation from climate model output (e.g., Beniston et al., 2007; Fowler et al., 2007; Kharin et al., 2013; Schliep et al., 2010; Wehner, 2013). We employ a non-stationary GEV model (Coles, 2001, Ch. 6) because the climate model runs are transient, and we wish to model how extreme precipitation changes as

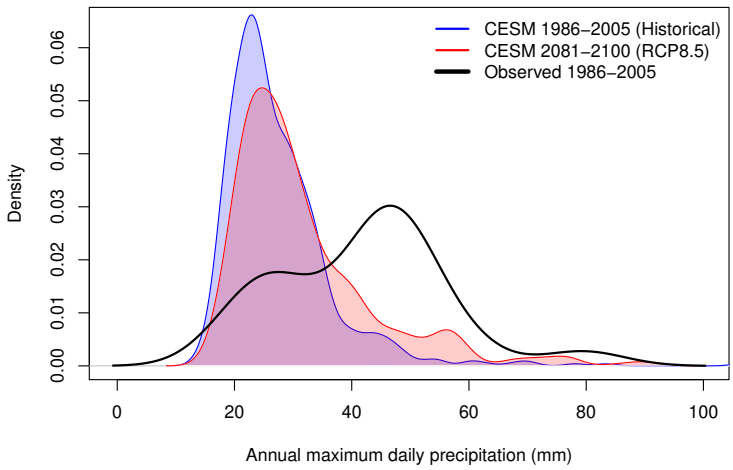

Figure 2.1: Comparison of kernel densities of annual maximum daily precipitation (mm) for 1986-2005 based on the 30-member CESM1(CAM5) initial condition ensemble (blue shaded) at the grid cell closest to Boulder, CO, and observations (black) from a weather station in Boulder, CO reveal a clear discrepancy between GCM and observed maxima. However, projections based on GCMs are still useful for describing how extreme precipitation may change in the future, as illustrated by the shifted kernel density of annual precipitation maxima for 2081-2100 based on the CESM RCP8.5 runs (red shaded).

climate changes. A simple approach to model climate trends in time is to allow the GEV parameters to be parametric functions of time (e.g., Fowler et al., 2010), although these functions may be too simple to accurately describe the time trend. Instead, we use a physical covariate of global mean temperature, a proxy for climate state, to implicitly model time dependence. There are several precedents for this approach. Brown et al. (2014), Hanel and Buishand (2011), Kharin et al. (2013), Westra et al. (2013) all employ a global temperature covariate to model a non-stationary GEV distribution.

The covariate of global mean temperature lends itself well to application of pattern scaling. Pattern scaling is a method of generating projections of future climate via a statistical model linking large-scale quantities (traditionally, global average temperature change) to local scale climate (Santer et al., 1990). Pattern scaling produces summary measures of future climate change at the regional scale without the need of running a fully coupled climate model for every scenario of interest. Pattern scaling is used widely by the impact and integrated assessment research communities, and Tebaldi and Arblaster (2014) provide a recent overview. Previously, pattern scaling has been primarily used to provide projections of mean behavior, with only

a limited number of investigations into pattern scaling for extremes (e.g., Brown et al., 2014; Pausader et al., 2012). In this study, we investigate a pattern scaling approach in which we fit a GEV model to only the CESM RCP8.5 runs, and then use this model to predict the distribution of annual maximum precipitation associated with the global mean temperatures provided by the RCP4.5 scenario. Having an ensemble of RCP4.5 runs allows us to evaluate this pattern scaling approach. We compare U.S. precipitation extremes projected by pattern scaling to those projected by a GEV model fitted directly to the RCP4.5 output.

The rest of the chapter is organized as follows. In Section 2.2, we describe our non-stationary GEV models, and summarize the initial condition ensemble output used to fit and validate these models. In Section 2.3, we present results from our fitted GEV models, which show that extreme precipitation levels tend to increase with global mean temperature across the contiguous U.S. Section 2.3.1 describes projections of future extreme precipitation under RCP8.5 based on CESM-LE, while Section 2.3.2 compares these to projections under RCP4.5 based on CESM-ME. In Section 2.3.3, we explore a pattern scaling approach to projecting extreme precipitation levels under RCP4.5, and compare these results to those obtained from the model fitted directly to the CESM-ME RCP4.5 output. Finally, Section 2.4 provides a summary and discussion.

2.2 Methods

2.2.1 Output from initial condition ensembles

Our statistical model is fit using precipitation and temperature output from two CESM1 (CAM5) initial condition ensembles. The 30-member CESM-LE uses historical (natural and anthropogenic) forcings for the years 1920-2005, followed by the RCP8.5 forcing scenario for the years 2006-2100. The 15-member CESM-ME uses the same historical forcings for the years 1920-2005 (and thus matches the first 15 members of CESM-LE during the historical period), followed by the RCP4.5 forcing scenario for the years 2006-2080. Let $y_{8.5}^{(i)}(s,t,d)$ and $y_{4.5}^{(i)}(s,t,d)$ denote the daily precipitation amount (mm) for grid cell s, year t, and day d for ensemble member i from CESM-LE and CESM-ME, respectively. For each grid cell and each year, we retain

the annual maximum daily precipitation amounts $m_{8.5}^{(i)}(s,t) = \max_d y_{8.5}^{(i)}(s,t,d)$ and $m_{4.5}^{(i)}(s,t) = \max_d y_{4.5}^{(i)}(s,t,d)$. Treating the ensemble members as independent replicates yields an 'artificially large' data set; for each year at each grid cell, we have thirty and fifteen realizations of annual maxima for CESM-LE and CESM-ME, respectively. We model extreme precipitation at 687 grid cells which are in the contiguous United States.

To obtain the annual global mean temperature (°C) covariate for each ensemble member, we calculate area-weighted global monthly average near-surface temperatures, then take the mean of these monthly averages within each year. We denote the global mean temperature for ensemble member i at year t as $x_{8.5}^{(i)}(t)$ and $x_{4.5}^{(i)}(t)$ for CESM-LE and CESM-ME, respectively. The ensemble average over all thirty CESM-LE members is given by $\bar{x}_{8.5}(t) = \frac{1}{30}\sum_{i=1}^{30} x_{8.5}^{(i)}(t)$, and similarly the ensemble average over all fifteen CESM-ME members is given by $\bar{x}_{4.5}(t) = \frac{1}{15}\sum_{i=1}^{15} x_{4.5}^{(i)}(t)$. The global mean temperature distributions produced by the ensembles under RCPs 8.5 and 4.5 diverge by 2050 (Sanderson et al., 2018). The ensemble average global mean temperature increases from approximately 14.5°C in 2005 to approximately 17.9°C in 2080 under RCP8.5, compared to 16.4°C under RCP4.5.

2.2.2 GEV modeling conditional on global mean temperature

Our statistical model is based on the GEV distribution (1.5) as it is the limiting distribution of sample maxima of stationary sequences of random variables which meet mild mixing conditions (Leadbetter, 1974). For the moment, we use generic notation as our model assumptions are the same for both the RCP8.5 and RCP4.5 projections. Let $M(s,t)$ be the random variable representing the annual maximum daily precipitation amount for grid cell s and year t. To model how the distribution of annual maximum precipitation changes with the climate, we assume

$$P(M(s,t) \leq y) = G_{s,x(t)}(y) = \exp\left[-\left(1+\xi(s)\frac{y-\mu(s,x(t))}{\sigma(s,x(t))}\right)^{-1/\xi(s)}\right] \tag{2.1}$$

defined on $\{y : 1+\xi(s)\frac{y-\mu(s,x(t))}{\sigma(s,x(t))} > 0\}$. The case of $\xi(s) = 0$ is interpreted as the limit of (2.1) as $\xi(s) \to 0$.

We allow the GEV location and scale parameters μ and $\sigma > 0$ to vary both with grid cell s and global mean temperature $x(t)$. Preliminary investigations provided little evidence for significant climate driven changes in the shape parameter ξ, thus we assume that the shape parameter ξ only varies by grid cell and does not change with climate. This is consistent with some previous studies (e.g., Brown et al., 2014; Zhang et al., 2004) but in contrast to others (e.g., Hanel and Buishand, 2011; Kharin and Zwiers, 2005). Further, we assume

$$\mu(s, x(t)) = \mu_0(s) + \mu_1(s)(x(t) - x(2005)), \tag{2.2}$$

and

$$\phi(s, x(t)) := \log(\sigma(s, x(t))) = \phi_0(s) + \phi_1(s)(x(t) - x(2005)), \tag{2.3}$$

where $x(2005)$ is the global mean temperature at year 2005. Thus the intercept parameters $\mu_0(s)$ and $\phi_0(s)$ are defined as the value of the location parameter and log-scale parameter at year 2005. The slope parameters $\mu_1(s)$ and $\phi_1(s)$ can be interpreted as the change in the location parameter and log-scale parameter associated with a 1°C increase in global mean temperature.

For inference, we use CESM initial condition ensemble output to estimate model parameters via numerical maximum likelihood independently at each grid cell s. We let the superscripts "8.5" and "4.5" denote the parameters associated with the RCP8.5 scenario and the RCP4.5 scenario, respectively. Our likelihood for CESM-LE is

$$\mathscr{L}(\mu_0^{8.5}(s), \mu_1^{8.5}(s), \phi_0^{8.5}(s), \phi_1^{8.5}(s), \xi^{8.5}(s)) = \prod_{t=1920}^{2100} \prod_{i=1}^{30} g_{s,x(t)}^{8.5}\big(m_{8.5}^{(i)}(s,t), x_{8.5}^{(i)}(t)\big), \tag{2.4}$$

where $g_{s,x(t)}^{8.5}$ is the density associated with $G_{s,x(t)}^{8.5}$. With 181 years and 30 members, each grid cell has a total of 181×30=5430 data points to estimate the model parameters. Similarly, our likelihood for CESM-ME is

$$\mathscr{L}(\mu_0^{4.5}(s), \mu_1^{4.5}(s), \phi_0^{4.5}(s), \phi_1^{4.5}(s), \xi^{4.5}(s)) = \prod_{t=1920}^{2080} \prod_{i=1}^{15} g_{s,x(t)}^{4.5}\big(m_{4.5}^{(i)}(s,t), x_{4.5}^{(i)}(t)\big), \tag{2.5}$$

where $g^{4.5}_{s,x(t)}$ is the density associated with $G^{4.5}_{s,x(t)}$. With 161 years and 15 members, each grid cell has a total of 161×15=2415 data points to estimate the model parameters. Standard error estimates for the model parameters are obtained from the inverse of the numerically-estimated Hessian of the negative log-likelihood surface at the maximum likelihood estimates. The analysis was conducted using the `R` package `extRemes` (Gilleland and Katz, 2016).

In traditional extremes studies, the quantity of interest is often a return level. An r-year return level is simply the $1-1/r$ quantile of the distribution of the annual maximum, and under stationarity, r is the expected number of years between exceedances of the corresponding return level. The term 'return level' becomes ambiguous under non-stationarity (Cooley, 2012; Rootzén and Katz, 2013). We will focus on the $1-1/100$ quantile of the distribution of the annual maximum, which is obtained by setting (2.1) equal to 0.99. Because this level changes with year, we will eschew the term 'return level', and instead refer to this quantile as the '1% annual exceedance probability (AEP) level'. To obtain a single AEP level from the ensemble, $\bar{x}(t)$ is plugged in for $x(t)$ in (2.2) and (2.3). Note that we focus on the 1%-probability event only for convenience, as we could use (2.1) to calculate any quantile of our estimated distribution.

2.2.3 Pattern scaling

The rationale for pattern scaling is that because our model (2.1) requires only a covariate of global mean temperature, conceptually it could be used to estimate the distribution of the annual maximum daily precipitation for any global mean temperature of interest. The CESM-ME RCP4.5 runs give us an opportunity to evaluate the skill of a pattern scaling approach for extreme precipitation. We construct predictive GEV distributions $\hat{G}_{s,x_{4.5}(t)}$ by letting the GEV parameters be given by

$$\hat{\mu}(s, x_{4.5}(t)) = \hat{\mu}_0^{8.5}(s) + \hat{\mu}_1^{8.5}(s)(x_{4.5}(t) - x_{4.5}(2005)), \tag{2.6}$$

$$\hat{\sigma}(s, x_{4.5}(t)) = \exp\{\hat{\phi}_0^{8.5}(s) + \hat{\phi}_1^{8.5}(s)(x_{4.5}(t) - x_{4.5}(2005))\}, \tag{2.7}$$

and

$$\hat{\xi}(s) = \hat{\xi}^{8.5}(s). \tag{2.8}$$

That is, we use the parameters estimated from the CESM-LE RCP8.5 runs and plug in the global mean temperatures from the CESM-ME RCP4.5 runs. As the global mean temperatures produced by RCP4.5 are within the range of those of RCP8.5, we are not extrapolating the model to temperatures outside the range to which the model was fit. We test the pattern scaling predictive distributions by comparing the annual maxima $m_{4.5}^{(i)}(s,t)$ to $\hat{G}_{s,x_{4.5}(t)}$.

2.3 Results

2.3.1 Estimates for CESM-LE (historical/RCP8.5)

Parameter estimates and standard errors for $\mu_0^{8.5}(s), \mu_1^{8.5}(s), \phi_0^{8.5}(s), \phi_1^{8.5}(s)$, and $\xi^{8.5}(s)$ based on CESM-LE are obtained for all grid cells in the contiguous United States (see Figure A.1). The map of estimates for $\mu_1^{8.5}(s)$ (Figure A.1(c)) indicates that the location parameter is most sensitive to changes in global mean temperature in the east and southeast U.S., and also in northern California. The map of estimates for $\phi_1^{8.5}(s)$ (Figure A.1(g)) shows that the log-scale parameter is most sensitive to changes in global mean temperature in the southeast and in a region of the west between the coastal mountain ranges and the Rocky Mountains. P-values associated with one-sided hypothesis tests of $\mu_1^{8.5}(s) \leq 0$ by grid cell (Figure A.2(a)) show that this null hypothesis is rejected (for $\alpha = 0.05$) at nearly all grid cells outside a small region near the Mexican border, indicating that the GEV location parameter μ for extreme precipitation increases with global mean temperature. Similarly, the null hypothesis of $\phi_1^{8.5}(s) \leq 0$ is rejected at all but two grid cells in the U.S. (Figure A.2(b)), indicating that the log-scale parameter ϕ also increases with global mean temperature. Standard model diagnostic plots revealed no issues with model fit. We also assessed goodness of fit by applying the Anderson-Darling (AD) test to each grid cell, and found that fewer than 4% of the grid cells reject at the $\alpha = 0.05$ level (even without accounting for multiple testing). (See Figure A.3.) Given that the GEV model is only

asymptotically correct, that we fit a rather simplistic trend model, and that the 5430 points give the AD test high power to reject, we find this model fit to be quite suitable.

Although the GEV parameter estimates completely determine the fitted distribution of annual maximum daily precipitation, the parameters themselves have limited interpretability. Using the parameter estimates, we are able to produce estimates of the 1% AEP level for any particular year. Figure 2.2 shows 1% AEP level estimates for the years 2005 and 2080. These estimates range between 26.8-144.8 mm in 2005, compared to 32.6-164.6 mm in 2080. Standard errors calculated via the delta method (Coles, 2001, Section 2.6.4) range from 0.5-4.6 mm in 2080 (see Figure A.4). Between the years 2005 and 2080, the 1% AEP level is projected to increase for all grid cells under RCP8.5, with a U.S. median percentage increase of 17%. Changes in magnitude are most noticeable in California and along the Eastern seaboard, but percentage changes (Figure 2.2(c)) are also more than 30% in the Basin and Range area of the west where current levels are quite low. Another way to understand the change in extreme precipitation is to consider the annual exceedance probability for a given level of daily precipitation, say the 1% AEP level in 2005. Under RCP8.5, the 2005 1% AEP level corresponds to a higher annual exceedance probability in 2080 across the contiguous U.S. (Figure 2.2(d)). In some areas such as southwestern Idaho and southern Appalachia, a 1-in-100-chance event in 2005 is projected to become a 1-in-15 or higher chance event in 2080.

2.3.2 Comparison with CESM-ME (RCP4.5)

Parameter estimates and standard errors for $\mu_0^{4.5}(s), \mu_1^{4.5}(s), \phi_0^{4.5}(s), \phi_1^{4.5}(s)$, and $\xi^{4.5}(s)$ based on CESM-ME are obtained for all grid cells in the contiguous United States (see Figure A.5). Maps of estimates for $\mu_0^{4.5}(s), \phi_0^{4.5}(s)$, and $\xi^{4.5}(s)$ (Figure A.5(a), (e), and (i)) are very similar to those estimated from CESM-LE. On the other hand, estimates of the slope parameters $\mu_1^{4.5}(s)$ and $\phi_1^{4.5}(s)$ differ somewhat from those based on CESM-LE. Plots of these differences are given in Figure A.6(c) and (d). Null hypotheses of $\mu_1^{4.5}(s) \leq 0$ and $\phi_1^{4.5}(s) \leq 0$ are still rejected in over 90% of the grid cells (Figure A.6(a) and (b)), agreeing with our finding from CESM-LE that the

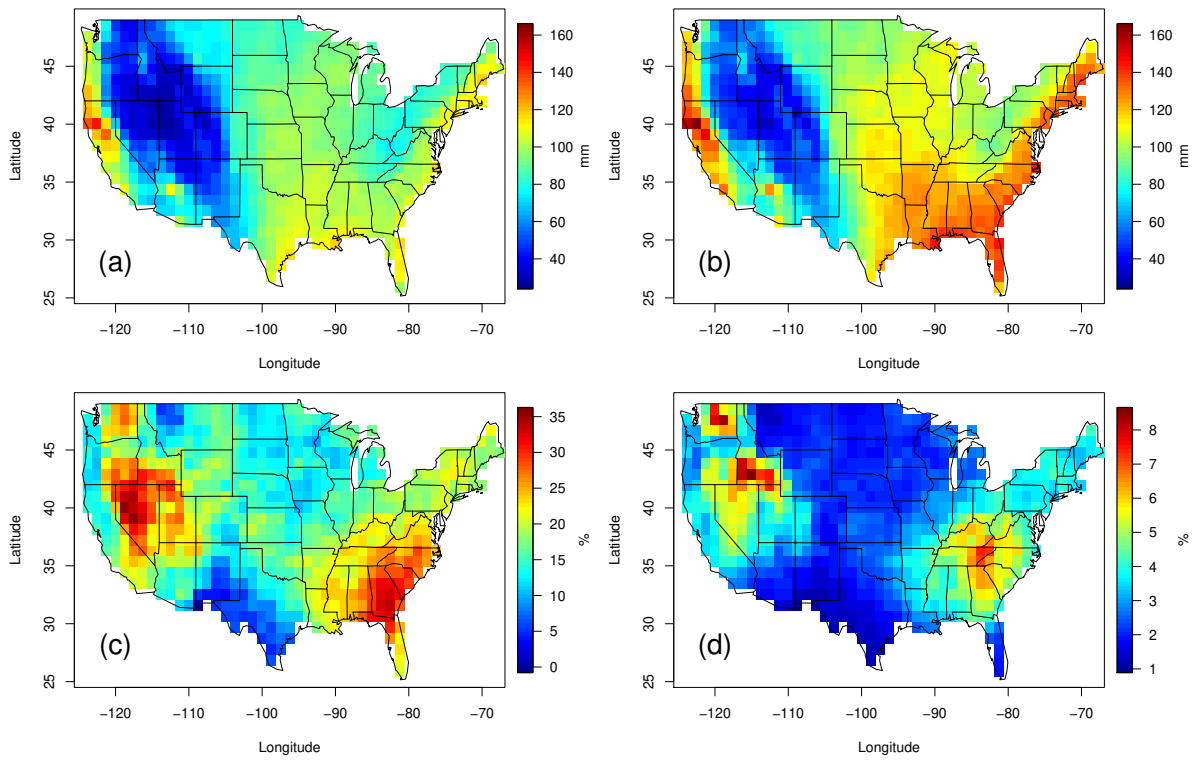

Figure 2.2: Point estimates for the 1% AEP level (in mm precipitation) for the years 2005 (a) and 2080 (b), based on CESM-LE simulations of annual maximum daily precipitation under RCP8.5. Percentage increase in the point estimate of 2080 from 2005 is shown in (c), while (d) maps the AEP (%) in 2080 corresponding to a 1% AEP level in 2005.

GEV location parameter and log-scale parameter tend to increase with global mean temperature. Standard model diagnostic plots revealed no issues with model fit. This was confirmed by applying the AD goodness of fit test to each grid cell (see Figure A.7), and only 2 of the grid cells rejected at the $\alpha = 0.05$ level.

Based on the GEV parameter estimates from CESM-ME, we can again produce estimates of the 1% AEP level for a given year. For the years 2006-2080, these 1% AEP level estimates correspond to projections under the RCP4.5 scenario. (For the historical period 1920-2005, note that the 1% AEP level estimates from CESM-ME do not perfectly match those from CESM-LE; however, they are not significantly different given the associated standard errors.) Spatial patterns of the 1% AEP level estimates under RCP4.5 are similar to those under RCP8.5, but with decreased magnitude (see Figure A.8(b)). Figure 2.3(a) shows the percentage change in 1% AEP level between 2005 and 2080 under RCP4.5, on the same scale as Figure 2.2(c). For 95% of the

grid cells in the U.S., the relative change from 2005 to 2080 is reduced under RCP4.5. Similarly, Figure A.8(c) shows the annual exceedance probability in 2080 under RCP4.5 for the 2005 1% AEP level, on the same scale as Figure 2.2(d).

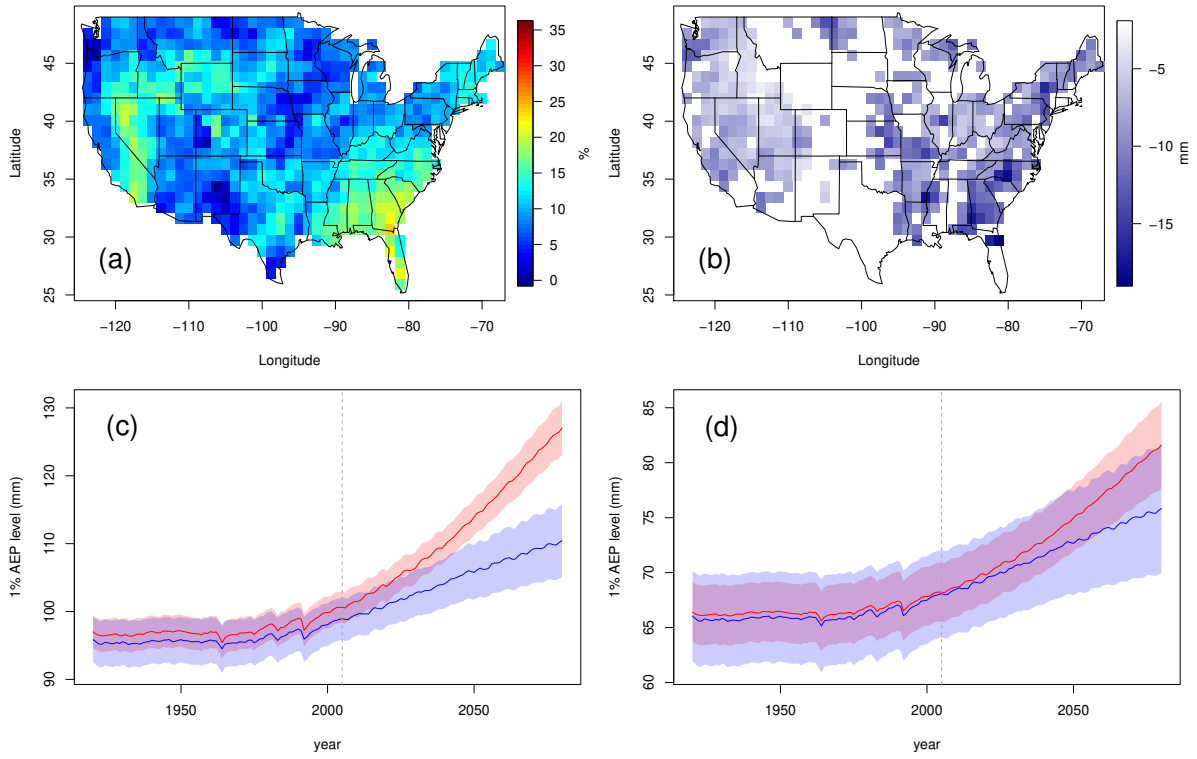

Figure 2.3: Projected U.S. 1% AEP levels under RCP4.5 (based on CESM-ME simulations of annual maximum daily precipitation) compared to RCP8.5. Percentage increase under RCP4.5 in the 1% AEP level of 2080 from 2005 is shown in (a), which can be compared to Figure 2.2(c). Reductions in magnitude of the projected 2080 1% AEP level under RCP4.5 relative to RCP8.5 are mapped in (b) for grid cells with a significant difference between scenarios. Example trajectories of the estimated 1% AEP level over time under RCP4.5 (blue) and RCP8.5 (red) are given for the grid cells closest to Charlotte, NC (c) and Fort Collins, CO (d), along with 95% confidence bands approximated via the delta method.

Figure 2.3(b) maps the difference in magnitude (mm) for the 2080 1% AEP level between the two scenarios, showing only the grid cells for which this difference is significant. One such grid cell is that corresponding to Charlotte, NC. As Figure 2.3(c) illustrates, Charlotte's RCP8.5 1% AEP level is projected to diverge from the RCP4.5 level by the mid-21st century. On the other hand, although Fort Collins, CO, shows a similar trend in 1% AEP level over time (see Figure 2.3(d)), the uncertainty around these estimates is too great to conclude significant dif-

ference. All significant differences in the 2080 1% AEP level represent reductions under RCP4.5. We note that for some areas such as the Basin and Range region of the west, the magnitude of this reduction is small, but corresponds to a high percentage reduction (see Figure A.8(d)). Across all grid cells in the contiguous U.S., we find a 7% median percentage reduction in the 2080 1% AEP level under RCP4.5 compared to RCP8.5, with reductions as large as 18% for some grid cells.

2.3.3 Results of pattern scaling

Given the parameter estimates from the CESM-LE RCP8.5 runs and the global mean temperatures from the CESM-ME RCP4.5 runs, predictive GEV distributions $\hat{G}_{s,x_{4.5}(t)}$ for RCP4.5 are produced by letting the GEV parameters be given by (2.6), (2.7) and (2.8). We evaluate the skill of this pattern scaling method by comparing the annual precipitation maxima $m_{4.5}^{(i)}(s,t)$ to the predictive distribution $\hat{G}_{s,x_{4.5}(t)}$ using AD tests at each grid cell. Results of the pattern scaling approach indicate reasonable fit, as for 84% of the grid cells in the contiguous U.S., the AD test fails to reject the null hypothesis that the annual maxima $m_{4.5}^{(i)}(s,t)$ come from the pattern-scaled GEV distribution. We again note that the sample size of 2415 gives the AD test prodigious power to reject the null hypothesis. Figure 2.4(a) maps locations with AD p-values less than $\alpha = 0.05$. For a grid cell in SE Colorado which had a particularly low p-value of < 0.001, Figure 2.4(b) compares the densities of the pattern-scaled GEV distribution and the GEV distribution fitted directly to the RCP4.5 maxima for a global mean temperature change of 1.5°C from 2005 (corresponding to the average global mean temperature change in 2080). Although the AD test soundly rejects these are the same distribution (with high power due to large sample size), we see that qualitative differences in the distributions are slight.

We also compare the 1% AEP levels projected by our pattern scaling method to those projected by the GEV model fitted directly to the CESM-ME RCP4.5 runs. For this same SE Colorado grid cell, the pattern-scaled point estimate of the 1% AEP level in 2080 is 92.6 mm with a delta-based 95% confidence interval of 89.4-95.7 mm, whereas fitting the RCP4.5 output di-

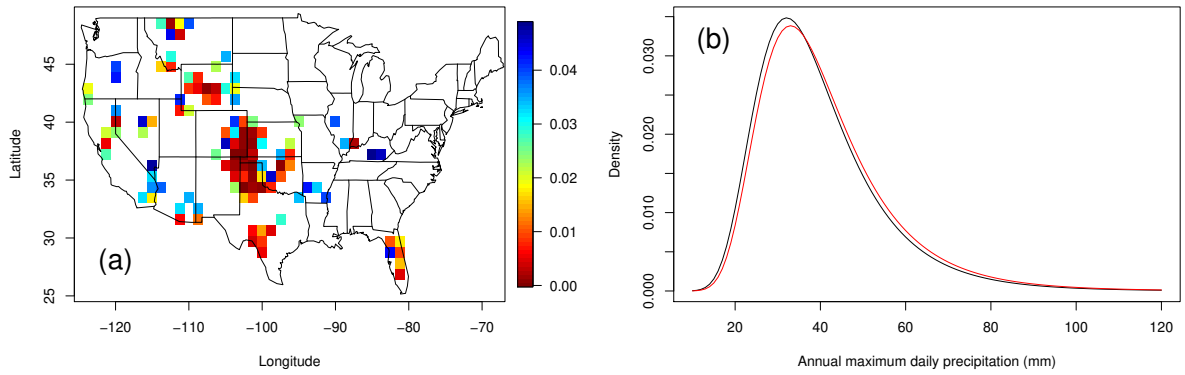

Figure 2.4: Evaluation of pattern scaling precipitation extremes for global mean temperatures under RCP4.5. P-values are shown in (a) for grid cells where the Anderson-Darling test rejects the null hypothesis that the annual precipitation maxima $m_{4.5}^{(i)}(s,t)$ come from the pattern-scaled GEV distribution $\hat{G}_{s,x_{4.5}(t)}$ at the $\alpha = 0.05$ level. For a grid cell in SE Colorado with an especially low p-value of < 0.001, (b) compares the densities of the pattern-scaled GEV distribution (black) and the GEV distribution fitted directly to the RCP4.5 precipitation maxima (red) for a global mean temperature change of 1.5°C from 2005.

rectly yielded a point estimate of 97.3 mm with confidence interval of 90.7-104 mm. These intervals overlap, indicating that even significant differences in distribution have relatively little consequence for the 1% AEP level. The differences between 1% AEP levels across the US are illustrated in Figure A.9. Compared to RCP8.5, results using the pattern-scaled distribution are similar to those produced by fitting directly to the RCP4.5 output: we still find an overall 7% median percentage reduction in the 2080 1% AEP level under RCP4.5 compared to RCP8.5 across all grid cells in the contiguous U.S.

2.3.4 Ensemble advantage for shape parameter estimation

For observational data and typical climate model output, each year yields only a single annual maximum. Here, CESM-LE provides 30 independent realizations of the annual maximum, $\{m^{(i)}(s,t)\}_{i=1}^{30}$, so we have the luxury of using an artificially large data set to estimate the GEV parameters. Of particular interest is the shape parameter $\xi(s)$ as this parameter is both influential for estimating high quantiles and particularly difficult to estimate. Figure 2.5 compares the point estimate for $\xi(s)$ using all 30 ensemble members versus an estimate from a single ensemble member. Qualitatively the maps show similar behavior, with higher values of ξ in the Great

Plains and Gulf Coast. However, due to the single ensemble member's limited information, the upper right panel shows much more spatial variability. This is also reflected in the standard errors (not shown) where the standard error for the single ensemble member's estimates were on average 5.97 times greater than those of the 30-member ensemble. We note this is quite close to $\sqrt{30} \approx 5.48$, which it should be theoretically.

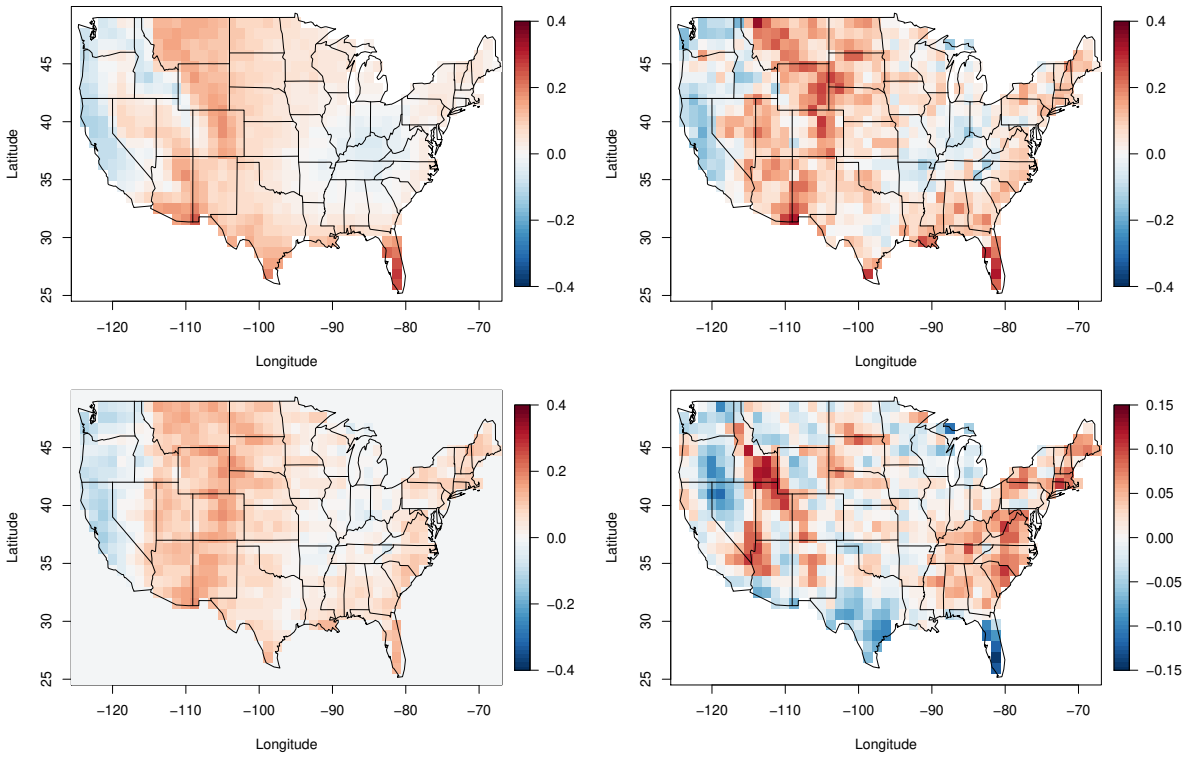

Figure 2.5: Comparison of $\hat{\xi}(s)$. Shape parameter estimates utilizing all 30 ensemble members (top left), and from a single ensemble member (top right). A spatially smoothed estimate of the single ensemble member is on the bottom left, and bottom right shows the "bias" from smoothing (smoothed estimate minus full ensemble estimate).

The spatial variability in $\hat{\xi}(s)$ from the single ensemble member likely arises from where this particular ensemble's most extreme events happened to occur, as Tye and Cooley (2015) have shown that estimates of ξ can be sensitive to the occurrence of an extreme event in the data record. To decrease spatial variability, it is common to "borrow strength" across spatial locations when estimating ξ and other GEV parameters and various methods have been suggested (Cooley et al., 2007; Hosking and Wallis, 1997; Sang and Gelfand, 2010). Borrowing strength has

the effect of spatially smoothing the estimates. The lower left panel of Figure 2.5 shows the result of applying a relatively simple smoothing method (Tye and Cooley, 2015) to the estimates from the single ensemble member. Qualitatively, the smoothed estimates seem similar to those of the full ensemble in the upper left panel. However, smoothing also has the effect of introducing bias. The lower right panel shows the difference between the estimates from the full ensemble and the spatially smoothed estimates. Compared to the estimates of the 30-member ensemble, the spatially smoothed estimate underestimates the shape parameter in Florida, Texas and Nevada, and overestimates this parameter in Idaho and much of the southeast US.

