Statistical Analysis of Extremes Motivated by Weather and Climate Studies: Applied and Theoretical Advances

by

Daniel S. Cooley

B.A., University of Colorado at Boulder, 1994

M.S., University of Colorado at Boulder, 2002

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Applied Mathematics

2005

This thesis entitled:
Statistical Analysis of Extremes Motivated by Weather and Climate Studies: Applied and Theoretical Advances
written by Daniel S. Cooley
has been approved for the Department of Applied Mathematics

Philippe Naveau

Prof. Douglas Nychka

Prof. Anne Dougherty

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Cooley, Daniel S. (Ph.D., Applied Mathematics)

Statistical Analysis of Extremes Motivated by Weather and Climate Studies: Applied and Theoretical Advances

Thesis directed by Prof. Philippe Naveau

The statistical study of extreme values has seen much growth since the field's beginnings about eighty years ago. Extreme value theory (EVT) is used in a wide range of applications from hydrology to engineering to economics. In this thesis, three EVT investigations related to weather and climate problems are presented. Two are novel applications of extremes techniques to weather or climate studies, and the other is an advancement of extremes theory motivated by issues arising in weather and climate studies.

The first investigation is a unique application of extreme value theory to lichenometry. The aim is to estimate the ages of paleo-climate events by employing lichen measurements on glacial features formed by the events. This application of extremes differs from most in that although the lichen data are extremes, a hidden covariate of the data is the main quantity of interest. Presented are two different models for estimating the ages of the climate events. The first model relies on the EM-algorithm to estimate the ages, while the second builds a hierarchical model which estimates the ages through a Bayesian framework. In addition to providing age estimates and confidence intervals, the models provide a comprehensive statistical model for lichenometry based on EVT, which had not been done previously.

Often in weather and climate studies, the data are recorded at specific locations, and the second and third investigations deal with spatial data. The second investigation advances the theory of extremes on spatial fields. One of the primary issues with spatial observations is quantifying their spatial dependence. The field of geostatistics is typically used to model spatial data, but it has its roots in studying the central tendencies of the distribution rather than the distribution's tails. The madogram, a new dependence measure for extremes, is introduced. The madogram has its roots in traditional geostatistics and is therefore well-suited for measuring spatial dependence. Estimators for the madogram are also presented and are compared those of other extremes dependence measures.

The goal of the third project is to produce a map describing potential extreme precipitation and uncertainty measures for a study region in Colorado. Like the previous project, it is necessary to characterize spatial dependence. However, here the goal is to model climatological precipitation, and thus it is necessary to model the spatial dependence between the distributions which characterize the extremes rather than the observations themselves. The methodology constructs a Bayesian hierarchical model to characterize the extreme precipitation for the region. The spatial analysis is performed in a space defined by climatological coordinates rather than in the traditional latitude/longitude space. Using this methodology, the precipitation map has strong ties to the orography of the region.

Dedication

To Tina

Acknowledgements

I have been very fortunate to have been advised by Philippe Naveau. Philippe has showed great skill in guiding my research and also great patience in working with me. Philippe did a tremendous job of overseeing my research, often from across the Atlantic. He also managed to have me visit Europe three times, something I never imagined doing while in graduate school.

I have been equally fortunate to have worked with Doug Nychka. His experience and expertise have been invaluble. The Geophysical Statistics Project (GSP) at NCAR has been a great way for me to expand my statistics contacts and knowledge. My thanks extend to all the GSP's post-docs, visitors, and collaborators including Reinhard Furrer, Uli Schneider, Dorin Drignei, Rick Katz, Tomoko Matsuo, Tim Hoar, and others.

The probability and statistics group in CU's Department of Applied Mathematics has been very supportive. Both Anne Dougherty and Jem Corcoran have served advisory roles for me and both were instrumental in convincing me that I should pursue a Ph.D. after completing my M.S. degree. I have also benefitted from discussions with other prob/stat students including Uli Schneider, Marcio Carvalho, David Beltran del Rio, Debbie Batista, and Betsy Siewert.

CU's Applied Math Department is outstanding. Dr. Segur and Dr. Curry have done outstanding jobs of overseeing the department serving as chairs during my time in the department. It is a remarkably collegial department. Tina and I have also enjoyed several fantastic dinners at the invitation of Dr. Curry.

I have had many collegues within the department whom I wish to thank: Scott Maclachlan, the Matts (Carroll and Nabity), Maribeth Oscamou, and Ryan Girard, among others. I look back at my n years of grad-school as a positive time, thanks largely in part to these people.

I was primarily supported by the department's VIGRE grant which allowed me tremendous freedom to work with Philippe and the GSP group at NCAR. I received summer support from the Weather and Climate Impacts Initiative at NCAR, and a portion of my travel was covered by grant NSF-ATM-0327936.

I also benefitted from a two-month visit to CNRS-Meudon/Bellevue. Many thanks to Vincent Jomelli, Delphine Grancher, and the entire Laboratoire Géographie Physique.

Most of all, I wish to thank Tina without whose support and patience none of this work would mean anything. I look forward to our next adventure.

Contents

Chapter

1 Introduction 1

1.1 Weather, climate and extreme value theory . . . 1

1.2 Background in extreme value theory . . . 2

1.3 Outline . . . 7

2 Extreme Value Models for Lichenometry 9

2.1 Background . . . 9

2.2 Bivariate lichenometry model . . . 12

2.2.1 Model structure . . . 12

2.2.2 Estimation procedure . . . 14

2.2.3 Results from both simulated data and Bolivian glacier data . . . 15

2.3 Bayesian hierarchical model . . . 17

2.3.1 Model structure . . . 18

2.3.2 Model inference . . . 24

2.3.3 Results . . . 25

2.3.4 Discussion . . . 32

3 Spatial Extreme Value Analysis for Weather and Climate 34

3.1 Weather, climate and geostatistics . . . 34

3.2 Modeling spatial dependence in extreme observations . . . 36

3.2.1 Multivariate extreme-value theory . . . 36

3.2.2 Characterizing dependence in extreme weather observations . . . 37

3.3 Modeling spatial dependence in extreme distributions . . . 39

3.4 Summary . . . 40

4 Madogram 41

4.1 Background . . . 41

4.1.1 Bivariate extreme value distributions . . . 41

4.1.2 Max-stable processes . . . 42

4.2 Existing pairwise dependence measures . . . 45

4.2.1 Extremal coefficient . . . 45

4.2.2 Pickands dependence function . . . 47

4.2.3 Estimation . . . 48

4.3 Madogram . . . 49

4.3.1 Madogram for bivariate extremes . . . 49

4.3.2 Madograms for spatial extremes . . . 53

4.3.3 Application: France precipitation data . . . 56

4.4 Extending the extremal coefficient and madogram . . . 57

4.4.1 Bivariate Case . . . 57

4.4.2 Spatial Case . . . 64

4.5 Conclusions . . . 67

5 An extreme precipitation return-levels map for Colorado's Front Range 68

5.1 Background . . . 68

5.1.1 A precipitation atlas for Colorado's Front Range . . . 68

5.1.2 Modeling distributional dependence . . . 70

5.1.3 Regional frequency analysis . . . 71

5.1.4 Outline . . . 72

5.2 Data . . . 72

5.2.1 Study region, weather stations, and covariates . . . 72

5.2.2 Threshold selection . . . 74

5.2.3 Declustering . . . 74

5.3 Models . . . 76

5.3.1 Hierarchical model for threshold exceedances . . . 77

5.3.2 Exceedance rate model . . . 81

5.3.3 MCMC structure . . . 82

5.3.4 Spatial interpolation and inference . . . 84

5.4 Model selection and results . . . 84

5.5 Conclusions and discussion 92

6 Conclusion and Future Work 95

Bibliography 99

Appendix

.1 Appendix for Lichenometry Chapter 103

.2 Appendix for Madogram Chapter 104

Tables

Table

2.1 Results for bivariate model on simulated data 16

2.2 Results for the bivariate method on the South Charquini Glacier data. 16

2.3 Priors in Bayesian model 22

2.4 Simulated data results for Bayesian model 26

2.5 Age estimates for unpooled data 27

2.6 Moraine/event alignment 27

2.7 Age and parameter estimates, linear growth curve 28

2.8 Age and parameter estimates, exponential growth curve 29

2.9 DIC results 31

4.1 Comparison of the estimators for θ 52

4.2 Comparison of estimators of $A(w)$ 60

5.1 Exceedance amounts models tested and corresponding DIC scores 86

5.2 Exceedance rate models tested and corresponding DIC scores 89

Figures

Figure

1.1 Histograms of sample means and sample maxima 4

1.2 Three types of max-stable distributions 6

2.1 GEV versus a Gaussian distribution for lichen diameters 11

2.2 Map of Cordillera Real region 12

2.3 New growth curve plot 14

2.4 Block structure of the covariance matrix in Bayesian lichen model. 21

2.5 Schematic diagram of Bayesian model 23

2.6 Box plots of pooled and unpooled data 28

2.7 Posterior densities 30

2.8 Growth curve models comparison 31

3.1 Examples of non-extreme geostatistics data 35

3.2 Example of spatial extreme observations 38

3.3 Example of return level estimates at spatial locations 39

4.1 Examples of simulated max-stable random fields 44

4.2 The range of values for the Pickands dependence function. The dashed line represents independence and the solid line represents $X = Y$ in probability. 48

4.3 Box plots of absolute error of dependence esimators 52

4.4 Madogram and θ estimates for simulated fields 54

4.5 France 30-year maximum precipitation data 56

4.6 Non-parametric and parametric madogram estimates for the France precipitation data 57

4.7 Range of values for the λ-madogram. 59

4.8 Simulation examples for estimator comparison 62

4.9 Comparison of estimators of $A(w)$ 63

4.10 The behavior of the spatial λ-madogram 65
4.11 Estimates of the spatial λ-madogram 66

5.1 Map of Colorado study region 73
5.2 Plots of the bias in parameter estimates due to precision level of the data 75
5.3 Shows the translation from the latitude/longitude space to the climate space 78
5.4 Effect of pooling the data 80
5.5 Emprical information for choosing priors for $\vec{\beta}_\phi$. 81
5.6 Return levels model schematic 83
5.7 Posterior density estimates 87
5.8 Point estimate for the log-transformed GPD scale parameter ϕ 88
5.9 Point estimate for the exceedance rate parameter ζ 89
5.10 Uncertainty estimates for the 25 year return level 90
5.11 Point estimate for the 25 year return level from the traditional and climate spaces 91
5.12 Madogram of the normalized annual maximum data 92
5.13 QQ plot of all threshold exceedances 93

Chapter 1

Introduction

1.1 Weather, climate and extreme value theory

Humankind is captivated (both physically and intellectually) by the weather. Weather is one of the few things we all share, one cannot escape it, and it is almost completely out of our control. Conversations are begun or voids are filled by interjections about the recent weather. We are inundated with weather forecasts on the radio, television, the internet, and newspapers. However, our fascination with the weather is not baseless. Unusual or extreme weather events do much more than just provide interesting kindling for conversation, as they can have enormous economic and/or human impact. Understanding and preparing for extreme weather events is essential for our society, whether this understanding enables us evacuate in advance of an oncoming hurricane or to construct storm sewers to adequately drain a city's streets.

Climate can be defined as the meteorological conditions that characteristically prevail over a certain region. Increasingly, we are aware that climate, too, is dynamic. The evidence is clear that the climate we currently observe has not remained unchanged throughout time. As with weather, there are many questions that remain unanswered about climate and climate change. What causes climate change and how rapidly it occurs are not fully understood. The question of how much anthropogenic forcings are effecting climate is a hot-button topic about which nearly everyone has an opinion and which science actively pursuing. Certainly, never before has there been a period of such intense climate research. Because of all the different research and its implications, the Intergovernmental Panel on Climate Change (IPCC) was established. A good summary of the scientific assessment of past, present, and future climate change was produced as part of the IPCC's Third Assessment Report (IPCC (2001b)), and the next IPCC report is due out in 2007.

Because of the dynamic nature of climate, a working definition used by atmospheric scientists for climate is the average weather (usually for a particular season) over a 30-year period (IPCC, 2001a). However, defining climate as an average is not completely adequate. Climate refers to the entire distribution of weather over a long period of time. If one were to completely describe a region's climate, one would not only talk of averages, but would also have to characterize the extremes. If the characteristics of the climate can change, so can the

extremes.

From a statistician's standpoint, weather and climate are wonderful areas in which to work. One would be hard-pressed to find a field where more data have been collected than in weather and climate studies. In addition to the long time series of weather data collected at locations scattered over the globe, we now benefit from data collected by satellites and other remote sensing techniques. Numerical models generate "data" which generate short-term weather forecasts and long-term climate change scenarios. Paleo-climatologists study proxy records such as ice cores and tree rings to gain insight about climates which pre-date historical records. From all this data arise interesting statistical questions, as the data have measurement and/or model error, have differences of support both spatially (point locations versus grid-cell measures) and temporally (instantaneous, hourly, or daily amounts), and have interesting and unresolved spatial and temporal dependencies.

Generally, statistical studies attempt to describe the behavior of the central tendency of the distribution function. A climate application might be to describe the accumulated precipitation of a particular month. In this chapter, we will assume that we have a sample $X_1, \ldots, X_n$ of iid random variables from an unknown distribution F. For our example, these data could be that month's precipitation recorded at a particular location over n years. To describe the distribution, one of the first tasks would be to calculate the sample mean and variance. Given that n were large enough and the variance was assumed to be finite, we could assume from the Central Limit Theorem that the distribution of sample means was approximately Gaussian allowing us to estimate confidence intervals for the mean. If more information than just the mean were desired, a parametric model of the distribution might be assumed and the parameters of that distribution estimated by maximum likelihood or method-of-moments techniques. The distribution might be described non-parametrically through the quantiles of the empirical distribution or by kernel density methods. When describing the main portion of the distribution, one uses most, if not all of the data, disregarding only those observations determined to be outliers which have undue influence on the summarizing statistics.

In weather and climate studies as well as in other fields, sometimes what one wants to characterize is not the usual behavior, but the extreme events. An extreme event is an event with a very high (or low) value and a low probability of occurrence. The extremes of most meteorological variables (e.g. wind speed, precipitation, temperature) are of interest because they can have tremendous human impact. For this reason, it is desired to characterize extremal behavior statistically.

1.2 Background in extreme value theory

Extreme value theory (EVT) is the branch of probability and statistics dedicated to characterizing the behavior of the extreme observations. EVT has its beginnings in the early to middle part of the last century. Much of the early work was done in hydrology, from which we also get the canonical example of an extreme

value statistic, the so-called "100-year flood". Other areas of application include finance where characterizing extremes is essential for insurance and financial risk analysis, and engineering where EVT is used for studies on fatigue, corrosion, and structural failure. There are many excellent books on the subject including Embrechts et al. (1997) which gives a comprehensive mathematical background of the theory, Beirlant et. al. (2004), and Coles (2001b) which focuses on applications and data analysis and whose notation is primarily followed in this thesis.

Contrary to the typical goal of statistics to describe the majority of the distribution, extreme value theory is interested only in describing the behavior found in the upper (or lower) tail. Therefore, it is specifically the outliers which are of interest, and usually in extreme value analyses, the data which describes the typical behavior of the random variable is discarded. Even with large data sets, when studying extremes, one is always data poor.

The familiar statistical techniques fail to give reliable information about a distribution's tail. The sample mean and variance are not helpful in describing extreme observations. Maximum likelihood techniques, if based on the entire data set, get overwhelmed by the non-extreme data and do not accurately model the tail. One cannot estimate empirical quantiles beyond the range of the data. To describe the tail of the distribution, different statistical techniques are needed.

Much of the theory underlying EVT techniques arises from studying block maxima. Intuitively, if given our sample $X_1, \ldots, X_n$, to understand extreme behavior, one would almost certainly inspect the value of the maximum of the sample $M_n = \max(X_1, \ldots, X_n)$. If one had repeated samples of n random variables, and one looked at the maxima of each of these samples, then one might begin to gain an understanding of the behavior found in the tail of the distribution. This idea leads to the notion of describing distributions of maxima which in turn leads to the idea max stability.

Fundamental to the study of extremes is the idea of max stability. Max stability is best understood in contrast to the more familiar idea of stability–that of sum stability. First, let us recall the idea of the distribution of a sum of random variables. As students learn in their first course in statistics, just as one can describe the distribution of the individual random variables X, one can describe the distribution of the sum of the n random variables, $S_n = \sum_{i=1}^{n} X_i$. For some distributions, the distribution of the sum resembles the distribution of the individual random variables, and this property is the idea of stability. A distribution is sum stable if the distribution of the sum S_n is of the same type as the distribution of the original random variables. More specifically, F is sum stable if there exist renormalizing constants $a_n > 0$ and b_n such that $\mathbb{P}\left(\frac{S_n - b_n}{a_n} \leq x\right) = \mathbb{P}\left(X \leq x\right)$. The most familiar sum stable distribution is the Gaussian distribution, but there exist other sum stable distributions such as the Cauchy distribution.