2.4 Discussion

This study takes advantage of two CESM initial condition ensembles to present a detailed analysis of U.S. precipitation extremes under two climate change scenarios, RCP8.5 and RCP4.5. Using the covariate of global mean temperature, we fit a non-stationary GEV model to annual precipitation maxima simulated from each ensemble at each grid cell. Our model shows that the 1% AEP level for precipitation increases with global mean temperature. Under RCP8.5, we find the projected median increase in the 1% AEP level is 17% between 2005 and 2080. Given that the change in global mean temperature is about 3.3°C over this time period, this is consistent with the 4-10%/°C range found by Kharin et al. (2013) for CMIP5 models, and is roughly in line with the Clausius-Clapeyron relationship that predicts an increase in moisture availability of about 6-7%/°C (Allen and Ingram, 2002). In terms of reducing impacts, we find that RCP4.5 reduces the 1% AEP level by 7% on average over the contiguous United States compared to levels predicted under RCP8.5, and projected reductions are as large as 18% in some grid cells.

We also investigate a pattern scaling approach in which we use our GEV model fit to annual precipitation maxima under the RCP8.5 scenario to create a predictive distribution of annual precipitation maxima for global mean temperatures under the RCP4.5 scenario. We find that these predictive distributions are well calibrated with the annual maxima produced by the CESM-ME RCP4.5 runs, and that estimates of tail quantities of interest such as the 1% AEP level

produced by the pattern-scaling approach are similar to those from directly modeling the GCM output. There are some slight differences, such as in the region just east of the Rocky Mountains, and Florida, where we surmise that the change in extreme precipitation is not fully described by the change in global mean temperature, or at least not by our statistical model's relatively simple linear functions in the GEV parameters. We expect scenarios driven mainly by a monotonic growth in greenhouse gases to be good candidates for the same type of approximations, while strongly mitigated scenarios where forcings plateau or decrease may present problems to the general pattern scaling approach (e.g., RCP2.6; Tebaldi and Arblaster, 2014). We would also caution against applying pattern scaling to a projection whose temperatures are out of the range of the fitted projection.

While the foci of this work have been to compare projected extreme precipitation under the two scenarios and to evaluate a pattern scaling approach for extremes, the Large Ensemble allows us a unique opportunity to investigate parameter uncertainty. Having multiple ensemble members allows us to more precisely estimate the GEV parameters than in the usual situation of a single realization or with observational data. In Section 2.3.4 we investigate the spatial behavior of the shape parameter ξ, which is a parameter of great interest in extremes and which is typically the most difficult GEV parameter to estimate. In the typical situation when one only has a single realization, we see that the estimates for ξ are highly variable in space. Spatial smoothing methods can achieve point estimates whose spatial characteristics resemble those we see with the ensemble, but we also find that spatial smoothing methods introduce bias into the estimates as expected.

There are several things to keep in mind with regards to this study and future impacts of extreme events. Like the other studies in the BRACE project, we address scenario differences accounting for the uncertainty that originates from different initial conditions, but do not address uncertainties due to different climate models' structural choices. The results here represent only one climate model's version of what the future climate could be, and are therefore necessarily affected by structural uncertainty (Knutti et al., 2010), i.e. the effects of a particu-

lar model's choices of what processes to represent or parameterize, its resolution, and its parameter settings. Further work is needed to explore the robustness of our findings when other climate models are used. Note, though, that recently Seneviratne et al. (2016) showed how linear relations between global average temperature and several aspects of climate extremes at regional levels apply across the whole CMIP5 ensemble, which leads us to believe that other models would be amenable to the same approach that we apply here to CESM. Another concern is that we analyze output with relatively low spatial resolution, thus assessing local impacts of these changes in extreme precipitation behavior would require downscaling. We also only consider annual maximum precipitation and do not consider seasonal effects. We summarize changes in extreme precipitation primarily in terms of the 1% AEP level; however, we point out that our methodology allows us to characterize changes in any tail quantity of interest. Although the precipitation AEP level by itself does not quantify risk say due to flooding, changes in AEP level are an important piece of assessing the changing risk of such events.

Chapter 3

Observed and Predicted Sensitivities of High and Extreme Surface Ozone to Meteorological Drivers in Three US Cities[4]

3.1 Introduction

Surface ozone (O_3) is one of the major air pollutants associated with adverse health effects. According to the US Environmental Protection Agency (EPA), current scientific evidence supports a causal relationship between short-term exposures to O_3 and respiratory health effects, and a likely to be causal association with total mortality (IHME, 2013). The O_3 health effects have been found to be non-linear, and may be especially detrimental at high levels of O_3 (Wilson et al., 2014). In addition, ambient air quality standards for “criteria" pollutants such as O_3 typically impose a penalty for exceeding a high concentration threshold. Thus for both air quality regulation and human health concerns, it is important to understand the conditions leading to the most extreme O_3 levels and to be able to reliably predict these extreme levels under present and future climate via atmospheric chemistry models.

Processes controlling O_3 concentrations are relatively well understood (Seinfeld et al., 1998). Surface O_3 is mostly a summertime pollutant produced by photochemical oxidation of volatile organic compounds (VOCs) by hydroxyl radical (OH) in the presence of nitrogen oxides (NO_x) and sunlight. Most efficient losses of surface O_3 include the removal by dry deposition uptake to vegetation, and its photolysis in the presence of water vapor which leads to the formation of OH. It is also well known that O_3 concentrations near the surface are strongly affected by meteorological parameters including (but not limited to) the boundary layer winds (mixing/dispersion), temperature which influences the emissions of biogenic precursors, and cloudiness which influences the radiation fluxes available for photolytic reactions.

[4]Fix, M. J., Cooley, D., Hodzic, A., Gilleland, E., Russell, B. T., Porter, W. C., & Pfister, G. G. (2018). Observed and predicted sensitivities of extreme surface ozone to meteorological drivers in three US cities. *Atmospheric Environment*, 176:292–300.

Accurate estimation of O_3 sensitivity to individual meteorological variables is challenging due to the complex interdependencies and processes at play. Research conducted across many settings, including both observational and model perturbation studies, suggests that elevated O_3 concentrations are most strongly linked with increases in temperature (Jacob and Winner, 2009; Pearce et al., 2011). Exceptionally high O_3 levels were observed in Europe in August 2003 associated with hot and dry heat-wave conditions (Vautard et al., 2007). In an analysis of covariance performed on observed daily O_3 maxima in Switzerland during the 1992-2002 period, Ordónez et al. (2005) found that temperature and global radiation accounted for most of the meteorological variability in summer O_3 concentration. In a model perturbation study over the eastern US during July 2001, Dawson et al. (2007) found that on average temperature had the largest (positive) effect on maximum daily 8-hour average (MDA8) O_3. Absolute humidity had a smaller but appreciable (negative) impact. Also focusing on the eastern US, Camalier et al. (2007) were able to explain up to 80% of the variability in observed MDA8 O_3 with a generalized linear model. They found regional variability in the prevailing meteorological parameters driving O_3 response, with temperature most dominant in the northeast US and relative humidity playing a more significant role in the southeast US. Transport distance and direction also had strong effects in some areas.

The studies referenced above focus on the average O_3 response. However, meteorological sensitivities at high quantiles of O_3 have been shown to differ from those of the overall median (Baur et al., 2004; Porter et al., 2015). In the present study we focus on high and extreme O_3 levels, thus requiring specialized tools such as quantile regression and extreme value analysis. Quantile regression is beginning to be recognized as a powerful tool in air pollution studies (Zhao et al., 2016). For instance, Otero et al. (2016) applied quantile regression to estimate the meteorological influence on the 0.95 quantile of MDA8 O_3 over Europe during 1998-2012. In summer months, they found that maximum temperature and southerly flow were selected as predictors in over 80% of the models, with relative humidity and surface solar radiation following closely behind. Porter et al. (2015) applied quantile regression to observed daily O_3 levels

across the US during 2004-2012, and found maximum temperature to be the dominant driver of 0.95 quantile MDA8 O_3 in the summer. Consistent with the analysis of Camalier et al. (2007), they also found a strong negative relationship of relative humidity with O_3 in many locations, especially in the southern US. For extremely high quantiles, quantile regression suffers from data scarcity and extreme value analysis is needed. Russell et al. (2016b) developed a method to optimize tail dependence between O_3 and a linear combination of meteorological drivers. Russell et al. (2016a) applied this method to a spatial study of extreme summer MDA8 O_3 in the southeast and mid-Atlantic region of the US, and similarly found that air temperature was more important in the northern portion of the region while low humidity was more influential in the southern portion of the region.

Atmospheric chemistry models are essential for making short-term predictions of air quality, as well as projections of future air quality under climate change. Reproducing observed sensitivities of pollutants to meteorology is needed for building confidence in such model projections, but evaluation of model performance is lacking for air quality at high and extreme levels. The goal of this chapter is to evaluate model skill in reproducing observed relationships between meteorology and O_3 extremes in the US, such as those explored in Porter et al. (2015) and Russell et al. (2016b). We utilize a set of high resolution, regional scale atmospheric chemistry model simulations by Pfister et al. (2014). Although our focus is on the relationship between high/extreme O_3 and meteorological predictors, it is also necessary to examine the marginal distributions of both response and predictor variables individually. Thus, as illustrated in Figure 3.1, our study framework includes three types of comparisons between simulated and observational data, comparing (1) the O_3 response variable, (2) the meteorological predictor variables, and (3) the sensitivities of high and extreme O_3 to meteorological predictors. The first two are comparisons of distributions, and for the O_3 response we largely focus on comparing the distributions' tails. The comparison of sensitivities is made using two methods: quantile regression and the tail dependence optimization method developed by Russell et al. (2016b). To our knowledge, this is the first study to apply these statistical methods to O_3 simulated from

an atmospheric chemistry model, as well as the first study to compare the meteorological sensitivities of high/extreme O_3 between simulated and observed O_3.

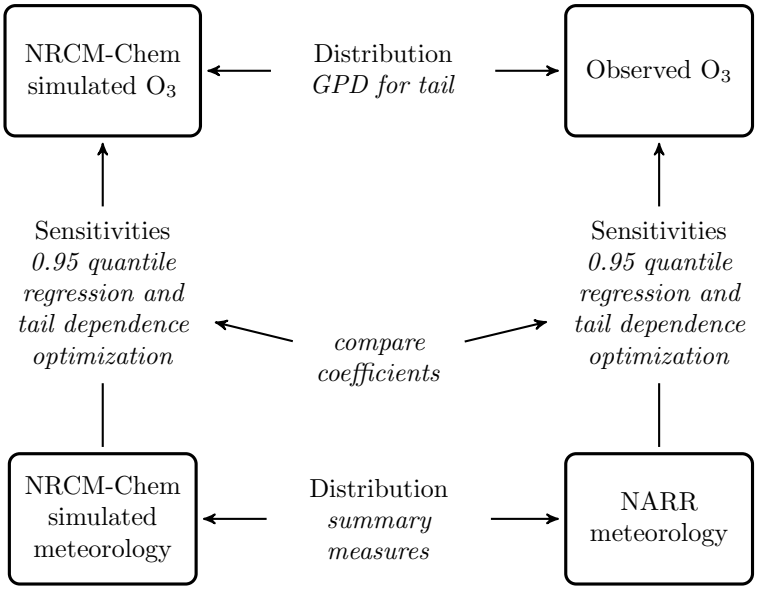

Figure 3.1: Illustration of the framework used in this study to compare simulated and observational data.

3.2 Inputs

3.2.1 Observations and NARR

We analyze surface O_3 measurements from the EPA's air quality system (AQS[5]) for summers (JJA) during the years 1996-2005. For consistency with the EPA's National Ambient Air Quality Standards (NAAQS), we extract MDA8 O_3 concentrations for our analysis. Because the statistical methodology is computationally costly, we focus on a case study of three AQS monitoring stations: station 13-121-0055 in Atlanta, station 48-201-0046 in Houston, and station 04-013-3002 in Phoenix. There were at most 5 days of data missing out of 920 days total at each of the stations. These three US cities have historically high levels of O_3, and fall within 8-hour O_3 nonattainment areas as designated by the EPA. Atlanta, Houston, and Phoenix represent a

[5]https://www.epa.gov/outdoor-air-quality-data

range of regional climates across the southern US, and belong to EPA regions 4, 6, and 9, respectively. However, we do not view this as a comprehensive study of these regions. These stations all reflect urban environments, however exploratory analysis found that a rural station in moderate proximity to Atlanta showed strong correlation to the urban Atlanta station, thus the sensitivities of high and extreme ozone to NARR meteorology would be very similar.

Following Porter et al. (2015) and Russell et al. (2016b), we obtain meteorological variables from the National Centers for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR) product (Mesinger et al., 2006), which combines model and assimilated observational datasets. NARR is a gridded product with a spatial resolution of 32 km and 8 output fields per day (representing 3-hour means). There is a spatial mismatch between the point-located O_3 observations and the gridded NARR meteorology. We use output from the NARR grid cell whose midpoint is closest to the AQS monitoring station of interest. NARR has been used previously to examine meteorological drivers of observed air pollution (e.g. Tai et al., 2010). In addition, the NARR output is complete and does not need additional quality control.

3.2.2 NRCM-Chem simulations

We utilize a set of climate simulations conducted by Pfister et al. (2014) using the nested regional climate model with chemistry (NRCM-Chem), which is based on the regional Weather Research and Forecasting model with chemistry (WRF-Chem, version 3.3). WRF-Chem is a fully coupled chemical transport model (Grell et al., 2005), which was run at a high spatial resolution of 12 km providing hourly outputs for the variables that we consider. We extract MDA8 O_3 concentrations from the NRCM-Chem gridpoint closest to each of the AQS stations. Daily meteorological variables (see Section 3.2.3) are also extracted from the NRCM-Chem simulations at these gridpoints.

We use the present time NRCM-Chem simulations for the 10 summers (1996-2005). Simulations are initialized each April, and we analyze output from June through August to allow for a 2 month spin-up phase. Meteorological initial conditions (IC) and boundary conditions

(BC) driving the NRCM-Chem simulations are provided by a NRCM 36 km domain simulation described in Done et al. (2015). Chemical IC and BC for trace gases and aerosols were taken from a global simulation with the Community Atmosphere Model with Chemistry (CAM-Chem V4) detailed in Lamarque et al. (2011). Each present time NRCM-Chem year uses the same chemical IC and BC based on the CAM-Chem output for the year 2000. More details about the simulations can be found in Pfister et al. (2014).

3.2.3 Selecting meteorological predictors

To compare the sensitivities to meteorology between observed and simulated O_3, we must choose meteorological predictor variables which are available both in NARR and NRCM-Chem output. Based on results from previous studies, we select five meteorological predictors of interest (see Table 3.1). These variables represent a subset of those found by Otero et al. (2016), Porter et al. (2015), and Russell et al. (2016b) to be key drivers of high or extreme observed summer O_3. To examine the relationship between meteorology and MDA8 O_3, which is a daily quantity, daily summary measures are chosen for each predictor variable. For consistency between NRCM-Chem and NARR output, which is available as 3-hour means, we first convert the NRCM-Chem output to 3-hour means before taking the daily maximum.

Table 3.1: Meteorological predictors and corresponding daily summary measures used in the analysis, for both NARR and NRCM-Chem outputs.

Meteorological predictor	**Abbreviation**	**Definition**
Air temperature at 2m	T	Daily maximum
Wind speed at 10m	WS	Daily mean
Relative humidity	RH	Daily mean
Height of the planetary boundary layer	HBL	Daily maximum
Downward shortwave radiation flux	DSR	Daily maximum

3.3 Statistical Methods

3.3.1 Marginal analysis of extreme O_3

In addition to using standard summary statistics to compare the distributions of MDA8 O_3 between observations and NRCM-Chem simulations, we employ extreme value theory to analyze the tails of these distributions. We use the generalized Pareto distribution (GPD) to model exceedances of a sufficiently high threshold u. The threshold exceedance approach is frequently applied because it offers greater efficiency of data usage over block-maxima approaches, and has been used previously to model the tail behavior of O_3 (e.g., Phalitnonkiat et al., 2016; Rieder et al., 2013). The GPD is parameterized by scale and shape parameters $\sigma_u > 0$ and ξ, and can be defined as in (1.6). When $\xi < 0$ there is an upper limit such that $u < x < u - \sigma_u/\xi$, i.e. the tail is bounded. The cases $\xi = 0$ and $\xi > 0$ correspond to light and heavy tails, respectively.

To maintain a consistent approach among our analyses, for each series we choose our threshold, u, such that approximately 5% of the O_3 values exceed it. Standard diagnostics such as the mean residual life plot (Coles, 2001) confirm that this threshold appears to be high enough that the limiting GPD is a good approximation for the exceedance distribution, while at the same time this threshold retains a reasonable number of exceedances for the analysis. As a result of emissions controls, concentrations of surface O_3 have been decreasing over much of the US in recent years (Lefohn et al., 2008). We see this downward trend in observed O_3 at the Atlanta and Houston stations (see Figure 3.2), and account for this non-stationarity by setting a linearly-varying threshold in time, u_y, via 0.95 quantile regression (Koenker and Bassett Jr, 1978). The quantile regression coefficient for year is significantly less than zero at Atlanta and Houston (point estimates and standard errors are given in Table 3.2). Because the NRCM-Chem simulations use anthropogenic emission inputs from the year 2000 for the entire time period, we do not observe the same downward trend as in the observations, and thus employ a constant threshold u for simulated O_3 which is the empirical 0.95 quantile over the entire series at a given location. Given the threshold estimate, GPD parameters are estimated by maximum likelihood,

and standard errors are obtained via standard likelihood-based procedures. These standard error estimates do not take into account threshold uncertainty.

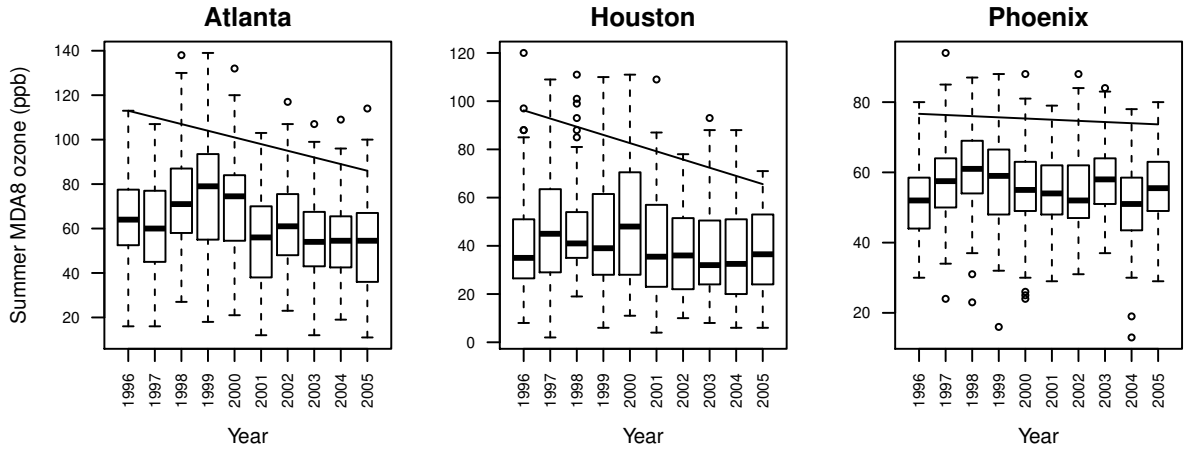

Figure 3.2: Distribution of observed summer (JJA) MDA8 O_3 by year at three AQS monitoring stations. The 0.95 quantile regression line represents the linearly-varying threshold in time used for the marginal analysis of extreme O_3. The trend in year is significantly less than zero for Atlanta and Houston stations, but not for the Phoenix station.

The usual likelihood formed by the product of GPD densities assumes independence of threshold excesses. However, initial examination of the O_3 series reveals short-term temporal dependence in the exceedances – if O_3 concentration exceeds the threshold today, it is more likely to exceed the threshold tomorrow compared to if it did not exceed today. Fitting the GPD to all exceedances using the usual likelihood in the presence of such serial correlation would result in underestimated standard errors. We avoid this issue by declustering the excesses prior to model fitting. We use the intervals method proposed by Ferro and Segers (2003) to estimate run length, and then apply runs declustering (Leadbetter et al., 1989) with clusters restricted to occur within the same year. Once a cluster is identified, it is replaced with the cluster maximum. The GPD is fit to the declustered series, with parameters computed via numerical maximum likelihood estimation. Sample sizes before and after declustering are given in Table 3.2. Analyses are done using the `extRemes` package (Gilleland and Katz, 2016) in `R` (R Core Team, 2015). Using the fitted GPD, we can estimate high quantiles of the O_3 distributions. In this study we

report estimates of the 0.99 quantile, with confidence intervals obtained by profile likelihood to account for asymmetry in the likelihood surface.

3.3.2 Relating high and extreme O_3 to meteorological drivers

We use two methods to examine the sensitivities of high or extreme O_3 to the selected meteorological predictors: quantile regression and the tail dependence optimization method developed by Russell et al. (2016b). The two frameworks are described below. In both approaches, we fit statistical models relating (a) NRCM-Chem O_3 to NRCM-Chem meteorology and (b) observed O_3 to NARR meteorology. The fitted models include the five meteorological predictor variables found in Table 3.1 for both NRCM-Chem and NARR, allowing us to compare the estimated model coefficients which represent the sensitivities of the O_3 response to the meteorological drivers.

Quantile regression

In contrast to ordinary least squares regression, which models the linear relationship between one or more predictor variables $\mathbf{X}$ and the conditional *mean* of a response variable Y given $\mathbf{X} = \mathbf{x}$, quantile regression (Koenker and Bassett Jr, 1978) extends the regression model to conditional *quantiles* of the response Y given $\mathbf{X} = \mathbf{x}$. For $\tau \in (0,1)$, we define the *τth conditional quantile* of Y by

$$Q_{Y|\mathbf{X}}(\tau) = \inf\{y : Pr(Y \leq y|\mathbf{X} = \mathbf{x}) \geq \tau\}.$$

Our model assumes a linear relationship between the conditional quantile and the p predictors, i.e.

$$Q_{Y|\mathbf{X}}(\tau) = \mathbf{x}^T \boldsymbol{\alpha}(\tau) = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_p x_p. \tag{3.1}$$

The coefficients $\boldsymbol{\alpha}(\tau) = (\alpha_0, \alpha_1, \ldots, \alpha_p)$ of the linear conditional quantile function can be estimated by solving

$$\hat{\boldsymbol{\alpha}}(\tau) = \underset{\boldsymbol{\alpha} \in \mathbb{R}^p}{\arg\min} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^T \boldsymbol{\alpha}), \tag{3.2}$$

where $\rho_\tau(\cdot)$ represents the check function $\rho_\tau(u) = u(\tau - \mathbb{I}(u < 0))$ and $\mathbb{I}(\cdot)$ is the indicator function. In this study, $\tau = 0.95$ because we are interested in a high level of ozone.

Because the distributions of O_3 and meteorology may differ between NRCM-Chem simulations and observational products (see Sections 3.4.1 and 3.4.2), we center and scale both the O_3 response and each of the meteorological predictors so as to be able to compare the estimated coefficients between the two analyses. We also center the year variable so that the intercept is at the year 2000. We implement quantile regression using the `quantreg` package (Koenker, 2016) in `R`, with standard errors obtained by paired bootstrap (Givens and Hoeting, 2013, §9.2.3). Specifically, we fit a model for the conditional 0.95 quantile with all five meteorological main effects. Note that these quantile regression models are different from the quantile regression used for threshold estimation in Section 3.3.1, which included only year as a predictor to account for non-stationarity in the tail.

Tail dependence optimization

Quantile regression is not well-suited to modeling extremely high quantiles for which there may be inadequate data above the desired quantile for quantile regression estimation methods to succeed. To understand the meteorological variables associated with the highest O_3 levels, we apply the method developed by Russell et al. (2016b) to find the linear combination of a set of meteorological predictors which has the strongest tail dependence with the O_3 response. This approach is based on the multivariate (in this case, bivariate) regular variation framework for multivariate extremes (see Section 1.4 for background).

The procedure of Russell et al. (2016b) aims to optimize a metric of tail dependence γ, where $\gamma = 0$ corresponds to perfect asymptotic dependence, while $\gamma = 1$ corresponds to asymptotic independence. Because the regular variation framework requires heavy-tailed marginals, the procedure requires transformation of both the response and predictor functional. Let Y_t be the random variable representing the response at time t, and let $X_{t,i}$ be value of the ith predictor at time t, for $i = 1, \ldots, k$. First we transform the response to be approximately unit Fréchet by letting $Y_t^{**} = G^{-1}[\hat{F}_Y(Y_t)]$ where G is the unit Fréchet distribution function and $\hat{F}_Y$ is an

estimate of the marginal distribution of Y_t. Next we apply a two-step transformation procedure to the predictors. In the first step, each predictor is transformed to the standard Gaussian scale using $X^*_{t,i} = \Phi^{-1}[\hat{F}_{X_i}(X_{t,i})]$ where Φ is the standard Gaussian distribution function. We consider linear combinations of the form $\boldsymbol{X}_{\boldsymbol{t}}^{*'}\boldsymbol{\beta} = \beta_1 X^*_{t,1} + \cdots + \beta_k X^*_{t,k}$, where $\mathbf{X}^*_{\mathbf{t}} = (X^*_{t,1}, \ldots, X^*_{t,k})$. For identifiability purposes, $\boldsymbol{\beta}$ is constrained such that $\boldsymbol{\beta}'\text{Cov}(\mathbf{X}^*_{\mathbf{t}})\boldsymbol{\beta} = 1$, and then $\boldsymbol{X}_{\boldsymbol{t}}^{*'}\boldsymbol{\beta}$ is assumed to be approximately standard Gaussian. In the second step, this linear combination is transformed back to be approximately unit Fréchet using $X^{**}_t(\boldsymbol{\beta}) = G^{-1}[\Phi(\boldsymbol{X}_{\boldsymbol{t}}^{*'}\boldsymbol{\beta})]$.

Our modeling framework assumes the random vector $(X^{**}_t(\boldsymbol{\beta}), Y^{**}_t)$ is bivariate regularly varying, and we seek the vector of coefficients $\tilde{\boldsymbol{\beta}}$ whose linear combination has the highest degree of tail dependence with the response. We find

$$\tilde{\boldsymbol{\beta}} = \underset{\{\boldsymbol{\beta}\in\mathbb{R}^k : \boldsymbol{\beta}'\text{Cov}(\mathbf{X}^*_{\mathbf{t}})\boldsymbol{\beta}=1\}}{\arg\min} \hat{\gamma}(\boldsymbol{\beta}),$$

where the estimator

$$\hat{\gamma}(\boldsymbol{\beta}) = \frac{\sum_{t=1}^{n} \delta(x^{**}_t(\boldsymbol{\beta}) + y^{**}_t)\frac{|x^{**}_t(\beta) - y^{**}_t|}{x^{**}_t(\beta) + y^{**}_t}}{\sum_{t=1}^{n} \delta(x^{**}_t(\boldsymbol{\beta}) + y^{**}_t)}, \tag{3.3}$$

and $\delta : \mathbb{R}_+ \to [0,1]$ is a non-decreasing weighting function. More details can be found in Russell et al. (2016b). Russell (2015) found that tail dependence optimization outperformed regression approaches, including quantile regression and logistic regression, as well as other extreme value approaches in terms of concordance in the upper tail. One disadvantage, however, is the large uncertainty in parameter estimates inherent to this and other extreme value methods.

We obtain 95% confidence intervals for parameter estimates using paired bootstrap and the percentile method (Givens and Hoeting, 2013). Model comparison can be achieved via cross-validation. Specifically, we use 10-fold cross-validation, in which the data is partitioned into 10 subsets. For each fold, the optimization is done on the training set (90% of the data) and $\hat{\gamma}$ is calculated for the test set (the remaining 10% of the data). The cross-validation score $\hat{\gamma}_{CV}$ is then the average over all 10 test sets.

3.4 Results

3.4.1 Comparing tails of O_3 response

Having implemented the procedure described in Section 3.3.1, Figure 3.3 compares the distribution of summer MDA8 O_3 between observations and NRCM-Chem simulations at our three study locations. In each panel between boxplots, the estimated 0.99 quantile for the year 2000 is shown with the corresponding 95% profile likelihood confidence interval. The 0.99 quantile roughly corresponds to the annual 4th highest MDA8, which forms the basis of the NAAQS for O_3. These extreme quantile estimates are made using the GPD fit to each series. The fitted GPD parameters are given in Table 3.2. In Atlanta and Phoenix, we see relatively good correspondence between observations and NRCM-Chem simulations, and 0.99 quantile estimates are not significantly different. In Houston, there is a noticeable difference in the upper tail, and the 0.99 quantile estimate is significantly lower for simulated O_3. This result is consistent with the tendency of regional air quality models to underpredict the high O_3 events, as found by Im et al. (2015) for example.

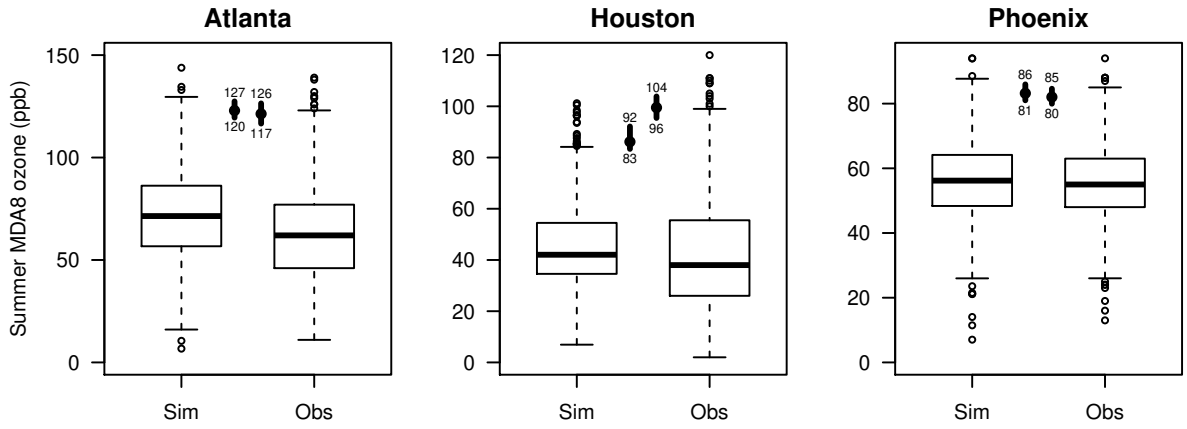

Figure 3.3: Boxplots of summer MDA8 O_3 during the years 1996-2005 from NRCM-Chem simulations (Sim) and AQS observations (Obs) at the three study locations. In each panel between the boxplots is the 0.99 quantile for the year 2000 estimated by fitting a GPD to threshold exceedances of simulations (left) and observations (right). Upper and lower limits are given for the corresponding 95% profile likelihood confidence intervals of each quantile estimate.

Table 3.2: GPD parameter estimates (standard errors in parentheses) for simulated (Sim) and observed (Obs) summer MDA8 O_3 at the three study locations. σ_u is the scale parameter, ξ is the shape parameter, u is the threshold, and n_{exc} is the number of exceedances after declustering (before declustering in parentheses). For Sim, u is set to the empirical 0.95 quantile. For Obs, u_y is a linearly-varying threshold in time with 0.95 quantile regression coefficients α_0 and α_1, where the intercept α_0 represents the threshold for the year 2000. Standard errors for threshold parameters are obtained via bootstrapping.

		Atlanta	**Houston**	**Phoenix**
σ_u	**Sim**	11.88 (2.50)	7.84 (2.56)	6.36 (1.70)
	Obs	23.91 (5.20)	17.09 (3.89)	6.03 (1.62)
ξ	**Sim**	-0.26 (0.13)	0.00 (0.30)	-0.19 (0.20)
	Obs	-0.66 (0.17)	-0.52 (0.18)	-0.20 (0.21)
u	**Sim**	109.74 (1.80)	76.62 (1.07)	76.11 (0.72)
	Obs α_0	101.00 (1.74)	82.60 (1.75)	75.33 (0.81)
	Obs α_1	-3.00 (0.56)	-3.40 (0.61)	-0.33 (0.34)
n_{exc}	**Sim**	34 (46)	31 (46)	32 (46)
	Obs	32 (43)	37 (45)	32 (42)

3.4.2 Comparing meteorological predictors

As in Section 3.4.1, we compare the *distributions* of meteorological variables. We do this because the NRCM-Chem simulations are not driven by reanalysis. We find that the distributions of the selected meteorological predictors differ considerably between NRCM-Chem and NARR output. NRCM-Chem tends to underestimate daily maximum air temperature and daily mean relative humidity, and exhibits much larger variability in these predictors than seen in the NARR product (Figure 3.4 top, center). In Atlanta, for example, the summer median for relative humidity according to NRCM-Chem is 55% compared to 74% based on NARR. In Phoenix, the summer median for relative humidity is 16% in NRCM-Chem vs. 23% in NARR, however NRCM-Chem records a summer maximum of 79% daily mean relative humidity compared to NARR which has a maximum value of 52%. NRCM-Chem also tends to underestimate daily maximum height of the planetary boundary layer compared to the NARR product (Figure 3.4 bottom).

The discrepancy between the meteorology in NRCM-Chem and NARR is not explained by their difference in spatial resolution. We explored taking the average of the nine NRCM-Chem grid cells surrounding each location, to obtain 36 km resolution similar to NARR's 32 km reso-

lution. However, the NRCM-Chem simulations are so strongly correlated between neighboring grid cells that the results are extremely similar to what is shown in Figure 3.4.

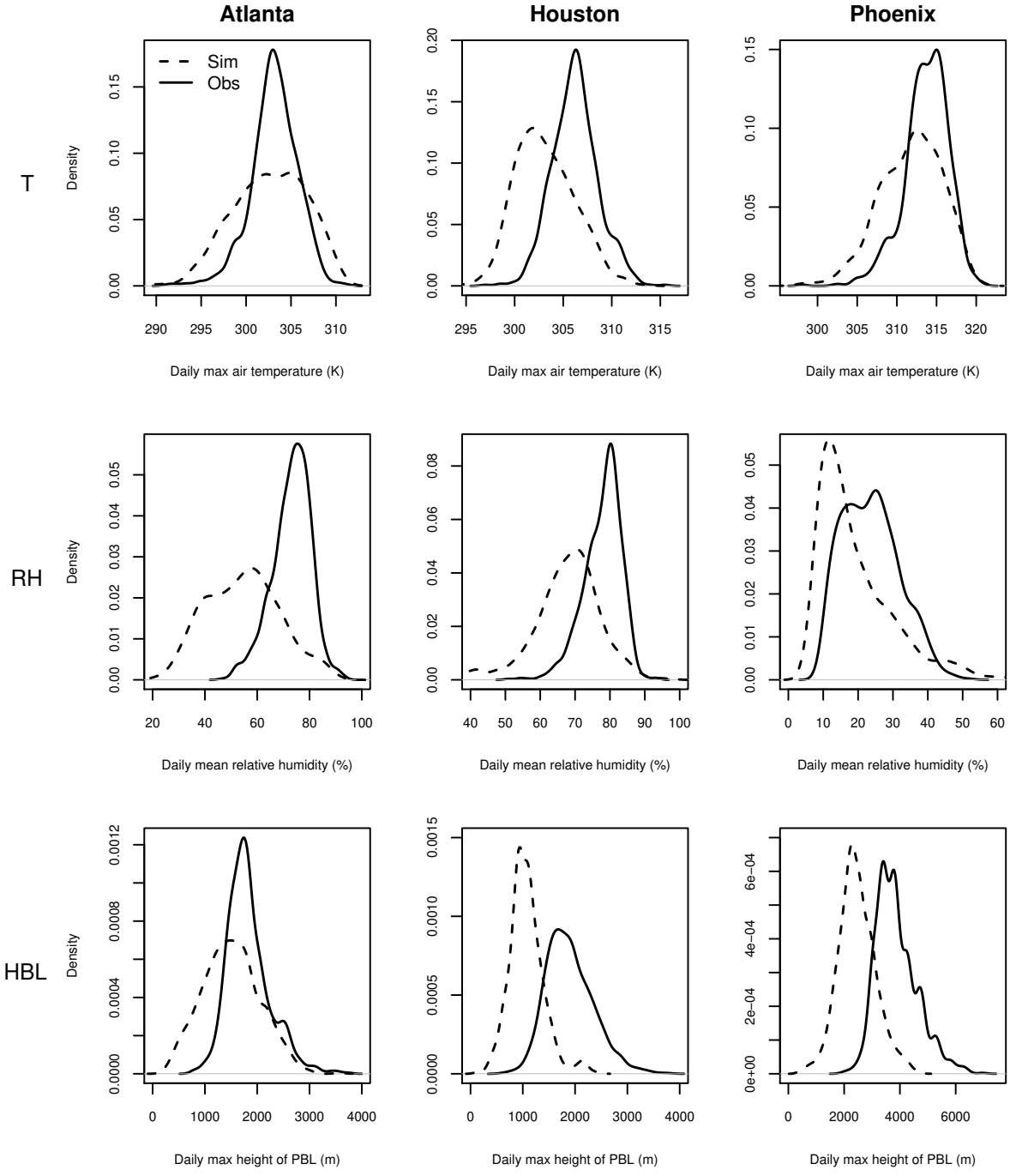

Figure 3.4: Kernel density plots of NRCM-Chem simulated (Sim, dashed lines) and NARR (Obs, solid lines) daily maximum air temperature at 2m (top row), daily mean relative humidity (center row), and daily maximum height of the planetary boundary layer (bottom row) for summers (JJA) during 1996-2005 at the three study locations.

3.4.3 Comparing relationships between O_3 and meteorology

Quantile regression

The left column of Figure 3.5 presents the estimated coefficients of the fitted 0.95 quantile regression models at each location for two analyses. The first analysis (in triangles) relates NRCM-Chem O_3 to NRCM-Chem meteorology, and the second (in circles) relates observed O_3 to NARR meteorology. Year is included as a predictor for observed O_3, to account for the downward trend observed in Figure 3.2. As expected, we find a significant negative year trend for observed O_3 in both Atlanta and Houston that is not present in the NRCM-Chem simulations (not shown). Coefficients are shown for the five meteorological predictors included as main effects in the full model. Some coefficients are not significantly different from zero, as indicated by the 95% confidence interval intersecting zero. We explored using backwards stepwise selection to remove nonsignificant predictors, however we found that in all cases the full model had the best (lowest) Akaike information criterion (AIC) value, so we report results for this model.

For both analyses across the three study locations, in most cases we see that daily mean wind speed (WS) and relative humidity (RH) have negative effects on the 0.95 quantile of MDA8 O_3. In Atlanta and Phoenix, daily maximum air temperature (T) has a positive effect. (The negative coefficient for T in Houston is evidence of multicollinearity, as a quantile regression model including only T results in a positive coefficient for T.) When the daily maximum height of the planetary boundary layer (HBL) is significant, it appears to have a positive effect. Daily maximum downward shortwave radiation flux (DSR) does not have a significant effect in any of the fitted models.

At each location, we find differences in the fitted full model for NRCM-Chem simulated vs. observed O_3. These differences are not consistent across study locations. In Atlanta, T has a significant (positive) effect on observed O_3, however it is borderline nonsignificant for NRCM-Chem simulations. WS has the strongest (negative) effect on simulated O_3, but not a significant effect on observed O_3. In Houston, in contrast, WS has a significantly more negative effect on observed O_3. Unlike at other locations, there is a similar (significant, positive) effect of HBL on

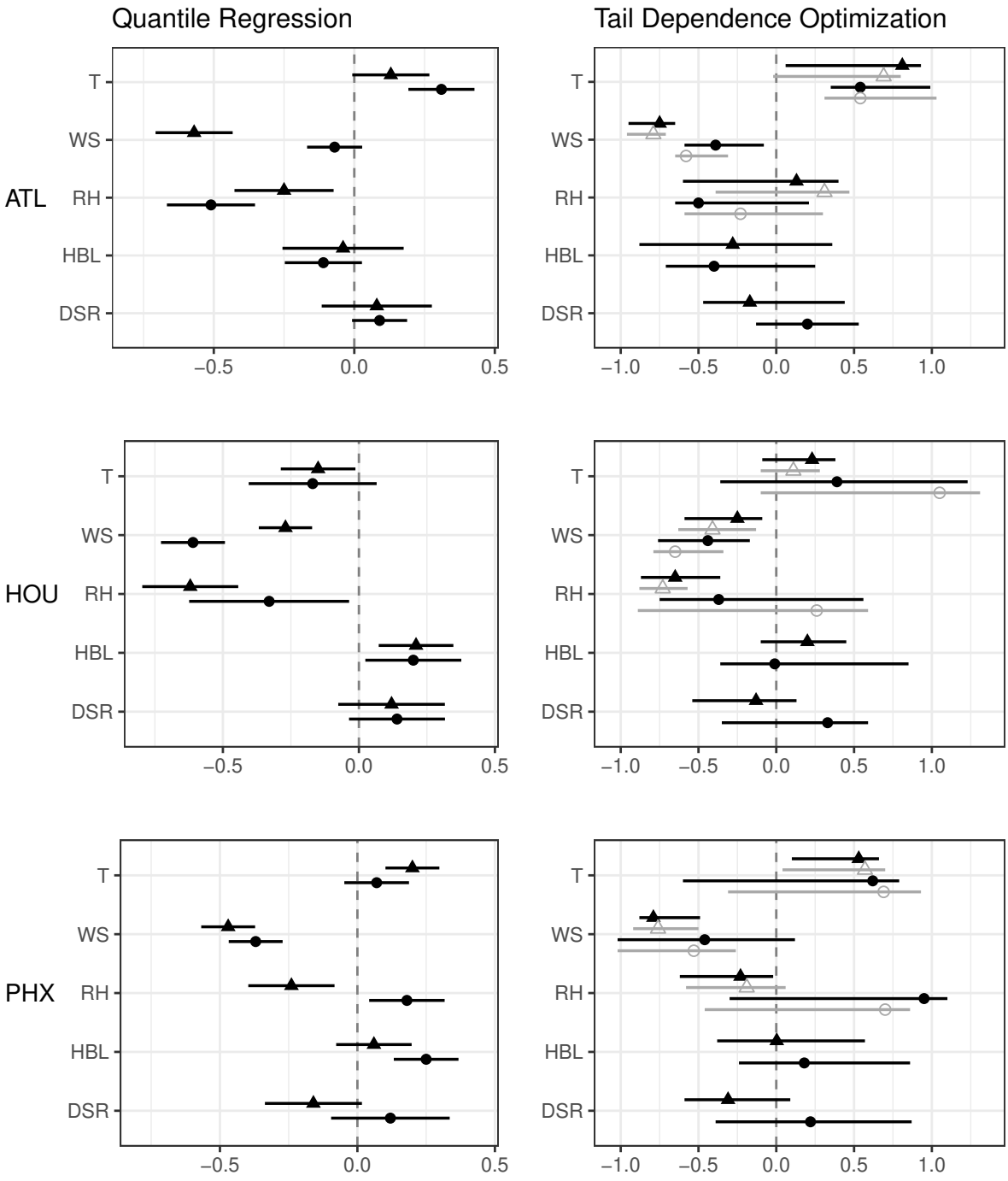

Figure 3.5: Parameter estimates with 95% confidence intervals for 0.95 quantile regression (left column) and tail dependence optimization (right column) at Atlanta (top), Houston (center), and Phoenix (bottom). Triangles indicate estimates from models relating NRCM-Chem O_3 to NRCM-Chem meteorology, while circles indicate estimates from models relating observed O_3 to NARR meteorology. Estimates in gray correspond to a reduced tail dependence model including only T, WS, and RH as predictors.

both simulations and observations in Houston. In Phoenix, the largest difference is in relative humidity: RH has a significant negative effect on simulated O_3, but a significant positive effect on observed O_3, conditional on the other predictors. T has a significant effect for simulations but not for observations, while HBL is significant for observations but not for simulations. In contrast to the other two locations, there is a similar negative effect of WS on both simulated and observed O_3 in Phoenix.

Tail dependence optimization

Unlike quantile regression, where we directly model the effect of year for observed O_3, the tail dependence optimization method of Russell et al. (2016b) requires stationary data. To account for non-stationarity in observed O_3, we transform the response variable by using 0.95 quantile regression to obtain the linearly-varying threshold in time as in Section 3.3.1. We then fit a gamma distribution to observations below and a GPD to observations above this year-varying 0.95 quantile, as explained in Russell et al. (2016b). This detrended response is then transformed to unit Fréchet as required by the method.

The right column of Figure 3.5 presents parameter estimates with bootstrap confidence intervals (based on 1000 bootstrap replicates) for tail dependence optimization applied to the two analyses at each of the three study locations. Similar to quantile regression, across all locations and analyses we see that T tends to have a positive relationship with extreme O_3, while WS tends to have a negative relationship. When RH is found to be significant, it has a negative relationship with extreme O_3. These three predictors have significant effects in at least some cases, while confidence intervals for HBL and DSR cover zero in all cases. Therefore, in addition to the full model with all five meteorological variables, we also fit a model with only T, WS, and RH as predictors (see Figure 3.5 results in gray). In all cases this improves (lowers) the cross-validation score $\hat{\gamma}_{CV}$.

Parameters obtained by tail dependence optimization are less straightforward to interpret than those obtained by quantile regression. For a given model fit, we can compare relative magnitudes and signs of the estimated parameters. Some differences are evident between model fits

for simulated vs. observed extreme O_3. In Atlanta, as was the case for quantile regression, there appears to be a more negative effect of WS on simulated than observed O_3. In Houston, RH has the strongest effect for simulated O_3, while T and WS appear to have stronger effects on observed O_3. In almost all cases, the point estimate for RH differs in sign between analyses for simulated and observed O_3, although the bootstrap confidence intervals are too wide to conclude any significant difference. In all three locations, DSR is estimated to have a negative effect on simulated O_3 and a positive effect on observed O_3, though again we are not able to conclude a significant difference.

3.5 Summary and Discussion

In this case study of summer surface O_3 in three US cities, we employ a set of high resolution NRCM-Chem simulations to make three types of comparisons between simulated and observational data, comparing (1) tails of the O_3 response, (2) distributions of meteorological predictor variables, and (3) sensitivities of high and extreme O_3 to meteorological predictors. This last comparison is made using both quantile regression, for the 0.95 quantile of O_3, and the tail dependence method of Russell et al. (2016b), which is used to investigate even higher O_3 extremes. To our knowledge, ours is the first study to apply quantile regression and tail dependence optimization to O_3 simulated from an atmospheric chemistry model. Additionally, this is the first study to compare the meteorological sensitivities of high/extreme O_3 between simulations and observational data.

Results from comparing the distributions of the O_3 response variable show that NCRM-Chem represents O_3 adequately overall, but underestimates extreme quantiles of O_3 in Houston. Results from comparing the distributions of meteorological predictors show clear discrepancies between the meteorology produced by NARR and that found in the NRCM-Chem simulations at all three locations. We recognize that NARR, being a reanalysis product, will not exactly match weather station data. There are further questions about NARR, for example the diagnostic parameter HBL is likely too high in NARR, e.g. a comparison with the MERRA re-

analysis found that NARR is more than 500m higher over the western US (McGrath-Spangler and Denning, 2012). However, we see surprising meteorology produced by NRCM-Chem, for instance the very low RH levels in Atlanta. These drier model conditions could increase the lifetime of O_3, as lower water vapor leads to reduced loss of O_3. The bias in the NRCM-Chem meteorology could be at least partially due to known SST errors in the model runs (Pfister et al., 2014).

Comparisons of the sensitivities of high and extreme O_3 to meteorological drivers also show clear differences between simulations and observational data. These differences are not consistent across the three study locations. For both high and extreme O_3 in Atlanta, simulations significantly overpredict the strength of the (negative) effect of WS. For the 0.95 quantile in Houston, we see the opposite, in that simulations significantly underpredict the effect of WS. In Phoenix, the quantile regression coefficient for RH is negative for simulated O_3, but positive for observed O_3. We also see a sign difference in the Phoenix point estimates for RH from the tail dependence method, however the confidence intervals are too large to conclude significance.

Differences in the sensitivities of observed vs. simulated O_3 could be driven by differences in how meteorological variables interact with O_3 formation and removal processes. For example, previous studies of average MDA8 O_3 over the eastern US have found that air quality models underpredict the strength of the effects of T and RH (Davis et al., 2011; Rasmussen et al., 2012). In Atlanta, we similarly find that our 0.95 quantile regression coefficients underestimate the effects of T and RH, though not significantly. Kavassalis and Murphy (2017) suggest that such a discrepancy may result from the lack of vapor pressure deficit-dependent dry deposition in the chemical transport model. Differences in Houston may be attributed to the difficulty in representing coastal dynamics such as recirculation patterns (e.g., Russo et al., 2016). Coastal areas often show a diurnal cycle in wind patterns, which in some cases can lead to either stagnancy, or the recirculation of polluted air away and then back to the original location. Poorly representing these coastal wind patterns, and how the observed recirculation or stagnancy affect O_3 levels, could explain the discrepancies in sensitivities for both WS and RH.

An important finding of this study is that the distribution of simulated O_3 matches observed O_3 quite well at two out of the three locations, despite rather large differences – and in some cases even sign reversal – in the meteorological sensitivities. It is possible that the differences in modeled and observed sensitivities are superficial, and that the underlying mechanisms leading to extreme O_3 formation and loss are still being represented, even if attribution is not identical between model and observation. It may be that the linearity assumption inherent to both quantile regression and tail dependence optimization methods is too simple to capture the complex relationship between O_3 and meteorology. In addition, models with multiple predictors face issues of collinearity in the predictors which increase the uncertainty. Future work could relax linearity assumptions or investigate interaction effects between predictor variables.

However, our finding raises a concern for modelers that the O_3 distributions are matching up well for the wrong reasons, due to parameter tuning within the model. Modeled chemistry related to O_3 formation, for example, has been steadily evolving and improving, but some of the improvements actually worsen agreement with observations because other processes are not included yet, or else have been misrepresented (see, e.g., Porter et al., 2017; Sherwen et al., 2016). If we seek modeling tools that can adapt to changing emissions and climatology, it is important to not only capture the current pollutant distribution, but also the relationships between the pollutant and its drivers. Our results suggest that, even in the locations where O_3 seems to be fairly well represented, NCRM-Chem may not be accurately representing the mechanisms behind O_3 formation or loss. Correctly describing current levels of O_3, while failing to capture the key mechanisms responsible, implies that our predictions will be unable to adapt to a changing climate. The poor agreement of meteorological sensitivities may evidence a need for mechanism improvement, either in terms of chemistry or physical dynamics.

We have proposed and applied a framework for comparing the meteorological sensitivities of high/extreme O_3 between observed data and simulated output. While this study analyzes only one atmospheric chemistry model, our methodology could be applied to any pairs of observational and simulated O_3-meteorology data. Despite having only 10 years of data which is

a very short record for an extreme value analysis, we find important differences between the observed and simulated O_3, the driving meteorology, and the sensitivities linking these. However, there are large uncertainties in parameter estimates, as evidenced by the wide confidence intervals in Figure 3.5. Such uncertainty is inherent to extremes approaches which focus on the most extreme values and thus use only a small subset of the data. In addition, this case study was a detailed analysis of only a few urban locations. Future work could consider aggregating results or conducting a spatial analysis over a larger region, as borrowing strength across locations could reduce uncertainties in parameter estimates (Russell et al., 2016a).

Chapter 4

Simultaneous Autoregressive Models for Spatial Extremes

4.1 Introduction

Natural hazards, such as floods and heatwaves, arise as extremes of physical processes that are inherently spatial. Spatial modeling of extremes can reduce uncertainty of risk estimates at a single location by borrowing strength across spatial locations, as well as capture spatial dependence between locations. When estimating joint risk over a region, e.g., for planning or insurance purposes, it is essential to account for spatial dependence in the tail. Classical spatial statistics is unsuited to extremal modeling because it is largely based on Gaussian distributions and its methods are best suited for describing mean behavior. It is well known that the tails of the multivariate Gaussian distribution lead to independent extremes (Sibuya, 1960), which can result in potentially catastrophic underestimation of joint risk. In recent years, there has been much interest in developing spatial models appropriate for extremes (see, e.g., Cooley et al., 2012a; Davison et al., 2012, for reviews).

In classical spatial statistics, two common approaches are geostatistical models and areal data models. Geostatistical models are process models designed for point-referenced data, where we assume we have sampled d of an infinite number of potential spatial locations. If isotropy is assumed, spatial correlation is modeled as a function of distance, and spatial prediction can be done using kriging. In areal data models, on the other hand, the entire domain is partitioned into d regions and values are associated with these regions. Areal data models, then, are not process models but rather multivariate models incorporating spatial (neighborhood) structure. As such, spatial prediction is not done for areal models. The simultaneous autoregressive (SAR; Whittle, 1954) and conditional autoregressive (CAR; Besag, 1974) are two well-known areal models.

Existing models for spatial extremes, such as max-stable process models (see Section 1.5), tend to be geostatistical in nature. Despite recent advances, current models are also compu-

tationally intensive and inference has generally been limited to a moderate number ($d \approx 30$) of spatial locations. Larger problems require methods such as composite likelihood (Padoan et al., 2010). Only very recently (Reich and Shaby, 2018) have areal models been considered for spatial extremes. Unlike the Reich and Shaby (2018) approach, which accounts for spatial dependence via latent clustering of neighboring regions, our approach is more directly analogous to classical areal models.

The goal of this work is to develop a simple spatial extremes model that is computationally feasible for high-dimensional areal data. We propose a new class of multivariate extremes models inspired by the simultaneous autoregressive (SAR) model in classical spatial statistics. Our extremal SAR models employ the framework of multivariate regular variation (MVRV; see Section 1.4) on the nonnegative orthant. Section 4.2 provides background from two disparate directions: Section 4.2.1 covers the classical SAR model and Section 4.2.2 reviews recent work on transformed-linear operations applied to regularly varying random vectors. On this foundation, Section 4.3 develops two versions of an extremal SAR model, the first closely connected to the classical SAR model, and the second an extension more amenable to extreme analysis. Section 4.4 discusses advantages and possible further extensions of the proposed models.

4.2 Background

4.2.1 Classical SAR model

First introduced by Whittle (1954), the simultaneous autoregressive (SAR) model is a simple model for areal data which can capture spatial dependence via a neighborhood structure. The classical SAR model can be specified by

$$\mathbf{Y} = S\mathbf{Y} + \boldsymbol{\epsilon}, \tag{4.1}$$

where S is a $d \times d$ matrix (not necessarily symmetric) whose diagonal elements are zero, and $\boldsymbol{\epsilon}$ is a $d \times 1$ vector of independent zero-mean errors. In general, it is assumed $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \Lambda)$ with Λ

diagonal. We can rewrite (4.1) as

$$(I - S)\mathbf{Y} = \boldsymbol{\epsilon}, \tag{4.2}$$

and, if $(I - S)$ is nonsingular, then

$$\mathbf{Y} = (I - S)^{-1}\boldsymbol{\epsilon}. \tag{4.3}$$

Note that we begin with an independent vector $\boldsymbol{\epsilon}$ and induce dependence in $\mathbf{Y}$, specifically $\mathbf{Y} \sim N\left[\mathbf{0}, (I-S)^{-1}\Lambda(I-S)^{-1^\top}\right]$. In the case of iid Gaussian errors, $\Lambda = \sigma^2 I$ and we have $\mathbf{Y} \sim N\left[\mathbf{0}, \sigma^2(I-S)^{-1}(I-S)^{-1^\top}\right]$.

To include spatial information in S, the typical approach (e.g., Anselin, 1988; Cliff and Ord, 1973) is to let $S = \rho W$, where ρ is a single spatial dependence parameter and W is a $d \times d$ spatial proximity matrix that specifies the neighborhood structure. Many options for spatial proximity measures can be considered. In a simple case, W has entries that are 1 or 0 according to whether or not unit i and unit j are classified as neighbors (with $w_{ii} = 0$). Two common neighborhood definitions are the rook's case, in which neighbors share a common edge, and the queen's case, in which neighbors share a common edge or a common vertex (see Figure 4.1 for an illustration on a regular grid). More complex definitions of W which extend non-zero values further can induce longer-range dependence.

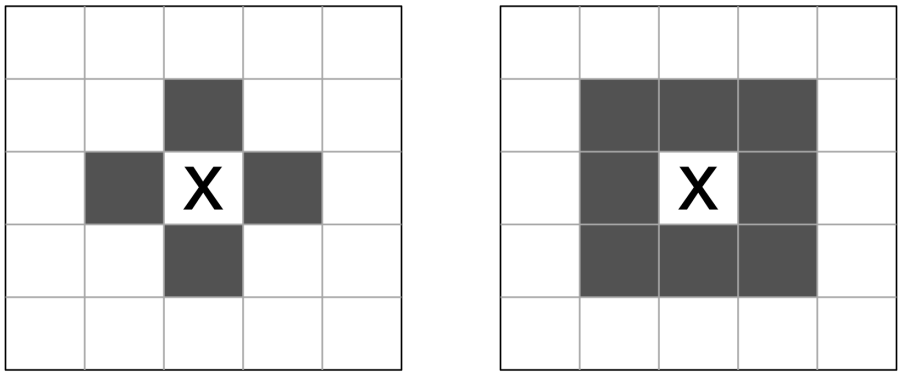

Figure 4.1: Two common neighborhood definitions for the classical SAR model: the rook's case (left) and the queen's case (right). X marks the grid cell under consideration and the shaded cells are its neighbors.

To ensure invertibility of $(I - \rho W)$, ρ is required to satisfy $\rho\lambda_i \neq 1$ for $i = 1,\ldots,d$, where $\lambda_1 < \lambda_2 < \cdots < \lambda_d$ are the ordered eigenvalues of W (Kelejian and Robinson, 1995). In practice, ρ is almost always restricted to the interval $(\lambda_1^{-1}, \lambda_d^{-1})$, or $(0, \lambda_d^{-1})$ when negative spatial autocorrelation is not of practical interest (Haining, 1990; Song and De Oliveira, 2012; Wall, 2004). In some cases, one may wish to adjust for the total number of neighbors to each areal region by using a row standardized matrix $\widetilde{W}$, where $\widetilde{w}_{ij} = w_{ij}/w_{i+}$ and $w_{i+} = \sum_{j=1}^{d} w_{ij}$. A sufficient condition for invertibility of $(I - \rho\widetilde{W})$ is $|\rho| < 1$ (Banerjee et al., 2014; Waller and Gotway, 2004).

4.2.2 Transformed-linear operations on regularly varying random vectors

The classical SAR model based on Gaussian errors is unsuitable for modeling extremal dependence. Instead, we work in the MVRV framework for multivariate threshold exceedances, as defined in Section 1.4.1. The classical SAR construction relies on linear operations. As we model in the nonnegative orthant to focus attention on upper tail, this naturally raises the question of how to define a vector space on the nonnegative orthant. To this end, we employ "transformed-linear" operations as defined by Cooley and Thibaud (2018).

Vector space via transformation

Let t be a bijection from $\mathbb{R}$ onto some open set $\mathbb{X}$, with t^{-1} its inverse. For vectors, we apply t and t^{-1} componentwise. Define vector addition in $\mathbb{X}^d$ by $\mathbf{x}_1 \oplus \mathbf{x}_2 = t\{t^{-1}(\mathbf{x}_1) + t^{-1}(\mathbf{x}_2)\}$. For any $a \in \mathbb{R}$, define scalar multiplication of a vector in $\mathbb{X}^d$ by $a \circ \mathbf{x} = t\{at^{-1}(\mathbf{x})\}$. The additive identity in $\mathbb{X}^d$ is defined by $\mathbf{0}_{\mathbb{X}^d} = t(\mathbf{0})$, and the additive inverse of any $\mathbf{x} \in \mathbb{X}^d$ is $-\mathbf{x} = t\{-t^{-1}(\mathbf{x})\}$. Then, Cooley and Thibaud (2018) show that $\mathbb{X}^d$ is a vector space under these operations, which we refer to as transformed-linear operations.

Notable for our purposes is the transformed-linear analogue of matrix multiplication. Let $A = (\mathbf{a}_1,\ldots,\mathbf{a}_q)$ be a $d \times q$ matrix of real numbers, and let $\mathbf{x} \in \mathbb{X}^q$. Then we can define $A \circ \mathbf{x} = \mathbf{a}_1 \circ x_1 \oplus \cdots \oplus \mathbf{a}_q \circ x_q = t\left\{At^{-1}(\mathbf{x})\right\} \in \mathbb{X}^d$. If $I_{d\times d}$ is the usual identity matrix, then $I \circ \mathbf{x} = t\left\{It^{-1}(\mathbf{x})\right\} = \mathbf{x}$. For a nonsingular matrix $B_{d\times d}$, we define the inverse operator B^{-1} to be a matrix such that $B^{-1} \circ (B \circ \mathbf{x}) = B \circ (B^{-1} \circ \mathbf{x}) = \mathbf{x}$. This coincides with the usual matrix inverse.

In the next subsection, we will apply these transformed-linear operations to regularly varying random vectors. We will use the specific transform $t : \mathbb{R} \to (0, \infty)$ defined by

$$t(v) = \log\{\exp(v) + 1\}, \tag{4.4}$$

with inverse $t^{-1}(x) = \log\{\exp(x) - 1\}$. This bijection has the important property that

$$\lim_{v \to \infty} \frac{t(v)}{v} = \lim_{x \to \infty} \frac{t^{-1}(x)}{x} = 1. \tag{4.5}$$

In other words, the transform and its inverse have a negligible effect on large values; they leave the upper tail alone (see Figure 4.2). Thus with this particular transform, regular variation will be preserved under transformed-linear operations. We extend the definition of t such that $t(-\infty) = 0$, $t^{-1}(0) = -\infty$, and $t(\infty) = t^{-1}(\infty) = \infty$. So $t : \bar{\mathbb{R}}^d \to \bar{\mathbb{X}}^d$, where $\bar{\mathbb{R}}^d = [-\infty, \infty]^d$ and $\bar{\mathbb{X}}^d = [0, \infty]^d$.

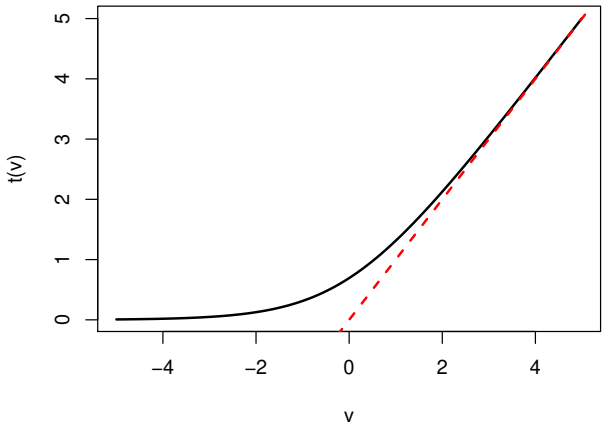

Figure 4.2: The specific transform t (black) given by (4.4) that is used for transformed-linear operations. The red dashed line is the 1:1 line.

Transformed-linear operations applied to multivariate regular varying random vectors

Suppose $\mathbf{X} \in RV_+^d(\alpha)$ as defined by (1.11) and (1.13). We say that $\mathbf{X}$ has limiting measure $\nu_{\mathbf{X}}$ and angular measure $H_{\mathbf{X}}$ when renormalized by $\{c_n\}$. Throughout this chapter, we assume

$\mathbf{X} \in RV_+^d(\alpha)$ verifies the lower tail condition

$$n\mathbb{P}\left\{X_i \leq \exp(-kc_n)\right\} \to 0, \qquad k > 0, \quad i = 1, \ldots, d, \tag{4.6}$$

as $n \to \infty$. Condition (4.6) is met by standard regular varying distributions such as the Pareto and the Fréchet. The following results from Cooley and Thibaud (2018) show that MVRV is preserved by transformed-linear operations with t given by (4.4). Furthermore, the limiting measures ν are known.

Proposition 4.2.1 (Cooley and Thibaud, 2018)**.** *Let* $\mathbf{X}_1 \in RV_+^d(\alpha)$ *such that* $n\mathbb{P}(c_n^{-1}\mathbf{X}_1 \in \cdot) \xrightarrow{v} \nu_{\mathbf{X}_1}(\cdot)$ *and* $\mathbf{X}_2 \in RV_+^d(\alpha)$ *such that* $n\mathbb{P}(c_n^{-1}\mathbf{X}_2 \in \cdot) \xrightarrow{v} \nu_{\mathbf{X}_2}(\cdot)$*, with* $\mathbf{X}_1$, $\mathbf{X}_2$ *independent. Then* $\mathbf{X}_1 \oplus \mathbf{X}_2 \in RV_+^d(\alpha)$*, and* $n\mathbb{P}\left\{c_n^{-1}(\mathbf{X}_1 \oplus \mathbf{X}_2) \in \cdot\right\} \xrightarrow{v} \nu_{\mathbf{X}_1}(\cdot) + \nu_{\mathbf{X}_2}(\cdot)$.

Proposition 4.2.2 (Cooley and Thibaud, 2018)**.** *Let* $\mathbf{X} \in RV_+^d(\alpha)$ *such that* $n\mathbb{P}(c_n^{-1}\mathbf{X} \in \cdot) \xrightarrow{v} \nu_{\mathbf{X}}(\cdot)$*. Then for* $a \in \mathbb{R}$*,* $n\mathbb{P}\left\{c_n^{-1}(a \circ X) \in \cdot\right\} \xrightarrow{v} a^\alpha \nu_{\mathbf{X}}(\cdot)$ *if* $a > 0$*, and* $n\mathbb{P}\left\{c_n^{-1}(a \circ X) \in \cdot\right\} \xrightarrow{v} 0$ *if* $a \leq 0$.

Prior to these results, traditional linear algebra operations had not previously been linked to methods for extremes. Max-linear approaches (e.g., Davis and Resnick, 1989; Strokorb and Schlather, 2015) had been explored, but these are not directly analogous because of the maximum operation. Propositions 4.2.1 and 4.2.1 pave the way for constructing new models for extremal dependence within the MVRV framework.

Construction method via the matrix A

In the non-extreme setting, one can construct a correlated Gaussian random vector by applying a matrix to a vector of iid standard Gaussians, i.e., by letting $\mathbf{Y} = A\mathbf{Z}$, where $A \in \mathbb{R}^{d \times q}$ and $\mathbf{Z} \sim N(\mathbf{0}, I_{q \times q})$. Propositions 4.2.1 and 4.2.2 imply a similar method for constructing regular varying random vectors by applying transformed-linear matrix multiplication to a vector of iid regularly varying random variables.

Corollary 4.2.1 (Cooley and Thibaud, 2018)**.** *Let* $A = (\mathbf{a}_1, \ldots, \mathbf{a}_q) \in \mathbb{R}^{d \times q}$ *with* $\max_{i=1,\ldots,d} a_{ij} > 0$ *for all* $j = 1, \ldots, q$*. Let* $\mathbf{Z} = (Z_1, \ldots, Z_q)^\top$ *be a vector of iid regular varying random variables*

with tail index α *such that* $n\mathbb{P}\left(Z_j > c_n z\right) \to z^{-\alpha}$ *and* $n\mathbb{P}\left\{Z_j \leq \exp(-kc_n)\right\} \to 0$ *for any* $k > 0$, $j = 1, \ldots, q$. *Then* $A \circ \mathbf{Z} \in RV_+^d(\alpha)$ *and when renormalized by* $\{c_n\}$ *has limiting angular measure*

$$H_{A\circ\mathbf{Z}}(\cdot) = \sum_{j=1}^{q} ||\mathbf{a}_j^{(0)}||^{\alpha} \delta_{\mathbf{a}_j^{(0)}/||\mathbf{a}_j^{(0)}||}(\cdot), \tag{4.7}$$

where δ *is the Dirac mass function, and* $\mathbf{a}_j^{(0)} = [\max(a_{1j}, 0), \ldots, \max(a_{dj}, 0)]^\top$, $j = 1, \ldots, q$.

Thus, by applying the matrix A to a vector of independent regularly varying random variables, we obtain a regularly varying random vector with extremal dependence structure described by $H_{A\circ\mathbf{Z}}$. The angular measure $H_{A\circ\mathbf{Z}}$ is discrete, with point masses corresponding to the normalized columns of A. In fact, it is the same limiting angular measure as for the max-linear construction $A \times_{\max} \mathbf{Z} = \left(\max_{j=1,\ldots,d} a_{1j} Z_j, \ldots, \max_{j=1,\ldots,d} a_{dj} Z_j\right)^\top$ (e.g., Fougères et al., 2013). Although the constructions $A \circ \mathbf{Z}$ and $A \times_{\max} \mathbf{Z}$ are similar, they differ in their realizations. Large realizations of the max-linear construction tend to have angular components which correspond exactly to the normalized columns of A. In contrast, large realizations of the transformed-linear construction have angular components near but not exactly at these discrete locations.

The construction method in Corollary 4.2.1 works for any $A \in \mathbb{R}^{d\times q}$ meeting $\max_{i=1,\ldots,d} a_{ij} > 0$ for all $j = 1, \ldots, q$. However, due to the zeroing operation, different matrices can result in the same limiting angular measure. Let $A^{(0)} = (\mathbf{a}_1^{(0)}, \ldots, \mathbf{a}_q^{(0)})$. If $A \neq A'$ but $A^{(0)} = A'^{(0)}$, then we still have $H_{A\circ\mathbf{Z}} = H_{A'\circ\mathbf{Z}}$. Thus it may be desirable to consider only matrices A with nonnegative entries, as we will do with the extremal SAR models. Further, Cooley and Thibaud (2018) show that the class of angular measures arising from the construction of Corollary 4.2.1 is dense in the class of possible angular measures, and this result only requires consideration of nonnegative matrices (see Proposition 4.2.3). The approximation to a continuous angular measure improves as the number of columns of A increases.

Proposition 4.2.3 (Cooley and Thibaud, 2018)**.** *Given any angular measure* H, *there exists a sequence of nonnegative matrices* $\{A_q\}, q = 1, 2, \ldots,$ *such that* $H_{A_q\circ\mathbf{Z}_q} \xrightarrow{w} H$.

Tail pairwise dependence matrix

The angular measure (4.7) completely characterizes the extremal dependence of $A \circ \mathbf{Z}$. We can also construct a matrix that summarizes this dependence via second-order properties of the angular measure. If we assume tail index $\alpha = 2$ and use the L_2 norm for the pseudo-polar decomposition, we can show that a suitably defined matrix has properties akin to those of a covariance matrix. Suppose $\mathbf{X} = (X_1, \ldots, X_d)^\top \in RV_+^d(2)$ such that $n\mathbb{P}\left(n^{-1/2}\mathbf{X} \in \cdot\right) \xrightarrow{v} \nu_{\mathbf{X}}(\cdot)$, where $\nu_{\mathbf{X}}(dr \times d\boldsymbol{\theta}) = 2r^{-3}drdH_{\mathbf{X}}(\boldsymbol{\theta})$ with $H_{\mathbf{X}}$ a Radon measure on the nonnegative L_2 unit sphere $\mathbb{S}_{+(2)}^{d-1} = \{\boldsymbol{\theta} \in \bar{\mathbb{X}}^d : ||\boldsymbol{\theta}||_2 = 1\}$ Then Cooley and Thibaud (2018) define the $d \times d$ *tail pairwise dependence matrix* (TPDM) as $\Sigma_{\mathbf{X}} = (\sigma_{\mathbf{X}ik})_{i,k=1,\ldots,d}$, where

$$\sigma_{\mathbf{X}ik} = \int_{\mathbb{S}_{+(2)}^{d-1}} \theta_i \theta_k dH_{\mathbf{X}}(\boldsymbol{\theta}) \tag{4.8}$$

is the extremal dependence measure defined in the bivariate case by Larsson and Resnick (2012).