As with the sum, one can also characterize the distribution the sample maximum of the n random

variables, $M_N = \max(X_1, \ldots, X_n)$ (Figure 1.1). A distribution G is max stable if the distribution of the sample maximum is of the same type as G, i.e. if there exist renormalizing constants a_n and $b_n > 0$ such that

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = \mathbb{P}\left(X \leq x\right). \tag{1.1}$$

If X_i has distribution G, then $\mathbb{P}(M_n \leq x) = G^n(x)$. Therefore a distribution F is max stable if

$$G^n\left(a_n x + b_n\right) = G(x). \tag{1.2}$$

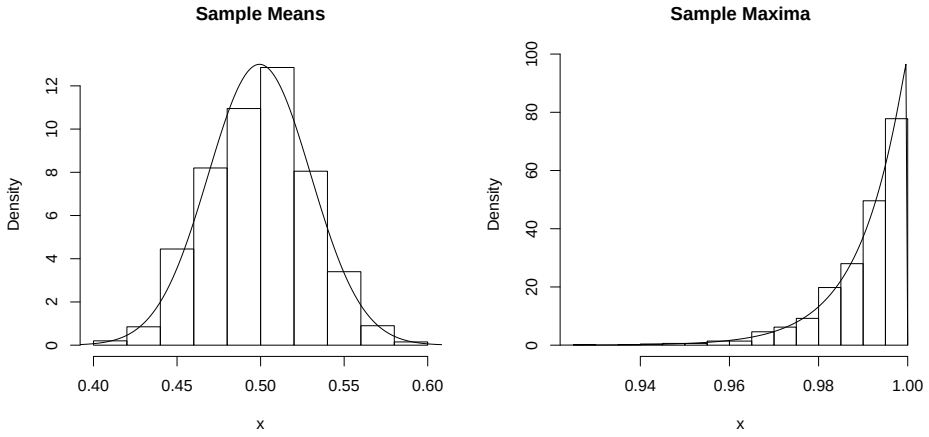

Figure 1.1: The empirical distributions of the mean and maximum of 100 Uniform[0,1] random variables. Histograms are of 1000 repeated experiments. Also plotted are parametric estimates of the distribution of means (Gaussian) and maxima (GEV).

The concept of stability becomes a powerful tool when combined with an asymptotic argument. First, review the case of sum stability. Let $X_1, X_2, \ldots$ be iid random variables with *any* distribution F. If there exist renormalizing constants $a_n > 0$ and b_n such that

$$\mathbb{P}\left(\frac{S_n - b_n}{a_n} \leq x\right) \to G(x),$$

then it can be shown that $G(x)$ must be a sum stable distribution (this result can be found in most theoretical probability books such as Lamperti (1996)). This property is most familiarly known in the specific case of the Central Limit Theorem, which states that if F has finite second moments, then S_n converges to a Gaussian distribution. Therefore, while assuming very little about the distribution of the individual random variables, the sum (or mean) n random variables (provided n is large) can be assumed to approximately follow a Gaussian distribution.

A similar asymptotic argument holds for max stability. If $\{X_n\}$ is a sequence of random variables from *any* continuous distribution F, and there exist renormalizing constants $a_n > 0$ and b_n such that

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

then it can be shown that $G(x)$ must be a max stable distribution (Lamperti, 1996). Hence, one can assume that the behavior of the maximum of a finite random sample $X_1, \ldots, X_n$ (provided n is large) must approximately follow a max-stable distribution, and this assumption can be made without knowing the distribution of the individual random variables.

Armed with the knowledge that the distribution of a maximum converges to a max stable distribution, one's attention turns to trying to characterize the max stable distributions. Fortunately, the class of max stable distributions is quite limited. Fisher and Tippett (1928) initially determined that there exist only three types of max stable distributions, and this fact was later rigorously proved by Gnedenko (1943). The three types of extremal distributions are the Gumbel, the Fréchet and the Weibull and their distribution functions are:

$$\text{Gumbel:} \qquad \Lambda(x) = \exp\left\{-\exp\left[-\left(\frac{x-b}{a}\right)\right]\right\}, -\infty < x < \infty$$

$$\text{Fréchet:} \qquad \Phi(x) = \begin{cases} 0, & x \le b, \\ \exp\left\{-\left(\frac{x-b}{a}\right)^{-\alpha}\right\}, & x > b \end{cases}$$

$$\text{Weibull:} \qquad \Psi(x) = \begin{cases} \exp\left\{-\left[-\left(\frac{x-b}{a}\right)\right]^{\alpha}\right\}, & x < b \\ 1, & x \ge b \end{cases},$$

where $\alpha > 0$ and $a > 0$. The three distributions describe very different limiting behavior in the tail of the distribution (Figure 1.2). The most familiar of the three is the Gumbel distribution, which has an unbounded tail that decreases exponentially. The Gumbel is said to have a light tail and all its moments are finite. Like the Gumbel, the Fréchet distribution has an unbounded tail, but unlike the Gumbel, there exists an N_α such that for all $n > N_\alpha$ all nth-order moments are infinite. Such behavior is termed a heavy tail. Finally, the Weibull distribution has a bounded upper tail.

Each of the aforementioned distributions is a limiting distribution. If $X_1, X_2, \ldots$ have distribution F such that $\mathbb{P}\left((M_n - b_n)/a_n \le x\right) \to \Lambda$, then F is said to be in the domain of attraction of the Gumbel distribution, and two such distributions are the Gaussian and Exponential distributions. Similar definitions hold for the domain of attraction of the Fréchet and Weibull distributions. With the knowledge of the three types theorem, if one knows the domain of attraction of the distribution of X_i, then one can use the appropriate extremal distribution to model the distribution of M_n.

However, it is not required that the domain of attraction of the distribution F is known to be able to model its maxima. This is because the three extremal types can be combined into one parametric family: the Generalized Extreme Value (GEV) family of distributions (Von Mises, 1954). The GEV has distribution function

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\} \tag{1.3}$$

where μ is the location parameter, $\sigma > 0$ is the scale parameter, ξ is the shape parameter, and $a_+ = a$ if $a > 0$

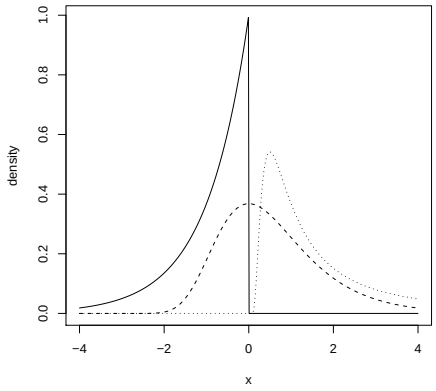

Figure 1.2: Shows the three types of max-stable distributions. Long dashed line is the standard $(a = 1, b = 0)$ Gumbel, short dashed line is the standard $(a = 1, b = 0, \alpha = 1)$ Frechet and solid is the standard $(a = 1, b = 0, \alpha = 1)$ Weibull.

and $a_+ = 0$ otherwise. The parameter ξ controls the tail behavior; if ξ is positive, then the tail is heavy, if $\xi = 0$ (taken as the limit) then the tail is light, and if ξ is negative then the tail is bounded.

The GEV distribution is a powerful tool for modeling extreme values. Just as the Gaussian distribution is used to model finite sums of iid random variables because of the asymptotic result of the Central Limit Theorem, the GEV can be employed to model maxima of a finite number of iid random variables because of the asymptotic result of the three types theorem. Additionally, the GEV is an easy tool to use for modeling. Although the asymptotic results rely on the renormalizing sequence of constants a_n and b_n, the parameters μ and σ alleviate the need to find these sequences (Coles, 2001b). Also, the distribution function, density function, and quantile function of the GEV distribution all have closed-form representations. This allows one to estimate parameters numerically using maximum likelihood methods, although other method-of-moments type estimators are preferred by some practitioners, especially in hydrology.

The three types theorem and GEV distribution provide the starting point for much of EVT. However, in terms of statistical practice, the GEV has one shortcoming. Because it models block maxima, use of the GEV can result in wasting data. For example, assume that one has 50 years of daily precipitation data. To use the GEV, the data must be divided into blocks large enough such that the model is appropriate for the block maxima, and common practice is to use year-long blocks. Hence, the GEV distribution would be fit using only the 50 annual maximum data points. It is not hard to imagine that among the discarded data would be precipitation events that could yield additional information about the distribution's tail.

The Generalized Pareto Distribution (GPD) is another distribution from EVT that addresses some of the shortcomings of the GEV. The GPD models exceedances over a threshold u and, like the GEV, has an asymptotic justification. Assume $X_1, X_2, \ldots$ has distribution F and $M_n = \max(X_1, \ldots, X_n)$ converges to a

GEV distribution with parameter vector (μ, σ, ξ). As $u \to F^*$ where F^* is the right-endpoint of the distribution function F, then

$$\mathbb{P}(X - u \leq x | X > u) \to H(x) = 1 - \left(1 + \frac{\xi x}{\tilde{\sigma}}\right)_+^{-1/\xi}, \tag{1.4}$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$ (Pickands, 1975). As with the GEV, few assumptions are made about the distribution function F, making the GPD a general tool for modeling large observations. The amounts that the observation exceeds the threshold $X_i - u$ are referred to as the exceedances. Statistical application of the GPD requires choosing an appropriate threshold. The choice of the threshold is a classical variance versus bias trade off. Since the GPD models threshold exceedances, choosing a lower threshold allows one to keep more data which results in smaller confidence intervals for the parameter estimates. However, since the distribution has an asymptotic justification, choosing a higher threshold makes the GPD model more appropriate for the tail of the distribution. There are several diagnostics for choosing an appropriate threshold such as the mean excess plot (Davison and Smith, 1990; Embrechts et al., 1997; Coles, 2001b), but the choice is often not obvious.

The derivations of both the GEV and the GPD assume that the data are iid. Weather data (and data in other fields) are often time series for which there is temporal dependence. If one wishes to apply the previously mentioned models, one must account for any temporal dependence in the data. Provided that the extreme observations are nearly independent if they occur sufficiently far apart in time, then it has been shown that the the distribution of the block maxima still converges to a GEV. These conditions are made explicit in Leadbetter (1983). Additionally, weather data is unlikely to be identically distributed. Weather data may have annual and diurnal cycles and may be effected by trends, such as long-term climate change.

In practice, since the GEV models block maxima, there is usually very little temporal dependence between the observations, and parameter estimation proceeds without concern. However, for GPD models, some work must be done to account for the fact that extremes may occur in clusters. Rather than trying to develop methods which specify the joint distribution of clusters of excesses, more often the data is simply declustered. Declustering involves filtering the exceedances such that they are approximately independent, generally keeping the highest observation of each cluster (Coles, 2001b).

1.3 Outline

In this thesis three analyses involving extremes in weather and climate studies are presented. Earlier it was said that extremes are usually of interest primarily because of their significant human and economic impact. Generally such studies try to quantify a worst-case scenario (or at least a really-bad-case scenario). For instance, in hydrology stream-flow measurements are used to estimate quantities such as the aforementioned "100-year flood". The 100-year flood is an example of a return-level, which is simply a high quantile. The n-th year return level is the quantile which has probability $1/n$ of being exceeded at least once in a given year. In

such studies, it is not uncommon (and the application often requires) extrapolation beyond the temporal range of the data, i.e. 50 years of data may be used to estimate the 100 year flood.

The first project, however, is not a worst-case-scenario type of application. The objective of the project is to use proxy data to estimate the ages of paleoclimate events. Extreme value techniques are used because the data are block maxima, but the quantities of interest are not the data values themselves. Hence, the project is an inverse problem, and, to our knowledge, is a unique application of EVT.

The second and third projects revolve around extremes on spatial domains. Central to each project is the need to model spatial dependence. However, because the goals of the two projects are distinct, the spatial dependence in each is modeled very differently. Chapter 3 provides some background in geostatistics and multivariate extremes and then outlines the different nature of the two projects. The projects are then presented in Chapters 4 and 5.

Chapter 6 concludes the thesis. In it is a brief summary of the work completed and some thoughts on future extensions of each of the projects.

Chapter 2

Extreme Value Models for Lichenometry

2.1 Background

Studying long-term climate change has been at the center stage of the Earth environment research for the last few decades. This scientific interest is often justified by the interaction between human activities and the Earth system. For example, the Little Ice Age (LIA), a period of several hundred years in which the winters were particularly severe in the Northern Hemisphere (around 1450-1850 AD), had important influences on human development. One may wonder if the climatic phenomenon was specific to Europe, or if the effect was worldwide and/or synchronous in time. This study can be viewed as a step in answering such questions. More precisely, our overall goal is to gain a better understanding of the timing of the LIA in South America. To reach this goal one obviously requires observations or measurement, but time series of temperatures or precipitation are not available from this tropical region during the LIA. As an alternative, one must rely upon indirect measurements that are classically called proxy records by the climate community.

One proxy record for climate change is glacier moraine formation. The moraines show a record of glacier advances during relatively cool periods and retreats during warmer periods. One can learn about the timing of the climatic events which formed the moraines by estimating their ages. Unfortunately, in the alpine and arctic environments where glaciers are found, there are few environmental recorders available to estimate the moraines' ages. Dendrochronology (tree-ring dating) cannot be used as there generally are no trees, and often the ages of interest are too recent for accurate radiocarbon dating. However, lichens are found in these environments, they live to be centuries old, and they grow throughout their lifetime. For these reasons, lichenometry, the study of using lichen measurements to determine the ages of glacial landforms, was introduced by Beschel in the 1960's (Beschel R.E., 1961). The basic premise of lichenometry is as follows. During relatively cool periods of time, a glacier advances forward forming a moraine. When warming occurs, the glacier recedes allowing lichens to colonize the moraine. A fundamental assumption of lichenometry is that the largest lichens found on a moraine are also the oldest lichens on the moraine. Consequently, lichenometery practitioners believe that the largest lichens provide the best information about the moraine's age, and they generally only record the measurements

of large lichens while in the field. Lichenometry field procedures differ with the practitioner; some record only the largest lichen on each moraine, some record the largest five or ten lichens and take the mean, and some record the largest lichen from a number of blocks on a given moraine. For an introduction to lichenometry, see Innes (1985).

To estimate a moraine's age from lichen measurements, lichen-size-to-age calibration must be performed. Lichen growth rates are affected by climate (Benedict J.B., 1967, 1990, 1991), lithology (Rodbell D.T. , 1992), and exposition (Pentecost A., 1979). This implies that the calibration must be done locally. To perform such a calibration, lichen measurements are taken from surfaces with known ages. These measurements generally come from nearby features such as bridges, tombstones, mine tailings, or natural features that have been dated by other methods. Classically, these measurements from dated features are used to construct a growth curve to which the moraines' lichen measurements are compared.

Two primary methodological problems have not been addressed in previous lichenometry studies. First, the majority of past statistical lichen analysis does not take advantage of the specificity of the data. The data of lichenometry are the extreme observations, yet EVT is generally not applied to the analysis of these data. Second, margins of error for the estimated dates are rarely present in lichen analysis. In the few studies where confidence intervals or p-values were computed, these calculations were based on the assumption of normality (O'Neal M. and Schoenenberger K. R., 2003). An assumption of a Gaussian distribution is justified for averages, but not maxima (Figure 2.1). The only two studies which have used EVT (Bull W.B. and Brandon M.T. (1998); Karlen W. and Black J. (2002)) arbitrarily assumed that the lichen measurements have a Gumbel distribution, and no statistical model based on the Gumbel was developed to derive estimation errors or confidence intervals. The difficulty of determining statistical uncertainties is that a reasonable statistical model must be developed. Although lichenometry has been used for over forty years, the practice has never had a statistical model at its foundation. This work represents the first comprehensive statistical models for lichenometry based in EVT.

In the following sections of this chapter, two different models are detailed which link the lichen diameter measurements to the known and unknown ages of the features from which the measurements are taken. As the data analyzed are block maxima, the GEV forms the foundation of both models. In Section 2.2 a likelihood-based model is described that treats lichen diameter and moraine age as a bivariate vector and employs the EM algorithm to fill in the unknown ages of the moraines. The second model, introduced in Section 2.3, constructs a Bayesian hierarchical model which treats the unknown ages as hidden covariates. This more complex model arose from a need to model spatial effects in a larger data set.

Both lichenometry models are unique to the extremes literature. The lichenometry application is novel because the quantities of interest (the ages of climatic events) are not the measured quantities (lichen diameters), i.e. it is a inverse problem. As explained in Section 1.3, extreme value theory is generally used to describe the

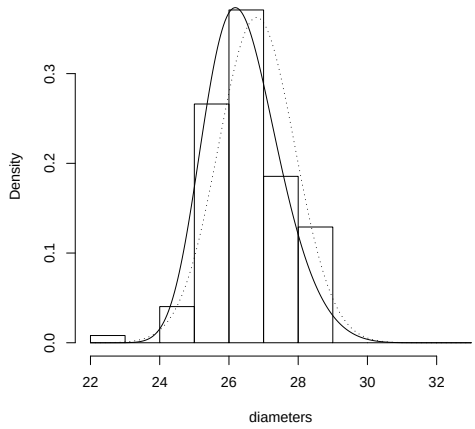

Figure 2.1: Histogram of lichen measurements from the fourth moraine of the Charquini glaciers. The solid line shows the GEV which was fit to the data as a result of the second lichenometry model. The dashed line shows a normal distribution fit to this data. Although the distributions have a similar appearance, the GEV is asymmetric and its shape parameter ξ is negative indicating a bounded tail.

extreme behavior of the measured quantity. In this application, one is not interested in predicting future observations in the data (the 100-year return level of a lichen?), but instead the interest lies in what the maximum lichen measurements can tell us about the ages of the moraines.