The TPDM $\Sigma_{\mathbf{X}}$ has many properties that are analogous to those of a covariance matrix. Like a covariance matrix, it can be shown that $\Sigma_{\mathbf{X}}$ is positive semidefinite. The proof starts with (4.8) and proceeds in a similar manner to that for a covariance matrix. Also similar to a covariance matrix, the diagonal entries of $\Sigma_{\mathbf{X}}$ provide information about the scale of the components of $\mathbf{X}$. Specifically, $\sigma_{\mathbf{X}ii}$ is equal to the square of the marginal scale parameter of X_i. We can see this by considering the following. Given $\alpha = 2$, we have

$$\lim_{n\to\infty} n\mathbb{P}\left(n^{-1/2}X_i > x\right) = \int_{\mathbb{S}_{+(2)}^{d-1}} \int_{x/\theta_i}^{\infty} 2r^{-3}drdH_{\mathbf{X}}(\boldsymbol{\theta}) = x^{-2}\int_{\mathbb{S}_{+(2)}^{d-1}} \theta_i^2 dH_{\mathbf{X}}(\boldsymbol{\theta}) = x^{-2}\sigma_{\mathbf{X}ii} \tag{4.9}$$

by definition (4.8). Suppose X_i has scale 1, defined as $n\mathbb{P}\left(n^{-1/2}X_i > x\right) = x^{-2}$. This implies $n\mathbb{P}\left(n^{-1/2}cX_i > x\right) = c^2x^{-2}$, so we can see from (4.9) that $\sigma_{\mathbf{X}ii}$ is the square of the scale of X_i. Because we use the L_2 norm, $\sum_{i=1}^d \sigma_{\mathbf{X}ii} = \int_{\mathbb{S}_{+(2)}^{d-1}} dH(\boldsymbol{\theta})$, i.e., the diagonal elements of $\Sigma_{\mathbf{X}}$ sum to the total mass of the angular measure. Another property of the TPDM is that the case of $\sigma_{\mathbf{X}ik} = 0$ corresponds to asymptotic independence of the components X_i and X_k. This is because $\sigma_{\mathbf{X}ik} = 0$ if and only if $H_{\mathbf{X}}(\{\boldsymbol{\theta} \in \int_{\mathbb{S}_{+(2)}^{d-1}} : \theta_i > 0, \theta_k > 0\}) = 0$.

Regular varying random vectors constructed in the manner of Corollary 4.2.1 have a special form of the TPDM. Let $A = (\mathbf{a}_1, \ldots, \mathbf{a}_q) \in \mathbb{R}^{d \times q}$ and $\mathbf{Z} = (Z_1, \ldots, Z_q)^\top$ satisfy the conditions of Corollary 4.2.1 with $\alpha = 2$ and $c_n = n^{1/2}$. We further assume $q \geq d$. Then, using the form of the angular measure (4.7), we can see that the (i,k)th entry of $\Sigma_{A \circ \mathbf{Z}}$ is

$$\sigma_{A \circ \mathbf{Z} ik} = \int_{\mathbb{S}^{d-1}_{+(2)}} \theta_i \theta_k \, dH_{A \circ \mathbf{Z}}(\boldsymbol{\theta}) = \sum_{j=1}^{q} \left(\frac{a_{ij}^{(0)}}{||\mathbf{a}_j^{(0)}||_2} \right) \left(\frac{a_{kj}^{(0)}}{||\mathbf{a}_j^{(0)}||_2} \right) ||\mathbf{a}_j^{(0)}||_2^2 = \sum_{j=1}^{q} a_{ij}^{(0)} a_{kj}^{(0)}, \tag{4.10}$$

so $\Sigma_{A \circ \mathbf{Z}} = A^{(0)} A^{(0)\top}$. Note that this bears a striking resemblance to the traditional covariance matrix of $A\mathbf{Z}$ in the case of independent $\mathbf{Z}$ with unit variance, especially if A is nonnegative.

4.3 Extreme SAR Models

In this section we propose multivariate extreme analogues of the classical SAR model presented in Section 4.2.1. These extremal models apply the transformed-linear operations from Section 4.2.2 to multivariate regularly varying random vectors in the nonnegative orthant.

4.3.1 A SAR-inspired model for areal extremes

Formulation

Let $\mathbf{Z} = (Z_1, \ldots, Z_d)^\top$ be a vector of iid regularly varying random variables with tail index α and meeting condition (4.6). Let S be a $d \times d$ matrix whose diagonal elements are zero, and satisfying $(I - S)$ nonsingular where I is the $d \times d$ identity matrix. Then, using transformed-linear operations, we can define the equivalent expressions:

$$\mathbf{Y} = S \circ \mathbf{Y} \oplus \mathbf{Z}, \tag{4.11}$$

$$(I - S) \circ \mathbf{Y} = \mathbf{Z}, \tag{4.12}$$

and

$$\mathbf{Y} = (I - S)^{-1} \circ \mathbf{Z}. \tag{4.13}$$

Below we show that (4.11)$\Leftrightarrow$(4.12)$\Leftrightarrow$(4.13) by performing the full expansions.

$$\mathbf{Y} = S \circ \mathbf{Y} \oplus \mathbf{Z} \tag{4.11}$$

$$\Leftrightarrow \mathbf{Y} = t\{St^{-1}(\mathbf{Y})\} \oplus \mathbf{Z}$$

$$\Leftrightarrow \mathbf{Y} = t\left(t^{-1}\left[t\{St^{-1}(\mathbf{Y})\}\right] + t^{-1}(\mathbf{Z})\right)$$

$$\Leftrightarrow t\{It^{-1}(\mathbf{Y})\} = t\{St^{-1}(\mathbf{Y}) + t^{-1}(\mathbf{Z})\}$$

$$\Leftrightarrow It^{-1}(\mathbf{Y}) = St^{-1}(\mathbf{Y}) + t^{-1}(\mathbf{Z})$$

$$\Leftrightarrow It^{-1}(\mathbf{Y}) - St^{-1}(\mathbf{Y}) = t^{-1}(\mathbf{Z})$$

$$\Leftrightarrow (I-S)t^{-1}(\mathbf{Y}) = t^{-1}(\mathbf{Z})$$

$$\Leftrightarrow t\{(I-S)t^{-1}(\mathbf{Y})\} = \mathbf{Z}$$

$$\Leftrightarrow (I-S) \circ \mathbf{Y} = \mathbf{Z} \tag{4.12}$$

$$\Leftrightarrow (I-S)^{-1} \circ \{(I-S) \circ \mathbf{Y}\} = (I-S)^{-1} \circ \mathbf{Z}$$

$$\Leftrightarrow t\left[(I-S)^{-1}t^{-1}\{(I-S) \circ \mathbf{Y}\}\right] = (I-S)^{-1} \circ \mathbf{Z}$$

$$\Leftrightarrow t\left((I-S)^{-1}t^{-1}\left[t\{(I-S)t^{-1}(\mathbf{Y})\}\right]\right) = (I-S)^{-1} \circ \mathbf{Z}$$

$$\Leftrightarrow t\{(I-S)^{-1}(I-S)t^{-1}(\mathbf{Y})\} = (I-S)^{-1} \circ \mathbf{Z}$$

$$\Leftrightarrow t\{t^{-1}(\mathbf{Y})\} = (I-S)^{-1} \circ \mathbf{Z}$$

$$\Leftrightarrow \mathbf{Y} = (I-S)^{-1} \circ \mathbf{Z}. \tag{4.13}$$

Following the usual single-parameter scheme for the classical SAR model (see Section 4.2.1), we choose to let $S = \rho W$, where ρ is a single spatial dependence parameter and W is a nonnegative spatial proximity matrix. This results in the model formulation

$$\mathbf{Y} = (I - \rho W)^{-1} \circ \mathbf{Z}. \tag{4.14}$$

For simplicity of notation, we denote the matrix $(I-\rho W)^{-1}$ by $A = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_d)$. Thus, we can write $\mathbf{Y} = A \circ Z$, matching the construction of Corollary 4.2.1.

Condition on ρ

Recall from our discussion of Corollary 4.2.1 that it is desirable for A to have nonnegative entries. Using results from nonnegative matrix theory, we can show that in the case of $A = (I - \rho W)^{-1}$, if ρ takes on a similar restriction as for the classical SAR, then A is both well-defined and nonnegative. The proof relies on the notion of M-matrices. The class of M-matrices is a proper subclass of the class of $d \times d$ Z-matrices (Plemmons, 1977), where a Z-matrix is simply a matrix with nonpositive off-diagonal entries (Fiedler and Ptak, 1962). An M-matrix can be defined as an $d \times d$ Z-matrix B that can be expressed in the form $B = sI - P$, where P is entrywise nonnegative and s is at least as large as the spectral radius of P, defined as the maximum of the absolute values of the eigenvalues of P (Johnson, 1982). Plemmons (1977) lists 40 equivalent characterizations of nonsingular M-matrices, including inverse-positivity. Below we specify a sufficient condition on ρ such that A is an inverse M-matrix and hence entrywise nonnegative.

Proposition 4.3.1. *Let W be a $d \times d$ spatial proximity matrix whose diagonal elements are zero and off-diagonal entries are nonnegative. Let $\lambda_1 < \lambda_2 < \cdots < \lambda_d$ denote the ordered eigenvalues of W. If*

$$\rho \in \left(0\ ,\ \min\left\{|\lambda_1|^{-1}, |\lambda_d|^{-1}\right\}\right), \tag{4.15}$$

then $A = (I - \rho W)^{-1}$ exists and is a nonnegative matrix.

Proof. Notice that if $\rho > 0$, then ρW is a nonnegative matrix. This implies that $I - \rho W$ is a $d \times d$ Z-matrix of the required form $B = sI - P$, where $s = 1$ and $P = \rho W$ is entrywise nonnegative. Since $\rho > 0$, the ordered eigenvalues of P are $\rho\lambda_1 < \rho\lambda_2 < \cdots < \rho\lambda_d$. If ρ further meets condition (4.15), then the spectral radius of P is strictly less than $s = 1$:

$$\begin{aligned}
\max_{i=1,\ldots,d} |\rho\lambda_i| &= \max\left\{\rho|\lambda_1|, \rho|\lambda_d|\right\} \\
&< \max\left\{\min\left\{|\lambda_1|^{-1}, |\lambda_d|^{-1}\right\}|\lambda_1|, \min\left\{|\lambda_1|^{-1}, |\lambda_d|^{-1}\right\}|\lambda_d|\right\} \\
&= 1.
\end{aligned}$$

By the above result, $B = I - \rho W$ satisfies the definition of an M-matrix, specifically a nonsingular M-matrix because the inequality is strict (Johnson, 1982; Plemmons, 1977). By equivalent characterization of nonsingular M-matrices, B is inverse-positive. That is, the inverse M-matrix $B^{-1} = (I - \rho W)^{-1} = A$ exists and A is entrywise nonnegative. □

Note that for square lattices with W a binary proximity matrix under the rook neighborhood specification (recall Figure 4.1), $\lambda_1 \downarrow -4$ and $\lambda_d \uparrow 4$ as $d \to \infty$ (Haining, 1990), so $\rho \in (0, 0.25)$ is a sufficient restriction to ensure A exists and is entrywise nonnegative. Note that this is the same restricted range for ρ that is commonly used for the traditional (non-extreme) SAR.

Intuition

In the context of areal data, the d-dimensional vector $\mathbf{Z}$ represents the values at d areal regions, and W encodes the neighborhood structure between these regions. Akin to the classical Gaussian case, given W, applying the matrix A to the independent vector $\mathbf{Z}$ induces spatial dependence in $\mathbf{Y}$ that is described by the single ρ parameter. Under usual matrix multiplication, $A \cdot \mathbf{Z}$ can be written as $\sum_{i=1}^{d} \mathbf{a}_i \cdot Z_i$, i.e., the sum of each column of A multiplied by each element of $\mathbf{Z}$. Although the transformed-linear matrix multiplication changes this slightly, it is still instructive to consider the columns of A to gain insight into how the extremal SAR model is inducing spatial dependence. The top row of Figure 4.3 shows some examples of the columns of A (in lattice form) when $\mathbf{Z}$ is on a 10×10 square lattice ($d = 100$) and W is the binary proximity matrix with rook structure. Each column $\mathbf{a}_i$ of A forms a kernel with its mode at the element i it corresponds to. The range or bandwidth of the kernel increases as ρ increases. If the corresponding element is large in the given realization of $\mathbf{Z}$, then the kernel will essentially smooth out the large spike at that location. The bottom row of Figure 4.3 shows (from left to right) a realization of the independent $\mathbf{Z}$ and the corresponding $\mathbf{Y} = A \circ \mathbf{Z}$ for $\rho = 0.2$ and $\rho = 0.24$.

Properties

By Proposition 4.3.1, we know $A = (I - \rho W)^{-1}$ exists and is entrywise nonnegative. By definition, A is nonsingular. If A contained a column of zeros, it would be singular, thus it must be

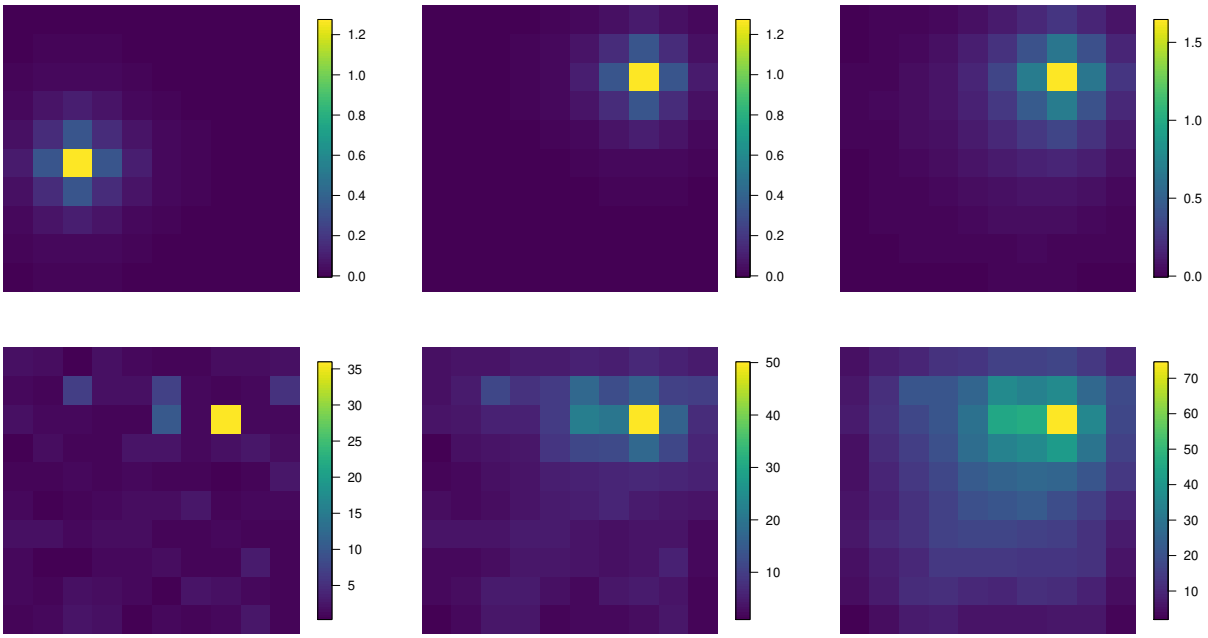

Figure 4.3: Illustration of how the extremal SAR model induces spatial dependence, shown on a 10×10 square lattice with rook neighborhood structure. Top (from left to right): column 43 of A for $\rho = 0.2$, column 78 of A for $\rho = 0.2$, and column 78 of A for $\rho = 0.24$. Bottom (from left to right): a realization of $\mathbf{Z} \overset{iid}{\sim} RV_+^{100}(\alpha = 2)$ with a large spike at element 78, $\mathbf{Y} = A \circ \mathbf{Z}$ for $\rho = 0.2$, and $\mathbf{Y} = A \circ \mathbf{Z}$ for $\rho = 0.24$.

true that $\max_{i=1,\ldots,d} a_{ij} > 0$ for all $j = 1, \ldots, d$. By Corollary 4.2.1, $\mathbf{Y} = A \circ \mathbf{Z} \in RV_+^d(\alpha)$. If ρ satisfies (4.15), then A is a nonnegative matrix. Therefore, when renormalized by $\{c_n\}$, $\mathbf{Y}$ has limiting angular measure

$$H_{\mathbf{Y}}(\cdot) = H_{A \circ \mathbf{Z}}(\cdot) = \sum_{j=1}^{d} ||\mathbf{a}_j||^{\alpha} \delta_{\mathbf{a}_j / ||\mathbf{a}_j||}(\cdot) \tag{4.16}$$

(see (4.7) for comparison). Recall that this is a discrete angular measure, with point masses corresponding to the columns of A, normalized by their norm. Models with similar angular measures have proven challenging to fit, as mentioned in Section 1.4.3. We will discuss these challenges more fully in Chapter 5.

In the case of $\alpha = 2$ and the L_2 norm, assuming $\mathbf{Z}$ meets the conditions of Corollary 4.2.1 with $c_n = n^{1/2}$, the TPDM simplifies to

$$\Sigma_{\mathbf{Y}} = \Sigma_{A \circ \mathbf{Z}} = AA^{\top} = (I - \rho W)^{-1} (I - \rho W)^{-1^{\top}}, \tag{4.17}$$

which is reminiscent of the covariance matrix in the classical Gaussian case.

It is well-known that the covariance structure implied by the classical SAR model is not marginally stationary even for square lattices. The diagonal elements (variances) of the covariance matrix are not identical, and there are unequal covariances between regions that are the same number of neighbors apart (Haining, 1990; Wall, 2004). This issue persists for the TPDM given in (4.17), in particular the marginal scale parameters given by the square-root of the diagonal elements of $\Sigma_{\mathbf{Y}}$ are not identical (see Figure 4.4, left) and depend on ρ. This dependence on ρ poses a challenge for the usual two-stage inference procedure for multivariate extremes, as described in Section 1.4.3, in which one first estimates the marginal effects and then estimates the tail dependence structure after marginal transformation.

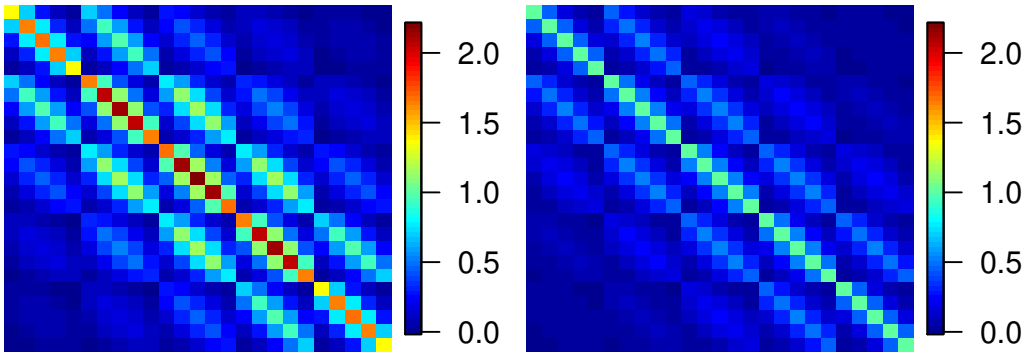

Figure 4.4: Tail pairwise dependence matrices for $\mathbf{Y} = A \circ \mathbf{Z}$ (left) and $\mathbf{Y} = \widetilde{A} \circ \mathbf{Z}$ (right), with $d = 25$, $\rho = 0.2$, and W a binary proximity matrix following the rook neighborhood structure. The diagonal elements $\sigma_{\mathbf{Y}ii}$ are equal to the square of the scale of Y_i. The case of $\mathbf{Y} = \widetilde{A} \circ \mathbf{Z}$ (right) has a common scale of 1.

4.3.2 An extreme SAR model with common scale

We have seen that the extremal SAR model proposed in Section 4.3.1 results in unequal marginal scale parameters that depend on the spatial dependence parameter ρ. If ρ was known, it would be possible to transform the marginals accordingly, but here ρ is unknown and must be estimated. To address this issue, we propose the following extension, which results in common

(unit) marginal scale parameters and is thus more amenable to the usual two-stage inference procedure for multivariate extremes.

Let $\mathbf{Z} = (Z_1, \ldots, Z_d)^\top$ be a vector of iid regularly varying random variables with tail index $\alpha = 2$ such that $n\mathbb{P}\left(n^{-1/2} Z_j > z\right) \to z^{-2}$ and $n\mathbb{P}\left\{Z_j \leq \exp(-kn^{1/2})\right\} \to 0$ for any $k > 0$, $j = 1, \ldots, d$. Let $A = (I - \rho W)^{-1}$ as before, and let $D = \text{diag}(||A_{1\cdot}||_2, \ldots, ||A_{d\cdot}||_2)$ denote the diagonal matrix of the L_2 norms of the rows of A. Then the common-scale model is defined as follows:

$$\mathbf{Y} = \widetilde{A} \circ \mathbf{Z} = D^{-1} A \circ \mathbf{Z}. \tag{4.18}$$

Note that D, like A, depends on the value of ρ.

If ρ meets condition (4.15), then A is nonnegative and has positive row norms. Thus $\widetilde{A}$ is also nonnegative, and the modified formulation (4.18) results in the TPDM

$$\Sigma_{\mathbf{Y}} = \Sigma_{\widetilde{A} \circ \mathbf{Z}} = \widetilde{A}\widetilde{A}^\top = D^{-1} A A^\top D^{-1}, \tag{4.19}$$

which has diagonal elements $\sigma_{\mathbf{Y}ii} = ||A_{i\cdot}||_2^{-1} ||A_{i\cdot}||_2^2 ||A_{i\cdot}||_2^{-1} = 1$ (Figure 4.4, right). This means the components of Y have a common marginal scale parameter of 1, i.e., $\lim_{n\to\infty} n\mathbb{P}\left(n^{-1/2} Y_i > y\right) = y^{-2}$, $i = 1, \ldots, d$. The limiting angular measure for this common-scale model is

$$H_{\mathbf{Y}}(\cdot) = H_{\widetilde{A} \circ \mathbf{Z}}(\cdot) = \sum_{j=1}^{d} ||\widetilde{\mathbf{a}}_j||_2^2 \, \delta_{\widetilde{\mathbf{a}}_j / ||\widetilde{\mathbf{a}}_j||_2}(\cdot), \tag{4.20}$$

where $\widetilde{\mathbf{a}}_j$ denotes the jth column of $\widetilde{A}$, for $j = 1, \ldots, d$. Again, this is a discrete angular measure, with point masses corresponding to the columns of $\widetilde{A}$ normalized by their L_2 norm.

4.4 Discussion

We have proposed a new model for extremes of areal data, with a common-scale extension. This multivariate model is inspired by the SAR model from classical spatial statistics, and utilizes recent results on transformed-linear operations applied to regularly varying random

vectors. Like the classical SAR under the single-parameter scheme, our spatial dependence parameter ρ is interpretable, as the degree of spatial dependence increases with ρ. Also similar to the classical case, given a spatial proximity matrix W that characterizes neighborhood structure, we can define restrictions on ρ such that the model has desirable properties. Given ρ and W, we can summarize the spatial tail dependence via the TPDM, which is similar to the covariance matrix in the Gaussian case. We can also write down the limiting angular measure, from which we could calculate the probability of landing in any risk region of interest.

This chapter has focused on the single-parameter scheme for the extreme SAR model, as is often done in the classical SAR literature. In addition, we have illustrated this simple model with the simplest possible binary proximity matrix W with rook neighborhood structure. This setup is more suitable for modeling shorter range dependence, however localized behavior is not appropriate for certain types of data, such as temperature extremes. Many extensions are possible. In the single-parameter scheme, W is flexible – its only restriction is to be a nonnegative matrix with zeros on the diagonal – and can be tailored to the application. In addition, higher-order SAR models could be used that include more than one spatial proximity or spatial weights matrix. Instead of $S = \rho W$, one could have $S = \rho_1 W_1 + \rho_2 W_2 + \cdots + \rho_k W_k$ with $\{W_i\}$ specifying neighbors at different distances (Haining, 1990).

Areal models have only recently become a topic of interest in the spatial extreme value analysis literature. Such models are needed, especially in atmospheric sciences, because much of the spatial data available for analysis are indexed by regular grids. For example, climate analyses often draw on gridded climate model output, or direct observations that are converted to gridded “data products" to, among other reasons, facilitate comparison to numerical climate model output. Recently, Reich and Shaby (2018) proposed a method for areal extremes that accounts for spatial dependence using latent clustering of neighboring regions. To our knowledge, our extreme SAR model is the only model proposed for areal extremes that is directly analogous to the areal models in classical spatial statistics. It is also to our knowledge the only areal model proposed in the MVRV framework.

A potential future use of the extreme SAR model could be in a hierarchical setting. In many extremes problems, marginal behavior is of primary interest, such as in estimating high quantiles (return levels) or occurrence probabilities at individual locations. In such cases, it may be effective and computationally efficient to use hierarchical approaches (e.g., Cooley et al., 2007) that spatially model the parameters of univariate extremes. Most such hierarchical models assume extremes at different spatial locations are independent conditional on the latent processes. This conditional independence assumption ignores the fact that multiple locations can be affected by the same event (e.g., storm). Replacing the conditional independence likelihood with one which accounts for spatial dependence is challenging, and the extreme SAR model could potentially fill this gap.

Chapter 5

Estimation and Inference for Extreme SAR Models

5.1 Introduction

Although our interest is in spatial applications, the extremal SAR models proposed in Chapter 4 are fundamentally multivariate. The current chapter focuses on using multivariate threshold exceedances to estimate and perform inference for ρ, the parameter that characterizes dependence in the joint upper tail. As discussed in Section 1.4.3, statistical modeling of multivariate extremes typically involves two stages: marginal estimation and dependence estimation. In the MVRV framework this involves transforming the univariate marginals to a convenient regularly varying distribution (e.g., Fréchet) with common tail index α. Such a transformation retains the tail dependence structure (Resnick, 1987, Proposition 5.10). To facilitate the two-stage inference procedure, as discussed in Section 4.3.2, we restrict attention to the common-scale formulation given by (4.18). Thus, the methods that we propose for dependence estimation assume regularly varying marginals with common $\alpha = 2$ and scale = 1.

The beauty of classical areal models such as the SAR and CAR is their simplicity: spatial dependence is induced by a single parameter ρ and pre-determined spatial proximity matrix W. However, this overly simplistic representation of dependence can lead to difficulties for inference. For example, Besag and Kooperberg (1995) note that a common issue with the CAR is that appreciable correlations between neighboring sites require parameter values extremely close to the boundary of the parameter space.

In the extremes setting, there are further challenges to estimation and inference. Likelihood-based procedures are common for estimating tail dependence parameters. However, likelihoods can be difficult to compute in higher dimensions, leading many current methods to be based on composite likelihoods (e.g., Davison and Gholamrezaee, 2012; Genton et al., 2011; Huser and Davison, 2013; Padoan et al., 2010). The discrete nature of the limiting angular measure of the extreme SAR model poses an additional challenge. To our knowledge, likelihood-

based inference has not been attempted for existing models with discrete angular measures, such as max-linear models (Fougères et al., 2013; Wang and Stoev, 2011). Instead, several minimum distance methods have been proposed. Yuen and Stoev (2014) propose an M-estimator based on the continuous ranked probability score of multivariate cumulative distribution functions. Einmahl et al. (2016) also propose an M-estimator, theirs based on the stable tail dependence function, which is related to the upper tail of the underlying cumulative distribution function. Einmahl et al. (2018) extend the previous approach with an adaptive weighted least-squares procedure. The max-linear examples examined in the aforementioned studies are of dimension $d = 4$, whereas in this chapter we consider an extreme SAR model of dimension $d = 400$.

The remainder of the chapter is organized as follows. Section 5.2 describes the simulations we use to test our inference procedures, and the gridded precipitation observations we use for the data application. We use simulations from the true extreme SAR model as well as from a Brown-Resnick process with similar pairwise extremal dependence, which we derive for $\alpha = 2$ and the L_2 norm. We investigate two approaches to estimation and inference. Section 5.3 presents a censored likelihood approach in which censored components below a threshold are imputed within a Bayesian framework. This method performs very well for simulations from the true model, but is not robust to model misspecification. Section 5.4 presents an alternative approach that matches an empirical estimate of the TPDM to the model-based TPDM, with respect to the Frobenius matrix norm. Development of this method requires mitigating bias inherent to finite-sample estimation of the TPDM. After addressing this challenge, we are able to produce reasonable estimates for ρ even in the case of model misspecification.

5.2 Data

We will investigate our inference procedures on three different categories of data, each described in a subsection below. In each category, we consider n realizations or observations of a spatial field with d areal regions (grid cells). Note that this differs from the typical setup in

spatial statistics, which includes only one observation of a spatial field, but is standard for the analysis of multivariate and spatial extremes.

5.2.1 Simulation from the true model

To demonstrate our methods for estimation and inference, we begin by simulating realizations from the true extreme SAR model (4.18). We select two values of ρ for simulations: $\rho = 0.1$ (weaker dependence) and $\rho = 0.2$ (stronger dependence). For each value of ρ, we simulate $n = 10,000$ independent realizations of a $d = 400$-dimensional random vector $\mathbf{Z} \overset{iid}{\sim}$ Fréchet($\alpha = 2$, scale = 1). Each realization of $\mathbf{Z}$ represents values found on a 20×20 square lattice. Let W be the 400×400 binary proximity matrix following the rook neighborhood specification. To each realization of $\mathbf{Z}$, we apply transformed-linear multiplication with the $\widetilde{A}$ matrix corresponding to W and the given value of ρ, to obtain n independent realizations of $\mathbf{Y}$. These realizations $\{\mathbf{y}_t\}_{t=1}^n$ are independent temporal replicates of heavy-tailed spatial fields whose spatial extremal dependence is specified by (4.18).

Since we do not believe that real world data will follow the true model exactly, it is important to investigate how our estimation and inference procedures react to deviations from the true model. A first step is to add light-tailed noise. For purposes of illustration, we add $\boldsymbol{\zeta}_t \overset{iid}{\sim} \text{Exp}(2)$ to each realization $\mathbf{y}_t$. At moderate thresholds we expect this independent noise to dilute the spatial dependence and lead to underestimation of ρ. Because the noise is light-tailed, we expect its effect to diminish as the threshold increases.

5.2.2 Simulation from a Brown-Resnick process

Our second category of data comprises simulations from a Brown-Resnick process (Brown and Resnick, 1977; Kabluchko et al., 2009). Brown-Resnick processes form a flexible class of max-stable processes constructed from intrinsically stationary Gaussian processes parameterized by variograms (see Section 1.5.1). As a more realistic deviation from the true model (4.18), we simulate $n = 10,000$ independent realizations from a Brown-Resnick process at the set of $d = 400$ grid cell centers. We use the variogram

$$2\gamma(\mathbf{s},\mathbf{s}') = 2(||\mathbf{s}-\mathbf{s}'||_2/\lambda)^{\kappa}, \tag{5.1}$$

where $\mathbf{s}$, $\mathbf{s}'$ denote locations of grid cell centers, and $\lambda > 0$ and $\kappa \in (0,2]$ control the range and the smoothness, respectively (Thibaud et al., 2016). Although the Brown-Resnick processes are max-stable, the processes observed at a finite number of locations are also multivariate regular varying, so long as their marginal distributions are GEV with $\xi > 0$.

To mimic the tail dependence structure of the true extreme SAR model at the settings described in Section 5.2.1, we choose λ and κ to minimize the squared error between TPDM values from the true model (as given by (4.19)) and pairwise extremal dependence measures for the Brown-Resnick model. We derive the latter below. We start with the known bivariate distribution function for the simple Brown-Resnick process ($\alpha = 1$ and L_1 norm) and from it obtain the corresponding bivariate angular density. We then make a conversion to the case with $\alpha = 2$ and the L_2 norm to enable comparison with the TPDM of the extreme SAR.

Case of $\alpha = 1$ and L_1 norm

Recall from Section 1.5.1 that the bivariate distribution function of a Brown-Resnick process $Z(\mathbf{s})$ with unit Fréchet margins at the pair of sites $\{\mathbf{s}_i, \mathbf{s}_j\}$ is $\exp\{-V(z_i, z_j)\}$, with exponent measure given by

$$V(z_i, z_j) = \frac{1}{z_i}\Phi\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{z_i}{z_j}\right)\right\} + \frac{1}{z_j}\Phi\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{z_j}{z_i}\right)\right\}, \tag{5.2}$$

where $z_i = z(\mathbf{s}_i)$, $z_j = z(\mathbf{s}_j)$, $a = \left\{2\gamma(\mathbf{s}_i, \mathbf{s}_j)\right\}^{1/2}$ with $2\gamma(\cdot)$ given in (5.1), and $\Phi(\cdot)$ is the standard normal distribution function (Huser and Davison, 2013). Following Coles and Tawn (1991), we can obtain the bivariate angular density by taking partial derivatives of V. Let $\phi(\cdot)$ be the standard normal density, and $\phi'(\cdot)$ be its first derivative. We find

$$\frac{\partial V}{\partial z_i \partial z_j} = -\frac{1}{a z_i^2 z_j}\phi\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{z_i}{z_j}\right)\right\} - \frac{1}{a z_i z_j^2}\phi\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{z_j}{z_i}\right)\right\}$$

$$
\begin{aligned}
&\quad - \frac{1}{a^2 z_i^2 z_j}\phi'\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{z_i}{z_j}\right)\right\} - \frac{1}{a^2 z_i z_j^2}\phi'\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{z_j}{z_i}\right)\right\} \\
&= -(z_i+z_j)^{-3} h_{L_1}\left(\frac{z_i}{z_i+z_j}, \frac{z_j}{z_i+z_j}\right) \\
&= -r^{-3} h_{L_1}(\theta_i, \theta_j),
\end{aligned}
$$

where

$$
\begin{aligned}
h_{L_1}(\theta_i,\theta_j) &= \frac{1}{a\theta_i^2\theta_j}\phi\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{\theta_i}{\theta_j}\right)\right\} + \frac{1}{a\theta_i\theta_j^2}\phi\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{\theta_j}{\theta_i}\right)\right\} \\
&\quad + \frac{1}{a^2\theta_i^2\theta_j}\phi'\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{\theta_i}{\theta_j}\right)\right\} + \frac{1}{a^2\theta_i\theta_j^2}\phi'\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{\theta_j}{\theta_i}\right)\right\} \\
&= \frac{1}{a\theta_i\theta_j}\left[\frac{1}{\theta_i}\left\{1 - \frac{f_1(a,\boldsymbol{\theta})}{a}\right\}\phi\left\{f_1(a,\boldsymbol{\theta})\right\} + \frac{1}{\theta_j}\left\{1 - \frac{f_2(a,\boldsymbol{\theta})}{a}\right\}\phi\left\{f_2(a,\boldsymbol{\theta})\right\}\right], \qquad (5.3)
\end{aligned}
$$

with $f_1(a,\boldsymbol{\theta}) = \frac{a}{2} - \frac{1}{a}\log\left(\frac{\theta_i}{\theta_j}\right)$ and $f_2(a,\boldsymbol{\theta}) = \frac{a}{2} - \frac{1}{a}\log\left(\frac{\theta_j}{\theta_i}\right)$. The simplification follows because $\phi'(y) = -y\phi(y)$.

Note that this bivariate angular density lives on the 1-dimensional L_1 simplex, so we could also write $\boldsymbol{\theta} = (\theta_i, \theta_j)$ as $(\theta, 1-\theta)$. If desired, we could then use the alternative notation

$$
h_{L_1}(\theta) = \frac{1}{a\theta(1-\theta)}\left[\frac{1}{\theta}\left\{1 - \frac{f_1(a,\theta)}{a}\right\}\phi\left\{f_1(a,\theta)\right\} + \frac{1}{1-\theta}\left\{1 - \frac{f_2(a,\theta)}{a}\right\}\phi\left\{f_2(a,\theta)\right\}\right], \qquad (5.4)
$$

where $f_1(a,\theta) = \frac{a}{2} - \frac{1}{a}\log\left(\frac{\theta}{1-\theta}\right)$ and $f_2(a,\theta) = \frac{a}{2} - \frac{1}{a}\log\left(\frac{1-\theta}{\theta}\right)$.

Case of $\alpha = 2$ and L_2 norm

The extremal dependence measure depends on the choice of norm and the angular measure associated with that norm (Larsson and Resnick, 2012). Therefore, to allow comparison to the TPDM values of the extreme SAR model, we must apply the necessary transformations to (5.3) to obtain the bivariate angular density for the case of $\alpha = 2$ and the L_2 norm.

First, we transform from L_1 polar coordinates $(r, \boldsymbol{\theta})$ back to Cartesian coordinates $\mathbf{z} = (z_i, z_j)$. By Proposition 1 in Cooley et al. (2012b), the Jacobian of the transformation is $|J_1| = ||\mathbf{z}||_1^{-1}$, such

that

$$\begin{aligned}\nu(dr \times d\boldsymbol{\theta}) &= r^{-2} h_{L_1}(\boldsymbol{\theta}) dr d\boldsymbol{\theta} \\ &= ||\mathbf{z}||_1^{-2} h_{L_1}\left(\mathbf{z}||\mathbf{z}||_1^{-1}\right) ||\mathbf{z}||_1^{-1} d\mathbf{z} \\ &= ||\mathbf{z}||_1^{-3} h_{L_1}\left(\mathbf{z}||\mathbf{z}||_1^{-1}\right) d\mathbf{z} = \nu(d\mathbf{z}).\end{aligned}$$

Next, to change the tail index from $\alpha = 1$ to $\alpha = 2$, let $\mathbf{y} = (y_i, y_j)$, where $y_i^2 = z_i$ and $y_j^2 = z_j$. The Jacobian of this transformation is $|J_2| = 4y_i y_j$, so we can write

$$\nu(d\mathbf{z}) = \left\{||\mathbf{y}||_2^2\right\}^{-3} h_{L_1}\left(\frac{y_i^2}{||\mathbf{y}||_2^2}, \frac{y_j^2}{||\mathbf{y}||_2^2}\right) 4y_i y_j d\mathbf{y} = \nu(d\mathbf{y}).$$

Finally, we transform to L_2 polar coordinates by letting $s = ||\mathbf{y}||_2$ and $\mathbf{v} = (v_i, v_j)$ with $v_i = y_i s^{-1}$, and $v_j = y_j s^{-1}$. By Lemma 1.1 in Song and Gupta (1997), the Jacobian is $|J_3| = s{v_j}^{-1}$, so we have

$$\begin{aligned}\nu(d\mathbf{y}) &= \left\{s^2\right\}^{-3} h_{L_1}\left(v_i^2, v_j^2\right) 4 s v_i s v_j s {v_j}^{-1} ds d\mathbf{v} \\ &= 2s^{-3} 2v_i h_{L_1}(v_i^2, v_j^2) ds d\mathbf{v} \\ &= 2s^{-3} h_{L_2}(\mathbf{v}) ds d\mathbf{v} = \nu(ds \times d\mathbf{v}).\end{aligned}$$

Recognizing that the angular density $h_{L_2}(\mathbf{v})$ lives on the 1-dimensional L_2 simplex $\mathbb{S}^1_{+(2)}$, so $\mathbf{v} = (v_i, v_j)$ can be written as $(v, \sqrt{1-v^2})$, we obtain

$$\begin{aligned}h_{L_2}(v) &= 2v \frac{1}{av^2(1-v^2)} \left[\frac{1}{v^2}\left\{1 - \frac{g_1(a,v)}{a}\right\}\phi\left\{g_1(a,v)\right\} + \frac{1}{1-v^2}\left\{1 - \frac{g_2(a,v)}{a}\right\}\phi\left\{g_2(a,v)\right\}\right] \\ &= \frac{2}{av(1-v^2)} \left[\frac{1}{v^2}\left\{1 - \frac{g_1(a,v)}{a}\right\}\phi\left\{g_1(a,v)\right\} + \frac{1}{1-v^2}\left\{1 - \frac{g_2(a,v)}{a}\right\}\phi\left\{g_2(a,v)\right\}\right], \quad (5.5)\end{aligned}$$

where $g_1(a,v) = \frac{a}{2} - \frac{1}{a}\log\left(\frac{v^2}{1-v^2}\right)$ and $g_2(a,v) = \frac{a}{2} - \frac{1}{a}\log\left(\frac{1-v^2}{v^2}\right)$. Note that $\int_0^1 h_{L_2}(v)dv = 2$. The angular density in (5.5) represents a projection from $\mathbb{S}^1_{+(2)}$, defined by $v_i^2 + v_j^2 = 1$, down to $[0,1]$.

Pairwise extremal dependence measure and Brown-Resnick parameters

Plugging in the angular density (5.5) derived above, we can now obtain the pairwise extremal dependence measure for each pair of grid cell centers $(\mathbf{s}_i, \mathbf{s}_j)$:

$$\begin{aligned}\sigma_{ij}^{BR} &= \int_0^1 v\sqrt{1-v^2}\, h_{L_2}(v)\, dv \\ &= \int_0^1 \frac{2}{a\sqrt{1-v^2}} \left[\frac{1}{v^2}\left\{1 - \frac{g_1(a,v)}{a}\right\}\phi\left\{g_1(a,v)\right\} + \frac{1}{1-v^2}\left\{1 - \frac{g_2(a,v)}{a}\right\}\phi\left\{g_2(a,v)\right\}\right] dv,\end{aligned} \tag{5.6}$$

where, as before, $a = \left\{2\gamma(\mathbf{s}_i, \mathbf{s}_j)\right\}^{1/2}$, $g_1(a,v) = \frac{a}{2} - \frac{1}{a}\log\left(\frac{v^2}{1-v^2}\right)$ and $g_2(a,v) = \frac{a}{2} - \frac{1}{a}\log\left(\frac{1-v^2}{v^2}\right)$.

Let $\Sigma^{BR} = \left(\sigma_{ij}^{BR}\right)_{i,j=1,\ldots,d}$. For a given value of ρ and corresponding TPDM $\Sigma_{\mathbf{Y}} = \widetilde{A}\widetilde{A}^{\top}$ for the true model, we can choose Brown-Resnick parameters that minimize the squared Frobenius norm $||\Sigma_{\mathbf{Y}} - \Sigma^{BR}||_F^2 = \sum_{i=1}^d \sum_{j=1}^d |\sigma_{\mathbf{Y}ij} - \sigma_{ij}^{BR}|^2$. When $\rho = 0.2$, we find $\lambda = 0.431, \kappa = 1.148$ to be optimal. When $\rho = 0.1$, we find $\lambda = 0.254, \kappa = 1.341$ to be optimal. Figure 5.1 compares the pairwise extremal dependence measures for a Brown-Resnick process with our chosen parameters to the TPDM values of the true extreme SAR model when $\rho = 0.2$ (left) and $\rho = 0.1$ (right).

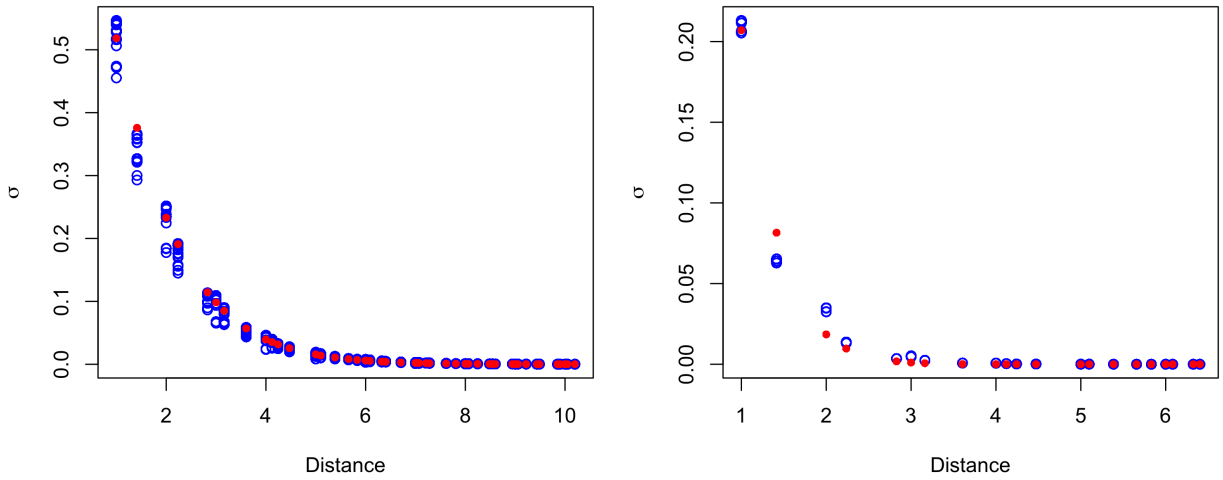

Figure 5.1: Derived pairwise extremal dependence measures $\{\sigma_{ij}^{BR}\}$, plotted against Euclidean distance, for the Brown-Resnick process (red) with range and smoothness parameters that minimize squared error with respect to the TPDM values $\{\sigma_{\mathbf{Y}ij}\}$ of the true extreme SAR model (blue) when $\rho = 0.2$ (left) and $\rho = 0.1$ (right). The range of distances along the x-axis has been selected for visual clarity.

5.2.3 Gridded precipitation observations

For our data application, we use the Climate Prediction Center (CPC) unified gauge-based analysis of daily precipitation provided by NOAA Earth System Research Laboratory's Physical Sciences Division (PSD).[6] This is a gridded product with 0.25° resolution over the contiguous United States and daily observations for the 59 years between 1948-2006. We choose a region of 20×20 grid cells containing northeastern Colorado, and restrict our attention to precipitation between May 1st and September 30th of each year.

Unlike the simulated data in Section 5.2.1 and Section 5.2.2, which already have appropriate regularly varying marginals, the precipitation data requires transforming the marginal distributions to be approximately Fréchet with $\alpha = 2$ and scale = 1 before fitting our dependence model. There are many possible ways to implement such a marginal transformation. Below we describe our chosen procedure, which uses a parametric form for the upper tail and a nonparametric form below a marginal threshold.

Denote our original observations by $\mathbf{x}_t = (x_{t1}, x_{t2}, \ldots, x_{td}), t = 1, \ldots, n$, where $d = 400$ grid cells and $n = 9027$ days. Let $\hat{u}_i$ be the empirical 0.96 quantile for the ith marginal, $i = 1, \ldots, d$, where the marginal threshold quantile is chosen using standard diagnostic techniques. Above this threshold, we first fit a GPD at each individual grid cell. We then smooth the individual GPD scale and shape parameter estimates using a thin plate spline regression, to obtain more spatially coherent GPD parameter estimates. Let $\hat{\psi}_{u_i}$ and $\hat{\xi}_i$, $i = 1, \ldots, d$ denote the smoothed GPD scale and shape estimates. To handle the issue of days with zero precipitation, we also add Unif$(0, 1 \times 10^{-5})$ noise to any zero observations. Next, we let

$$\hat{F}_i(x) = \begin{cases} (n+1)^{-1} \sum_{t=1}^{n} \mathbb{I}(x_{ti} \leq x), & x \leq \hat{u}_i, \\ 1 - 0.04 \left(1 + \hat{\xi}_i (x - \hat{u}_i)/\hat{\psi}_{u_i}\right)^{-1/\hat{\xi}_i}, & x > \hat{u}_i. \end{cases} \tag{5.7}$$

[6] https://www.esrl.noaa.gov/psd/

Let G denote the standard Fréchet distribution function (1.3) with $\alpha = 2$. Then our transformed observations are given by $\mathbf{y}_t = \left(G^{-1}(\hat{F}_1(x_{t1})), G^{-1}(\hat{F}_2(x_{t2})), \ldots, G^{-1}(\hat{F}_d(x_{td}))\right)$, $t = 1, \ldots, n$. As an example, Figure 5.2 shows a day with heavy precipitation over northeastern Colorado on the original scale (left) as well as the transformed data on the Fréchet scale (right).

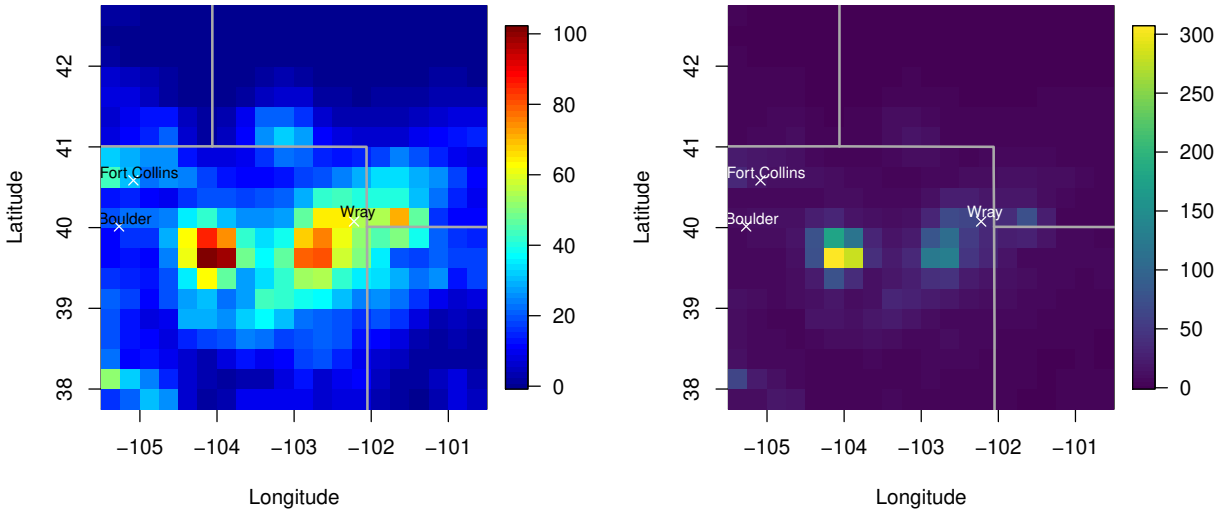

Figure 5.2: Example of daily precipitation data from the CPC unified gauge-based analysis on the original scale (left) and after transformation to Fréchet($\alpha = 2$, scale=1) marginals (right). Gray lines indicate state borders, and white X's mark three cities in Colorado.

5.3 Censored Likelihood with Bayesian Data Augmentation

5.3.1 Methods

Density for a model constructed by $\widetilde{A} \circ \mathbf{Z}$

If we assume a specific regularly varying distribution for $\mathbf{Z}$, it is straightforward to derive the joint density of $\mathbf{Y} = \widetilde{A} \circ \mathbf{Z}$. Here we assume the distribution is Fréchet. In the tail, the density obtained under this distributional assumption for $\mathbf{Z}$ is essentially equivalent to the density with any other regularly varying random vector with the same tail index and scale.

Consider for the moment a single replicate of a spatial field with d areal regions. Let $\mathbf{Z} \in \mathbb{R}_+^d$ be iid Fréchet($\alpha = 2$, scale = 1), such that

$$f_{\mathbf{Z}}(z_1,\ldots,z_d) = \prod_{i=1}^{d} \left\{2\ z_i^{-3}\ \exp(-z_i^{-2})\right\}, \qquad z_i > 0, \quad i = 1,\ldots,d.$$

Now let $\mathbf{Y} = \widetilde{A} \circ \mathbf{Z} = \widetilde{B}^{-1} \circ \mathbf{Z} = t(\widetilde{B}^{-1} t^{-1}(\mathbf{Z}))$, where $B = (I - \rho W)$ and $\widetilde{B} = BD$. Then we have $\mathbf{Z} = \widetilde{B} \circ \mathbf{Y} = t(\widetilde{B}(t^{-1}(\mathbf{Y})))$. Recall that the transformations t and t^{-1} are applied componentwise, where $t(v) = \log\{\exp(v) + 1\}$ and $t^{-1}(x) = \log\{\exp(x) - 1\}$.

We derive the density of $\mathbf{Y}$ in three steps:

1. Let $\mathbf{U} = t^{-1}(\mathbf{Z}) \Leftrightarrow \mathbf{Z} = t(\mathbf{U})$. Then the Jacobian J_1 of this transformation is the determinant of the diagonal matrix $\text{diag}(t'(u_1),\ldots,t'(u_d))$, where $t'(u_i) = \frac{\exp(u_i)}{1+\exp(u_i)}$. Therefore $|J_1| = \prod_{i=1}^{d} |t'(u_i)|$, and

$$f_{\mathbf{U}}(u_1,\ldots,u_d) = \prod_{i=1}^{d} \left\{2\ t(u_i)^{-3}\ \exp\left(-t(u_i)^{-2}\right)\ |t'(u_i)|\right\}.$$

2. Now let $\mathbf{V} = \widetilde{B}^{-1}\mathbf{U} \Leftrightarrow \mathbf{U} = \widetilde{B}\mathbf{V}$. Then $f_{\mathbf{V}}(\mathbf{v}) = f_{\mathbf{U}}(\widetilde{B}\mathbf{v}) \frac{1}{|\det(\widetilde{B}^{-1})|} = f_{\mathbf{U}}(\widetilde{B}\mathbf{v})|\det(\widetilde{B})|$, i.e.,

$$f_{\mathbf{V}}(v_1,\ldots,v_d) = |\det(\widetilde{B})| \prod_{i=1}^{d} \left\{2\ t(\widetilde{B}_{i\cdot}\mathbf{v})^{-3}\ \exp\left(-t(\widetilde{B}_{i\cdot}\mathbf{v})^{-2}\right)\ |t'(\widetilde{B}_{i\cdot}\mathbf{v})|\right\},$$

   where $\widetilde{B}_{i\cdot}$ denotes the ith row of $\widetilde{B}$.

3. Finally, let $\mathbf{Y} = t(\mathbf{V}) \Leftrightarrow \mathbf{V} = t^{-1}(\mathbf{Y})$. Then the Jacobian J_3 of this transformation is the determinant of the diagonal matrix $\text{diag}(t^{-1'}(y_1),\ldots,t^{-1'}(y_d))$, where $t^{-1'}(y_i) = \frac{\exp(y_i)}{\exp(y_i)-1}$. So $|J_3| = \prod_{i=1}^{d} |t^{-1'}(y_i)|$, and our final density is

$$f_{\mathbf{Y}}(y_1,\ldots,y_d) = |\det(\widetilde{B})| \prod_{i=1}^{d} \left\{2\ t(\widetilde{B}_{i\cdot}.t^{-1}(\mathbf{y}))^{-3}\ \exp\left(-t(\widetilde{B}_{i\cdot}.t^{-1}(\mathbf{y}))^{-2}\right)\ |t'(\widetilde{B}_{i\cdot}.t^{-1}(\mathbf{y}))|\ |t^{-1'}(y_i)|\right\}. \tag{5.8}$$

Censored likelihood approach

Following the usual philosophy of extremes, we wish to fit our model using only data which are large, and here we will use threshold exceedances. There are many ways to define a multi-

variate threshold exceedance (see Section 1.4.3 for example). Here, we will consider threshold exceedances defined in terms of the marginals. Suppose our complete data $\{\mathbf{y}_t\}_{t=1}^n$ has appropriate regularly varying marginals with common $\alpha = 2$ and scale = 1. An observation will be considered a threshold exceedance if at least one of its d components is above a high marginal threshold T. We denote these exceedances by $\{\mathbf{y}_{t_k}\}_{k=1}^{n_T}$, where n_T is the number of exceedances and n_T/n is small. We assume the threshold exceedances are independent in time.

We adopt a censored likelihood approach to fitting our model to the threshold exceedances. Under this approach, components above the marginal threshold T are taken at face value. Components falling below T are censored at T, i.e., we assume only that they fall somewhere in the interval $(0, T)$. Fitting a censored likelihood is a common practice to prevent low values from affecting estimation of the extremal dependence structure. Censored likelihood approaches were first proposed by Ledford and Tawn (1996) and Smith et al. (1997), then more recently extended to the spatial framework by Wadsworth and Tawn (2014) and Thibaud and Opitz (2015). It is well-established that censored approaches reduce bias in the estimation of dependence parameters by accounting for misspecification below the marginal threshold (Huser et al., 2016).

Let $\mathscr{D} := \{1, \ldots, d\}$ and $C_k \subset \mathscr{D}$ be the subset of indices denoting which components of $\mathbf{y}_{t_k}$ fall below T, i.e., $y_{t_k,j} \leq T$ for $j \in C_k$, and $y_{t_k,j} > T$ for $j \in E_k := \mathscr{D} \setminus C_k$, for $k = 1, \ldots, n_T$. In other words, we can split each threshold exceedance $\mathbf{y}_{t_k}$ into a vector $\mathbf{y}_{E_k} := \{y_{t_k,j} : j \in E_k\}$ of "exact" components above T and a vector $\mathbf{y}_{C_k} := \{y_{t_k,j} : j \in C_k\}$ of censored components belonging to the interval $(0, T)$. We assume that the density $f_{\mathbf{Y}}(\mathbf{y})$ is a suitable model for $y_j > T$, $j = 1, \ldots, d$. Suppose we reorder the arguments of $f_{\mathbf{Y}}$ to be $(\mathbf{y}_{C_k}, \mathbf{y}_{E_k})$. Let $\mathbf{y}$ be a dummy variable with dimension equal to the cardinality of the set C_k, denoted by $\text{card}(C_k)$. Then for each partially censored observation, the likelihood contribution will be

$$p_T(\mathbf{y}_{t_k}; \rho) = \underbrace{\int_0^T \cdots \int_0^T}_{\text{card}(C_k)} f_{\mathbf{Y}}(\mathbf{y}, \mathbf{y}_{E_k}; \rho) d\mathbf{y}. \tag{5.9}$$

For n_T independent threshold exceedances, the overall censored likelihood is then

$$L_T(\rho) = \prod_{k=1}^{n_T} p_T(\mathbf{y}_{t_k}; \rho). \tag{5.10}$$

The likelihood in (5.10) is not available analytically. Furthermore, due to the relatively localized behavior of our model, we anticipate that card(C_k) will be much larger than card(E_k) for all k. Thus, numerical integration will be computationally impractical, especially for large d. Rather than working with the censored likelihood directly, as most previous studies have done, we instead propose a data augmentation method via Bayesian Markov chain Monte Carlo (MCMC).

Data augmentation via Bayesian MCMC

We use Bayesian MCMC methods to fit partially censored threshold exceedances and estimate the posterior distribution of ρ. Let $\mathbf{Y} = \widetilde{A} \circ \mathbf{Z}$, where $\mathbf{Z} \overset{iid}{\sim}$ Fréchet($\alpha = 2$, scale = 1). To complete the Bayesian model specification, we use the proper prior $\rho \sim \text{Unif}(0, \rho_{\max})$, where $\rho_{\max} = \min\left\{|\lambda_1|^{-1}, |\lambda_d|^{-1}\right\}$. For square lattices with the rook neighborhood structure, we use $\rho_{\max} = 0.25$ as explained in Section 4.3.1.

As discussed earlier, the overall censored likelihood (5.10) is not available either analytically or via numerical integration. Instead, we use data augmentation (e.g., De Oliveira, 2005; Morris et al., 2017; Tanner and Wong, 1987) to incorporate information from censored components. In the data augmentation step of the MCMC algorithm detailed below, we impute censored values below the threshold T such that the estimation of ρ can be based on the "complete" vector $(\mathbf{Y}_{C_k}, \mathbf{y}_{E_k})$, where $\mathbf{Y}_{C_k}$ is drawn from the distribution of censored values given the observed exact components $\mathbf{y}_{E_k}$.

We implement our model using a Gibbs sampler with Metropolis-Hastings steps (Hastings, 1970; Metropolis et al., 1953). The sampler alternates between updating $\mathbf{Y}_{C_k}$ via the data augmentation step and updating ρ. At each iteration $(m+1)$, after imputing the censored components for each of the n_T partially censored threshold exceedance observations, we sample

$\rho^{(m+1)} \sim p(\rho|\mathbf{y}_{t_1}^{(m+1)},\ldots,\mathbf{y}_{t_{n_T}}^{(m+1)})$. The ρ update is accomplished with a random walk Metropolis-Hastings step on the logit scale (e.g., Givens and Hoeting, 2013, §7.1.2).

At the data augmentation step, for each partially censored d-dimensional threshold exceedance $\mathbf{y}_{t_k}$ $(k = 1,\ldots,n_T)$, we sample $\mathbf{Y}_{C_k}^{(m+1)} \sim p(\mathbf{Y}_{C_k}|\mathbf{y}_{E_k},\rho^{(m)})$ using a Metropolis-Hastings update. Our Metropolis-Hastings proposal for $\mathbf{Y}_{C_k}$ exploits the transformed-linear relationship between $\mathbf{Y}$ and $\mathbf{Z}$. Let $\mathbf{Z}_{t_k}$ denote a vector of latent iid Fréchet($\alpha = 2$, scale $= 1$) random variables which could generate $\mathbf{y}_{t_k}$ given the current state of the matrix $\widetilde{A}^{(m)}$. As with $\mathbf{y}_{t_k}$, we can split $\mathbf{Z}_{t_k}$ into $\mathbf{Z}_{E_k} := \{Z_{t_k,j} : j \in E_k\}$ and $\mathbf{Z}_{C_k} := \{Z_{t_k,j} : j \in C_k\}$. For any $d \times d$ matrix M, denote the submatrix of M lying in rows $\mathscr{I} \subset \mathscr{D}$ and columns $\mathscr{J} \subset \mathscr{D}$ by $M[\mathscr{I},\mathscr{J}]$. We can write

$$\mathbf{Z}_{E_k} = t\left\{\widetilde{A}^{(m)}[E_k,E_k]^{-1}\left(t^{-1}(\mathbf{y}_{E_k}) - \widetilde{A}^{(m)}[E_k,C_k]\; t^{-1}(\mathbf{Z}_{C_k})\right)\right\}, \tag{5.11}$$

which in turn allows us to write $\mathbf{Y}_{C_k}$ in terms of $\mathbf{y}_{E_k}$, $\widetilde{A}^{(m)}$, and $\mathbf{Z}_{C_k}$:

$$\begin{aligned}
\mathbf{Y}_{C_k} &= \widetilde{A}^{(m)}[C_k,E_k] \circ \mathbf{Z}_{C_k} \oplus \widetilde{A}^{(m)}[C_k,C_k] \circ \mathbf{Z}_{E_k} \\
&= t\Big\{\widetilde{A}^{(m)}[C_k,E_k]\widetilde{A}^{(m)}[E_k,E_k]^{-1} t^{-1}(\mathbf{y}_{E_k}) \\
&\qquad + \left(\widetilde{A}^{(m)}[C_k,C_k] - \widetilde{A}^{(m)}[C_k,E_k]\widetilde{A}^{(m)}[E_k,E_k]^{-1}\widetilde{A}^{(m)}[E_k,C_k]\right) t^{-1}(\mathbf{Z}_{C_k})\Big\}.
\end{aligned} \tag{5.12}$$

Note that $\det\left(\widetilde{A}[E_k,E_k]\right) = \det\left(D^{-1}[E_k,E_k]\; A[E_k,E_k]\right) = \left\{\prod_{i\in E_k} \frac{1}{||A_{i\cdot}||_2}\right\} \det\left(A[E_k,E_k]\right) > 0$ since A is an inverse M-matrix and thus has positive principal minors (Johnson and Smith, 2011). Therefore $\widetilde{A}^{(m)}[E_k,E_k]^{-1}$ always exists. To generate a proposal candidate for the censored components at iteration $(m+1)$, we sample $\mathbf{Z}_{C_k}^* \overset{iid}{\sim}$ Fréchet($\alpha = 2$, scale $= 1$), and use (5.12) to generate $\mathbf{Y}_{C_k}^*$. The proposal procedure does not guarantee $\mathbf{Y}_{C_k}^* \leq \mathbf{T} := (T,\ldots,T)^\top \in \mathbb{R}_+^{\text{card}(C_k)}$, so we also reject any proposals such that $\mathbf{Y}_{C_k}^* \not\leq \mathbf{T}$.

In many cases $B := I - \rho W$ is a sparse matrix, as $B_{ij} = 0$ whenever components i and j are not neighbors, and consequently $\widetilde{B} := BD$ inherits this sparsity. We are able to speed up computation for high dimension d by implementing sparse matrix operations (e.g., multiplication and inversion) via the `spam` package (Furrer and Sain, 2010) in `R`. In addition, instead of directly

calculating the determinant of $\widetilde{B}$ within the Metropolis-Hastings ratios, we use the property $\det(B) = \prod_{i=1}^{d}(1-\rho\lambda_i)$ (Ord, 1975), which implies $\det(\widetilde{B}) = \det(B)\det(D) = \prod_{i=1}^{d}\{(1-\rho\lambda_i)||A_{i\cdot}||_2\}$. This is advantageous because we only need to calculate the eigenvalues $\{\lambda_i\}_{i=1}^{d}$ of W once for the entire sampler.

Algorithm

- Initialize $\rho^{(0)}$, $\mathbf{y}$
- At iteration $m+1$, given $\rho^{(m)}$

  1. For $k = 1, \ldots, n_T$

     Sample $Z_j \overset{iid}{\sim}$ Fréchet($\alpha = 2$, scale $= 1$), $j = 1, \ldots,$ card(C_k). Call this $\mathbf{Z}_{C_k}$.

     Generate a candidate $\mathbf{Y}^*_{C_k}$ using Equation (5.12), with $A^{(m)} = (I-\rho^{(m)}W)^{-1}$ and $\widetilde{A}^{(m)} = \text{diag}\left(1/||A^{(m)}_{1\cdot}||_2\ ,\ \ldots\ ,\ 1/||A^{(m)}_{d\cdot}||_2\right)A^{(m)}$.

     Take

$$\mathbf{Y}^{(m+1)}_{C_k} = \begin{cases} \mathbf{Y}^*_{C_k} \text{ with probability } \min\left\{R\left(\mathbf{y}^{(m)}_{C_k}, \mathbf{Y}^*_{C_k}\right), 1\right\}, \\ \mathbf{y}^{(m)}_{C_k} \text{ otherwise,} \end{cases}$$

     where $R\left(\mathbf{y}^{(m)}_{C_k}, \mathbf{Y}^*_{C_k}\right) = \frac{f\left(\mathbf{Y}^*_{C_k}|\mathbf{y}_{E_k},\rho^{(m)}\right)}{f\left(\mathbf{y}^{(m)}_{C_k}|\mathbf{y}_{E_k},\rho^{(m)}\right)} \frac{g\left(\mathbf{y}^{(m)}_{C_k}|\mathbf{y}^{(m)}_{E_k},\rho^{(m)}\right)}{g\left(\mathbf{Y}^*_{C_k}|\mathbf{y}_{E_k},\rho^{(m)}\right)} = \frac{f\left((\mathbf{Y}^*_{C_k},\mathbf{y}_{E_k})|\rho^{(m)}\right)}{f\left((\mathbf{y}^{(m)}_{C_k},\mathbf{y}_{E_k})|\rho^{(m)}\right)} \frac{g\left(\mathbf{y}^{(m)}_{C_k}|\mathbf{y}_{E_k},\rho^{(m)}\right)}{g\left(\mathbf{Y}^*_{C_k}|\mathbf{y}_{E_k},\rho^{(m)}\right)}$, and the proposal density g can be derived in a similar manner as (5.8).

  2. Generate a candidate ρ^* such that $\text{logit}\left(\frac{\rho^*}{\rho_{\max}}\right) \sim N\left(\text{logit}\left(\frac{\rho^{(m)}}{\rho_{\max}}\right), \sigma^2\right)$.

     Take

$$\rho^{(m+1)} = \begin{cases} \rho^* \text{ with probability } \min\left\{R\left(\rho^{(m)}, \rho^*\right), 1\right\}, \\ \rho^{(m)} \text{ otherwise,} \end{cases}$$

     where $R\left(\rho^{(m)}, \rho^*\right) = \frac{\prod_{k=1}^{n_T} f\left((\mathbf{Y}^{(m+1)}_{C_k}, y_{E_k})|\rho^*\right)}{\prod_{k=1}^{n_T} f\left((\mathbf{Y}^{(m+1)}_{C_k}, y_{E_k})|\rho^{(m)}\right)} \frac{\rho^*\left(1-\frac{\rho^*}{\rho_{\max}}\right)}{\rho^{(m)}\left(1-\frac{\rho^{(m)}}{\rho_{\max}}\right)}$.

5.3.2 Results

Simulation from the true model

Table 5.1 shows results from applying the our censored likelihood approach with Bayesian data augmentation to data simulated from the true extreme SAR model (see Section 5.2.1). We start with the base case of $\rho = 0.2$, a threshold T such that $1 - n_T/n = 0.975$, and no added noise (first row). Each subsequent row of the table represents a change to either the true ρ, the threshold, or the presence of noise, from this base case. Parameter estimates are taken to be posterior means. In all cases, the 95% highest posterior density interval (HPDI) covers the true value of ρ.

Table 5.1: Results from the censored likelihood approach for data simulated from the true model. Columns represent (from left to right): the true value of ρ, threshold proportion, presence of light-tailed noise, posterior mean, and 95% highest posterior density interval (HPDI) for ρ.

ρ	$1 - n_T/n$	noise	$\hat{\rho}$	95% HPDI
0.2	0.975	no	0.20002	(0.19991, 0.20012)
0.1	0.975	no	0.09997	(0.09974, 0.10018)
0.2	0.950	no	0.19996	(0.19986, 0.20006)
0.2	0.975	yes	0.19998	(0.19987, 0.20007)

MCMC chains were run for 31,000 iterations with a burn-in of 1,000 iterations, then thinned to 10,000 final iterations. Acceptance rates for ρ proposals ranged from 40% to 44%, and inspection of trace plots indicated that the ρ chains were mixing well. Some example diagnostic plots are given in Figure 5.3. In addition, Figure 5.4 illustrates mixing of the censored components below the threshold that are sampled in the data augmentation step. As desired, the sampled components cover a large range of values below the threshold.