The models have several advantages over previous lichenometry studies. They address the two methodological problems mentioned above: they each have a theoretical framework based in EVT, which, in turn, appropriately accounts for the uncertainty associated with the estimates of moraine ages. Additionally, both models can accommodate different growth curve models and the models' fits can be compared. Uncertainty estimates are also provided for the parameters which determine the growth curve function. Finally, our models analyze the data from both dated features and the undated moraines of interest with a one-step procedure instead of treating the data from the dated features such as bridges and tombstones separately from that of the undated moraines.

Our data comes from the Cordillera Real region of Bolivia (Figure 2.2). Very few lichenometry studies have been undertaken in Bolivia and a growth curve has not been previously constructed for this region. On each moraine, a number of blocks were identified, and the largest lichen measurement was recorded for each block. For diameter-to-age calibration purposes, measurements were taken from features with known dates (e.g. mine tailings, tombstones and bridges) in a similar manner to the moraines. The first model analyzes lichen measurements from ten moraines of the South Charquini glacier. The second model analyzes measurements from four additional glaciers, all of which surround on Mount Charquini.

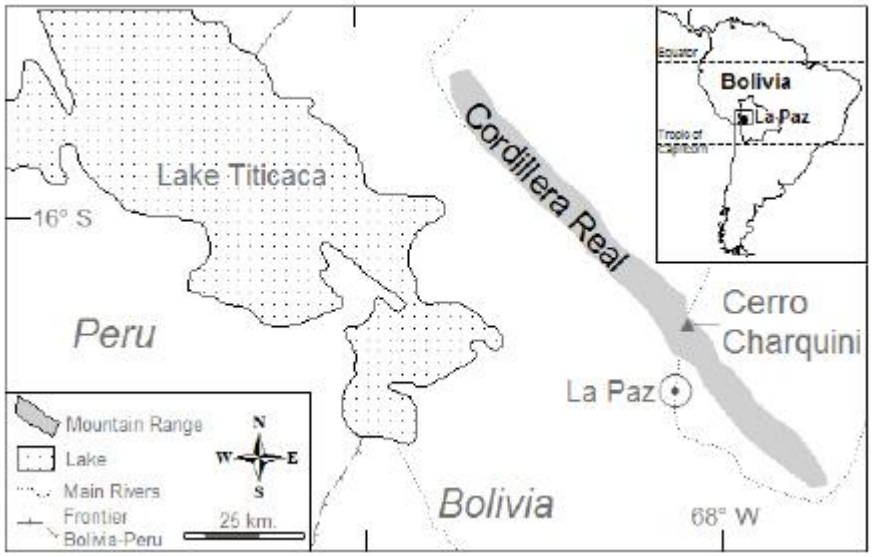

Figure 2.2: Map of Cordillera Real region of Bolivia.

2.2 Bivariate lichenometry model

The inverse nature of this application is that the data (lichen measurements) indirectly relate to the quantities of interest (moraine ages). To develop a statistical model for analyzing lichen diameters, one first needs to identify the main sources of randomness. There are two sources of uncertainty; the error in measuring lichen diameters and the error when associating a date with a lichen. The first approach models a bivariate random vector, $(\boldsymbol{X}_i, T_i)$; the vector of maximum diameter measurements $\boldsymbol{X}_i = [X_{i,1}, \ldots, X_{i,n_i}]$ and the moraine's age T_i , where the moraine index i varies from $i = 1$ to m, the total number of features (dated and undated) and n_i is the number of blocks found on moraine i. Apart from a few cases for which the date of the feature is known, the variable T_i is not observed and has to be determined. How to estimate this date is the most important question in the analysis of lichen diameters.

2.2.1 Model structure

Both the lichen diameter measurements and moraine ages are assumed to be random variables, so an appropriate distribution for each must be chosen and a link between the two distributions must be constructed. When modeling the diameters, the observations are maxima

$$X_{i,j} = \max_{k=1,\ldots,n_{i,j}} \{X_{i,j,k}\}, \text{for } i = 1, \ldots, m,\ j = 1, .., n_i$$

where $X_{i,j,k}$ are the $n_{i,j}$ lichen diameters on the jth block from the ith moraine and m is the number of moraines. The lichen diameters inside a block are assumed independent and the measurements for all blocks on a moraine

are assumed to follow the same distribution. Between moraines, the lichens are assumed to be independent but they do not have the same distribution. This assumption is needed in order to find a temporal growth curve. Although the value of $n_{i,j}$ and $X_{i,j,k}$ are not necessarily known, it is assumed that the maximum $X_{i,j}$ comes from a sample of lichen diameters whose size is relatively large.

Because the data are block maxima, it is natural to assume that the $X_{i,j}$ are distributed approximately GEV whose parameters can vary in time through the index of the moraine: $GEV(\mu_i, \sigma_i, \xi_i)$. The location μ_i describes the size of the lichen measurements found on moraine i, which the study of lichenometry assumes changes with the age of the moraine. We denote by $\vec{\mu}$ the vector $[\mu_1, \dots \mu_m]$. The scale parameter σ_i might also vary with the age of the moraine; perhaps the maximum lichen measurements of older moraines show more variability. However, for our initial studies, we will assume that σ_i is a constant σ_0. In practice, we model $\log \sigma_0$ since the scale parameter must be positive. The shape parameter ξ_i controls tail behavior and represents the large scale behavior of the lichen distributions. It is more likely that the shape parameter is controlled by the biology of the lichens than by effects which have varied over the last few centuries. Furthermore, since the shape parameter is the most difficult to estimate, we borrow strength from all the observations and assume that ξ_i is a constant ξ_0.

We likewise need to define a reasonable distribution function for the random variable T_i. We assume that T_i follows a Gaussian variable with mean α_i and standard deviation β_i for $i = 1, ..., m$. We choose a symmetric distribution because the lichen ages on a given moraine, say i, are centered around a mean value, say α_i, that represents the age of the moraine i. Interpreting the standard deviation is not as straightforward. In lichenometry, there are two categories of dated surfaces. The first one is characterized by the presence of a confidence interval. For example, a date computed with a ^{14}C carbon dating procedure is classically associated with a lower and upper bound. The second category corresponds to the situation in which the estimated moraine date is not associated with any error analysis, such as the date of a man-made structure. In this case, we have two options, either we assume that the estimated date is perfect, or we arbitrarily choose to fix a small standard deviation β that characterizes the a priori error. Because we prefer to overestimate rather than underestimate the error for our undated surfaces, we opt for the introduction of an apriori error.

The last step in our construction is to combine the distribution of the maximum lichens with the moraine dates (both dated and undated). Let $\boldsymbol{X} = [X_{1,1}, \dots X_{1,n_1}, X_{2,1}, \dots X_{2,n_2}, \dots, X_{m,1}, \dots X_{,,n_m}]^T$, and $\boldsymbol{T} = [T_1, \dots T_m]$. We assume that the joint distribution of $(\boldsymbol{X}_i, T_i)$ can be written as the product of the GEV and Gaussian distributions where the parameters are linked through the relationship $\alpha_i = g(\mu_i)$. In other words, we assume that the lichen diameters and the moraine dates are independent in terms of distribution but are linked

through their parameters. By differentiating (1.3) one obtains the likelihood function:

$$
\begin{aligned}
l\left(\vec{\mu}, \sigma_0, \xi_0 | \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{T} = \boldsymbol{t}\right) &= \prod_{i=1}^{m} \left\{ \prod_{j=1}^{n_i} \left\{ \exp\left[-\left(1 + \xi_0 \left(\frac{x_{i,j} - \mu_i}{\sigma_0}\right)\right)^{-\frac{1}{\xi_0}} \right] \right.\right. \\
&\quad \left. \times \left[\frac{1}{\sigma_0} \left(1 + \xi_0 \left(\frac{x_{i,j} - \mu_i}{\sigma_0}\right)\right)^{-\frac{1}{\xi_0} - 1} \right] \right\} \\
&\quad \left. \times \frac{1}{\sqrt{2\pi\beta_i^2}} \exp\left[-\frac{1}{2\beta_i^2} (t_i - g(\mu_i))^2 \right] \right\}, \qquad (2.1)
\end{aligned}
$$

where $\boldsymbol{x} = [x_{1,1}, \ldots x_{1,n_1}, x_{2,1}, \ldots x_{2,n_2}, \ldots, x_{m,1}, \ldots x_{,,n_m}]^T$, and $\boldsymbol{T} = [t_1, \ldots t_m]$. For our Bolivian lichen data, we chose $\beta_i = 3$, for all moraines.

The function g takes the role of the growth curve in this model. As a consequence, the notion of a growth curve takes on a new meaning (Figure 2.3). Rather than being a simple plot between the age of the feature and the lichen diameters directly, the growth curve is now a relationship between the age of the feature and the GEV parameter μ. For this preliminary study we model the growth curve with a linear model letting $\alpha_i = a + b\mu_i$. However, the growth curve can easily be changed to accommodate different growth models with respect to time.

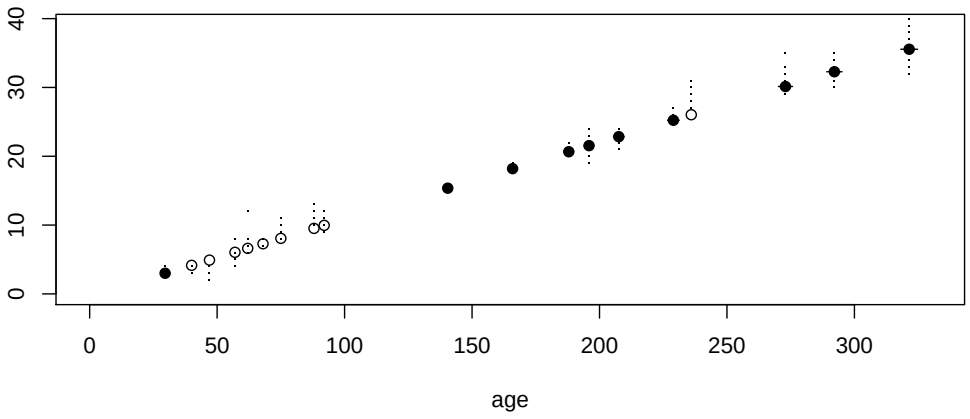

Figure 2.3: A new type of growth curve plot which relates each feature's GEV location parameter μ to the feature's age (empty circles correspond known ages; filled circles to estimated ages). Also plotted are a sample of the lichen measurements from the feature (dots).

2.2.2 Estimation procedure

Given the joint density (Equation (2.1)) and the random vectors $(\mathbf{X_i}, T_i)$, we could estimate the parameters $\vec{\mu}$, σ_0, ξ_0, a, and b using numerical maximum likelihood methods. However, the variable T_i is only observed for the features which have already been dated, and this random variable must be estimated for the moraines which have no dates. To deal with the missing dates, we resort to the EM algorithm. The recursive

algorithm uses the current parameter estimates to find the conditional expected values of the missing dates, and then obtains the best parameter estimates given those dates via maximum likelihood (Dempster, et al. 1977, McLachlan and Thriyambakam 1997). The algorithm is allowed to run until the likelihood converges, yielding both point estimates for the parameters and the ages of the undated moraines.

We rely on asymptotic results to obtain confidence intervals for our parameters and age estimators. Let $\boldsymbol{\theta} = [\mu_1, \ldots, \mu_m, \log \sigma_0, \xi_0, a, b]^T$. It is well known that the distribution of the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ is asymptotically

$$N\left(\boldsymbol{\theta}, \mathcal{I}^{-1}(\boldsymbol{\theta})\right) \tag{2.2}$$

where $\mathcal{I}$ is the Fisher information matrix. To calculate the confidence intervals of $\boldsymbol{\theta}$ we replace $\mathcal{I}(\boldsymbol{\theta})$ in (2.2) with $\hat{\mathcal{I}}(\hat{\boldsymbol{\theta}})$, the numerically approximated empirical information matrix evaluated at the maximum likelihood estimates.

We also wish to obtain confidence intervals for the age estimates $\hat{\alpha}_i$. For the Bolivian study, we used the function

$$\hat{\alpha}_i = \hat{a} + \hat{b}\hat{\mu}_i$$

to relate the feature's age to the GEV location parameter, but our method could be used for other simple growth curve models. To obtain confidence intervals for the age estimates, we resort to the delta method which uses a first-order Taylor approximation of the growth curve function to derive the asymptotic variance of $\hat{\alpha}_i$. Let $\boldsymbol{\theta}_i = [\mu_i, a, b]^T$. Then according to the delta method, the asymptotic distribution of $\hat{\alpha}_i$ is

$$\hat{\alpha}_i \tilde{\sim} N(\alpha_i, \bigtriangledown \alpha_i^T V_i \bigtriangledown \alpha_i)$$

where $\bigtriangledown$ represents the gradient operator and V_i is the sub-matrix of $\mathcal{I}^{-1}$ for the parameters $[\mu_i, a, b]$. For more background on the delta method see Casella and Berger (2002).

2.2.3 Results from both simulated data and Bolivian glacier data

Before embarking on analyzing the data from the South Charquini Glacier, we first test the method's ability to estimate parameters and unknown ages by running it on simulated data. We simulate data with known parameter values and dates and then hide some of the dates to create a data set similar to the one we are given. The simulation whose results are shown in Table 2.1 was for eight dated features and four undated moraines.

The method recovered accurate values for the hidden parameters, and yields relatively narrow confidence intervals for the GEV parameters $\vec{\mu}$ (not shown), σ_0, ξ_0, and the slope of the growth curve b, indicating that the data provides good information about these parameters. The point estimates for the estimated dates are close to the hidden values as well, and not surprisingly, the confidence intervals are wider as the dates are further

removed from the data than are the other parameters. Notice, however, that the confidence intervals for one of the moraines does not include the actual simulated age indicating that our asymptotically-justified confidence intervals may be too narrow.

	GEV parameters		linear parameters	
	$\log(\sigma_0)$	ξ_0	a	b
Simulated Value	0.5	-0.2	-20	10
Point Estimates	0.50	-0.18	-17.90	9.93
0.90 Conf Intvl	(0.43, 0.57)	(-0.23, -0.13)	(-21.26, -14.54)	(9.78, 10.10)
Moraine Ages				
Simulated Age	70	60	280	120
Point Estimates	75.6	60.5	281.6	124.1
0.90 Conf Intvl	(71.1, 80.1)	(56.2, 64.7)	(273.7, 289.4)	(118.8, 129.4)

Table 2.1: Results for the bivariate method on simulated data with eight dated features and four undated features.

The method was then applied to the South Charquini Glacier data. As with the simulated data, the method achieves narrow confidence intervals for the GEV parameters and the slope of the growth curve. The confidence intervals for the most recent events (M10 in Table 2.2, for example) are quite narrow, but the delta method causes the uncertainty to grow with the age of the moraine.

	GEV parameters		linear parameters	
	$\log(\sigma)$	ξ	a	b
Point Estimates	0.368	-0.184	2.69	8.97
0.90 Conf Intvl	(0.30, 0.43)	(-0.23, -0.14)	(0.29, 5.09)	(8.87, 9.07)

Moraine Ages					
	M1	M2	M3	M4	M5
Point Estimates	321.5	292.1	272.9	229.0	207.6
0.90 Conf Intvl	(318.2, 324.8)	(289.1, 295.1)	(270.1, 275.7)	(226.7, 231.3)	(205.5, 209.7)
	M6	M7	M8	M9	M10
Point Estimates	195.8	188.0	165.9	140.5	29.6
0.90 Conf Intvl	(194.0, 197.8)	(186.1, 189.8)	(164.3, 167.5)	(139.1, 141.8)	(29.1, 30.1)

Table 2.2: Results for the bivariate method on the South Charquini Glacier data.

This model has been effective in inferring moraine age from lichen measurements. Its age estimates were accepted as reasonable by the field scientists. It additionally provides confidence intervals for the parameter estimates and by extension, the estimated ages. The model has an appropriate foundation based in extreme value theory, and from an extremes perspective, the results show clear evidence of bounded tail behavior ($\xi_0 < 0$), implying that the Gumbel model suggested by other researchers is incorrect.

Several issues suggest the need for a more accurate statistical formulation. First, we feel that setting the variance parameter of the Gaussian distribution for the ages at $\beta = 3$ is an arbitrary assumption, and that it would be better to estimate this parameter. Unfortunately, using the bivariate model, this is not possible. Each feature's age is represented by a Gaussian distribution, and unfortunately, we have only one age observation for each feature, making it possible to only estimate one parameter (the mean). When using the EM algorithm, the Gaussian distributions converge to a point mass at the mean value if we attempt to also estimate the variance.

Second, we feel that treating the ages as random variables is not completely natural. If all of the ages were known, these would be treated as covariates rather than as components of a bivariate random variables. Third, the confidence intervals for the ages of the features are perhaps too narrow. Relying on asymptotic results and fixing the variance parameter of the Gaussian distribution have probably caused the confidence interval to be a bit narrow. It is hard to believe that we could find the age of a 300-year old feature with a 90% confidence interval spanning less than seven years. Finally, we were given a new data set which consisted of data from multiple glaciers. In addition to the South Charquini glacier, lichen measurements from the moraines of four other glaciers surrounding Mount Charquini were obtained. Knowing that lichen growth can differ by location, we hoped to construct a model which could model possible spatial variability. The shortcomings of the bivariate model and the desire to model spatial variability led us to a second and very different statistical model for lichenometry: a Bayesian hierarchical model.