Simulation from a Brown-Resnick process

While Table 5.1 indicates that the censored likelihood approach works quite well for data simulated from the true model, the method is not robust to model misspecification. Applying our MCMC algorithm to simulations from a Brown-Resnick process (see Section 5.2.2) results in

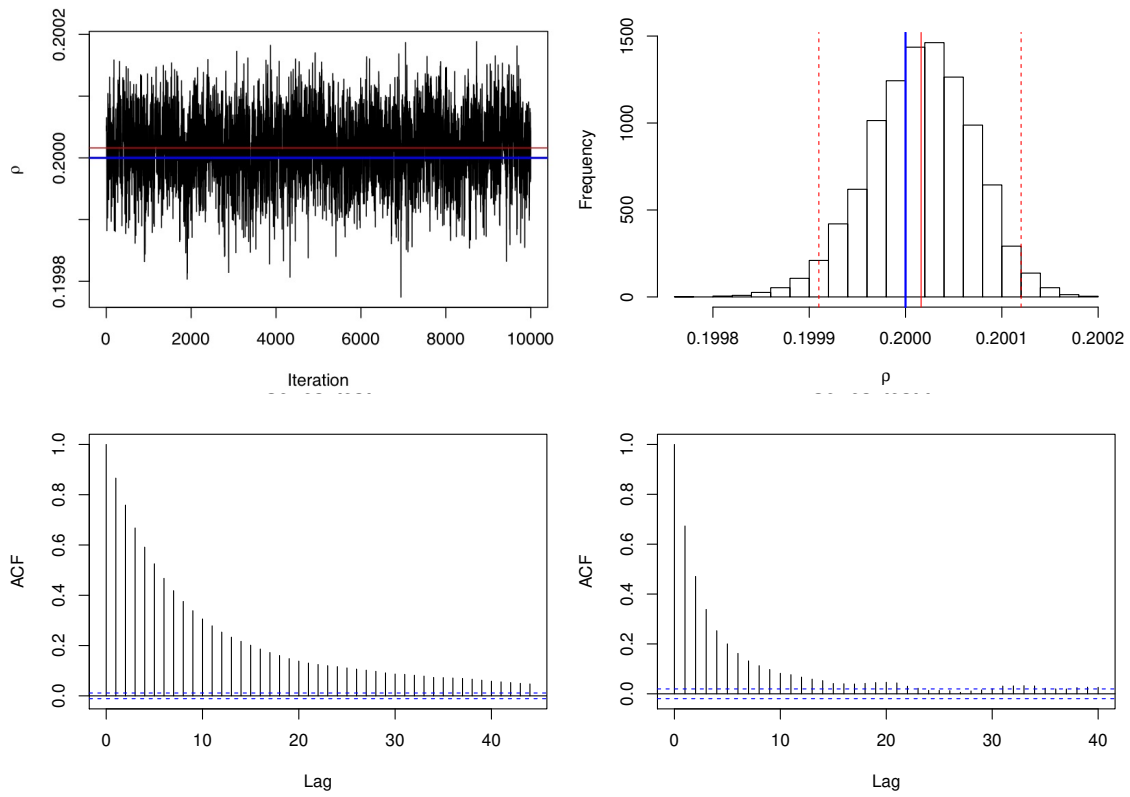

Figure 5.3: Example MCMC diagnostic plots for the ρ parameter chain for data simulated from the true model in the base case with true $\rho = 0.2$, $1 - n_T/n = 0.975$, and no added noise. Diagnostic plots for other cases are very similar. Top: trace plot (left) and histogram (right) corresponding to the thinned chain. Blue lines indicate the true value of ρ, while red lines indicate the posterior mean (solid) and 95% HPDI (dotted). Bottom: autocorrelation plots before (left) and after (right) thinning.

$\hat{\rho}$ being driven to zero. A likely explanation is that the extreme SAR model is not flexible enough. Even though realizations from the extreme SAR model do not have to correspond exactly to the point masses in the discrete angular measure, we hypothesize that the likelihood decays too rapidly away from these point masses. The Brown-Resnick simulations deviate too much from the extreme SAR model, so the likelihood does not agree with the data.

In diagnosing this issue, we examined differences in likelihood contributions by threshold exceedance between a starting value ρ_0 and a proposed value $\rho^* > \rho_0$, where each threshold exceedance is a 20×20 regular grid. Certain grids exhibit extremely large likelihood differences (see Figure 5.5 for example). Some grids with large likelihood differences contain a single component that exceeds the threshold by a large margin. For these grids, our intuition is that the

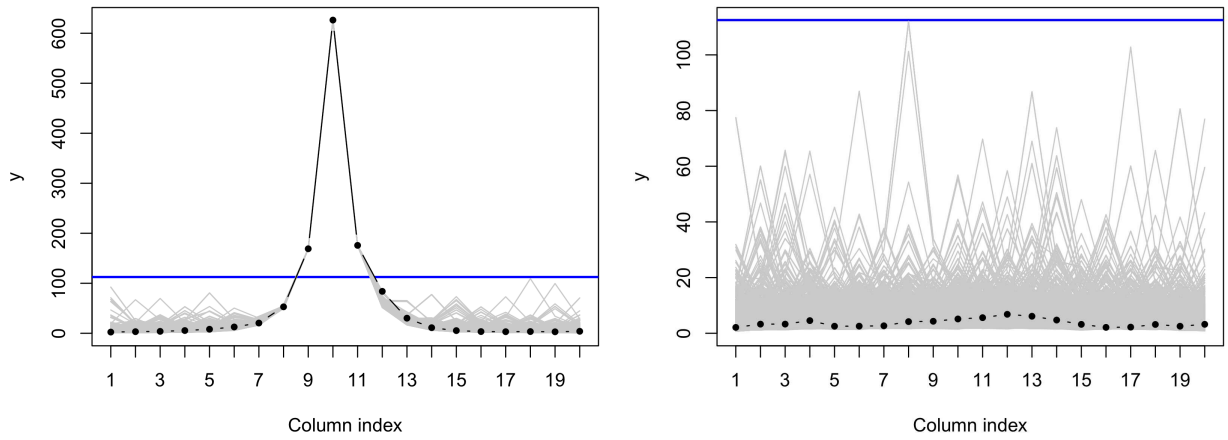

Figure 5.4: Illustration of **y** sampling for data simulated from the true model in the base case with true $\rho = 0.2$, $1 - n_T/n = 0.975$, and no added noise. The left panel corresponds to a row of a single 20×20 grid containing three "exact" components above the threshold (blue line), while the right panel corresponds to another row of the same grid with all components below the threshold. Black points are the original observations for each row, while gray lines show accepted **y** proposals over the final 10,000 iterations.

model wants ρ as small as possible because otherwise we should have observed some neighboring grid cells that exceed the threshold as well. Other grids with large likelihood differences contain large, irregular groupings of cells which exceed the threshold (e.g., Figure 5.5). Again, the best SAR approximation for these grids is obtained for ρ near zero, which corresponds to the observed pattern being obtained purely by chance.

Gridded precipitation observations

Results for the gridded precipitation observations described in Section 5.2.3 are similar to those for the Brown-Resnick simulations, and we omit them here.

In the next section, we propose an alternative inference procedure for the extreme SAR model. This method is not likelihood-based, and instead approaches model fitting via the TPDM. Using this approach, we are able to obtain reasonable estimates of dependence even in the case of model misspecification.

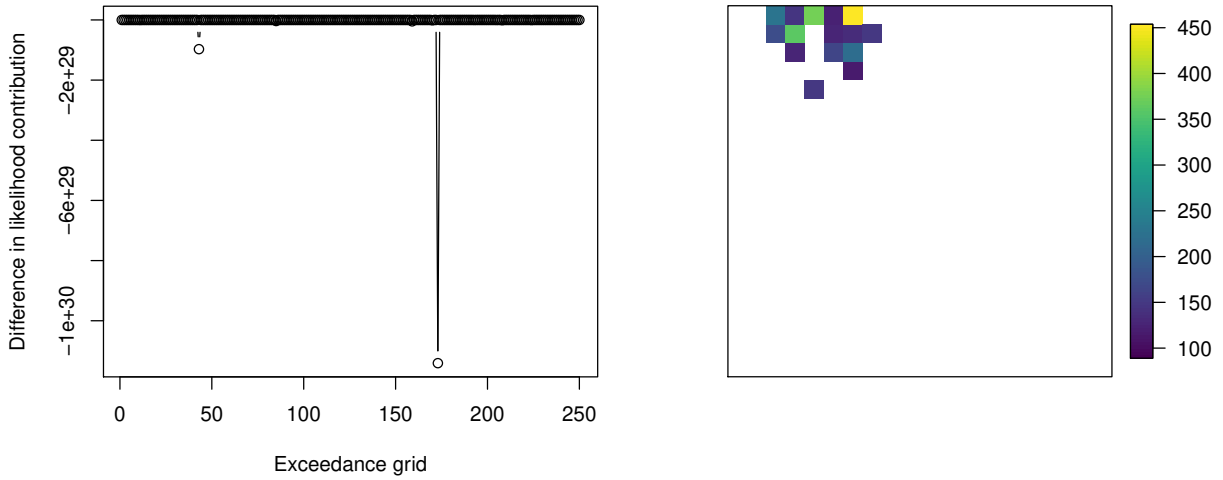

Figure 5.5: Example of issues with the censored likelihood approach when applied to data simulated from a Brown-Resnick process ($\lambda = 0.431$, $\kappa = 1.148$). Left: differences in likelihood contributions by threshold exceedance grid between a proposed value of $\rho = 0.1750$ and starting value of $\rho = 0.1749$, after one full iteration of sampling components below the threshold. Right: components above the threshold for grid 173, which corresponds to the most negative difference in likelihood contribution in this case.

5.4 Fitting via the TPDM

5.4.1 Methods

The TPDM $\Sigma_{\mathbf{Y}}$ is a useful summary of the extremal dependence of $\mathbf{Y}$. Hence, in this section, we propose an approach to model fitting that is based on the TPDM. The goal of this approach is to find the value of ρ whose corresponding theoretical TPDM most closely matches the observed tail pairwise dependence structure with respect to the Frobenius norm. Implementing this method involves several considerations. First, we must obtain an estimate of the TPDM from the data. We will consider two threshold-based estimators for the TPDM, one defined using a d-dimensional vector norm and the other a pairwise estimator. It is well-known that in cases of moderate to weak extremal dependence, threshold-based estimators tend to overestimate dependence (Huser et al., 2016). To mitigate this bias, we propose a simple method to obtain a bias-corrected estimate of the TPDM, $\widetilde{\Sigma}_{\mathbf{Y}}$. Finally, we find $\hat{\rho}$ such that we minimize the squared Frobenius norm of $\Sigma_{\mathbf{Y}} - \widetilde{\Sigma}_{\mathbf{Y}}$, and estimate uncertainty via bootstrapping. Below we describe each of these steps in more detail.

Estimation of the TPDM

We consider two estimators of the TPDM, each of which replaces the true angular measure with an empirical estimate. The first estimator was proposed by Cooley and Thibaud (2018), and extends an estimator given by Larsson and Resnick (2012) to dimension $d > 2$. Let $\mathbf{y}_t = (y_{t1}, y_{t2}, \ldots, y_{td}), t = 1, \ldots, n$, denote our observations with the appropriate regularly varying marginals. Note that this may require first performing a suitable marginal transformation such as the one described in Section 5.2.3. Let the radial component $r_t = ||\mathbf{y}_t||_2$, and the angular component $\boldsymbol{\theta}_t = \mathbf{y}_t r_t^{-1}$. Set a suitably high threshold r_0 for the radial components, and let $n_{\text{exc}} = \sum_{t=1}^{n} \mathbb{I}(r_t > r_0)$ be the number of threshold exceedances. In practice, it is convenient to define r_0 in terms of an empirical quantile of the data, which we will denote by q_{r_0}. We define

$$\hat{\sigma}_{\mathbf{Y}ij} = \hat{m} \int_{\mathbb{S}_{+(2)}^{d-1}} \theta_i \theta_j d\hat{N}_{\mathbf{Y}}(\boldsymbol{\theta}) = \frac{\hat{m}}{n_{\text{exc}}} \sum_{t=1}^{n} \theta_{ti} \theta_{tj} \mathbb{I}(r_t > r_0), \tag{5.13}$$

where $N_{\mathbf{Y}}(\cdot) = m^{-1} H_{\mathbf{Y}}(\cdot)$, and $\hat{m}$ is an estimate of $H_{\mathbf{Y}}(\mathbb{S}_{+(2)}^{d-1})$. In application, because we either simulate or pre-process our data to have common unit scale, $m = d$ and does not need to be estimated.

Let $\hat{\Sigma}_{\mathbf{Y}} = \left(\hat{\sigma}_{\mathbf{Y}ij}\right)_{i,j=1,\ldots,d}$. Note that $\hat{\Sigma}_{\mathbf{Y}}$ is a symmetric matrix by definition. $\hat{\Sigma}_{\mathbf{Y}}$ is also guaranteed to be positive semidefinite, as $\hat{\Sigma}_{\mathbf{Y}} = n_{\text{exc}}^{-1} \hat{m} \hat{\boldsymbol{\Theta}}^{\top} \hat{\boldsymbol{\Theta}}$, where $\hat{\boldsymbol{\Theta}}$ is a matrix whose rows are the vectors $\boldsymbol{\theta}_t$ for which $r_t > r_0$. The estimator (5.13) is of a form widely used in extreme analyses. Resnick (2004) and Larsson and Resnick (2012) show consistency of such estimators using standard intermediate asymptotic arguments.

Since the TPDM summarizes pairwise dependence, heuristically it is conceivable that estimation of $\Sigma_{\mathbf{Y}}$ could be improved by focusing on extreme instances of each pair of components. Thus we also consider a pairwise estimator of the TPDM. For each pair (i, j) of component indices ($i = 1, \ldots, d,\ j = 1, \ldots, d$), we denote $\mathbf{y}_t(i, j) = (y_{ti}, y_{tj})$ for $t = 1, \ldots n$. Let $r_t(i, j) = ||\mathbf{y}_t(i, j)||_2$, and $\boldsymbol{\theta}_t(i, j) = \left(\theta_{ti}(i, j), \theta_{tj}(i, j)\right) = \mathbf{y}_t(i, j) r_t(i, j)^{-1}$. We choose a high pairwise norm threshold $r_0(i, j)$ and let $n_{\text{exc}}(i, j) = \sum_{t=1}^{n} \mathbb{I}\left(r_t(i, j) > r_0(i, j)\right)$. It is convenient to set $r_0(i, j)$ to a

common empirical quantile q_{r_0} for all pairs (i, j). We define

$$\hat{\sigma}^{\text{pair}}_{\mathbf{Y}ij} = \frac{2}{n_{\text{exc}}(i,j)} \sum_{t=1}^{n} \theta_{ti}(i,j)\theta_{tj}(i,j) \mathbb{I}\big(r_t(i,j) > r_0(i,j)\big). \tag{5.14}$$

The factor of two arises because the data are simulated or pre-processed to have a common unit scale and we are considering pairs.

Let $\hat{\Sigma}^{\text{pair}}_{\mathbf{Y}} = \left(\hat{\sigma}^{\text{pair}}_{\mathbf{Y}ij}\right)_{i,j=1,\ldots,d}$. Note that $\hat{\Sigma}^{\text{pair}}_{\mathbf{Y}}$ is symmetric with ones on the diagonal, as $\hat{\sigma}^{\text{pair}}_{\mathbf{Y}ij} = \hat{\sigma}^{\text{pair}}_{\mathbf{Y}ji}$ for all (i, j), and $\hat{\sigma}^{\text{pair}}_{\mathbf{Y}ii} = 2n_{\text{exc}}(i,i)^{-1} \sum_{t=1}^{n} \frac{y_{ti}^2}{2y_{ti}^2} \mathbb{I}(r_t(i,i) > r_0(i,i)) = 1$ for all i. Unlike $\hat{\Sigma}_{\mathbf{Y}}$, the pairwise estimator $\hat{\Sigma}^{\text{pair}}_{\mathbf{Y}}$ is not guaranteed to be positive semidefinite. Other theoretical properties of $\hat{\Sigma}^{\text{pair}}_{\mathbf{Y}}$ are under development and require understanding heretofore unexplored bivariate (and more generally, lower-dimensional) angular measures of higher-dimensional regularly varying random vectors. This work is outside the scope of this dissertation.

Bias correction of the estimated TPDM

We anticipate that finite-sample estimates of $\Sigma_{\mathbf{Y}}$ will be biased. In particular, threshold-based estimators are known to overestimate dependence when true dependence is moderate to weak (Huser et al., 2016), such as is the case when the distance between grid cell pairs is moderate to large for the extreme SAR. Naively minimizing $||\Sigma_{\mathbf{Y}}(\rho) - \hat{\Sigma}_{\mathbf{Y}}||_F^2$ or $||\Sigma_{\mathbf{Y}}(\rho) - \hat{\Sigma}^{\text{pair}}_{\mathbf{Y}}||_F^2$ would result in an overestimate of ρ. Instead, we propose an intermediate step to obtain a bias-corrected estimate of the TPDM, denoted by $\widetilde{\Sigma}_{\mathbf{Y}}$. Our bias correction approach makes use of our knowledge that the true pairwise tail dependence should be close to zero at large distances. We provide more details on our proposed approach in Section 5.4.2.

Estimation of ρ

Recall from Section 4.3.2 that the theoretical TPDM for the common-scale formulation of the extreme SAR model is

$$\Sigma_{\mathbf{Y}} = \widetilde{A}\widetilde{A}^{\top} = \text{diag}(1/||A_{1\cdot}||_2, \ldots, 1/||A_{d\cdot}||_2) AA^{\top} \text{diag}(1/||A_{1\cdot}||_2, \ldots, 1/||A_{d\cdot}||_2), \tag{5.15}$$

where $A = (I - \rho W)^{-1}$. We will use the notation $\Sigma_{\mathbf{Y}}(\rho)$ to emphasize the dependence of the TPDM on the parameter ρ. After estimating $\hat{\Sigma}_{\mathbf{Y}}$ and performing bias correction to obtain $\widetilde{\Sigma}_{\mathbf{Y}}$, the final step is to find

$$\hat{\rho} = \underset{\rho \in (0\,,\, \min\{|\lambda_1|^{-1}, |\lambda_d|^{-1}\})}{\arg\min} ||\Sigma_{\mathbf{Y}}(\rho) - \widetilde{\Sigma}_{\mathbf{Y}}||_F^2, \tag{5.16}$$

where $\lambda_1 < \cdots < \lambda_d$ are the ordered eigenvalues of W, and the restriction on ρ follows from Proposition 4.3.1. We find $\hat{\rho}$ using numerical optimization in `R`.

We estimate uncertainty via a nonparametric bootstrap procedure. Specifically, we first sample with replacement from the original n observations (each of which is d-dimensional), ensuring the preservation of the dependence structure. This step would occur after marginal transformation if necessary. Then for each bootstrap replicate, we perform the inference procedure described above, starting with estimating the TPDM and ending with estimating ρ. Theoretical properties of the bootstrap are just beginning to be examined by the extremes community (Zhou, 2018).

5.4.2 Results

Below we investigate fitting via the TPDM for our three categories of data: simulations from the true extreme SAR model, simulations from a Brown-Resnick process with similar pairwise tail dependence, and gridded precipitation observations.

Simulation from the true model

For simulations from the true extreme SAR model, we know both the true ρ and the true $\Sigma_{\mathbf{Y}}(\rho)$. Below, we use the base case of data simulated from the true model with $\rho = 0.2$ with no added noise to illustrate our modeling choices in each of the steps described in Section 5.4.1. Estimation of $\Sigma_{\mathbf{Y}}$ and hence of ρ depends on the threshold quantile. We choose the 0.99 quantile for our initial illustration. We then explore the effect of varying the threshold quantile, and present results for various combinations of the true ρ, threshold quantile, and presence of light-tailed noise.

Estimation of the TPDM

Figure 5.6 shows results of estimating the TPDM for the base case using the two estimators defined in Section 5.4.1. As expected, both estimators tend to overestimate the tail pairwise dependence, especially at moderate to large distances when true dependence is weak. $\hat{\Sigma}_{\mathbf{Y}}$ exhibits more variability than $\hat{\Sigma}_{\mathbf{Y}}^{\text{pair}}$, but has substantially lower bias on average. For this reason, and because $\hat{\Sigma}_{\mathbf{Y}}$ has nice theoretical properties as outlined in Section 5.4.1, we choose to use this estimator for the remainder of the analysis.

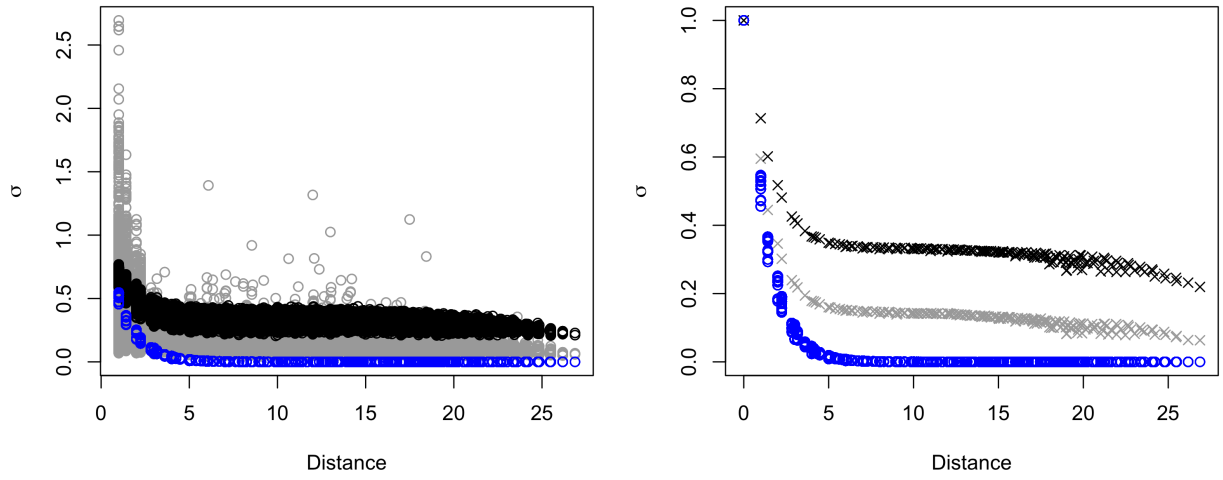

Figure 5.6: Comparison of two estimators of $\Sigma_{\mathbf{Y}}$ for data simulated from the true extreme SAR model with $\rho = 0.2$, $q_{r_0} = 0.99$, and no added noise. True values $\{\sigma_{\mathbf{Y}ij}\}$ (blue) and estimates $\{\hat{\sigma}_{\mathbf{Y}ij}\}$ (gray) and $\{\hat{\sigma}_{\mathbf{Y}ij}^{\text{pair}}\}$ (black) are plotted against pairwise Euclidean distance. The right plot shows means of the estimates at each distance, revealing that $\hat{\Sigma}_{\mathbf{Y}}$ is less biased on average than $\hat{\Sigma}_{\mathbf{Y}}^{\text{pair}}$.

Bias correction of the estimated TPDM

Based on the results of estimating $\Sigma_{\mathbf{Y}}$, we propose a simple bias correction method which reflects our knowledge that the true pairwise extremal dependence should be close to zero when the distance is large enough between grid cell centers. We temporarily view $\hat{\sigma}_{\mathbf{Y}ij}$ as a function of pairwise distance, and fit the curve $\hat{\sigma} = \beta_0 e^{-\beta_1 * \text{distance}} + \beta_2$ to the estimated TPDM values (see left panel of Figure 5.7). Then $\hat{\beta}_2$ is our estimate for the horizontal asymptote. In the base case of $\rho = 0.2$ and $q_{r_0} = 0.99$, we obtain $\hat{\beta}_2 = 0.134$. Note that $\sigma_{\mathbf{Y}ii}$ is unbiased on average (see right panel of Figure 5.6), so it is only necessary to bias-correct the off-diagonal entries of $\hat{\Sigma}_{\mathbf{Y}}$.

We make a simplifying assumption of constant bias for for pairwise distances greater than zero. Our bias-corrected estimate of the TPDM is then $\widetilde{\Sigma}_{\mathbf{Y}} = \left(\widetilde{\sigma}_{\mathbf{Y}ij}\right)_{i,j=1,\ldots,d}$, where

$$\widetilde{\sigma}_{\mathbf{Y}ij} = \begin{cases} \hat{\sigma}_{\mathbf{Y}ij}, & i = j, \\ \left(\hat{\sigma}_{\mathbf{Y}ij} - \hat{\beta}_2\right)_+, & i \neq j. \end{cases} \tag{5.17}$$

We take the positive part in the latter case because we know that $\Sigma_{\mathbf{Y}}$ is entrywise nonnegative for the true extreme SAR under our model specifications. The right panel of Figure 5.7 compares the bias-corrected $\widetilde{\Sigma}_{\mathbf{Y}}$ to the true $\Sigma_{\mathbf{Y}}$ for the base case.

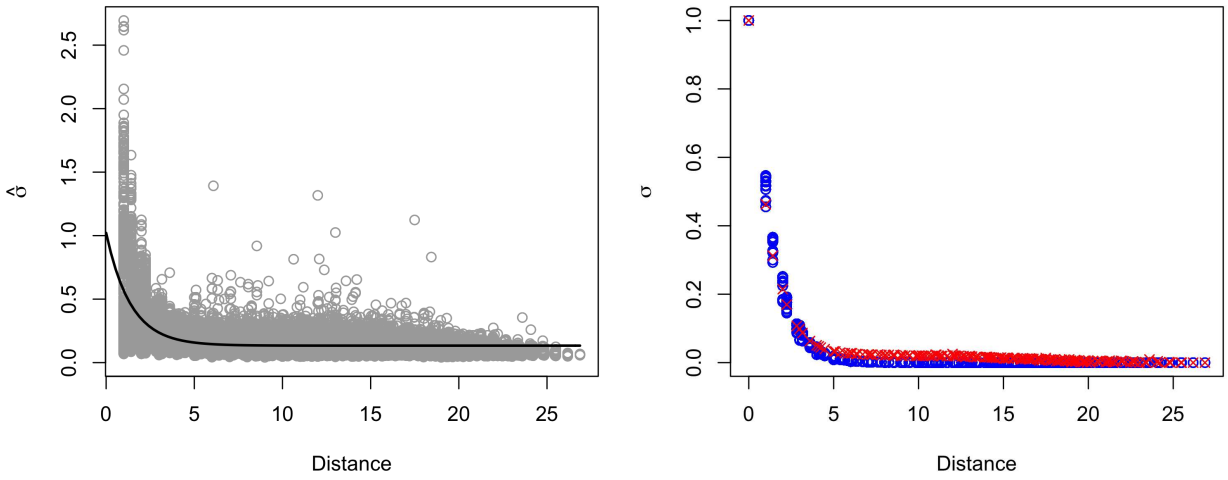

Figure 5.7: Illustration of the proposed approach to bias correction of $\hat{\Sigma}_{\mathbf{Y}}$ for data simulated from the true model. Left: exponential curve (black line) fitted to $\{\hat{\sigma}_{\mathbf{Y}ij}\}$ as a function of pairwise Euclidean distance. Right: means of bias-corrected $\{\widetilde{\sigma}_{\mathbf{Y}ij}\}$ at each distance (red) compared to the true TPDM values $\{\sigma_{\mathbf{Y}ij}\}$ (blue) for the extreme SAR model when the true $\rho = 0.2$.

Better results could potentially be obtained by letting the bias correction vary by distance. We also explored bias-correction via shrinkage or tapering, but there was not a clear choice of shrinkage target or taper function. These extensions could be avenues of future investigation.

Estimation of ρ

For the base case, minimizing the squared Frobenius norm $||\Sigma_{\mathbf{Y}}(\rho) - \widetilde{\Sigma}_{\mathbf{Y}}||_F^2$ results in an estimate of $\hat{\rho} = 0.197$, which is quite close to the true value of $\rho = 0.2$. The corresponding 95%

bootstrap confidence interval based on 1,000 bootstrap replicates is (0.192, 0.202), which covers the true value.

Estimation of a risk region occurrence probability

We can also compare the probability of an event occurring in a given risk region between the true and fitted extreme SAR models. Let us define a risk region by

$$\mathscr{R} = \{\mathbf{y} \in \mathbb{R}_+^d : \max_{i \in \mathscr{D}_{\mathscr{R}}} y_i > y_0\}, \tag{5.18}$$

where $\mathscr{D}_{\mathscr{R}} \subset \mathscr{D} = \{1, \ldots, d\}$. Our goal is to estimate $p_{\mathscr{R}} = \mathbb{P}(\mathbf{Y}_t \in \mathscr{R})$. We choose this particular form of the risk region for ease of calculation, but note that occurrence probabilities can be estimated in a similar manner for any risk region of interest. For the extreme SAR model, we first derive

$$\begin{aligned}
\nu(\mathscr{R}) &= \int_{(r,\boldsymbol{\theta}) \in \mathscr{R}} 2r^{-3} dr dH(\boldsymbol{\theta}) \\
&= \int_{\boldsymbol{\theta} \in \mathbb{S}_{+(2)}^{d-1}} \int_{r=\min_{i \in \mathscr{D}_{\mathscr{R}}} \frac{y_0}{\theta_i}}^{\infty} 2r^{-3} dr dH(\boldsymbol{\theta}) \\
&= \int_{\boldsymbol{\theta} \in \mathbb{S}_{+(2)}^{d-1}} \max_{i \in \mathscr{D}_{\mathscr{R}}} \left(\frac{\theta_i}{y_0}\right)^2 dH(\boldsymbol{\theta}) \\
&= \int_{\boldsymbol{\theta} \in \mathbb{S}_{+(2)}^{d-1}} \max_{i \in \mathscr{D}_{\mathscr{R}}} \frac{\theta_i^2}{y_0^2} \, ||\widetilde{\mathbf{a}}_j||_2^2 \, \mathbb{I}\left(\boldsymbol{\theta} = \frac{\widetilde{\mathbf{a}}_j}{||\widetilde{\mathbf{a}}_j||_2}\right) \\
&= \sum_{j=1}^{d} \max_{i \in \mathscr{D}_{\mathscr{R}}} \frac{\left(\frac{\widetilde{A}_{ij}}{||\widetilde{\mathbf{a}}_j||_2}\right)^2}{y_0^2} \, ||\widetilde{\mathbf{a}}_j||_2^2 \\
&= \frac{1}{y_0^2} \sum_{j=1}^{d} \max_{i \in \mathscr{D}_{\mathscr{R}}} \widetilde{A}_{ij}^2.
\end{aligned} \tag{5.19}$$

Next we obtain our occurrence probability estimate. We assume n is fixed and large enough such that the convergence (1.11) implies the approximate equality

$$n\mathbb{P}\left(\frac{\mathbf{Y}}{c_n} \in S\right) \approx \nu(S),$$

where the appropriate normalizing constant in this case is $c_n = n^{1/2}$. Letting $S = n^{-1/2}\mathscr{R}$, we get

$$\begin{aligned} n\mathbb{P}\left(n^{-1/2}\mathbf{Y}_t \in n^{-1/2}\mathscr{R}\right) &\approx \nu(n^{-1/2}\mathscr{R}) \\ \Rightarrow \mathbb{P}\left(\mathbf{Y}_t \in \mathscr{R}\right) &\approx n^{-1}\nu(n^{-1/2}\mathscr{R}) \\ \Rightarrow \mathbb{P}\left(\mathbf{Y}_t \in \mathscr{R}\right) &\approx \nu(\mathscr{R}), \end{aligned} \tag{5.20}$$

where the last line follows from the scaling property (1.12) of ν. Since the risk region occurrence probability depends on $\widetilde{A}$, which in turn depends on ρ, we will use the notation $p_{\mathscr{R}}(\rho)$.

As an example, suppose $\mathscr{D}_{\mathscr{R}} = \{174, 193, 194, 195, 214\}$, corresponding to a plus-shaped region of grid cells within the 20×20 grid, and let $y_0 = 30$. For the true extreme SAR model with $\rho = 0.2$, we can use (5.19) and (5.20) to obtain $p_{\mathscr{R}}(\rho) \approx 4.57 \times 10^{-3}$. Our fitted model with $\hat{\rho} = 0.197$ yields the estimate $\hat{p}_{\mathscr{R}} = p_{\mathscr{R}}(\hat{\rho}) \approx 4.63 \times 10^{-3}$, which is very similar. Plugging in our 1,000 bootstrapped estimates for ρ, we obtain a corresponding 95% confidence interval for the probability estimate of $(4.52 \times 10^{-3}, 4.70 \times 10^{-3})$, which covers the true $p_{\mathscr{R}}(\rho)$. For reference, the empirical occurrence probability based on our simulation of 10,000 grids was 5.70×10^{-3}.

Sensitivity to threshold selection, true dependence, and noise

Here, we investigate the sensitivity of our estimation method to threshold, different levels of true dependence, and presence of light-tailed noise. Estimation of $\Sigma_{\mathbf{Y}}$, and therefore of ρ, depends on the choice of threshold. Figure 5.8 illustrates the effect of threshold selection on estimates of ρ. We let the threshold r_0 be the empirical q_{r_0}-quantile of data simulated from the true model, for $q_{r_0} = 0.95$, 0.975, 0.99, 0.995, and 0.999. Figure 5.8 shows the corresponding estimates for ρ using our chosen bias-corrected method (in red) as well as the estimates we would have obtained if we did not bias correct $\hat{\Sigma}_{\mathbf{Y}}$ and instead minimized $||\Sigma_{\mathbf{Y}}(\rho) - \hat{\Sigma}_{\mathbf{Y}}||_F^2$ (in black). We see that as q_{r_0} increases, point estimates from both methods get closer to the true value of ρ. Also in both cases, the uncertainty increases with q_{r_0} since the sample size (number of exceedances) decreases. As we would expect based on the consistency of $\hat{\Sigma}_{\mathbf{Y}}$, omitting bias correction of $\hat{\Sigma}_{\mathbf{Y}}$ leads to accurate estimates of ρ when q_{r_0} is high enough (e.g., when $q_{r_0} =$

0.999). However, at less extreme threshold quantiles corresponding to a reasonable number of exceedances in practice, our bias-corrected method performs much better and is effective in reducing the bias in estimation of ρ.

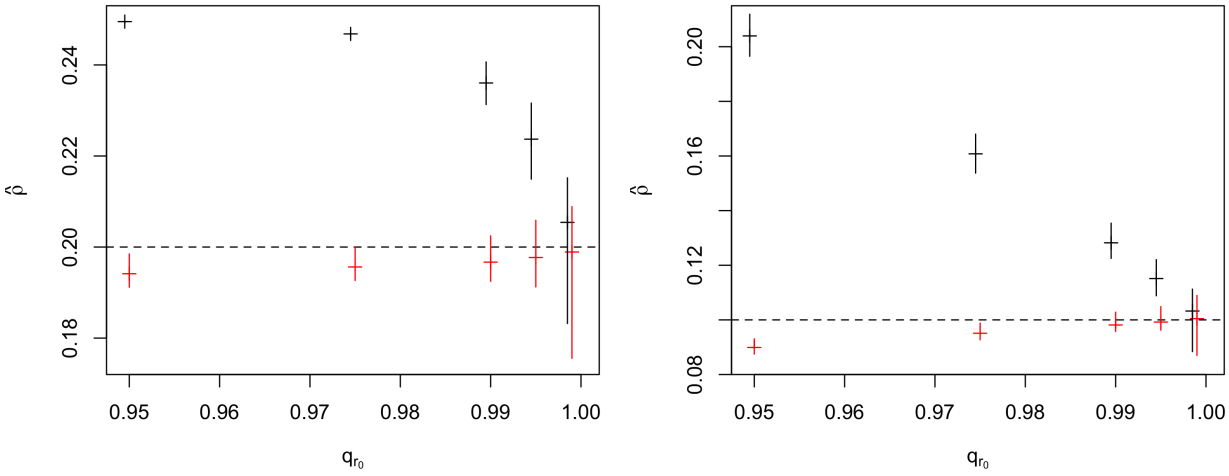

Figure 5.8: Effect of varying threshold quantile q_{r_0} on estimates of ρ using our proposed bias-corrected method (red) and estimates without bias-correction (black) for data simulated from the true extreme SAR model with $\rho = 0.2$ (left) and $\rho = 0.1$ (right). Dashed lines indicate the true value of ρ. Solid vertical lines represent 95% bootstrap confidence intervals.

In addition to varying the threshold quantile, we also explored different values of the true ρ and the effect of adding light-tailed noise. Similar to Table 5.1, Table 5.2 gives point estimates for ρ along with 95% confidence intervals based on 1,000 bootstrap replicates for four cases. The first row starts with the base case of $\rho = 0.2$, $q_{r_0} = 0.990$, and no added noise. Each subsequent row of the table represents a change to either the true ρ, the norm threshold quantile, or the presence of light-tailed noise, from this base case. We see that the estimates for ρ are slightly biased low in all cases, and as expected the bias increases when the threshold is lowered or noise is added. However, the 95% bootstrap confidence interval covers the true value of ρ in all cases.

Simulation from a Brown-Resnick process

In contrast to simulations from the true extreme SAR model, there is no true ρ for simulations from a Brown-Resnick process. However, given the range and smoothness parameters of

Table 5.2: Results from fitting via the TPDM for data simulated from the true model. Columns represent (from left to right): the true value of ρ, norm threshold quantile, presence of light-tailed noise, point estimate, and 95% bootstrap confidence interval for ρ.

ρ	q_{r_0}	noise	$\hat{\rho}$	95% CI
0.2	0.990	no	0.197	(0.192, 0.202)
0.1	0.990	no	0.098	(0.097, 0.103)
0.2	0.975	no	0.196	(0.193, 0.200)
0.2	0.990	yes	0.196	(0.191, 0.202)

the Brown-Resnick process, we can derive the true Σ^{BR}, whose entries σ_{ij}^{BR} are given by (5.6). (Note that we know σ_{ij}^{BR} as a function of the distance between grid cell centers, whereas for the extreme SAR model we know $\sigma_{\mathbf{Y}ij}$ for each pair of grid cells.) We can also calculate the theoretical occurrence probability for any risk region of interest.

Below, we use the base case of data simulated from a Brown-Resnick process with $\lambda = 0.431$, $\kappa = 1.148$ to illustrate the steps of our inference procedure. Recall from Section 5.2.2 that these particular Brown-Resnick parameters were chosen to match as closely as possible the pairwise tail dependence of an extreme SAR model with $\rho = 0.2$. We choose the 0.99 quantile for the threshold in our initial illustration, then explore the effect of varying the threshold quantile.

Estimation of the TPDM

Figure 5.9 shows results of estimating the TPDM for the base Brown-Resnick case using the two estimators defined in Section 5.4.1. As with data simulated from the true extreme SAR model (recall Figure 5.6), we see that $\hat{\Sigma}_{\mathbf{Y}}$ is more variable but less biased on average than $\hat{\Sigma}_{\mathbf{Y}}^{\text{pair}}$, and choose to use $\hat{\Sigma}_{\mathbf{Y}}$ for the rest of the analysis.

Bias correction of the estimated TPDM

Using the same bias correction procedure as for simulations from the true model, we obtain a bias-corrected estimate $\widetilde{\Sigma}_{\mathbf{Y}}$ with entries given by (5.17). The left panel of Figure 5.10 shows the exponential curve fitted to the estimated tail pairwise dependence measures. Our resulting estimate for the horizontal asymptote is $\hat{\beta}_2 = 0.045$, and we subtract this constant bias correction

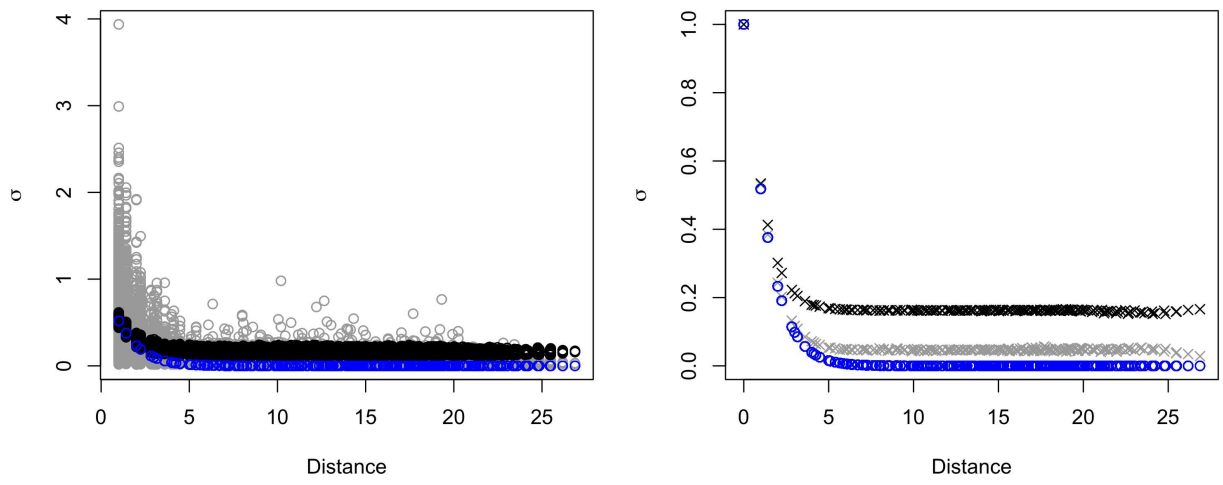

Figure 5.9: Comparison of two estimators of $\Sigma_{\mathbf{Y}}$ for data simulated from a Brown-Resnick process with $\lambda = 0.431$, $\kappa = 1.148$, and $q_{r_0} = 0.99$. True pairwise extremal dependence measures $\{\sigma_{ij}^{BR}\}$ (blue) and estimates $\{\hat{\sigma}_{\mathbf{Y}ij}\}$ (gray) and $\{\hat{\sigma}_{\mathbf{Y}ij}^{\text{pair}}\}$ (black) are plotted against pairwise Euclidean distance. The right plot shows means of the estimates at each distance, revealing that $\hat{\Sigma}_{\mathbf{Y}}$ is less biased on average than $\hat{\Sigma}_{\mathbf{Y}}^{\text{pair}}$.

factor from the off-diagonal entries of $\hat{\Sigma}_{\mathbf{Y}}$, taking the positive part as before. The right panel of Figure 5.10 compares the bias-corrected $\widetilde{\Sigma}_{\mathbf{Y}}$ to the true Σ^{BR} for the base case.

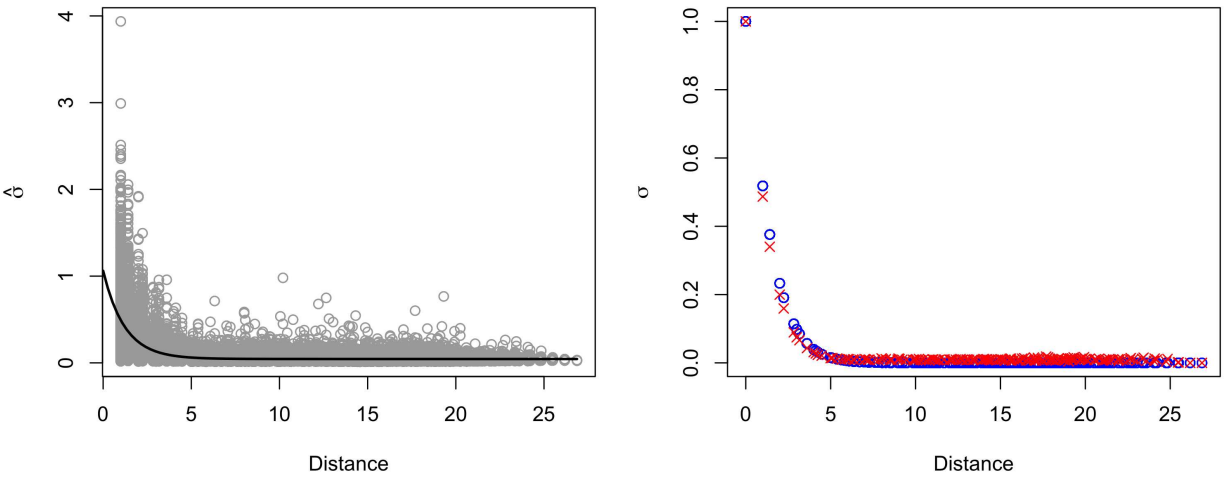

Figure 5.10: Illustration of bias correction of $\hat{\Sigma}_{\mathbf{Y}}$ for data simulated from a Brown-Resnick process. Left: exponential curve (black line) fitted to $\{\hat{\sigma}_{ij}^{BR}\}$ as a function of pairwise Euclidean distance. Right: means of bias-corrected $\{\widetilde{\sigma}_{\mathbf{Y}ij}\}$ at each distance (red) compared to the true pairwise extremal dependence measures $\{\sigma_{ij}^{BR}\}$ (blue) for a Brown-Resnick process with parameters $\lambda = 0.431$ and $\kappa = 1.148$.

Estimation of ρ

For the base Brown-Resnick case, minimizing $||\Sigma_{\mathbf{Y}}(\rho) - \widetilde{\Sigma}_{\mathbf{Y}}||_F^2$ results in an estimate of $\hat{\rho} = 0.192$ with a corresponding 95% bootstrap confidence interval of $(0.188, 0.199)$. Although there is no true ρ in the case of the Brown-Resnick simulations, we can assess our performance based on the known pairwise tail dependence. Figure 5.11 compares the theoretical TPDM $\Sigma_{\mathbf{Y}}(\hat{\rho})$ based on the fitted extreme SAR model to the true pairwise extremal dependence measures $\{\sigma_{ij}^{BR}\}$ of the Brown-Resnick process that we simulated from. Even though we are fitting the wrong model to the data, we are able to estimate dependence relatively well. Compared to the true pairwise dependence, the fitted extreme SAR model slightly underestimates dependence, especially at shorter distances.

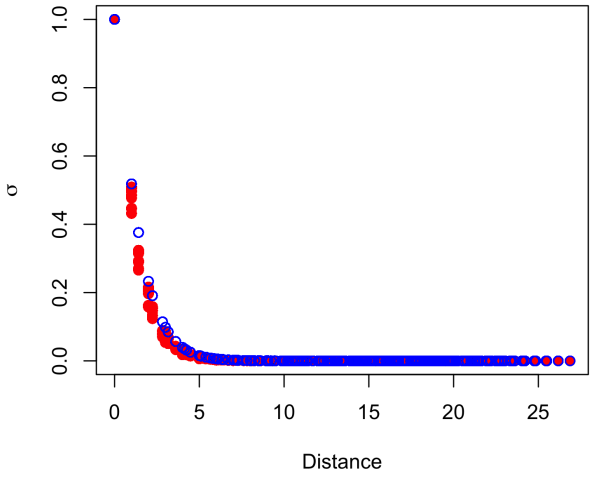

Figure 5.11: Comparison of the TPDM $\Sigma_{\mathbf{Y}}(\hat{\rho})$ for the fitted extreme SAR model (red) and the true Σ^{BR} of the Brown-Resnick process with $\lambda = 0.431$ and $\kappa = 1.148$ (blue).

Estimation of a risk region occurrence probability

We can also compare the theoretical occurrence probability for a given risk region to the estimated probability based on the fitted extreme SAR model. Again, we define a risk region $\mathscr{R}$ by (5.4.2). For the Brown-Resnick process, we can use the known exponent measure to calculate $p_{\mathscr{R}}^{BR} = \mathbb{P}(\mathbf{Y}_t \in \mathscr{R})$. Specifically,

$$\begin{aligned}\mathbb{P}(\mathbf{Y}_t \in \mathscr{R}) &= 1 - \mathbb{P}(Y_{ti} \leq y_0 \text{ for all } i \in \mathscr{D}_{\mathscr{R}}) \\ &= 1 - \exp\{-V(\mathbf{y_0})\}, \end{aligned} \tag{5.21}$$

where $\mathbf{y_0} = (y_0, \ldots, y_0)$ is a vector of dimension card$(\mathscr{D}_{\mathscr{R}})$. For the same plus-shaped example of $\mathscr{D}_{\mathscr{R}} = \{174, 193, 194, 195, 214\}$ and $y_0 = 30$ as before, we obtain $p_{\mathscr{R}}^{BR} = 4.18 \times 10^{-3}$. In comparison, the probability estimate for the fitted extreme SAR model is $p_{\mathscr{R}}(\hat{\rho}) = 4.71 \times 10^{-3}$, with a corresponding 95% bootstrap confidence interval of $(4.60 \times 10^{-3}, 4.77 \times 10^{-3})$. By fitting the extreme SAR model to simulations from a Brown-Resnick process, we slightly overestimate the occurrence probability. For our particular form of risk region, this corresponds to slightly underestimating the tail dependence, as we already saw from Figure 5.11. For reference, the empirical occurrence probability for our 10,000 simulated Brown-Resnick fields was 4.30×10^{-3}.

Sensitivity to threshold selection and true dependence

We investigate the effect of threshold selection on our estimation and inference of ρ in the same manner as we did for simulations from the true model. We let the threshold r_0 be the empirical q_{r_0}-quantile of data simulated from a given Brown-Resnick process, for $q_{r_0} = 0.95$, 0.975, 0.99, 0.995, and 0.999. Figure 5.12 shows the estimates for ρ at each threshold quantile using our chosen bias-corrected method (in red) as well as the estimates we would have obtained if we did not bias correct $\hat{\Sigma}_{\mathbf{Y}}$ and instead minimized $||\Sigma_{\mathbf{Y}}(\rho) - \hat{\Sigma}_{\mathbf{Y}}||_F^2$ (in black). The left panel shows results for data simulated from the base case, with Brown-Resnick parameters chosen to mimic pairwise tail dependence of the extreme SAR model with $\rho = 0.2$. The right panel presents analogous results for data simulated from a Brown-Resnick process with parameters chosen to match the pairwise tail dependence of the extreme SAR model with $\rho = 0.1$. We see that as q_{r_0} increases, the point estimates from the two methods get closer to each other, while the uncertainty of these estimates increases.

For easy comparison to Table 5.2, Table 5.3 gives point estimates for ρ along with 95% confidence intervals based on 1,000 bootstrap replicates for three cases. The first row starts with the base case of $q_{r_0} = 0.990$ and Brown-Resnick parameters $\lambda = 0.431$ and $\kappa = 1.148$, which give rise

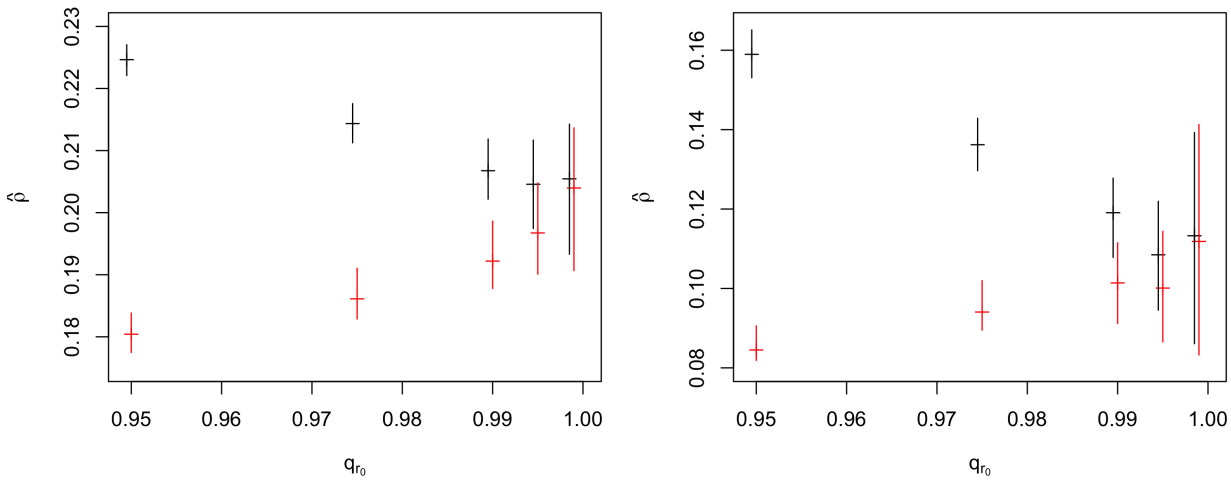

Figure 5.12: Effect of varying threshold quantile q_{r_0} on estimates of ρ using our proposed bias-corrected method (red) and estimates without bias-correction (black) for data simulated from a Brown-Resnick process with parameters $\lambda = 0.431$, $\kappa = 1.148$ (left) and $\lambda = 0.254$, $\kappa = 1.341$ (right). Solid vertical lines represent 95% bootstrap confidence intervals.

to the TPDM values closest to the extreme SAR model with $\rho = 0.2$ (see Section 5.2.2). Each subsequent row of the table represents a change to either the Brown-Resnick parameters (to those which minimize the squared error to TPDM values for the extreme SAR model with $\rho = 0.1$) or the norm threshold quantile from this base case.

Table 5.3: Results from fitting via the TPDM for data simulated from a Brown-Resnick process. Columns represent (from left to right): the range and smoothness parameters of the Brown-Resnick simulations, norm threshold quantile, point estimate, and 95% bootstrap confidence interval for ρ.

λ	κ	q_{r_0}	$\hat{\rho}$	95% CI
0.431	1.148	0.990	0.192	(0.188, 0.199)
0.254	1.341	0.990	0.101	(0.091, 0.112)
0.431	1.148	0.975	0.186	(0.183, 0.191)

Gridded precipitation observations

Here we apply our inference procedure to the CPC unified gauge-based analysis of daily precipitation introduced in Section 5.2.3. Recall we have $n = 9,027$ daily precipitation observations on a 20×20 grid that contains northeastern Colorado. As described in Section 5.2.3, the orig-

inal observations have been transformed to have approximately Fréchet marginals with $\alpha = 2$ and scale = 1. We select $q_{r_0} = 0.99$ as our threshold quantile. Following the same procedure as for simulations from the true model and simulations from a Brown-Resnick process, we first estimate the TPDM from the data, then bias-correct to obtain $\widetilde{\Sigma}_{\mathbf{Y}}$. The left panel of Figure 5.13 shows the fitted exponential curve used for bias correction, with estimated horizontal asymptote $\hat{\beta}_2 = 0.165$. Minimizing $||\Sigma_{\mathbf{Y}}(\rho) - \widetilde{\Sigma}_{\mathbf{Y}}||_F^2$ leads to an estimate of $\hat{\rho} = 0.236$, with corresponding 95% bootstrap confidence interval $(0.230, 0.241)$. The right panel of Figure 5.13 compares fitted TPDM $\Sigma_{\mathbf{Y}}(\hat{\rho})$ (in pink) to the means at each distance of the bias-corrected $\{\widetilde{\sigma}_{\mathbf{Y}ij}\}$ (in purple). The discrepancy shown here, as well as the fact that our estimate for ρ is near its upper bound of 0.25, indicate that our model is not able to full capture the longer-range dependence (i.e., larger storm sizes) apparent in the data. This is likely due, at least in part, to the rook neighborhood specification that we use for W.

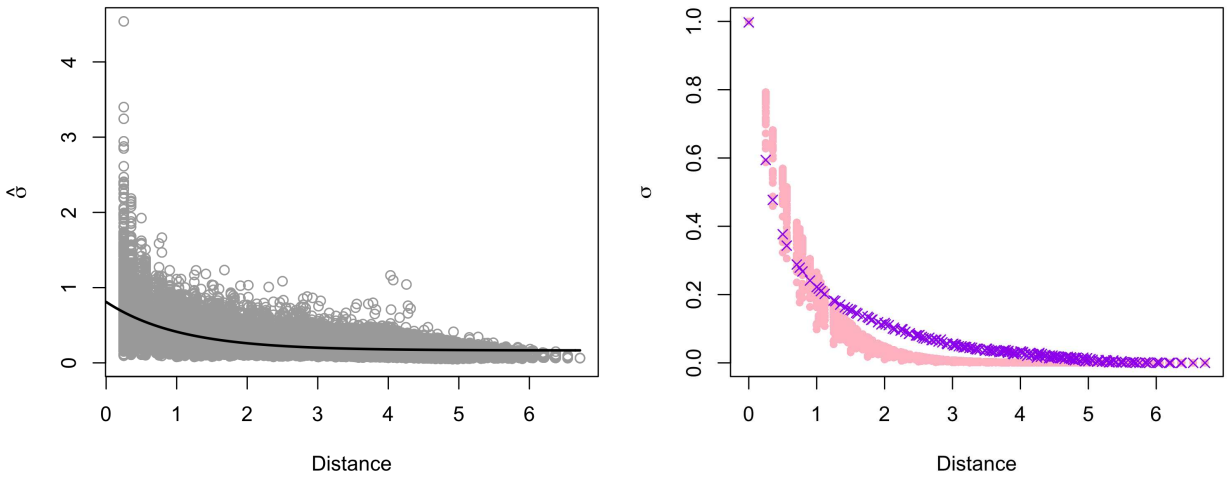

Figure 5.13: Results from fitting via the TPDM for the CPC unified gauge-based analysis of daily precipitation. Left: exponential curve (black line) fitted to $\{\hat{\sigma}_{\mathbf{Y}ij}\}$ (gray points) as a function of pairwise Euclidean distance. Right: The TPDM $\Sigma_{\mathbf{Y}}(\hat{\rho})$ for the fitted extreme SAR model (pink) compared to means of bias-corrected $\{\widetilde{\sigma}_{\mathbf{Y}ij}\}$ at each distance (purple).

As previously done with the simulated data, we consider a risk region $\mathscr{R}$ defined by (5.4.2). Now the set of component indices $\mathscr{D}_{\mathscr{R}} = \{174, 193, 194, 195, 214\}$ corresponds to the grid cell containing Wray, Colorado, and its four nearest neighbors. As before, we let $y_0 = 30$. Our risk region

occurrence probability can be interpreted as the probability of daily precipitation exceeding 30 mm in at least one of the five grid cells under consideration. Based on the fitted extreme SAR model, we estimate this probability to be $p_{\mathscr{R}}(\hat{\rho}) = 3.4 \times 10^{-3}$ with bootstrap confidence interval $(3.0 \times 10^{-3}, 3.7 \times 10^{-3})$. Although we don't know the true occurrence probability, we can compare to the empirical probability 3.1×10^{-3}, which falls within the confidence interval for our fitted model. Thus, even in the face of obvious model misspecification, it seems that we can do a reasonable job of assessing risk.

5.5 Discussion

In this chapter, we investigated two approaches to estimation and inference for the common-scale extreme SAR model proposed in Chapter 4. The first method we explored took a censored likelihood approach, and was implemented using Bayesian MCMC with a data augmentation step that imputes censored components below a marginal threshold. To our knowledge, no previous studies have attempted likelihood-based inference for extremal dependence models with discrete limiting angular measures. Due to the transformed-linear construction and unlike max-linear construction, realizations from our model have angular components close but not equal to the the discrete locations for which the angular measure has mass, thus allowing us to perform likelihood inference in a manner unachievable by these other extremes models with discrete angular measure. Our approach was successful at fitting to data arising from the model, even when lighter-tailed noise was added. However, we learned that our model is not flexible enough, in that the censored likelihood approach was not robust to model misspecification.

As an alternative, we developed a method to fit the extreme SAR model via the TPDM. This non-likelihood-based approach first estimates the TPDM from data, then performs a simple bias correction to the estimated TPDM. Finally, we estimate ρ by minimizing the squared Frobenius norm of the difference between the model-based TPDM $\Sigma_{\mathbf{Y}}(\rho)$ and the bias-corrected estimate $\widetilde{\Sigma}_{\mathbf{Y}}$. Fitting via the TPDM is able to produce reasonable estimates of extremal dependence and corresponding risk region occurrence probabilities even in the case of model misspecifica-

tion. We are able to asses this using simulations from a Brown-Resnick process, for which we derive the true pairwise extremal dependence measures. An additional advantage is that fitting via the TPDM is much faster than running our MCMC algorithm. Still, inference questions remain. Although not directly related to our SAR model, estimates of extremal dependence in cases of weak dependence exhibit bias, and we developed a simple but effective method for correcting for this bias. While the bias of dependence estimates is well known in the extremes community, the level of bias we found to be surprising.

We also explore fitting the extreme SAR model via the TPDM to a set of gridded precipitation observations. We find that these data imply longer-range dependence than our model can achieve, at least under the simple rook neighborhood specification. Similar issues arise in traditional non-extreme areal models (Besag and Kooperberg, 1995). Future work could define a neighborhood structure with longer range dependence, as well as consider model selection between different forms of W.

It is worthwhile to point out that the spatial data sets upon which we illustrate our two inference methods are very large by extremes standards. Our methods were quite feasible on a set of 400 locations, and there is no obvious reason that they could not be applied to problems of much greater size.

There are other possible inference methods for the extreme SAR model beyond the two which we explored in depth in this chapter. We also conducted preliminary investigations of several other approaches. For example, we tried fitting our model using the approach of Einmahl et al. (2018), who minimize a functional of the distance between the empirical and model exponent measure functions. Einmahl et al. (2018) used this approach to fit a max-linear model of dimension $d = 4$, but in our attempt to apply their algorithm to our extreme SAR model, we found that this approach was computationally infeasible for $d = 100$. We also considered a kernel-based method to make the limiting angular measure more flexible. This approach would place kernels at each of the discrete locations for which our SAR model's angular measure has mass, essentially smoothing out the point masses. However, developing appropriate kernels

on the L_2 nonnegative unit sphere proved challenging. Instead, we explored a version of the extremal SAR model with $\alpha = 1$ and L_1 norm, applying standard L_1 Dirichlet kernels. Although results were promising for lower dimensions, we encountered severe numerical problems with the Dirichlet kernels in higher dimensions. Lastly, we also considered another minimum distance method in which we first assign observed angular components to clusters and then minimize a distance between weighted cluster centroids to the normalized columns of $\widetilde{A}$. One issue with this approach was the ad-hoc nature of cluster assignment, however this idea could potentially be made more formal by treating cluster assignment in an EM-like manner. Other methods for inference warrant future investigation, and we comment on some additional ideas in Chapter 6.

Chapter 6

Conclusions and Future Work

This dissertation presents applied and methodological advances in the statistical analysis and modeling of extreme values. Our contributions are largely motivated by the types of data found in the atmospheric sciences, including gridded observational data products and numerical model output from both climate models and atmospheric chemistry models. Employing theory from both univariate and multivariate extremes, we develop new models and propose new inference methods for the univariate, multivariate, and (finite-dimensional) spatial settings. To model tail dependence, both between an extreme response and covariates as well as in a spatial context, we use the framework of multivariate regular variation (MVRV). This probabilistic framework characterizes the joint tail and is useful for modeling multivariate threshold exceedances.

In Chapter 2, we used large initial-condition ensembles of global climate model output to compare future precipitation extremes under two climate change scenarios. Our study revealed the advantage of using such ensembles to estimate parameters of extreme value distributions, especially the shape parameter which is often difficult to estimate but influential in quantifying high quantiles (return levels). Thus, our study provides important insight about how internal variability affects inference of extreme values. We also demonstrated the value of a pattern scaling approach for extremes. Pattern scaling techniques like ours, which assume that large-scale patterns of regional change will scale with global temperature change, can be used to generate regional projections for scenarios that are not readily available in databases of climate model simulations. Future work could examine pattern scaling extremes using a larger suite of climate models. It could also be interesting to investigate applying a similar pattern scaling approach to combine models and data in simulations of future climate extremes, for example following the philosophy of Leeds et al. (2015) and Poppick et al. (2016). These works advocate the “delta method," in which one models differences between past and future model runs, then

applies that to observations such that future simulations preserve some of the characteristics of the observations.

In Chapter 3, we extended methodology for quantile regression and in particular tail dependence optimization in a novel study in atmospheric science. For the first time, a framework was proposed and applied for comparing the meteorological sensitivities of high and extreme O_3 between simulated and observational data. We found substantial differences between NRCM-Chem-modeled and observed sensitivities of high levels of O_3 to meteorological drivers that were not consistent among the three study locations. These results raise concerns for modelers that NCRM-Chem may not be accurately representing the mechanisms behind O_3 formation or loss. In fact, a recent multi-model evaluation conducted by Otero et al. (2018) found considerable differences in the sensitivities of MDA8 O_3 (via multiple linear regression) to the different meteorological factors among five different chemical transport models. It seems further investigation into potential sources of variation is needed. In addition, future work could extend Chapter 3 by conducting a spatial analysis over a larger region, as borrowing strength across spatial locations could reduce uncertainties in parameter estimates.

Motivated by the prevalence of gridded climate model output and other gridded data products in the atmospheric sciences, Chapter 4 proposed a new model for extremes of areal data, with a common-scale extension. The proposed model extends recent work on transformed-linear operations applied to regularly varying random vectors, and is unique in being tied very directly to areal models from classical spatial statistics. Specifically, our model is a multivariate extreme analogue of the simultaneous autoregressive (SAR) model. In Chapter 4 we specified a sufficient condition on the spatial dependence parameter ρ such that the extreme SAR model has desirable properties. We also described the limiting angular measure and corresponding tail pairwise dependence matrix (TPDM) for the model.

In Chapter 5, we investigated two approaches to estimation and inference for the common-scale extreme SAR model proposed in Chapter 4. First, we considered a censored likelihood approach, implemented using Bayesian MCMC with a data augmentation step. We found that

this method was not robust to model misspecification, as evidenced by a poor fit to simulations from a Brown-Resnick process with similar pairwise tail dependence to an extreme SAR with known dependence parameter. As a result, we developed an alternative method to model fitting based on the TPDM. This method is able to produce reasonable estimates of extremal dependence even in the case of model misspecification. This is an especially noteworthy accomplishment considering that our example datasets are of dimension $d = 400$, which is an order of magnitude greater than what most existing methods for spatial extremes can handle.

In Chapter 5, we briefly described several other inference procedures that we explored. We consider some additional ideas here. Likelihood-based methods are desirable because they can be used within a Bayesian hierarchical framework, and computationally efficient Bayesian hierarchical models that account for spatial dependence are largely missing in the extremes literature. Our censored likelihood approach struggled with general data because our extreme SAR model was too rigid. Intuitively, we could think about adding noise to make the model more flexible. A natural approach to adding noise would be to let $\mathbf{Y} = A \circ \mathbf{Z}_1 \oplus \mathbf{Z}_2$. However, instead of smoothing the point masses, this approach would just add more point masses at the vertices, and the likelihood would not agree better with the data. Possible ways forward include introducing noise in the matrix A, or potentially adapting the sum characterization of hidden regular variation developed by Weller and Cooley (2014).

More generally, transformed-linear operations applied to regularly varying random vectors have potential to characterize and construct many new models for extremal dependence within the MVRV framework. Our extreme SAR model is but one example. Transformed-linear analogues can also be defined for many time series models, including moving average models and autoregressive models, since these are just linear combinations of previous observations or white noise terms. Tail pairwise dependence could also be examined between lagged observations from such models. However, inference for such models will undoubtedly be challenging just as we found for the SAR. Further research is needed in this area.

References

Allen, M. R. and Ingram, W. J. (2002). Constraints on future changes in climate and the hydrologic cycle. *Nature*, 419(6903):224–232.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Springer Science & Business Media.

Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 2(5):792–804.

Ballani, F. and Schlather, M. (2011). A construction principle for multivariate extreme value distributions. *Biometrika*, 98(3):633–645.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. CRC Press.

Baur, D., Saisana, M., and Schulze, N. (2004). Modelling the effects of meteorological variables on ozone concentration—a quantile regression approach. *Atmospheric Environment*, 38(28):4689–4699.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.

Beniston, M., Stephenson, D. B., Christensen, O. B., Ferro, C. A. T., Frei, C., Goyette, S., Halsnaes, K., Holt, T., Jylhä, K., Koffi, B., Palutikof, J., Schöll, R., Semmler, T., and Woth, K. (2007). Future extreme events in European climate: an exploration of regional climate model projections. *Climatic Change*, 81(Suppl 1):71–95.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236.

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82(4):733–746.

Brown, B. M. and Resnick, S. I. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14(4):732–739.

Brown, S. J., Murphy, J. M., Sexton, D. M. H., and Harris, G. R. (2014). Climate projections of future extreme events accounting for modelling uncertainties and historical simulation biases. *Climate Dynamics*, 43(9-10):2681–2705.

Camalier, L., Cox, W., and Dolwick, P. (2007). The effects of meteorology on ozone in urban areas and their use in assessing ozone trends. *Atmospheric Environment*, 41(33):7127–7137.

Chen, C.-T. and Knutson, T. (2008). On the verification and comparison of extreme rainfall indices from climate models. *Journal of Climate*, 21(7):1605–1621.

Cliff, A. and Ord, J. K. (1973). *Spatial autocorrelation*. Pion Limited.

Coles, S. G. (2001). *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag London Ltd., London.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 377–392.

Cooley, D. (2012). Return periods under climate change. In AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., and Sorooshian, S., editors, *Extremes in a changing climate: detection, analysis and uncertainty*, Vol. 65, pages 97–114. Springer, Netherlands, Dordrecht.

Cooley, D., Cisewski, J., Erhardt, R. J., Jeon, S., Mannshardt, E., Omolo, B. O., and Sun, Y. (2012a). A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects. *Revstat*, 10(1):135–165.

Cooley, D., Davis, R. A., and Naveau, P. (2010). The pairwise beta distribution: A flexible parametric multivariate model for extremes. *Journal of Multivariate Analysis*, 101(9):2103–2117.

Cooley, D., Davis, R. A., and Naveau, P. (2012b). Approximating the conditional density given large observed values via a multivariate extremes framework, with application to environmental data. *The Annals of Applied Statistics*, pages 1406–1429.

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.

Cooley, D. and Thibaud, E. (2018). Decompositions of dependence for high-dimensional extremes. *arXiv preprint arXiv:1612.07190*.

Cooley, D. S., Hunter, B. D., and Smith, R. L. (2018). Univariate and multivariate extremes for the environmental sciences. In *Handbook of environmental and ecological statistics*. Taylor and Francis.

Davis, J., Cox, W., Reff, A., and Dolwick, P. (2011). A comparison of CMAQ-based and observation-based statistical models relating ozone to meteorological parameters. *Atmospheric Environment*, 45(20):3481–3487.

Davis, R. A. and Resnick, S. I. (1989). Basic properties and prediction of max-arma processes. *Advances in Applied Probability*, 21(4):781–803.

Davison, A. C. and Gholamrezaee, M. M. (2012). Geostatistics of extremes. *Proceedings of the Royal Society A*, 468(2138):581–608.

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, 27(2):161–186.

Dawson, J. P., Adams, P. J., and Pandis, S. N. (2007). Sensitivity of ozone to summertime climate in the eastern USA: A modeling case study. *Atmospheric Environment*, 41(7):1494–1511.

de Haan, L. (1984). A spectral representation for max-stable processes. *The Annals of Probability*, pages 1194–1204.

de Haan, L. (1990). Fighting the arch-enemy with mathematics. *Statistica Neerlandica*, 44(2):45–68.

de Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction*. Springer Series in Operations Research and Financial Engineering.

De Oliveira, V. (2005). Bayesian inference and prediction of Gaussian random fields based on censored data. *Journal of Computational and Graphical Statistics*, 14(1):95–115.

Done, J. M., Holland, G. J., Bruyère, C. L., Leung, L. R., and Suzuki-Parker, A. (2015). Modeling high-impact weather and climate: lessons from a tropical cyclone perspective. *Climatic Change*, 129(3-4):381–395.

Einmahl, J. H., Kiriliouk, A., Krajina, A., and Segers, J. (2016). An M-estimator of spatial tail dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1):275–298.

Einmahl, J. H., Kiriliouk, A., and Segers, J. (2018). A continuous updating weighted least squares estimator of tail dependence in high dimensions. *Extremes*, pages 1–29.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2012). *Modelling extremal events for insurance and finance*. Springer Science & Business Media.

Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2015). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):239–265.

Ferro, C. A. and Segers, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):545–556.

Fiedler, M. and Ptak, V. (1962). On matrices with non-positive off-diagonal elements and positive principal minors. *Czechoslovak Mathematical Journal*, 12(3):382–400.

Fisher, R. and Tippett, L. (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.

Fougères, A.-L., Mercadier, C., and Nolan, J. (2013). Dense classes of multivariate extreme value distributions. *Journal of Multivariate Analysis*, 116:109–129.

Fowler, H., Cooley, D., Sain, S., and Thurston, M. (2010). Detecting change in UK extreme precipitation using results from the climateprediction.net BBC climate change experiment. *Extremes*, 13(2):241–267.

Fowler, H., Ekström, M., Blenkinsop, S., and Smith, A. (2007). Estimating change in extreme European precipitation using a multimodel ensemble. *Journal of Geophysical Research: Atmospheres*, 112(D18).

Furrer, E. M. and Katz, R. W. (2008). Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resources Research*, 44(12).

Furrer, R. and Sain, S. R. (2010). spam: A sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields. *Journal of Statistical Software*, 36(10):1–25.

Genton, M. G., Ma, Y., and Sang, H. (2011). On the likelihood function of Gaussian max-stable processes. *Biometrika*, pages 481–488.