2.3 Bayesian hierarchical model

Hierarchical models allow statistical modeling of data that have many parameters which are connected by the structure of the problem. These models have the flexibility to fit a large data set of related but not identically-distributed data. Bayesian hierarchical models build up a complicated statistical model through a set of conditional relationships that are more simple in form. Recent computing advances and Markov Chain Monte Carlo (MCMC) techniques have allowed parameter esimation to be feasible for more complex models. MCMC takes advantage of the conditional structure to generate approximate samples from a posterior distribution. For an introduction to Bayesian hierarchical models see Gelman, et al. (2003) or Carlin and Louis (2000).

Our data set of block lichen measurements fits quite well into a hierarchical model. The lichens have a common growth behavior, allowing all the data to be modeled together. However the lichen measurements are not identically distributed due to the differences in the feature's ages and location. These differences can be accounted for by varying the GEV distribution's parameters.

The Bayesian hierarchical model is more flexibile than the bivariate model. First, the model can account for differences in lichen growth between the measurements of the different glaciers. Additionally, the hierarchy handles the ages in a more intuitive manner. Rather than treating them as a component of a bivariate vector, the ages are treated as covariates, i.e. there is a conditional relationship between the lichen measurements and the ages. Because we now approach the problem from a Bayesian perspective, these "covariates" are treated as random variables and are assigned prior distributions.Furthermore, the Bayesian structure allows us to build prior information into the model. We know that the age of a downslope moraine must be older than those further uphill and closer to the remaining glacier. This information is not used in the likelihood-based model, but can be built into the Bayesian model by constraining the order of the moraine ages.

For this analysis, we augment the South Charquini data set used in the previous study with data from the four other glaciers (Northeast, Southeast, West, and North) surrounding Mount Charquini. On four of these glaciers, measurements were taken on ten different moraines which are numbered M1-M10 with M1 being the oldest and furthest from the existing glacier. On the Northeast glacier, measurements were taken only on eight moraines as two moraines (M2 and M5) were not distinguishable. From a statistical and climatic point of view, we assume that cooling and warming events are widespread, causing all of the glaciers to advance and recede at basically the same time. A strong climatic event causes each of the five glaciers to create a moraine, hence moraines from different glaciers have corresponding ages. It follows that our dating problem corresponds to finding the timing of each climatic event.

Like the first model, this analysis is unique to the extremes literature because its inverse nature: the quantities of interest (ages) are not the measured quantities (lichen diameters). This model is constructed such that the moraine ages are treated as hidden covariates. Extremes models with *known* covariates have been studied in the past. Davison and Smith (1990) described a threshold exceedance model with known covariates, and examples appear in Coles (2001b). Recently, Wang et.al. (2004) modeled wave heights with known atmospheric covariates using a GEV distribution. However, to our knowledge no one has attempted to estimate parameters related to latent variables in the context of extremes. Additionally, this work is novel as there are very few examples of hierarchical models in the extremes literature. We build a hierarchical structure to model data which varies due to both location and age. One earlier study applied a Bayesian hierarchical model to model only spatial dependency in simulated wind speed data (Casson and Coles, 1999).

The section is structured as follows. In Sections 2.3.1 and 2.3.2 we present the GEV-based hierarchical model and our method for model inference. In Section 2.3.3 we show results from a simulated data set and the Mount Charquini data set, including a comparison between two different growth curve models. We conclude in Section 2.3.4 with a discussion.

2.3.1 Model structure

To begin constructing our model, we assume the distribution of diameter measurements varies from glacier to glacier and moraine to moraine. Let $X_{i,j,k}$ be the maximum lichen diameter for glacier i, climatic event j, and block k. We assume that on the moraine from glacier i and corresponding to event j, the measurements $X_{i,j,k}$ are iid for $k = 1, \ldots n_{i,j}$. As the data are block maxima, we model the data using a $GEV(\mu_{i,j}, \sigma_{i,j}, \xi_{i,j})$ distribution where the parameters vary with moraine location and age.

Since we assume that the glaciers all advance together due to a widespread climatic event, moraines from different glaciers correspond to the same events. By indexing on climatic events, we pool the data from corresponding moraines of different glaciers. This allows us to borrow strength from all the corresponding

moraines when estimating the age of the event. Additionally, since moraines formed by the same climatic event may have slightly different ages due to variation in glacier response times, pooling corresponding moraines provides a more accurate estimate of the timing of the climatic event. We focus on the location parameter $\mu_{i,j}$ which describes the size of the lichens found on glacier i and the moraine formed by climatic event j. We denote our vector $\vec{\mu}$ as follows,

$$\vec{\mu} = \Bigg[\underbrace{\mu_{0,1}, \ldots, \mu_{0,n_0}}_{\text{dated features}}, \underbrace{\mu_{1,1}, \ldots, \mu_{1,n_e}}_{\text{glacier 1}}, \underbrace{\mu_{2,1}, \ldots, \mu_{2,n_e}}_{\text{glacier 2}}, \ldots, \underbrace{\mu_{n_g,1} \ldots, \mu_{n_g,n_e}}_{\text{glacier g}}\Bigg]^T, \tag{2.3}$$

where $i = 0, \ldots, n_g$ with n_g is the number of glaciers. The glacier index $i = 0$ corresponds to the location parameters of the features with known dates like bridges, while $i = 1, 2, \ldots, 5$ correspond to the South, Northeast, Southeast, West, and North glaciers respectively. For the dated features, $j = 1, \ldots, n_0$ where n_0 is the number of features observed. For the undated features, $i > 0$ and $j = 1, \ldots, n_e$ where n_e is the number of moraine-forming climatic events. Therefore $\mu_{1,j}$ and $\mu_{2,j}$ represent the same event j for glaciers 1 and 2, and they share the same unknown covariate date. However $\mu_{0,j}$ and $\mu_{1,j}$ do not correspond to the same event as the ages of the jth dated feature should not be the same as the age of the jth climatic event. Also, the vector $\vec{\mu}$ may not be complete. A glacier may be missing a moraine which corresponds to a certain event. This occurs when moraines become unidentifiable over time due to erosion as is the case with our data where the Northeast glacier did not have moraines M2 or M5.

In our Bayesian hierarchical model, we treat $\vec{\mu}$ as a random variable with an assumed prior distribution and build the model structure to link it to the age and location of the feature. We assume a multivariate normal prior distribution where the mean vector and covariance matrix of $\vec{\mu}$ are functions of age and location. We let

$$\vec{\mu} \sim MVN\left(f(\vec{\alpha}, \vec{\delta}, \vec{\gamma}, \vec{z}), g(\vec{\beta}, \vec{\delta}, \vec{\eta}, \vec{z})\right) \tag{2.4}$$

where f is a vector-valued function, g is a positive definite matrix, $\vec{\alpha}$, $\vec{\gamma}$, $\vec{\beta}$ and $\vec{\eta}$ are second-level parameters, $\vec{z}$ is the glacier location and $\vec{\delta}$ is the vector of known and hidden covariate dates. Since the vector of dates $\vec{\delta}$ is only partially known, it too is a second-level parameter. We let

$$\vec{\delta} = \Bigg[\underbrace{d_1, \ldots, d_{n_0}}_{\text{known ages}} \underbrace{\delta_1, \ldots, \delta_{n_e}}_{\text{glacier 1}} \underbrace{\delta_1, \ldots, \delta_{n_e}}_{\text{glacier 2}} \ldots \underbrace{\delta_1, \ldots, \delta_{n_e}}_{\text{glacier g}}\Bigg]^T \tag{2.5}$$

where d_k is the known fixed date of the kth dated feature, and δ_j is the unknown date of climatic event j. For computational purposes, the date vector $\vec{\delta}$ has the same number of elements as the GEV parameter vector $\vec{\mu}$. However, $\vec{\delta}$ has fewer unique elements than $\vec{\mu}$ as the unknown dates $\delta_1, \ldots, \delta_{n_e}$ get repeated for each glacier. Therefore parameters $\mu_{i,j}$ and $\mu_{i',j}$ from corresponding moraines on different glaciers share the same covariate date δ_j.

We specify the functions f and g which give the mean vector and covariance matrix for the multivariate normal distribution of $\vec{\mu}$. We first examine the mean function $f(\vec{\alpha}, \vec{\delta}, \vec{\gamma}, \vec{z}) \rightarrow \Re^{(\dim(\vec{\mu}))}$. For the mean function, we assume an additive model between the temporal and spatial pieces

$$f(\vec{\alpha}, \vec{\delta}, \vec{\gamma}, \vec{z}) = f^{(1)}(\vec{\alpha}, \vec{\delta}) + f^{(2)}(\vec{\gamma}, \vec{z}). \tag{2.6}$$

The function $f^{(1)}$ relates the size of the lichens to the age of the event which formed the moraine. This term takes the place of the growth curve in traditional lichenometry. The second term $f^{(2)}$ relates the size of the lichens to the location of the glacier, allowing us to model how lichen growth varies over a geographic region.

We compare two different growth curve models, although it would be easy to test other growth curve functions. Our first growth curve is simply a linear function of age.

$$f_a^{(1)}(\vec{\alpha}, \vec{\delta}) = \alpha_1 \vec{1} + \alpha_2 \vec{\delta} \text{ with } \vec{\alpha} = [\alpha_1, \alpha_2]^T \text{ and } \vec{1} = [1, ..., 1]^T. \tag{2.7}$$

Based on the bivariate model, a linear growth curve seemed to fit the Charquini South lichen data reasonably well. However, it is well known by geomorphologists that the lichen growth is better modeled by an increasing and concave down function than a simple linear curve. This behavior is due to the faster growth of younger lichens and the slower rate for older lichens whose diameters tend to plateau as their ages increases (Innes J. L., 1985). To accommodate this physical phenomenon, we propose a second growth curve model

$$f_b^{(1)}(\vec{\alpha}, \vec{\delta}) = \alpha_1(\vec{1} - \alpha_2 \exp(-\alpha_3 \vec{\delta})) \text{ with } \vec{\alpha} = [\alpha_1, \alpha_2, \alpha_3]^T. \tag{2.8}$$

For this study, the spatial term $f^{(2)}$ of the mean function is quite simple. We let

$$f^{(2)}(\vec{\gamma}, \vec{z}) = \left[\underbrace{0, 0, \ldots, 0}_{\text{dated features}}, \underbrace{\gamma_1, \gamma_1, \ldots, \gamma_1}_{\text{glacier 1}}, \underbrace{\gamma_2, \gamma_2, \ldots, \gamma_2}_{\text{glacier 2}}, \ldots, \underbrace{\gamma_{n_g}, \gamma_{n_g}, \ldots, \gamma_{n_g}}_{\text{glacier g}} \right]^T. \tag{2.9}$$

Similar to $\vec{\delta}$, for computational purposes the above vector has the same number of elements as $\vec{\mu}$, but now the number of unique elements is equal to the number of glaciers. The spatial term $f^{(2)}$ is not fully spatial, as it does not take into account the location of the glacier. The location $\vec{z}$ is used only to indicate from which glacier the data come. There are two reasons for this. First, we do not have precise location information for the moraines of the five glaciers surrounding Mount Charquini. Second, the five glaciers from which the data come are very close, and the a priori belief was that there should be no spatial difference between these glaciers. In fact, this is the hypothesis which the function $f^{(2)}$ as defined tests. If location information were available, the function $f^{(2)}$ could be changed to include a spatial trend.

Just as we did with the mean, we make the covariance matrix of the multivariate normal a function of the age and location of the moraine. Due to the structure of $\vec{\mu}$, its covariance matrix S has a block structure (Figure 2.4). Because we assume that the moraines of different glaciers correspond to the same climate events,

blocks B, C, and D in the matrix get repeated. Block A is the covariance matrix of the dated features with themselves, and blocks B are the covariance matrices of the dated features with the undated moraines. Blocks C are the covariance of the undated moraines of a glacier with those of the same glacier, and blocks D model the covariance of the undated moraines between two glaciers.

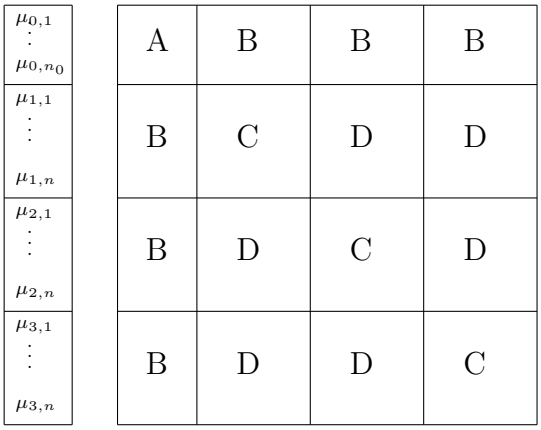

Figure 2.4: Shows the structure of the covariance matrix. Block A models the covariance of the dated features (bridges, etc) with themselves. Blocks B are the covariance matrices of the dated features and the moraines. Blocks C model the covariance of the moraines within the same glacier and blocks D model the covariance of the moraines between different glaciers.

We specify g, the function which relates the covariance matrix of $\vec{\mu}$ to the age and location of the moraine. As with f, we separate the temporal and spatial parts of g, letting

$$S = g(\vec{\beta}, \vec{\delta}, \vec{\eta}, \vec{z}) = U \circ V \tag{2.10}$$

where $U = g^{(1)}(\vec{\beta}, \vec{\delta})$ is the temporal piece, $V = g^{(2)}(\vec{\eta}, \vec{z})$ is the spatial piece, $g^{(i)} \rightarrow \Re^{(\dim(\vec{\mu}))} \times \Re^{(\dim(\vec{\mu}))}$ and $U \circ V$ denotes the Hadamard (component-wise) product. The elements of both matrices U and V are positive, as we assume there are no negative covariances.

For the Bolivian lichen data, we model the temporal piece with the function

$$g^{(1)}(\vec{\beta}, \vec{\delta}) = \beta_1 \exp\{-\beta_2 \Delta\}, \tag{2.11}$$

where Δ is a matrix of absolute age differences from $\vec{\delta}$. Thus, the covariance between moraines decreases as their age difference increases.

As with the mean structure, the spatial piece of the covariance matrix is again quite simple because we lack location information. Blocks A and C of the covariance matrix have no spatial piece as those blocks contain the covariance within a single glacier. Blocks B and D contain the covariance between glaciers. Since we have

no distance information, we simply let

$$g^{(2)}(\vec{\eta},\vec{z}) = \begin{cases} 1 & \text{for elements in Blocks A \& C} \\ \eta & \text{for elements in Blocks B \& D} \end{cases} \tag{2.12}$$

where $\eta \in [0,1]$. This allows the covariance between moraines of different glaciers to be weaker than the covariance between moraines of the same glacier. As with the mean function $f^{(2)}$, if location information were available, one could change the model such that $g^{(2)}(\vec{\eta},\vec{z})$ would take the information into account when computing the covariance.

Because the ages of the climatic events get repeated for the five glaciers, the temporal matrix U is only positive-semidefinite. The spatial matrix V for our application is also only positive-semidefinite because columns in the matrix get repeated since we have incomplete location information. However, since both U and V are positive-semidefinite and $U \circ V$ can be shown to be non-singular, then $U \circ V$ is positive-definite (Appendix .1).

Having developed a hierarchy for the GEV location parameter μ, we could likewise do the same for the scale and shape parameters. Since we believe the tail behavior of the distribution of maximum lichen measurements is due to biological factors, we choose to model the shape parameter as a constant value ξ_0. However, it is quite possible that the scale parameter of the distribution varies with the age of the feature. For now, we model the scale parameter as a constant σ_0, but this could be investigated in future work.

Priors

Set 1	Set 2
Alpha	
Linear Growth Curve	
$\alpha_1 \sim N(0,4)$	$\alpha_1 \sim N(0,25)$
$\alpha_2 \sim Gamma(2,10)$	$\alpha_1 \sim Beta(2,10)$
Exponential Growth Curve	
$\alpha_1 \sim N(40,100)$	$\alpha_1 \sim N(50,400)$
$\alpha_1 \sim N(0,4)$	$\alpha_2 \sim N(0,16)$
$10\alpha_3 \propto Beta(2,2)$	$10\alpha_3 \sim Beta(1,1)$
Other Second-Level Parameters	
$\beta_1/20 \propto Gamma(2,10)$	$\beta_1/10 \propto Beta(4,10)$
$\beta_2 \sim Gamma(2,10)$	$\beta_2 \sim Beta(4,10)$
$\eta \sim Beta(2,2)$	$\eta \sim Beta(1.2,1.2)$
$\gamma_i \sim N(0,4)$	$\gamma_i \sim N(0,9)$
Dates	
$\vec{\delta} \propto Unif(0,\infty)$ provided $\delta_i > \delta_j$ for $i < j$	
GEV Parameters	
$\log\sigma \propto Unif(-\infty,\infty)$	$\xi \propto Unif(-\infty,\infty)$

Table 2.3: Priors assigned to the parameters for the simulations presented.