Gilleland, E. and Katz, R. W. (2016). extRemes 2.0: An extreme value analysis package in R. *Journal of Statistical Software*, 72(8):1–39.

Givens, G. H. and Hoeting, J. A. (2013). *Computational statistics*. John Wiley & Sons.

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

Gochis, D., Schumacher, R., Friedrich, K., Doesken, N., Kelsch, M., Sun, J., Ikeda, K., Lindsey, D., Wood, A., Dolan, B., Matrosov, S., Newman, A., Mahoney, K., Rutledge, S., Johnson, R., Kucera, P., Kennedy, P., Sempere-Torres, D., Steiner, M., Roberts, R., Wilson, J., Yu, W., Chandrasekar, V., Rasmussen, R., Anderson, A., and Brown, B. (2015). The great Colorado flood of September 2013. *Bulletin of the American Meteorological Society*, 96(9):1461–1487.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B. (2005). Fully coupled "online" chemistry within the WRF model. *Atmospheric Environment*, 39(37):6957–6975.

Haining, R. (1990). *Spatial data analysis in the social and environmental sciences*. Cambridge University Press.

Hanel, M. and Buishand, T. A. (2011). Analysis of precipitation extremes in an ensemble of transient regional climate model simulations for the Rhine basin. *Climate Dynamics*, 36(5-6):1135–1153.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

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

Hosking, J. R. M. and Wallis, J. R. (1997). *Regional Frequency Analysis: An approach based on L-Moments*. Cambridge, University Press, Cambridge., U.K.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J., and Marshall, S. (2013). The community earth system model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, 94(9):1339–1360.

Huser, R. and Davison, A. C. (2013). Composite likelihood estimation for the Brown-Resnick process. *Biometrika*, 100(2):511–518.

Huser, R., Davison, A. C., and Genton, M. G. (2016). Likelihood estimators for multivariate extremes. *Extremes*, 19(1):79–103.

IHME (2013). Global burden of disease study 2010 results by cause 1990-2010 - country level. Technical report, Institute for Health Metrics and Evaluation, Seattle, United States.

Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A., Balzarini, A., Baró, R., Bellasio, R., Brunner, D., Chemel, C., Curci, G., Flemming, J., Forkel, R., Giordano, L., Jimenez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., and Galmarini, S. (2015). Evaluation of operational on-line-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part I: Ozone. *Atmospheric Environment*, 115:404–420.

IPCC (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation: Special report of the intergovernmental panel on climate change, summary for policymakers*. Cambridge University Press, Cambridge, UK.

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

Johnson, C. R. (1982). Inverse M-matrices. *Linear Algebra and its Applications*, 47:195–216.

Johnson, C. R. and Smith, R. L. (2011). Inverse M-matrices, II. *Linear Algebra and its Applications*, 435:953–983.

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.

Katz, R. W. (2013). Economic impact of extreme events: An approach based on extreme value theory. *Extreme events: Observations, modeling and economics*.

Kavassalis, S. and Murphy, J. G. (2017). Understanding ozone-meteorology correlations: a role for dry deposition. *Geophysical Research Letters*, 44(6):2922–2931.

Kay, J., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J., Bates, S., Danabasoglu, G., Edwards, J., and Vertenstein, M. (2015). The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, 96(8):1333–1349.

Kelejian, H. H. and Robinson, D. P. (1995). Spatial correlation: a suggested alternative to the autoregressive model. In *New directions in spatial econometrics*, pages 75–95. Springer.

Kharin, V. V. and Zwiers, F. W. (2005). Estimating extremes in transient climate change simulations. *Journal of Climate*, 18(8):1156–1173.

Kharin, V. V., Zwiers, F. W., Zhang, X., and Wehner, M. (2013). Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, 119(2):345–357.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A. (2010). Challenges in combining projections from multiple climate models. *Journal of Climate*, 23(10):2739–2758.

Koenker, R. (2016). *quantreg: Quantile Regression*.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, pages 33–50.

Lamarque, J.-F., Kyle, G. P., Meinshausen, M., Riahi, K., Smith, S. J., van Vuuren, D. P., Conley, A. J., and Vitt, F. (2011). Global and regional evolution of short-lived radiatively-active gases and aerosols in the representative concentration pathways. *Climatic Change*, 109(1-2):191–212.

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

Leadbetter, M. (1974). On extreme values in stationary processes. *Z. Wahrsch. verw. Gebiete*, 28(4):289–303.

Leadbetter, M., Weissman, I., de Haan, L., and Rootzén, H. (1989). On clustering of high values in statistically stationary series. *Proc. 4th Int. Meet. Statistical Climatology*, 16:217–222.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

Leeds, W. B., Moyer, E. J., and Stein, M. L. (2015). Simulation of future climate under changing temporal covariance structures. *Advances in Statistical Climatology, Meteorology and Oceanography*, 1(1):1–14.

Lefohn, A. S., Shadwick, D., and Oltmans, S. J. (2008). Characterizing long-term changes in surface ozone levels in the United States (1980–2005). *Atmospheric Environment*, 42(35):8252–8262.

Lukas, J., Barsugli, J., Doesken, N., Rangwala, I., and Wolter, K. (2014). Climate change in Colorado: a synthesis to support water resources management and adaptation. *University of Colorado, Boulder, Colorado*.

McGrath-Spangler, E. L. and Denning, A. S. (2012). Estimates of North American summertime planetary boundary layer depths derived from space-borne lidar. *Journal of Geophysical Research: Atmospheres*, 117(D15).

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

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

Morris, S. A., Reich, B. J., Thibaud, E., and Cooley, D. (2017). A space-time skew-t model for threshold exceedances. *Biometrics*.

O'Neill, B. and Gettelman, A. (2018). An introduction to the special issue on the Benefits of Reduced Anthropogenic Climate changE (BRACE). *Climatic Change*, 146(3-4):277.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70(349):120–126.

Ordónez, C., Mathis, H., Furger, M., Henne, S., Hüglin, C., Staehelin, J., and Prévôt, A. (2005). Changes of daily surface ozone maxima in Switzerland in all seasons from 1992 to 2002 and discussion of summer 2003. *Atmospheric Chemistry and Physics*, 5(5):1187–1203.

Otero, N., Sillmann, J., Mar, K. A., Rust, H. W., Solberg, S., Andersson, C., Engardt, M., Bergström, R., Bessagnet, B., Colette, A., et al. (2018). A multi-model comparison of meteorological drivers of surface ozone over europe. *Atmospheric Chemistry and Physics*, 18(16):12269–12288.

Otero, N., Sillmann, J., Schnell, J. L., Rust, H. W., and Butler, T. (2016). Synoptic and meteorological drivers of extreme ozone concentrations over Europe. *Environmental Research Letters*, 11(2).

Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277.

Pausader, M., Bernie, D., Parey, S., and Nogaj, M. (2012). Computing the distribution of return levels of extreme warm temperatures for future climate projections. *Climate Dynamics*, 38(5-6):1003–1015.

Pearce, J. L., Beringer, J., Nicholls, N., Hyndman, R. J., and Tapper, N. J. (2011). Quantifying the influence of local meteorology on air quality using generalized additive models. *Atmospheric Environment*, 45(6):1328–1336.

Pfister, G., Walters, S., Lamarque, J.-F., Fast, J., Barth, M., Wong, J., Done, J., Holland, G., and Bruyère, C. (2014). Projections of future summertime ozone over the US. *Journal of Geophysical Research: Atmospheres*, 119(9):5559–5582.

Phalitnonkiat, P., Sun, W., Grigoriu, M. D., Hess, P., and Samorodnitsky, G. (2016). Extreme ozone events: Tail behavior of the surface ozone distribution over the US. *Atmospheric Environment*, 128:134–146.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.

Plemmons, R. J. (1977). M-matrix characterizations. I—nonsingular M-matrices. *Linear Algebra and its Applications*, 18(2):175–188.

Poppick, A., McInerney, D. J., Moyer, E. J., and Stein, M. L. (2016). Temperatures in transient climates: improved methods for simulations with evolving temporal covariances. *The Annals of Applied Statistics*, 10(1):477–505.

Porter, W., Heald, C., Cooley, D., and Russell, B. (2015). Investigating the observed sensitivities of air-quality extremes to meteorological drivers via quantile regression. *Atmospheric Chemistry and Physics*, 15(18):10349–10366.

Porter, W. C., Safieddine, S. A., and Heald, C. L. (2017). Impact of aromatics and monoterpenes on simulated tropospheric ozone and total OH reactivity. *Atmospheric Environment*, 169:250–257.

Prescott, P. and Walden, A. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, 67(3):723–724.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rasmussen, D., Fiore, A., Naik, V., Horowitz, L., McGinnis, S., and Schultz, M. (2012). Surface ozone-temperature relationships in the eastern US: A monthly climatology for evaluating chemistry-climate models. *Atmospheric Environment*, 47:142–153.

Reich, B. J. and Shaby, B. A. (2018). A spatial Markov model for climate extremes. *Journal of Computational and Graphical Statistics*, (just-accepted).

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

Resnick, S. I. (1987). *Extreme values, regular variation, and point processes*. Springer-Verlag, New York.

Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.

Ribatet, M. (2013). Spatial extremes: Max-stable processes at work. *Journal de la Société Française de Statistique*, 154(2):156–177.

Rieder, H. E., Fiore, A. M., Polvani, L. M., Lamarque, J.-F., and Fang, Y. (2013). Changes in the frequency and return level of high ozone pollution events over the eastern United States following emission controls. *Environmental Research Letters*, 8(1).

Rootzén, H. and Katz, R. W. (2013). Design life level: Quantifying risk in a changing climate. *Water Resources Research*, 49(9):5964–5972.

Rootzén, H. and Tajvidi, N. (1997). Extreme value statistics and wind storm losses: a case study. *Scandinavian Actuarial Journal*, 1997(1):70–94.

Russell, B. T. (2015). *Understanding extreme behavior by optimizing tail dependence with application to ground level ozone via data mining and spatial modeling*. PhD thesis, Colorado State University. Libraries.

Russell, B. T., Cooley, D. S., Porter, W. C., and Heald, C. L. (2016a). Modeling the spatial behavior of the meteorological drivers' effects on extreme ozone. *Environmetrics*, 27(6):334–344.

Russell, B. T., Cooley, D. S., Porter, W. C., Reich, B. J., and Heald, C. L. (2016b). Data mining to investigate the meteorological drivers for extreme ground level ozone events. *The Annals of Applied Statistics*, 10(3):1673–1698.

Russo, A., Gouveia, C., Levy, I., Dayan, U., Jerez, S., Mendes, M., and Trigo, R. (2016). Coastal recirculation potential affecting air pollutants in Portugal: The role of circulation weather types. *Atmospheric Environment*, 135:9–19.

Sanderson, B. M., Oleson, K. W., Strand, W. G., Lehner, F., and O'Neill, B. C. (2018). A new ensemble of GCM simulations to assess avoided impacts in a climate mitigation scenario. *Climatic Change*, 146(3-4):303–318.

Sang, H. and Gelfand, A. E. (2010). Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(1):49–65.

Santer, B., Wigley, T., Schlesinger, M., and Mitchell, J. (1990). Developing climate scenarios from equilibrium GCM results. Technical Report 47, Max Planck Institute, Hamburg.

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. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44.

Schliep, E., Cooley, D., Sain, S. R., and Hoeting, J. A. (2010). A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. *Extremes*, 13(2):219–239.

Seinfeld, J. H., Pandis, S. N., and Noone, K. (1998). *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons, Inc.

Seneviratne, S. I., Donat, M. G., Pitman, A. J., Knutti, R., and Wilby, R. L. (2016). Allowable CO2 emissions based on regional and impact-related climate targets. *Nature*, 529:477–483.

Sherwen, T., Schmidt, J. A., Evans, M. J., Carpenter, L. J., Großmann, K., Eastham, S. D., Jacob, D. J., Dix, B., Koenig, T. K., Sinreich, R., and Ordónez, C. (2016). Global impacts of tropospheric halogens (Cl, Br, I) on oxidants and composition in GEOS-Chem. *Atmospheric Chemistry and Physics*, 16(18):12239–12271.

Sibuya, M. (1960). Bivariate extreme statistics, i. *Annals of the Institute of Statistical Mathematics*, 11(3):195–210.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90.

Smith, R. L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript, University of Surrey, England. http://www.stat.unc.edu/postscript/rs/spatex.pdf.

Smith, R. L., Tawn, J. A., and Coles, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika*, 84(2):249–268.

Song, D. and Gupta, A. K. (1997). L_p-norm uniform distribution. *Proceedings of the American Mathematical Society*, 125(2):595–601.

Song, J. J. and De Oliveira, V. (2012). Bayesian model selection in spatial lattice models. *Statistical Methodology*, 9(1):228–238.

Stephenson, A. and Tawn, J. (2005). Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika*, 92(1):213–227.

Strokorb, K. and Schlather, M. (2015). An exceptional max-stable process fully parameterized by its extremal coefficients. *Bernoulli*, 21(1):276–302.

Tai, A. P., Mickley, L. J., and Jacob, D. J. (2010). Correlations between fine particulate matter (PM 2.5) and meteorological variables in the United States: Implications for the sensitivity of PM 2.5 to climate change. *Atmospheric Environment*, 44(32):3976–3984.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540.

Tebaldi, C. and Arblaster, J. (2014). Pattern scaling: Its strengths and limitations, and an update on the latest model simulations. *Climatic Change*, 122(3):459–471.

Thibaud, E., Aalto, J., Cooley, D. S., Davison, A. C., and Heikkinen, J. (2016). Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures. *The Annals of Applied Statistics*, 10(4):2303–2324.

Thibaud, E. and Opitz, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, 102(4):855–870.

Tye, M. and Cooley, D. (2015). A spatial model to examine rainfall extremes in Colorado's Front Range. *Journal of Hydrology*, 530:15–23.

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., and Rose, S. K. (2011). The representative concentration pathways: an overview. *Climatic Change*, 109:5–31.

Vautard, R., Beekmann, M., Desplat, J., Hodzic, A., and Morel, S. (2007). Air quality in Europe during the summer of 2003 as a prototype of air quality in a warmer climate. *Comptes*

Rendus Geoscience, 339(11):747–763.

Wadsworth, J. L. and Tawn, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, 101(1):1–15.

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121(2):311–324.

Waller, L. A. and Gotway, C. A. (2004). *Applied spatial statistics for public health data*, volume 368. John Wiley & Sons.

Wang, Y. and Stoev, S. A. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Advances in Applied Probability*, 43(2):461–483.

Wehner, M. F. (2013). Very extreme seasonal precipitation in the NARCCAP ensemble: model performance and projections. *Climate Dynamics*, 40(1-2):59–80.

Weller, G. B. and Cooley, D. (2014). A sum characterization of hidden regular variation with likelihood inference via expectation-maximization. *Biometrika*, 101(1):17–36.

Westra, S., Alexander, L. V., and Zwiers, F. W. (2013). Global increasing trends in annual maximum daily precipitation. *Journal of Climate*, 26(11):3904–3918.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, pages 434–449.

Wilson, A., Rappold, A. G., Neas, L. M., and Reich, B. J. (2014). Modeling the effect of temperature on ozone-related mortality. *The Annals of Applied Statistics*, 8(3):1728–1749.

Yuen, R. and Stoev, S. (2014). CRPS M-estimation for max-stable models. *Extremes*, 17(3):387–410.

Zhang, X., Zwiers, F. W., and Li, G. (2004). Monte Carlo experiments on the detection of trends in extreme values. *Journal of Climate*, 17(10):1945–1952.

Zhao, W., Fan, S., Guo, H., Gao, B., Sun, J., and Chen, L. (2016). Assessing the impact of local meteorological variables on surface ozone in Hong Kong during 2000–2015 using quantile and multiple line regression models. *Atmospheric Environment*, 144:182–193.

Zhou, C. (2018). The bootstrap in extreme value theory. Workshop on Rare events, extremes, and machine learning, Télécom ParisTech.

Appendix A

Supplementary Material for Chapter 2

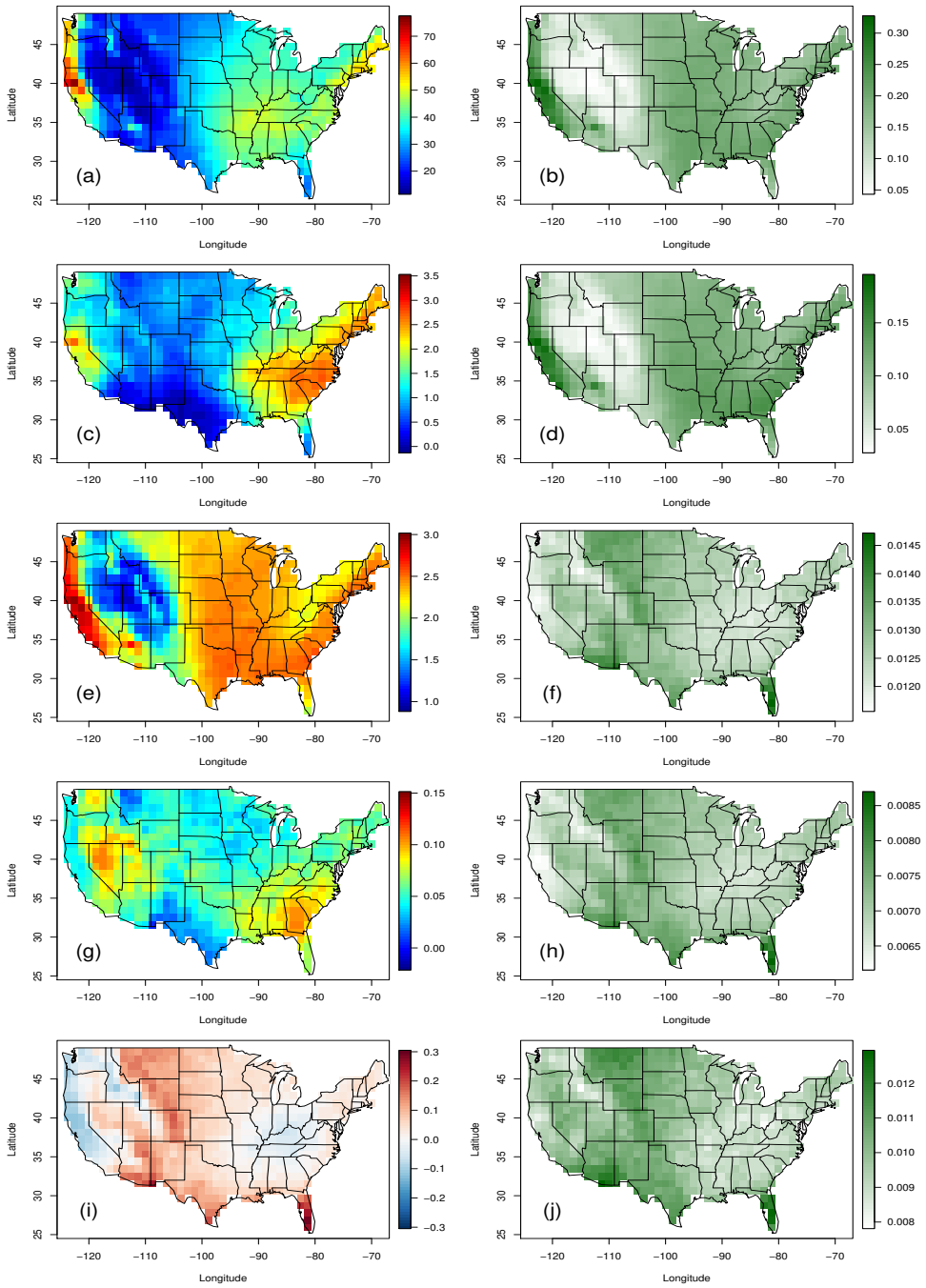

Figure A.1: Parameter estimates (left) and standard errors (right) for the GEV parameters (from top to bottom) $\mu_0^{8.5}(s), \mu_1^{8.5}(s), \phi_0^{8.5}(s), \phi_1^{8.5}(s)$, and $\xi^{8.5}(s)$, based on CESM-LE simulations of annual maximum daily precipitation under historical and RCP8.5 forcings over the contiguous United States.

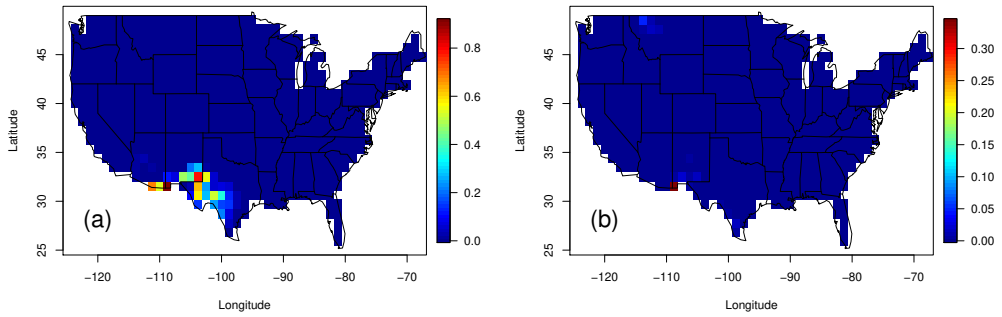

Figure A.2: Normal-based p-values associated with one-sided hypothesis tests of $\mu_1^{8.5}(s) \leq 0$ (a) show that this null hypothesis is rejected at the $\alpha = 0.05$ level for nearly all grid cells outside a small region near the Mexican border. Similarly, p-values associated with the one-sided hypothesis tests of $\phi_1^{8.5}(s) \leq 0$ (b) show that this null hypothesis is rejected at all but two grid cells in the contiguous U.S.

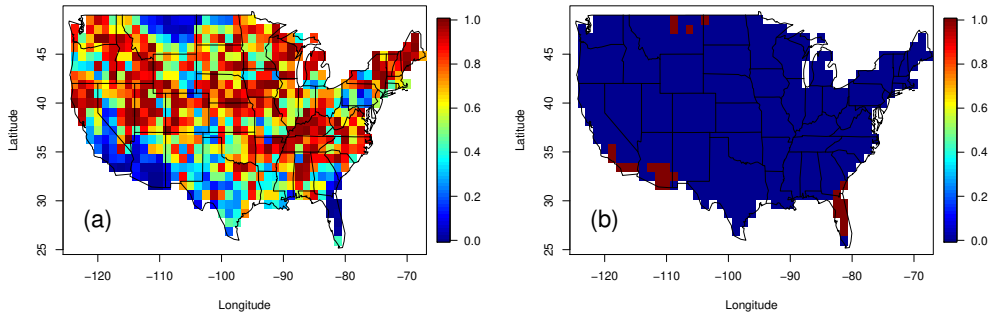

Figure A.3: Results of the Anderson-Darling (AD) tests for goodness of fit at each grid cell for the GEV model fitted to CESM-LE annual precipitation maxima. AD p-values are given in (a), and (b) shows which p-values are less than 0.05 in red (less than 4% of the grid cells reject at the $\alpha = 0.05$ level).

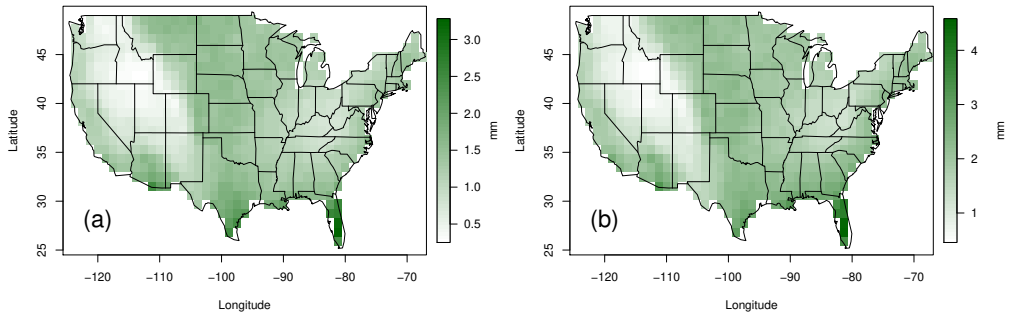

Figure A.4: Delta-method-based standard error estimates in 2005 (a) and 2080 (b) for the 1% AEP level estimated from CESM-LE annual precipitation maxima. These correspond to Figure 2.2(a) and (b) in the main text.

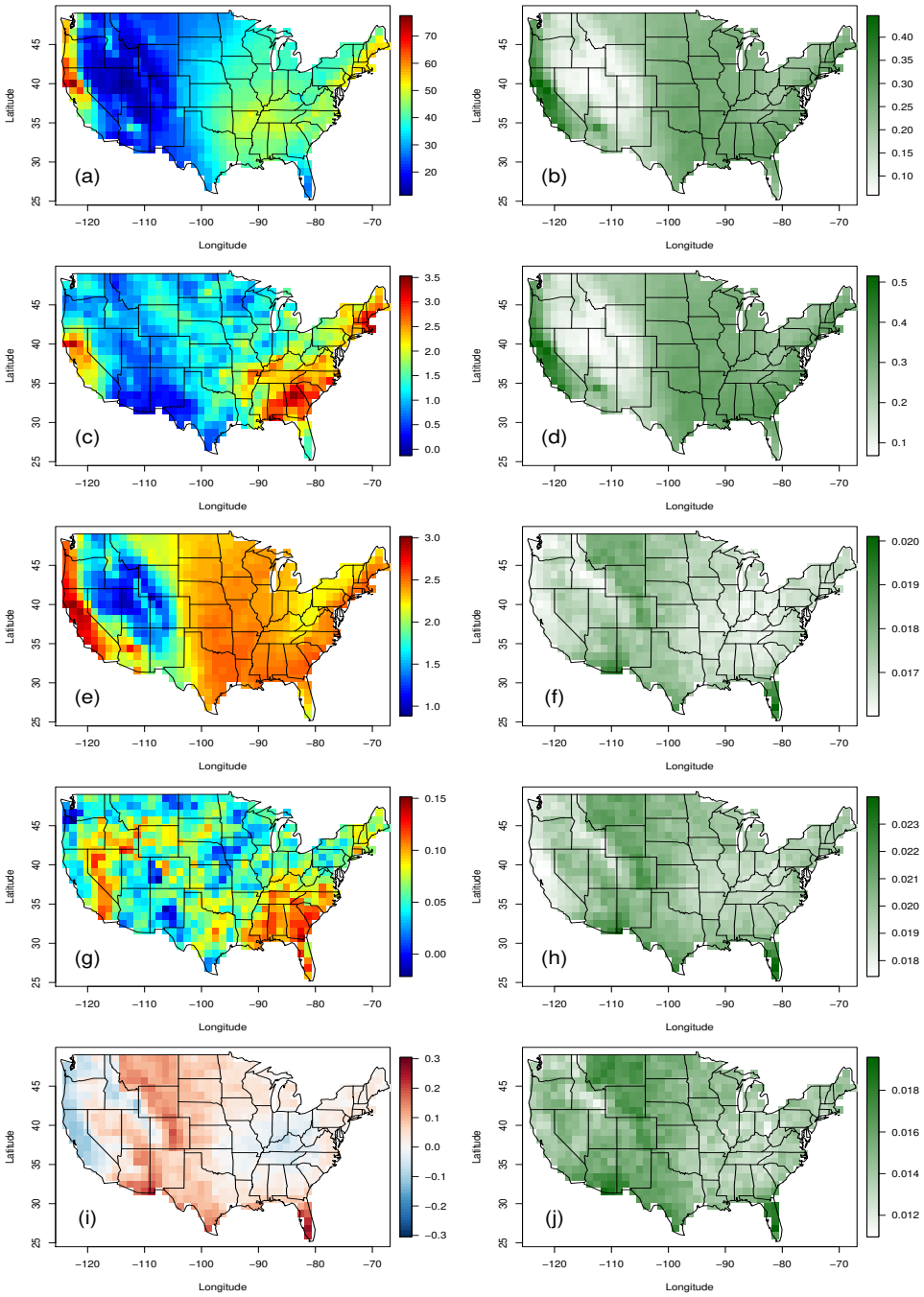

Figure A.5: Parameter estimates (left) and standard errors (right) for the GEV parameters (from top to bottom) $\mu_0^{4.5}(s), \mu_1^{4.5}(s), \phi_0^{4.5}(s), \phi_1^{4.5}(s)$, and $\xi^{4.5}(s)$, based on CESM-ME simulations of annual maximum daily precipitation under historical and RCP4.5 forcings over the contiguous United States. Parameter estimates (left column) are on the same scales as their counterparts in Figure A.1 for easy comparison.

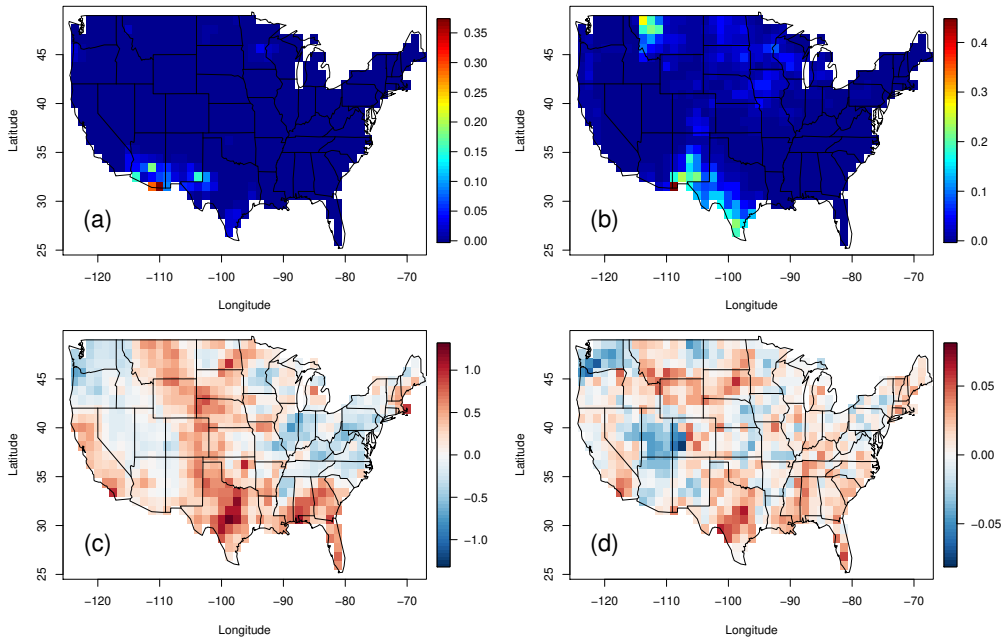

Figure A.6: Top: Normal-based p-values for one-sided hypothesis tests of $\mu_1^{4.5}(s) \leq 0$ (a) and $\phi_1^{4.5} \leq 0$ (b). Bottom: Comparison of slope parameters estimated from CESM-ME vs. CESM-LE. The difference between estimates of $\mu_1^{4.5}(s)$ and $\mu_1^{8.5}(s)$ is shown in (c), where red cells indicate $\hat{\mu}_1^{4.5}(s) > \hat{\mu}_1^{8.5}(s)$. Similarly, the difference between estimates of $\phi_1^{4.5}$ and $\phi_1^{8.5}$ is shown in (d).

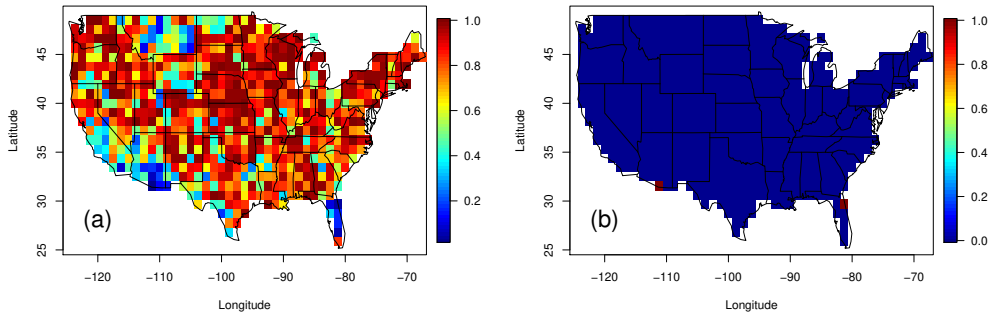

Figure A.7: Results of the Anderson-Darling (AD) tests for goodness of fit at each grid cell for the GEV model fitted to CESM-ME annual precipitation maxima. AD p-values are given in (a), and (b) shows which p-values are less than 0.05 in red (only 2 grid cells reject at the $\alpha = 0.05$ level).

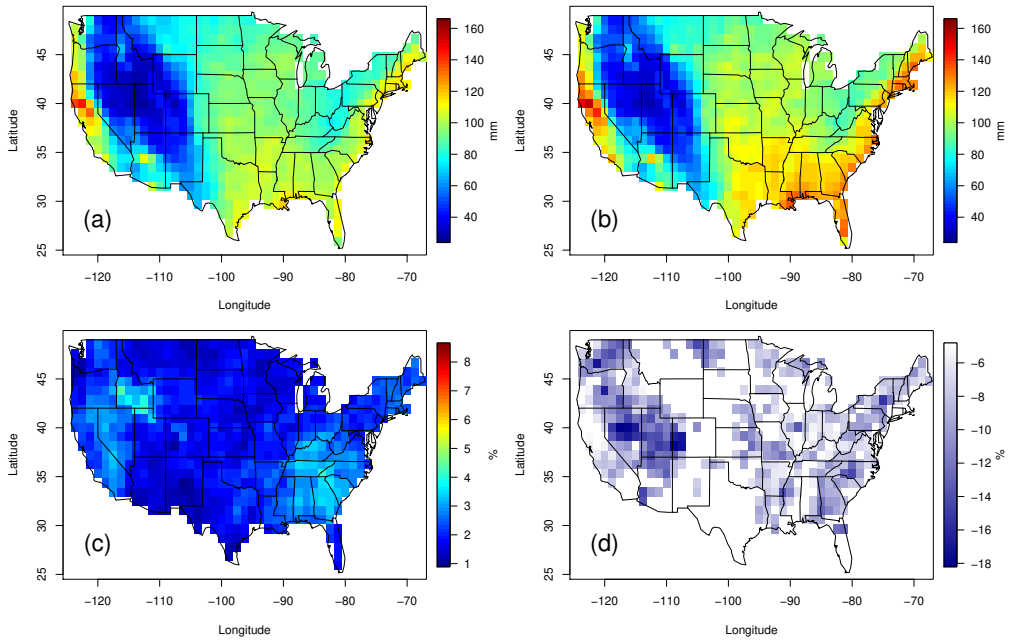

Figure A.8: Point estimates for the 1% AEP level (in mm precipitation) for the years 2005 (a) and 2080 (b), based on the CESM-ME simulations of annual maximum daily precipitation under RCP4.5. These are on the same scale as Figure 2.2(a) and (b) in the main text for easy comparison. The annual exceedance probability (%) in 2080 under RCP4.5 corresponding to a 1% AEP level in 2005 is shown in (c), on the same scale as Figure 2.2(d). Percentage reductions in the 2080 1% AEP level under RCP4.5 compared to RCP8.5 are mapped in (d) for grid cells with a significant difference between scenarios.

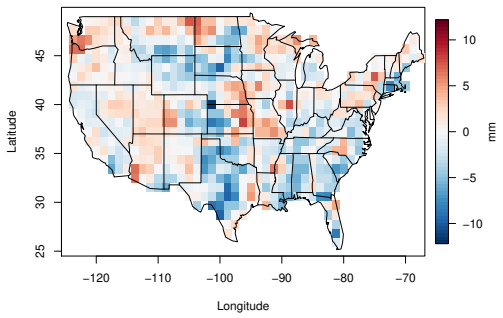

Figure A.9: Difference in the year 2080 between pattern-scaled projections of the 1% AEP level and the projected 1% AEP level based on the GEV model fitted to CESM-ME RCP4.5 annual precipitation maxima. Red cells indicate the 1% AEP level is overestimated by pattern scaling, while blue cells indicate the 1% AEP level is underestimated by pattern scaling, relative to the direct model fit.