In the hierarchical structure for $\vec{\mu}$, we introduce several new second-level parameters: $\vec{\alpha},\vec{\beta},\vec{\gamma}$, and $\vec{\eta}$, and the parameter for the unknown covariate dates $\vec{\delta}$. Since we are using a Bayesian framework, we treat

each of these parameters and the remaining GEV parameters ξ_0 and σ_0 as random variables and give each a prior distribution. For the GEV parameters, we know little about what to expect for the scale and shape of the distribution, so we put non-informative priors on these parameters. However, better prior information is available for the second-level parameters. The parameters $\vec{\alpha}$ determine the growth curve of the lichens through time. From previous lichenometry studies, we have a good idea of how the growth curve should behave, and we can adopt priors for $\vec{\alpha}$ which reflect this knowledge. Our priors for $\vec{\gamma}$ reflect our knowledge that although lichen growth varies due to geography, we expect to see only small geographical differences among our data which comes from a small region. Our priors for $\vec{\beta}$ and η are designed to reflect a reasonable and flexible covariance structure for data with values in the observed ranges. As for the dates, we must again assume an uninformative prior distribution. However, we do know the order in which the moraines formed. A moraine further from the current glacier must pre-date any moraine closer to the glacier, and we enforce this moraine order on the dates. To test the priors' effect on the results, MCMC simulations were run with two different sets of priors and results were similar. A list of the prior distributions is given in Table 2.3.

The schematic diagram in Figure 2.5 summarizes the model. As stated in Section 2.1 this is a unique application of extreme value theory in that the quantities of interest are not the lichen measurements, but instead are the ages and other second-level parameters. These quantities are quite removed from the data in this problem. As such, we expect the Bayesian credible intervals for the dates and other second-level parameters to be fairly wide.

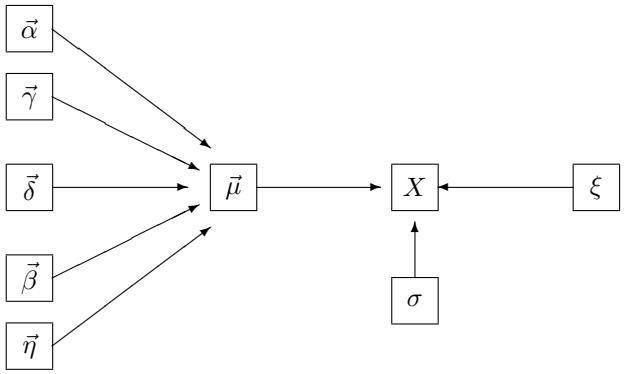

Figure 2.5: Schematic diagram of the model showing the data X is described by the GEV parameters $(\vec{\mu}, \sigma, \xi)$, and the second-level parameters $\vec{\alpha}$ (which determines lichen growth in time), $\vec{\gamma}$ (lichen growth spatially), $\vec{\beta}$ (covariance matrix in time), $\vec{\eta}$ (covariance matrix in space), and $\vec{\delta}$ (vector of known and unknown ages).

2.3.2 Model inference

The model described in the previous section provides a convenient link between the age and geographic location of the moraine and its lichen diameters through the parameters of the GEV distribution. Given the complexity of the model, standard parameter estimation procedures such as maximum-likelihood are difficult because the complete likelihood is unweildy. Fortunately, the Bayesian hierarchical framework is easily amenable to Markov Chain Monte Carlo (MCMC) methods (Robert and Casella 1999). A Markov Chain is constructed such that the equilibrium distribution is the desired posterior distribution. We run a Monte Carlo simulation of the chain, and, after convergence, obtain estimates of the posterior distribution for each parameter.

Let p represent a probability density. We wish to estimate the distribution $p(\vec{\theta}|\vec{x})$. According to Bayes' rule, the posterior is proportional to the likelihood times the prior:

$$p(\vec{\theta}|\vec{x}) \propto p(\vec{x}|\vec{\theta}) \times p_\theta(\vec{\theta}). \tag{2.13}$$

For the model described in Section 2.3.1 the right side of the relation becomes

$$g(\vec{x}|\vec{\mu},\sigma,\xi) \times \phi(\vec{\mu}|\vec{\alpha},\vec{\gamma},\vec{\beta},\vec{\eta},\vec{\delta}) \times p_\theta(\vec{\theta}) \tag{2.14}$$

where g and ϕ are the density functions of a GEV and multivariate normal random variable respectively, $\vec{\theta} = (\vec{\alpha},\vec{\gamma},\vec{\beta},\vec{\eta},\vec{\delta},\vec{\xi},\vec{\sigma})$ and p_θ is the joint prior distribution of $\vec{\theta}$. We assume that all the second-level parameters in $\vec{\theta}$ are independent, and the unnormalized density that we wish to draw from becomes

$$g(\vec{x}|\vec{\mu},\sigma,\xi) \times \phi(\vec{\mu}|\vec{\alpha},\vec{\gamma},\vec{\beta},\vec{\eta},\vec{\delta}) \times p_\alpha(\vec{\alpha}) \times p_\gamma(\vec{\gamma}) \times p_\beta(\vec{\beta}) \times p_\eta(\vec{\eta}) \times p_\delta(\vec{\delta}) \times p_\xi(\vec{\xi}) \times p_\sigma(\vec{\sigma}), \tag{2.15}$$

where $p_\cdot(.)$ is the prior density for the particular parameter (Table 2.3).

The MCMC method we use is a Metropolized Gibbs Sampler. The Gibbs sampler allows us to update the parameters one-at-a-time, and we use a Metropolis-Hastings step to update each parameter. The MCMC method is similar to the method presented by Casson and Coles (1999) and the method is explained in detail in Gelman et al.(2003). The advantage of the Metropolized Gibbs Sampler is that it is easy to implement, but it may take a long time to converge.

We wish to sample realizations of $\vec{\theta}$ from the distribution $p(\vec{\theta}|\vec{x}) \propto p(\vec{x}|\vec{\theta}) \times p(\theta)$. The Metropolis-Hastings algorithm requires that a transition density $q(\vec{\theta}^*|\vec{\theta})$ be specified, from which, given the current parameter estimate $\vec{\theta}$, a candidate estimate $\vec{\theta}^*$ can be simulated. The algorithm then either accepts or rejects the candidate with a probability which insures that the equilibrium distribution is the desired posterior. The probability with which a candidate is accepted is given by the ratio

$$\frac{p(\vec{\theta}^*|\vec{x}) \times q(\vec{\theta}|\vec{\theta}^*)}{p(\vec{\theta}|\vec{x}) \times q(\vec{\theta}^*|\vec{\theta})} = \frac{p(\vec{x}|\vec{\theta}^*) \times p_\theta(\theta^*) \times q(\vec{\theta}|\vec{\theta}^*)}{p(\vec{x}|\vec{\theta}) \times p_\theta(\theta) \times q(\vec{\theta}^*|\vec{\theta})}. \tag{2.16}$$

With the Gibbs sampler, we only need to update one parameter at a time which simplifies the acceptance ratio. For instance, assume that we are updating the parameter $\vec{\alpha}$. After cancellation, from (2.15) and (2.16) we get

$$\frac{\phi(\vec{\mu}|\vec{\alpha}^*,\vec{\gamma},\vec{\beta},\vec{\eta},\vec{\delta}) \times p^*_\alpha(\vec{\alpha}^*) \times q(\vec{\alpha}|\vec{\alpha}^*)}{\phi(\vec{\mu}|\vec{\alpha},\vec{\gamma},\vec{\beta},\vec{\eta},\vec{\delta}) \times p_\alpha(\vec{\alpha}) \times q(\vec{\alpha}^*|\vec{\alpha})}. \tag{2.17}$$

For our simulations, a uniform random walk is our transition density. Given the current parameter value (for instance $\vec{\alpha}$) we obtain our candidate ($\vec{\alpha}^*$) by $\vec{\alpha}^* = \vec{\alpha} + \vec{a} \otimes \vec{u}$, where $\vec{u}$ is a vector of independent uniform [-1,1] random variables of the same dimension as $\vec{\alpha}$, $\vec{a}$ is a vector of scaling constants, and $\otimes$ is component-wise multiplication. Then for $\vec{\alpha}$ and most of the other parameters, $q(\vec{\alpha}|\vec{\alpha}^*) = q(\vec{\alpha}^*|\vec{\alpha})$, and the acceptance ratio (2.17) further simplifies to

$$\frac{\phi(\vec{\mu}|\vec{\alpha}^*,\vec{\gamma},\vec{\beta},\vec{\eta},\vec{\delta}) \times p^*_\alpha(\vec{\alpha}^*)}{\phi(\vec{\mu}|\vec{\alpha},\vec{\gamma},\vec{\beta},\vec{\eta},\vec{\delta}) \times p_\alpha(\vec{\alpha})}. \tag{2.18}$$

However, the metropolis step becomes a bit more complicated for the dates $\vec{\delta}$, as $q(\vec{\alpha}|\vec{\alpha}^*)$ does not necessarily equal $q(\vec{\alpha}^*|\vec{\alpha})$ because moraine order is enforced.

After obtaining acceptance ratios for all of the parameter updates, the Gibbs sampler allows us to cycle through each of the parameters individually. We choose to update in the order $\vec{\alpha} \to \vec{\beta} \to \vec{\gamma} \to \vec{\eta} \to \vec{\delta} \to \vec{\mu} \to \vec{\phi} \to \vec{\xi}$.

For all of the results reported in Section 2.3.3, simulations were run for 200,000 iterations. Every 10th iteration was kept to reduce dependence, and we allowed a burn in of 5000 iterations. To test for convergence, we use the test criterion $\hat{R}$ as suggested by Gelman (1996). Several simulations were run from over-dispersed starting values for each set of simulations. These parallel simulations are used to check for convergence, and with one exception (which will be noted in Section 2.3.3) the value of $\hat{R}$ was well below the critical value of 1.1 for all parameters.

2.3.3 Results

2.3.3.1 Simulated Data

To test the model, we simulated data according to the linear growth curve model with known dates and parameters, and then ran MCMC simulations to obtain estimated distributions for the parameters. Data was created for 8 dated features and four glaciers with 6 climate events. Two of the glaciers were simulated to have only 5 moraines from the 6 events, and all the moraines had 20-35 block maxima each. Table 2.4 shows the results from one of our simulations, and all the other simulations have very similar results.

The results obtained by the model agreed with the values of the parameters and ages that had been hidden in the test. In Table 2.4, we see that for all ages and parameters, the simulated value falls within the credibility interval. As expected, the credibility intervals for the GEV parameters are quite tight, whereas the credibility intervals for the second-level parameters and ages display more uncertainty.

Ages of Undated Features						
Simulated Age	321	290	211	150	92	38
Posterior Mean	322.2	296.5	209.5	149.6	76.2	28.3
95% Cred. int.	(306.2,341.3)	(279.7,313.9)	(192.9,225.3)	(133.3,165)	(56.7,93.7)	(7.5,46.6)

Parameter	α_1	α_2	β_1	β_2	η	
Simulated Value	-0.5	0.1	1.76	0.22	0.5	
Posterior Mean	0.53	0.1	1.04	0.12	0.22	
95% Cred. Int.	(-0.97,2.06)	(0.09,0.11)	(0.46,2.25)	(0.02,0.42)	(0.02,0.61)	
Parameter	γ_1	γ_2	γ_3	γ_4	$\log \sigma$	ξ
Simulated Value	-0.75	-0.25	0.25	0.75	0	-.2
Posterior Mean	-1.11	0.19	0.48	0.44	-0.01	-0.2
95% Cred. Int.	(-1.82,-0.37)	(-0.51,0.79)	(-0.24,1.17)	(-0.29,1.17)	(-0.06,0.05)	(-0.23,-0.15)

Table 2.4: Results from simulated data. The credibility intervals contain the hidden parameter values with which the data were simulated.

2.3.3.2 Charquini Data – Individual Glaciers

We then turned our attention to the Mount Charquini data. The field researchers believed that the moraines M1-M10 on each glacier corresponded to ten climatic events as ordered. However, rather than simply assuming agreement between the corresponding moraines, we first ran the data from each glacier individually to estimate the age of each glacier's moraine. Although there was general agreement between the glaciers, the alignment of the moraines was not as obvious as was thought (Table 2.5). The estimated moraine ages of the South glacier did not correspond to the moraine ages of the other four. The youngest moraine (M10) on the South glacier was much younger than the youngest moraines on the other four glaciers, leading us to assume that it was formed by a different event. Also the oldest moraine (M1) on the South glacier had a much younger age than its counterparts on the other glaciers. However, it was strongly believed by the observers in the field that these oldest moraines were formed by the same event, and that the difference was due to unexplained variability. Perhaps most perplexing, the South glacier appeared to have three events in the 150-190 year-old range, whereas the other glaciers only had two. This caused alignment of events 4-8 to be difficult. It is worth noting that the measurements for the South glacier were taken by a different team than those of the other glaciers. Ultimately, two different moraine/event alignments were tested. The first alignment pools the moraines with the closest estimated dates together, while the second alignment pools the moraines as they were associated by the field researchers (Table 2.6).

2.3.3.3 Charquini Data – All Glaciers, Linear Growth Curve

The goal in pooling the data from corresponding moraines was to obtain more precise point estimates and tighter credibility intervals for our age and parameter estimates. We pooled the data according to the alignments specified in Section 2.3.3.2, and ran the model. As seen in Figure 2.6, we do in fact shrink the credibility intervals by pooling the data.

Ages					
Moraine	M1	M2	M3	M4	M5
South	319.8 (284.1,351.4)	286.1 (254.6,320.3)	263.6 (235.1,293)	224.7 (193.1,256.2)	204.2 (172.7,230)
Northeast	348.7 (313.1,394.2)	– –	279 (245.3,319.7)	251.8 (223.6,297)	– –
Southeast	350.5 (316.4,391.4)	309 (274.8,340.9)	269.7 (235.8,303.5)	249.6 (219.5,280.9)	234.2 (206,262.9)
West	347 (309.5,384.6)	305.2 (271.5,336.5)	271.2 (240.4,307.9)	249.9 (222.1,280.7)	234.6 (209,263)
North	347.2 (313.4,381.8)	305.2 (274.9,342)	275.4 (248.9,311.5)	254 (228.6,287.3)	235.1 (210.4,264.1)

	M6	M7	M8	M9	M10
South	188.8 (157,215.5)	175.9 (140.5,202.5)	154.2 (119.2,184.4)	128.7 (99.4,159.9)	27.5 (-16.2,57)
Northeast	215.2 (186.4,245.3)	185.8 (157.6,215.3)	157.4 (125.3,194.6)	132.3 (104.7,163.5)	91.8 (62.4,120.1)
Southeast	208.6 (182.7,230.9)	184.2 (158.9,209.4)	152.9 (124.6,179.6)	133.2 (106.6,161.8)	93.2 (64.5,123.2)
West	210.8 (182.3,241.8)	184.6 (153.1,211.3)	151.6 (121.2,180.9)	129.6 (100.7,160.4)	92.7 (61.5,122.3)
North	213.4 (189.6,242.1)	188.5 (163.4,213.9)	159.8 (135.3,185.5)	134.6 (107.7,160)	93.7 (64.8,121.3)

Table 2.5: Shows the age estimates (95% credible intervals) of the moraines when each glacier is run individually. Used to align moraines with climatic events. Notice the results for the South glacier do not correspond as well as the results from the other four.

Moraine/Event Alignment 1												
Event	1	2	3	4	5	6	7	8	9	10	11	12
South	M1	M2	M3	–	M4	M5	M6	M7	M8	M9	–	M10
Northeast	M1	–	M3	M4	–	M6	M7	–	M8	M9	M10	–
Southeast	M1	M2	M3	M4	M5	M6	M7	–	M8	M9	M10	–
West	M1	M2	M3	M4	M5	M6	M7	–	M8	M9	M10	–
North	M1	M2	M3	M4	M5	M6	M7	–	M8	M9	M10	–

Moraine/Event Alignment 2											
Event	1	2	3	4	5	6	7	8	9	10	11
South	M1	M2	M3	M4	M5	M6	M7	M8	M9	–	M10
Northeast	M1	–	M3	M4	–	M6	M7	M8	M9	M10	–
Southeast	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	–
West	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	–
North	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	–

Table 2.6: Shows the alignment of the moraines and events used in the MCMC simulations of the Charquini data

Table 2.7 gives the results for the pooled data with the first moraine alignment. Looking at the parameter estimates, we see that the estimates for the GEV parameters $\vec{\mu}$, σ, and ξ all have very tight credibility intervals. The second-level parameters have a bit wider intervals, as expected. However the credibility interval for α_2 (the slope of the linear growth curve) is quite narrow, meaning that the trend of the growth curve is well defined. Finally, the results for the glacier effect as shown by the parameters γ_i show that there is a slight glacier effect if any with this moraine/event alignment. The parameter γ_1 has the strongest signal and this parameter corresponds to the South glacier, where we have already noted a difference in results compared to the other glaciers. However, the credibility intervals for γ all include zero, which supports the hypothesis that there is

only a slight difference between glaciers on such a small spatial scale.

Ages of Events						
	Event 1	Event 2	Event 3	Event 4	Event 5	Event 6
Post. Mean	341.8	304.4	261.6	239.9	223.1	207
95% CI	(313.9,366.9)	(277.2,331)	(236.5,289.7)	(218.4,269.4)	(204,244.2)	(187.7,223.4)
	Event 7	Event 8	Event 9	Event 10	Event 11	Event 12
Post. mean	184.5	175.2	153.6	135.9	93.8	35.9
95% CI	(162.6,205.2)	(150,197.5)	(129.7,177.2)	(112.8,160.6)	(76.8,119.3)	(3,55.5)

Parameter	α_1	α_2	β_1	β_2	η
Post. Mean	0.25	0.11	2.53	0.04	0.8
95% CI	(-1.96,2.42)	(0.1,0.12)	(1.01,5.51)	(0.01,0.11)	(0.54,0.93)

γ_1	γ_2	γ_3	γ_4	γ_5	$\log \sigma$	ξ
-0.5	0.04	0.15	0.17	0.12	-0.03	-0.17
(-1.15,0.09)	(-0.62,0.69)	(-0.43,0.82)	(-0.42,0.81)	(-0.45,0.72)	(-0.07,0.01)	(-0.19,-0.15)

Table 2.7: Shows the age and parameter estimates and 95% credible intervals for the model with a linear growth curve model with the moraines/event alignment 1.

This model was also run with the second moraine/event alignment. Results were almost identical, with only two exceptions. The age estimates for the middle events differed, but the point estimates only differed by no more than three years. The biggest difference was in the γ parameters. With the second alignment, γ_1, the parameter which measured the glacier effect of the South glacier had a point estimate of -0.95 and a credibility interval of (-1.60, -0.39). With this alignment, there appears to be a significant glacier effect that the lichen measurements of the South glacier differ from those on the other glaciers.

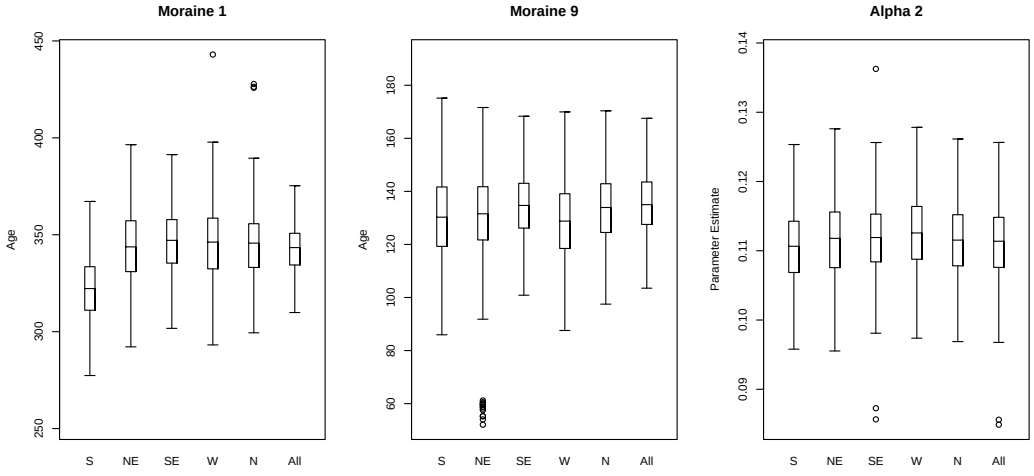

Figure 2.6: Box plots compare the results obtained by running the data from each glacier individually with the results from the pooled data. The left plot shows the age estimates for the oldest event, the center plot shows the age estimates for a relatively young event, and the right plot shows the parameter estimate for α_2, the linear coefficient in the linear growth curve model. From left to right, box plots are for the South, Northeast, Southeast, West, North, and Pooled glacier data.

2.3.3.4 Charquini Data – All Glaciers, Exponential Growth Curve

As stated in Section 2.3.1, one of our goals is to allow different growth curve models to be tested. In this section, we investigate the benefit of the exponential growth function defined by (2.8). Table 2.8 gives the results with the first moraine alignment.

Ages of Events						
	Event 1	Event 2	Event 3	Event 4	Event 5	Event 6
Post. Mean	368.9	300.3	250.5	225.1	206.7	184.6
95% CI	(322.4,438.4)	(267.7,332.8)	(224.8,280)	(199,252.2)	(181.4,227.6)	(161.4,207.3)
	Event 7	Event 8	Event 9	Event 10	Event 11	Event 12
Post. Mean	160.3	152.3	132.2	117.4	87.6	45.6
95% CI	(138.3,181.8)	(131.1,172.8)	(116,150.7)	(104,133.9)	(76.2,99.8)	(31.6,56.1)

Parameter	α_1	$log(\alpha_2)$	α_3	β_1	β_2	η
Posterior Mean	48.95	0.13	0.0045	1.82	0.08	0.7
95% Cred. Int.	(38.8,60.87)	(0.04,0.24)	(0.0029,0.0068)	(0.5,5.49)	(0.01,0.29)	(0.33,0.92)

γ_1	γ_2	γ_3	γ_4	γ_5	$\log \sigma$	ξ
-0.57	0.07	0.17	0.19	0.14	-0.03	-0.17
(-1.17,-0.07)	(-0.41,0.61)	(-0.36,0.79)	(-0.43,0.68)	(-0.36,0.68)	(-0.07,0.01)	(-0.19,-0.15)

Table 2.8: Shows the age and parameter estimates for the model with an exponential growth curve model with the moraines/event alignment 1.

The GEV parameter estimates $\hat{\vec{\mu}}$, $\hat{\sigma}$, and $\hat{\xi}$, are practically identical for the linear and exponential models. However, the age estimates obtained from the two models differ somewhat. Most of the estimated dates obtained by the exponential model are younger than the corresponding estimated dates from the linear model; with the exception of the youngest and oldest events (Event 12 and Event 1). The biggest difference is in the oldest event (Event 1), for which the point estimate given by the exponential model is nearly thirty years older than that given by the linear model. In addition to the shift in estimated dates, there is also a change in the credibility intervals. The credibility intervals of the young events are more narrow using the exponential growth curve, but the intervals of the old events are wider. The differences in point estimates and credibility intervals is a result of the changing slope of the exponential growth curve function. As Figure 2.7 shows, the μ values calculated by the different models are practically identical. But these same μ values result in different age point estimates and credibility intervals. In the exponential model, a large μ value gets associated with an older age because of the flatter slope in the exponential graph (see Figure 2.8). Furthermore, because of the nearly flat slope, a small change in the estimated value of μ results in a larger change in the estimated age. This results in both the older point estimate and the wider confidence interval for Event 1. The steeper slope of the exponential curve for young ages, results in the more narrow confidence intervals for the young events (see Figure 2.7).

When testing convergence, the $\hat{R}$ value for the age of Event 1 was not below the critical value of 1.1. On one of the parallel simulations, the age of Event 1 wandered off to very large values for a period of the simulation, as the exponential curve became very flat. This may illustrate the need for data obtained from an

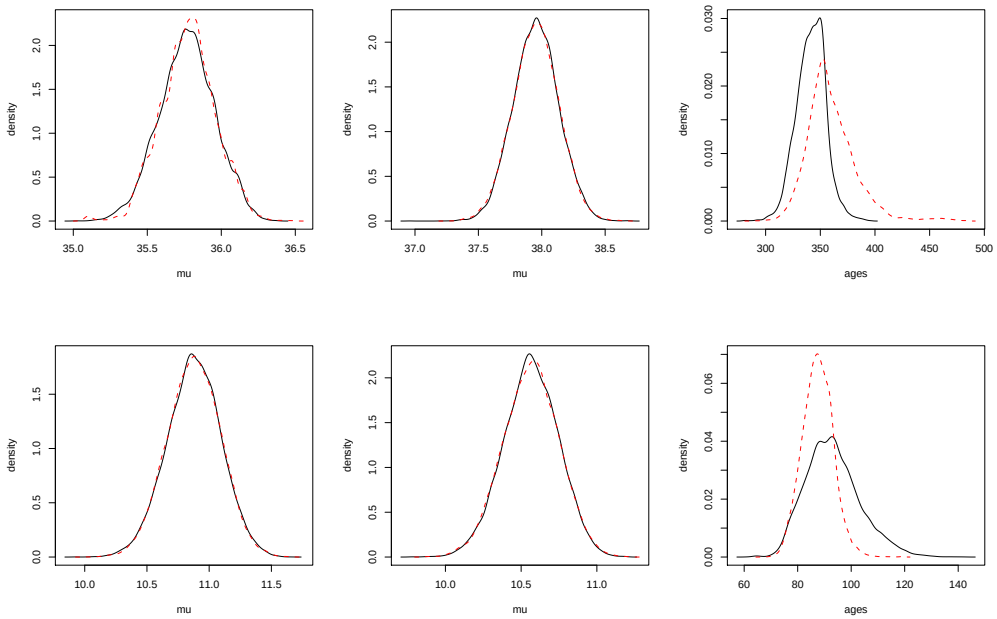

Figure 2.7: Compares the estimated posterior densities of the linear growth curve model (solid line) to the exponential growth curve model (dashed line). Top row shows parameters related to the oldest event. Left top and center top compare μ parameters for the oldest event for the South and Northeast glaciers respectively, while the top right graph compares the estimated age of the event by the two models. Bottom row shows parameters related to event 11, a relatively recent event. Bottom left and bottom center compare the model estimates for μ parameters for the Southeast and North glaciers respectively, and the bottom right compares the estimated ages of the event.

older dated feature, which would help pin down the behavior of the upper end of the growth curve.

Obviously, the growth curve model we assume will have a significant effect on our estimated ages. We wish to have a method for comparing models with different growth curves to determine which model fits best. Spiegelhalter et al (2002) proposed the deviance information criterion (DIC) to compare Bayesian models. The DIC is especially useful for comparing models whose inference is obtained via MCMC methods. The deviance score is a function of the likelihood $D(\vec{x},\theta) = -2\log p(\vec{x}|\vec{\theta})$. To estimate the DIC for an MCMC simulation, the average deviance is calculated for MCMC iteration (indexed by l, after convergence), which yields a value for the fit of the model $\bar{D} = 1/L\sum_{l=1}^{L} D(\vec{x},\theta_l)$. The deviance $D(\vec{x},\hat{\theta})$ is also calculated, where $\hat{\theta}$ is a point estimate for θ (posterior means were used for these calculations). The effective number of model parameters p_D can be estimated by $\bar{D} - D(\vec{x},\hat{\theta})$. The overall DIC score is simply the sum $\bar{D} + p_D$. The model with the lowest criterion value is judged to be best.

When comparing the linear and exponential growth curve models by sight, it would appear that the exponential model is superior, especially in capturing the behavior of the data for which dates are known (Figure 2.8). However, the DIC seems to show that the linear growth curve model is slightly better (Table 2.9).

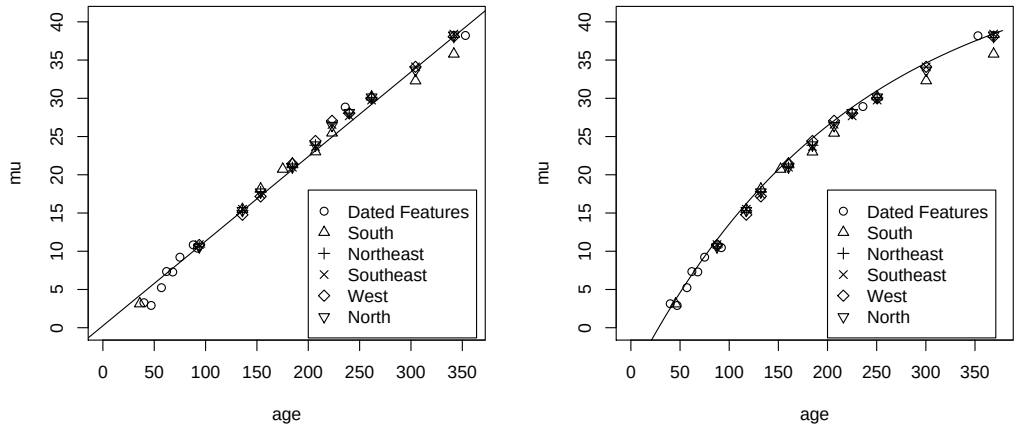

Figure 2.8: Compares the two growth curve models tested. Plotted are the posterior mean age verses the posterior mean μ parameter. Both growth curve models appear to fit the data well, but the exponential growth curve appears to fit the dated features a bit better. The exponential growth curve better models the known behavior of lichen growth. Also notice the exponential growth curve estimates the age of the oldest feature to be much older than the linear growth curve. These growth curve plots differ from past lichenometry studies in that we are plotting the location parameter μ verses age rather than the lichen diameters verses age.

The DIC measures how well the model fits the data, but only uses the first-level of the hierarchy; that is, it includes only the parameters $\vec{\mu}$, σ, and ξ. The two models have virtually the same first-level parameter estimates; this explains the very similar values for the fit portion ($\bar{D}$) of the DIC. The p_D value for the exponential model is higher, as we have added another parameter α_3. This brings up an important issue in model selection. The difference between the fits of the two models clearly lies in the ages, a second-level parameter which is not taken into account by the usual DIC. Because of the the inverse nature of our application, the usual DIC may not be an adequate tool for choosing the "best" model. However, the $\bar{D}$ portion of the DIC can be extended logically to measure the deviance of the first two levels of the hierarchy. When this is done, the linear model produces an estimate for $\bar{D}$ of 3757.9 and the exponential model produces an estimate of 3749.5, indicating that the exponential model fits the data better. Furthermore, from a geomorphological point of view, the exponential curve makes more sense. It is widely accepted that younger lichens grow at a faster rate than older lichens yielding an increasing and concave down growth curve.

	D	p_D	DIC
Linear Growth Curve Model	3598.57	55.45	3654.02
Exponential Growth Curve Model	3599.07	55.74	3654.81

Table 2.9: The DIC table for the linear and exponential growth curve models. $\bar{D}$ measures the fit, p_D measures the effective number of parameters, and their sum yields the DIC criterion measure.

We also examined Bayesian p-values for both the means and variances of the Bolivian data to replicate data simulated from the model (Gelman et al., 2003). P-values are generally used for model validation rather than model selection, and the p-values for the linear growth curve model and the exponential growth curve were almost identical. However, the p-values for the means of the data from each moraine were all between .05 and .95 except for M2 from the south glacier and M1 for the Northeast and West glaciers. It is not surprising that the means of the actual data would differ slightly from the model for these particular moraines as there was disagreement among the glaciers as to the estimated age of these oldest events. The p-values for variance showed that our model tended to underestimate the variance of the data for the old events and overestimate the variance for the young events. This would support the idea of expanding the model to allow the GEV scale parameter to vary with the age of the feature as we have done for the location parameter.

2.3.4 Discussion

This model presents an improved method for analyzing the data from lichenometry by providing a comprehensive statistical model. The model is built upon the specificity of the data, the largest lichen observations, and thus the model's foundation is based in extreme value theory. The Bayesian hierarchical structure ties the observed lichen measurements to the hidden covariate ages of interest. The MCMC method of inference naturally provides credibility intervals to assess the uncertainty associated with age and other parameter estimates. And the flexibility of the basic model allows for different growth curves and spatial relationships to be fitted.

In building our hierarchical structure, we have concentrated our efforts on the location parameter μ. It would be possible to build a hierarchical structure for the GEV shape parameter ξ and scale parameter σ. However, we believe that the shape parameter ξ should not vary with age nor location. Whether the distribution of the largest lichen has a bounded, light, or heavy tail should be determined by the biology of the lichen, and should not be affected by moraine location or age. Furthermore, from a practical standpoint, ξ is the hardest GEV parameter to estimate and we wish to pool as much data as possible to estimate it. On the other hand, it is conceivable that the scale parameter would vary with age, as is suggested by the Bayesian p-value analysis of variance. Likewise it might vary with geographic location. A hierarchical structure similar to the one for μ could be put into the model for σ, but we leave this for future work.

One area that warrants more investigation is model selection. As discussed in Section 2.3.3.4, the usual model selection tool for Bayesian models, the DIC, was not effective in selecting the growth curve, since it involves only the immediate parameters (μ, σ, and ξ in our case) when assessing fit. To select between models with different growth curves, the criterion must take into account the covariate ages which is where the model results differ most. Along the same lines, work could be done in developing a criterion for selecting moraine/event alignment. Our data showed that moraine/event alignment may not be obvious based on age estimates obtained

from individual glaciers. At the current time, we must rely on the expert opinion of the field researchers when choosing both the best growth curve and moraine/event alignment.

Our analysis would benefit from more lichen measurements from dated features. As seen in Figure 2.8, we have only two dated features over 100 years of age. This makes determining the appropriate growth curve more difficult. Of course, finding features with old and known dates in the region of interest is difficult or impossible. In the end, lichenometry is used specifically because there is little usable material in alpine environments.

Chapter 3

Spatial Extreme Value Analysis for Weather and Climate

3.1 Weather, climate and geostatistics

Unlike the lichenometry project in Chapter 2 whose spatial component was relatively minor, the central focus of the upcoming projects revolves around modeling extremes on spatial fields. In order to explain the upcoming investigations, some background material must be provided. The goal of this chapter is to extend the material from Chapter 1 to provide an adequate framework for the upcoming studies. This chapter will provide some background in spatial statistics and multivariate EVT. It will also explain the different nature of the two upcoming projects and why spatial dependence is modeled differently in each.

In a weather or climate study, one often wants to analyze data from more than one location and/or model over an areal study region. In this and the upcoming chapters, we will model extremes at point locations, although other location types (e.g. grid cells) are sometimes studied. The branch of spatial statistics that works with data gathered from point sources is geostatistics. In geostatistics, the data may come from point locations on a regular grid, but generally it is assumed that the data stations are irregularly located in the spatial domain. Not surprisingly, geostatistics has its roots in studying the central tendencies of the distribution rather than the distribution's tails.

An example of weather data from a point location might be the time series of daily precipitation for Denver, or more specifically, at the weather station located at the city's airport. A climate example might be the mean precipitation for Ft. Collins for June or some measure of the climatological extreme precipitation such as a return level. To illustrate common geostatistics methods, two toy examples of data that are not extremes (Figure 3.1) are given. On the left is a weather example, the recorded precipitation on a particular day for five cities in northeast Colorado. On the right is a climate example, the mean annual precipitation for the same locations.

One of the first questions that would have to be answered for each example would be to determine the data's spatial dependence. Geostatistics characterizes spatial dependence via the variogram, one of the cornerstones of geostatistics (Cressie, 1993; Chiles and Delphiner, 1999). Typically in geostatistics, one assumes

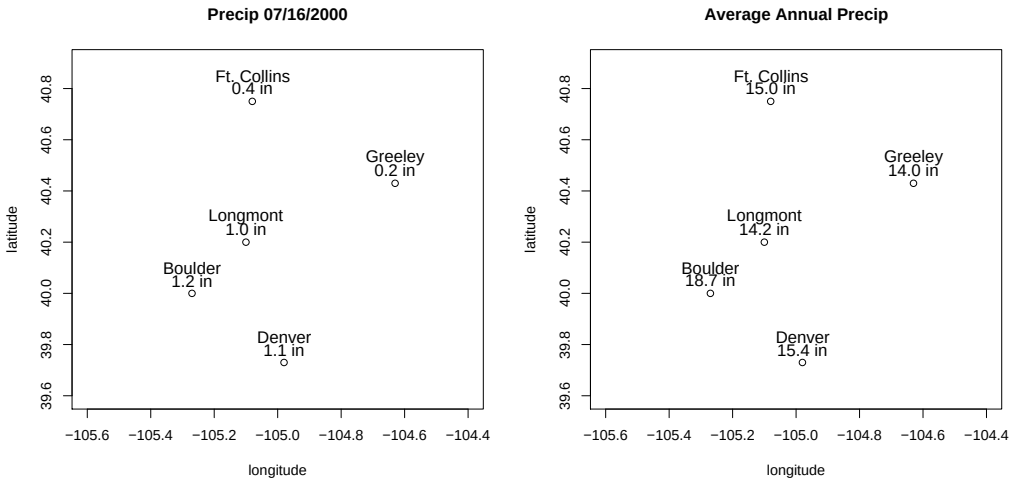

Figure 3.1: Toy examples of non-extreme geostatistics data. Left is a weather example, precipitation for a given day, and right is a climate example, average annual precipitation.

a stationary, ergodic spatial process $Z(\boldsymbol{x})$ that has a well-defined covariance function, where $\boldsymbol{x}$ denotes a location in some domain, say $\mathbb{R}^2$. The variogram is defined as $\gamma(\boldsymbol{h}) = \frac{1}{2}\mathbb{E}\left[(Z(\boldsymbol{x}) - Z(\boldsymbol{x}+\boldsymbol{h}))^2\right]$. In this work, it will be further assumed that the spatial process is isotropic, meaning that $\gamma(\boldsymbol{h})$ is a function only of $||\boldsymbol{h}||$, not its direction, and $||\boldsymbol{h}||$ will be denoted as h. A straightforward estimator of the variogram is $\hat{\gamma}(h) = \frac{1}{2|\mathcal{N}_h|}\sum_{(x_i,x_j)\in\mathcal{N}_h}\left(Z(x_i) - Z(x_j)\right)^2$, where $\mathcal{N}_h$ is the set of all sample pairs for which $||x_i - x_j|| \in [h-\delta, h+\delta)$. Often, the spatial process is further assumed to be Gaussian. Since covariance can be expressed as a function of γ, the variogram is especially useful. Under the Gaussian assumption, the dependence structure is completely described by pairwise covariances, thus the variogram and mean completely determine the distribution of the process.

Let us assume that the data in both examples come from a stationary and isotropic field. Then, the variogram could be used to estimate the spatial dependence for both the weather and climate data (of course, many more locations would be needed to apply the estimator). The variograms would probably differ between the two cases, as the weather data set would presumably exhibit shorter range dependence than the climate data set. This is because daily precipitation varies over space much more than the climatological precipitation.However, there is a more subtle and more important difference between two the examples. In the weather example, the variogram is being used to measure the spatial dependence between the *observations*. In the climate example, the "data" are actually a summary statistic, and thus the variogram is measuring the spatial dependence of this statistic which characterizes the *distributions*. This difference between describing the spatial dependence in the observations as opposed to the distributions is an essential difference between the projects in Chapters 4 and 5.

The projects presented in Chapters 4 and 5 both focus on spatial extremes and central to each is modeling

spatial dependence. The goal of the project in Chapter 4 is to develop a dependence measure similar to the variogram, but which measures the dependence among extreme observations (specifically, observations from a multivariate max-stable distribution or max-stable process). Section 3.2 provides background for this project including a brief introduction to multivariate extreme value theory. The relevance of such an extreme dependence measure is illustrated using a toy weather example (although the measure could certainly be applied outside of weather and climate studies). The goal of the project in Chapter 5 is to develop a methodology which allows one to describe the climatological extremes over a spatial field. To do so, a model is constructed which models the spatial dependence in the distributions which characterize the extremes, rather than in the extreme observations themselves. Section 3.3 illustrates the difference with another toy example.

3.2 Modeling spatial dependence in extreme observations

3.2.1 Multivariate extreme-value theory

To understand the challenges in describing spatial dependence between extreme observations, one must first have some background in multivariate EVT. Univariate EVT is based on the idea of the distribution of the sample maxima. To expand the theory to the multivariate setting, it must first be determined what is meant by a maximum of a sequence of random vectors $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})$ for $i = 1, 2, \ldots$. Multivariate EVT is based on the vector of componentwise maxima $\boldsymbol{M}_n = (M_1, \ldots . M_d)$ where $M_j = \max_{i=1,\ldots,n} X_{i,j}$ for $j = 1, \ldots d$. Interestingly, $\boldsymbol{M}_n$ does not usually correspond to any observation $\boldsymbol{X}_i$, as the maxima of each component can occur at different observation times. An excellent review of multivariate EVT is given by Fougéres (2004), which includes many of the historical developments, or an alternate source is Beirlant et al. (2004).

In Chapter 1 the notion of (univariate) max-stability was characterized as the property that the distribution of the sample maximum is of the same type as the distribution of the individual random variables. A similar property can be defined for multivariate distributions. Let $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d}),\ i = 1, 2, \ldots$ be iid random vectors with distribution function G and $\boldsymbol{M}_n$ be the vector of componentwise maxima as above. Then, analogous to 1.1 and 1.2, if there exist $\mathbb{R}^d$ sequences $\boldsymbol{a}_n > 0$ and $\boldsymbol{b}_n$ such that

$$\mathbb{P}\left\{\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{x}\right\} = \mathbb{P}\left\{\boldsymbol{X} \leq \boldsymbol{x}\right\}, \tag{3.1}$$

or equivalently

$$G^n(\boldsymbol{a}_n \boldsymbol{x} + \boldsymbol{b}_n) = G(\boldsymbol{x}), \tag{3.2}$$

then G is a d-dimensional max-stable distribution.

The significance of the multivariate max-stable distributions arises from an asymptotic argument for component-wise block maxima. Define $\boldsymbol{M}_n$ as above where $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d}),\ i = 1, 2, \ldots$ is now a random

vector from *any* continuous distribution function F. If there exist normalizing sequences $\boldsymbol{a}_n$ and $\boldsymbol{b}_n$ such that

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

then G is a multivariate max-stable distribution (multivariate EVD) (Resnick, 1987).

In the univariate case, one is able to characterize all of the max-stable distributions with the GEV family. The multivariate max-stable distributions can also be characterized, but not as easily. From the results of univariate EVT, the marginal distribution of the jth component of the multivariate EVD must be $GEV(\mu_j, \sigma_j, \xi_j)$. Letting

$$u_j(x) = \left[1 + \xi_j\left(\frac{x - \mu_j}{\sigma_j}\right)\right]^{1/\xi_j},$$

and denoting the set E as $[0, \infty)^d \setminus \mathbf{0}$, then if G is a multivariate EVD, there exists a finite measure H on $\mathcal{B} = \{\boldsymbol{y} \in E : ||\boldsymbol{y}|| = 1\}$ such that for each $\boldsymbol{x} = (x_1, \ldots, x_d) \in E$,

$$G(\boldsymbol{x}) = \exp\left\{-\int_{\mathcal{B}} \max_{j=1,\ldots,d} \frac{w_j}{u_j(x_j)} dH(\boldsymbol{w})\right\}, \tag{3.3}$$

where

$$\int_{\mathcal{B}} w_j dH(\boldsymbol{w}) = 1, \text{ for all } j = 1, \ldots, d \tag{3.4}$$

(Fougères, 2004).

In the univariate case, the asymptotic distribution (the GEV) could be written parametrically, but that is not the case for the multivariate EVD. Any finite measure H which meets the mean constraint (3.4) can give rise to a valid multivariate EVD. As there is no parametric representation, there is no obvious way to represent dependence between the vector components.

3.2.2 Characterizing dependence in extreme weather observations

The goal of the project presented in Chapter 4 is to develop a measure of dependence for extremes analogous to the variogram. Specifically, pairwise dependence measures for multivariate max-stable distributions are investigated. These measures are extended to the spatial case by applying them to two-dimensional max-stable processes. A max-stable process in $\mathbb{R}^2$ is a random field whose higher-order marginal distributions are multivariate EVDs as defined in (3.3), and it will be more rigorously defined in Chapter 4. In addition to reviewing two previously suggested extremes dependence measures, a new measure, the madogram, is presented. The madogram conveniently links geostatistical ideas to existing measures of dependence for extremes.

The madogram is a general measure and it will be shown that it can be used to quantify pairwise dependence in observations from bivariate or multivariate EVDs. However, because of its ties to geostatistics, it is particularly well-suited to measuring dependence in max-stable spatial fields. The toy example in Figure 3.2 is used to illustrate a possible application of the madogram, and used to contrast with the distributional

dependence to be described in Section 3.3. The data are the maximum daily precipitation values occurring between April and October of the year 2000 for the same northeast Colorado cities as before. Notice that the dates of the maximum precipitation are not the same for all the observations. As the observations are componentwise annual maxima, the distribution underlying these observations should be approximately max-stable.

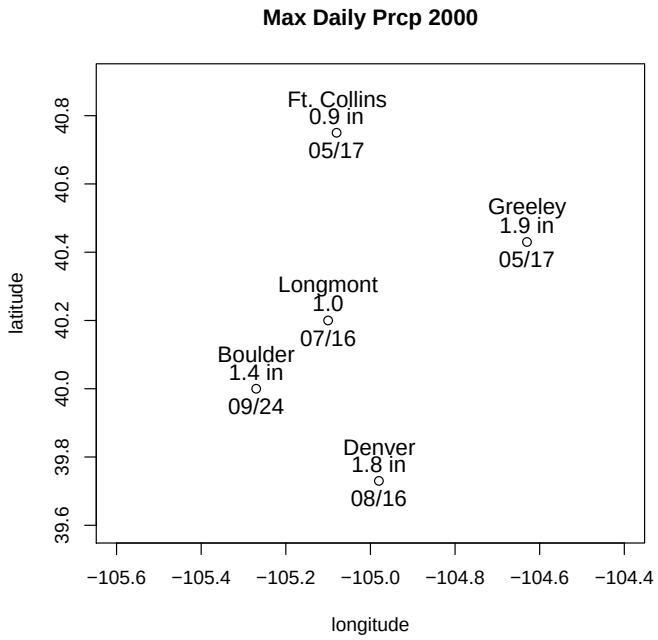

Figure 3.2: The maximum daily precipitation measurements for the year 2000. Such data could be assumped to approximately follow a max-stable distribution. Notice that the day the maxima occurred (month/day) below each location do not coincide. The goal of Chapter 4 is to develop dependence measures for similar situations.

Even though they occur on different days, the annual maximum observations are still likely to have spatial dependence, as an unusually high extreme in one location might yield information that the annual max at a nearby location should be at or above a certain level. Assuming the field were isotropic (and further assuming there were an adequate number of observations to apply an estimator), the madogram could be used to model the spatial dependence as a function of distance, much like the variogram described earlier.

There are other scenarios where one might wish to describe spatial dependence of extreme observations that fall outside the realm of max stability and componentwise block maxima. For instance, one might wish to characterize the spatial dependence of precipitation observations for a particular day in which at least one location receives extreme rainfall. Others have suggested dependence measures for extreme data which are only in the domain of attraction of a bivariate or multivariate extreme value distribution (Heffernan and Tawn (2004); Coles et al. (1999); Capéraà and Fougéres (2000)), but none of these have been extended to the spatial

case. Furthermore, any extreme dependence measure must first work in the max stable case, and this case is where the theory for extremes is best developed. It is hoped that measures of dependence in the max stable context can be extended to data which is only in the domain of attraction of max stable distributions.

3.3 Modeling spatial dependence in extreme distributions

In contrast to describing the dependence in observations as in the last section, consider Figure 3.3 which shows estimated return levels for the cities. One can imagine wanting to spatially model the return levels and one can posit a surface of return values. In fact, the National Weather Service maintains maps of estimated precipitation return level surfaces which are used by engineers and hydrologists for use in flood planning [1] .

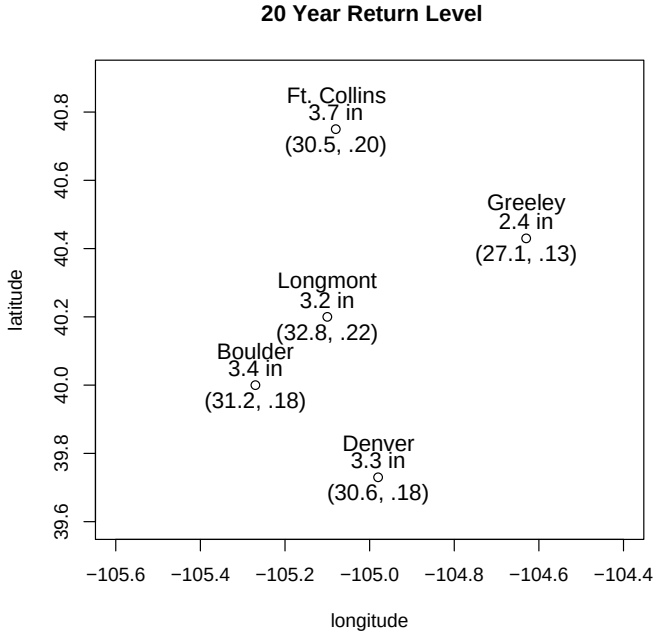

Figure 3.3: Point estimates of return levels calculated from the maximum likelihood estimated GPD parameters (σ, ξ) at the locations. The goal of Chapter 5 is to produce a methodolgy for spatially modeling return levels.

Return levels are not observations, they are high quantiles (Section 1.3). Return levels can be obtained via functions of the parameters that characterize the extreme observations. Hence, the spatial dependence among return levels must be modeled differently from that among observations (see Equation 5.1). To model climatological quantities such as return levels, one must model how the distributions, rather than the observations, vary over the study region.

The return level point estimates in Figure 3.3 were obtained from maximum-likelihood estimates of the GPD parameters $(\hat{\sigma}, \hat{\xi})$ at each station (shown in parentheses for each city). Notice that the shape parameter

[1] See `hdsc.nws.noaa.gov/hdsc/pfds/pfds_maps.html`

estimate $\hat{\xi}$ varies somewhat between the different locations. This is not surprising as the shape parameter is often difficult to estimate. But one is left to wonder how much of this variation is simply due to sampling error, especially since one is always data-poor when studying extremes.

In Chapter 5, a climatological extremes model which characterizes the spatial behavior of the GPD parameters is presented. It employs a Bayesian methodology which treats the parameters themselves as random variables, to which traditional geostatistics methods are applied. The model also has the advantage of pooling the observational data in order to better estimate parameters for the region. By modeling the GPD parameters spatially, return levels are estimated for the region. The idea of modeling parameters spatially is not unlike the non-extremes climate example in Figure 3.1. In that example, the mean precipitation was modeled, which could be viewed as a parameter which describes the central part of the distribution.

3.4 Summary

The difference between the spatial methods described may appear to be subtle, but is important. One might classify Figure 3.2 as a weather problem, and Figure 3.3 as a climate problem. Describing the first example as a weather problem and the second as a climate problem may be useful for illustration purposes, but it is an over-simplification. It would be wrong to assume that all climate problems should be modeled in the same way (the same is true for weather). In fact, one of the data sets to which the dependence measure for extreme observations is applied is not annual maximum data as illustrated in Figure 3.2, but is rather the 30-year maximum precipitation for locations in France. As such, it might be considered a climate data set rather than a weather one, but since the goal is to measure the spatial dependence in the observations we use the "weather" method as described above. Ultimately, the difference is not between climate and weather, but whether one needs to model spatial dependence in the observations or distributions.

Chapter 4

Madogram

4.1 Background

The goal of this chapter is to develop geostatistical tools, analogous to the variogram, for measuring spatial dependence in extreme observations. It is natural to begin in the max-stable case. Since the max-stable distributions can be summarized parametrically only in the univariate case, there is no obvious way to measure dependence. The fact that there is no dependence parameter suggests that non-parametric dependence measures should be developed.

Several non-parametric dependence measures for max-stable distributions have been suggested. In this chapter, two previously introduced dependence measures are reviewed, and another is proposed which has ties to geostatistics methods. The ultimate goal is to define and study measures (and their estimators) which effectively describe dependence between spatial observations of extremes. However, any good spatial measure must first be effective in measuring bivariate dependence. Since the bivariate case is more easily analyzed, each dependence measure is reviewed and compared for bivariate extremes before being extended to the multivariate and spatial cases.

4.1.1 Bivariate extreme value distributions

Before the various measures of dependence are discussed, the bivariate max-stable distribution and the max-stable process are defined. These are employed when developing and comparing the measures. In Section 3.2.1, the multivariate EVD was described; however, in the bivariate case, the distribution (3.3) can be simplified considerably. Let (X, Y) be a bivariate max-stable random variable with common GEV(μ, σ, ξ) marginal distributions. The joint distribution of X and Y can be written as

$$G(x, y) = \exp\left\{-V(u_\beta(x), u_\beta(y)\right\}, \text{ with } V(u, v) = \left[\int_0^1 \max\left(\frac{w}{u}, \frac{(1-w)}{v}\right) dH(w)\right], \tag{4.1}$$

where $u, v > 0$ and H is a finite measure on $[0, 1]$ such that

$$\int_0^1 w dH(w) = 1, \tag{4.2}$$

and the function $u_\beta(\cdot)$ is defined by

$$u_\beta(x) = \left(1 + \xi \tfrac{x-\mu}{\sigma}\right)_+^{1/\xi} \tag{4.3}$$

for the parameter vector $\beta = (\mu, \sigma, \xi)$ (Coles, 2001b). We refer to such distributions as bivariate EVD.

Usually in this chapter, it will usually be assumed that both X and Y have unit Fréchet marginal distributions, i.e. that the marginal distribution function is $F(x) = \exp(-1/x)$ where $x > 0$. There is no loss of generality in assuming specific margins (Resnick, 1987). The joint distribution function assuming Fréchet marginals is given by:

$$G(x, y) = \exp\{-V(x, y)\} \tag{4.4}$$

Henceforth, assume Fréchet marginals unless otherwise noted.

First, the joint bivariate distribution for the cases of independence and complete dependence are presented. If H is a measure that places point masses of 1 on $w = 0$ and $w = 1$, then (4.2) is trivially satisfied and $G(x, y) = \exp[-(x^{-1} + y^{-1})]$, which is factorizable and therefore corresponds to the independent case. If H places a point mass of 2 on $w = .5$, then $G(x, y) = \exp[-(\max(x^{-1}, y^{-1}))]$ which implies that $X = Y$ with probability 1.

Although the bivariate EVD is not parametric, several parametric subfamilies have been suggested (Tawn, 1988; Coles and Tawn, 1991; Joe, 1990; Tawn, 1990). A classical parametric bivariate extreme value family is the logistic model, proposed by Gumbel (1960), and defined by

$$G_\alpha(x, y) = \mathbb{P}(X \le x, Y \le y) = \exp\left\{-\left[\left(\frac{1}{x}\right)^{\frac{1}{\alpha}} + \left(\frac{1}{y}\right)^{\frac{1}{\alpha}}\right]^\alpha\right\}, \tag{4.5}$$

for $\alpha \in (0, 1]$. The case $\alpha = 1$ implies independence between the variables, and complete dependence ($X = Y$ in probability) is achieved in the limit as $\alpha \to 0$. There is a variant of the logistic model which allows for an asymmetrical dependence structure. The so-called asymmetric logistic (Tawn, 1988) is given by

$$G_{\alpha,t_1,t_2}(x, y) = \exp\left\{-\left[\frac{1-t_1}{x} + \frac{1-t_2}{y} + \left[\left(\frac{t_1}{x}\right)^{\frac{1}{\alpha}} + \left(\frac{t_2}{y}\right)^{\frac{1}{\alpha}}\right]^\alpha\right]\right\}, \tag{4.6}$$

where $\alpha \in (0, 1]$, and $t_1, t_2 \in [0, 1]$. The model is asymmetric because $G(x, y) = G(y, x)$ only if $t_1 = t_2$. When $t_1 = t_2 = 1$ the model is equivalent to the logistic model. Independence is attained when $\alpha = 1$, $t_1 = 0$ or $t_2 = 0$. Complete dependence is achieved in the limit when $t_1 = t_2 = 1$ and $\alpha \to 0$. In this study, the above parametric models are employed to simulate data with known dependence structures.

4.1.2 Max-stable processes

When studying spatial dependence measures, it will be assumed that data come from a spatial max-stable process. Let $\boldsymbol{x}$ be in some index set $\mathcal{X}$. Similar to (1.1) and (3.1), a stochastic process $Z(\boldsymbol{x})$ is called max-stable

if there exist constants $A_n(\boldsymbol{x}) > 0$ and $B_n(\boldsymbol{x})$ such that the following property holds. If $Z^{(1)}(\boldsymbol{x}), \ldots, Z^{(n)}(\boldsymbol{x})$ are n independent copies of the process, and $M_n(\boldsymbol{x}) = \max_{i=1,\ldots,n} Z^{(i)}(\boldsymbol{x})$, then

$$\mathbb{P}\left(\frac{M_n(\boldsymbol{x}) - B_n(\boldsymbol{x})}{A_n(\boldsymbol{x})} \leq z(\boldsymbol{x})\right) = \mathbb{P}\left(Z(\boldsymbol{x}) \leq z(\boldsymbol{x})\right). \tag{4.7}$$

The theory for max-stable processes was developed by de Haan (1984, see also Resnick (1987); Smith (2004)) who gave a spectral representation for the processes. Smith (1990) extended de Haan's characterization from a one-dimensional index set (e.g. time) to a spatial process (e.g $\mathcal{X} \subseteq \mathbb{R}^2$ or $\mathbb{R}^3$).

In this work, it will be assumed that $Z(\boldsymbol{x})$ is a stationary max-stable process for locations $\boldsymbol{x} \in \mathbb{R}^2$. It will further be assumed that the process is isotropic and the marginal distribution at each location $\boldsymbol{x}$ is unit Fréchet, which allows one to write the process as (Schlather, 2002; Schlather and Tawn, 2003; Davis and Resnick, 1993)

$$\mathbb{P}\left(Z(\boldsymbol{x}) \leq z(\boldsymbol{x}), \text{ for all } \boldsymbol{x} \in \mathbb{R}^2\right) = \exp\left[-\int_S \max_{\boldsymbol{x} \in \mathbb{R}^2}\left\{\frac{g(s,\boldsymbol{x})}{z(\boldsymbol{x})}\right\}\delta(ds)\right], \tag{4.8}$$

where the spectral function $g(.)$ is a nonnegative function which is measurable in the first argument, upper semi-continuous in the second and has to satisfy $\int g(s,\boldsymbol{x})\delta(ds) = 1$. Notice the common structure with equations (3.3) and (4.1). The focus of this chapter is quantifying the pairwise spatial dependence in max-stable random fields. The pairwise joint distribution function of $(Z(\boldsymbol{x}), Z(\boldsymbol{x}+\boldsymbol{h}))$ can be simplified from (4.8) to

$$\mathbb{P}\left(Z(\boldsymbol{x}) \leq z_1, Z(\boldsymbol{x}+\boldsymbol{h}) \leq z_2\right) = \exp\left[-\int_S \max\left(\frac{g(s,\boldsymbol{x})}{z_1}, \frac{g(s,\boldsymbol{x}+\boldsymbol{h})}{z_2}\right)\delta(ds)\right]. \tag{4.9}$$

As in the bivariate case, spatial fields of data with known dependence must be simulated to compare the spatial dependence measures. Two methods for simulating max-stable random fields have been proposed, and both arise from point processes which, in a sense, give the number of events over which the maximum is taken. The first class was developed by Smith (1990), where points are simulated according to a Poisson point process Π where Π is a Poisson point process on $\mathbb{R}^2 \times (0,\infty)$ with intensity $s^{-2}dyds$ and f is a non-negative function such that $\int f(x)dx = 1$. The max-stable process is given by

$$Z(\boldsymbol{x}) = \sup_{(\boldsymbol{y},s)\in\Pi}\left[s\, f(\boldsymbol{x}-\boldsymbol{y})\right], \text{ with } \boldsymbol{x} \in \mathbb{R}^2, \tag{4.10}$$

In our simulations, we let f be a Gaussian density function with covariance matrix $\Sigma = I$. The second type proposed by Schlather (2002) arise from Gaussian random fields which are then scaled by the realization of a point process on $(0,\infty)$:

$$Z(\boldsymbol{x}) = \max_{s\in\Pi} sY_s(x), \tag{4.11}$$

where Y_s are i.i.d. stationary standard Gaussian processes for all $s \in (0,\infty)$ and Π is a Poisson point process on $(0,\infty)$ with intensity $\sqrt{2\pi}r^{-2}dr$. The Schlather field requires that a variogram model be given for the underlying Gaussian process, and both exponential ($\gamma(h) = c_e \exp(-h/a_e)$) and Gaussian ($\gamma(h) = c_g \exp(-h/a_g)^2$) variogram models are used to simulate fields in this work. Figure 4.1 shows realizations of both the Smith and the Schlather processes employed in this study.

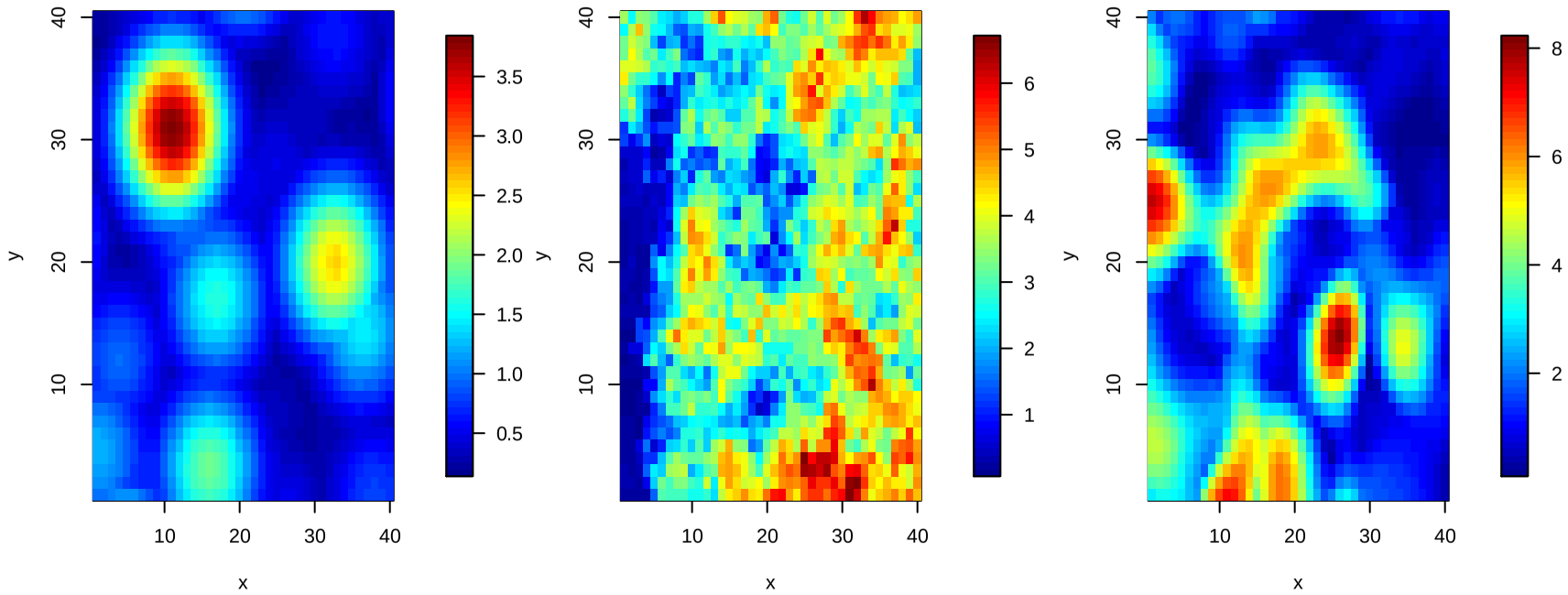

Figure 4.1: Examples of simulated max-stable random fields, left is a realization of the field proposed by Smith (1990), center is a realization of the Schlather (2002) field with an exponential variogram on the underlying Gaussian process, and the right is a realization of the Schlather field with a Gaussian variogram (2002).

4.2 Existing pairwise dependence measures

4.2.1 Extremal coefficient

Perhaps the simplest of the dependence measures for max-stable distributions, the extremal coefficient, θ, is due to de Haan (1985) and was more recently studied by Schlather and Tawn (2002; 2003). Given (X, Y) comes from a bivariate EVD defined by (4.4) the extremal coefficient is defined by

$$\theta = V(1,1) = \int_0^1 \max(w, 1-w)\, dH(w). \tag{4.12}$$

From (4.12) a relation is derived which yields an intuitive understanding of this dependence measure.

$$\begin{aligned} \mathbb{P}(\max(X,Y) < x) &= \mathbb{P}(X \le x, Y \le x) \\ &= \exp\left\{-\frac{1}{x}\int_0^1 \max(w,(1-w))\, dH(w)\right\} \\ &= \exp\left\{-\frac{\theta}{x}\right\} \\ &= [\mathbb{P}(X \le x)]^{\theta}. \end{aligned} \tag{4.13}$$

Hence, the extremal coefficient can be interpreted as the effective number of independent variables from which the maximum is drawn and $\theta \in [1,2]$. Complete dependence implies $\theta = 1$, and independence implies $\theta = 2$. For a bivariate logistic EV random variable with dependence parameter α, $\theta = 2^{\alpha}$, and for an asymmetric logistic, $\theta = (1-t_1) + (1-t_2) + [t_1^{1/\alpha} + t_2^{1/\alpha}]^{\alpha}$.

The definition of the extremal coefficient is easily extended to provide a measure of pairwise dependence for a spatial process. Given a process as in (4.8), Schlather and Tawn (2003) define the pairwise extremal coefficient function as

$$\theta(h) = \int \max\{g(s,\boldsymbol{x}), g(s,\boldsymbol{x}+\boldsymbol{h})\}\delta(ds), \tag{4.14}$$

Like the bivariate extremal coefficient, it is easily shown that $1 \le \theta(h) \le 2$, and $\mathbb{P}(\max(Z(\boldsymbol{x}), \quad Z(\boldsymbol{x}+\boldsymbol{h})) \le z) = [\mathbb{P}(Z(\boldsymbol{x}) \le z)]^{\theta(h)}$.

Unlike a Gaussian process, the dependence structure of a max-stable process is not completely characterized by the pairwise dependence structure. Schlather and Tawn (2003) also extend the definition of the extremal coefficient to the multivariate setting of any dimension. Given a random vector $(X_1, \ldots, X_n)$ from a multivariate EVD, then, analogous to (4.12), the multivariate extremal index is given by

$$\theta_A = \int_{S_n} \max_{i \in A} w_i dH_n(w_1, \ldots w_n),$$

where A is an index set of the elements of the random vector whose dependence is being quantified. It can be shown that $1 \le \theta_A \le |A|$, and that $\mathbb{P}(\max_{i\in A} X_i \le x) = [\mathbb{P}(X_1 \le x)]^{\theta_A}$ so the multivariate extremal coefficient shares the same interpretation as the bivariate measure.

Schlather and Tawn (2003) go on to give bounds which must be met for extremal coefficients of different orders to be mutually consistent. For example, given a multivariate distribution for (X_1, X_2, X_3), then

$$\max(\theta_{12}, \theta_{13}, \theta_{23}, \theta_{12} + \theta_{13}, \theta_{23} - 3) \leq \theta_{123} \leq \theta_{12} + \theta_{13} + \theta_{23} - 1 - \max(\theta_{12}, \theta_{23}, \theta_{23}).$$

However, characterizing dependence of higher orders quickly becomes unwieldy, and Schlather and Tawn (2003) concede that in practice the pairwise dependence measures are the most useful. They go on to prove necessary and sufficient conditions for a matrix of pairwise extremal coefficients to be self-consistent. These conditions can be translated into the spatial case as conditions on the extremal coefficient function. Among the results given by Schlather and Tawn (2003) are, given that $\theta(h)$ is an extremal coefficient function, then

$$(1) \qquad 2 - \theta(h) \text{ is a positive semidefinite function, and} \tag{4.15}$$

$$(2) \qquad 1 + |\theta(h) - \theta(k)| \leq \theta(h + k) \leq \min(2, \theta(h) + \theta(k) - 1). \tag{4.16}$$

The function $\theta(h)$ is illustrated by computing it for the max-stable random field models. The bivariate distribution for the max-stable process model proposed by Smith (1990) is equal to

$$\mathbb{P}(Z(\boldsymbol{x}) \leq z_1, Z(\boldsymbol{x} + \boldsymbol{h}) \leq z_2) = \exp\left\{-\left[\frac{1}{z_1}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_2}{z_1}\right) + \frac{1}{z_2}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_1}{z_2}\right)\right]\right\}$$

where Φ is the cdf of a normal(0,1) and $a^2 = \boldsymbol{h}^T\Sigma^{-1}\boldsymbol{h}$ ($= h^2$, since isotropy and $\Sigma = I$ are assumed in this work). In this case, the extremal coefficient is equal to

$$\theta(h) = 2\Phi\left(\frac{a}{2}\right).$$

The scale parameter is set to $a = 10$ in Figure 4.1. For the Schlather (2002) model the bivariate distribution corresponds to

$$\begin{aligned}\mathbb{P}(Z(\boldsymbol{x}) \leq z_1, Z(\boldsymbol{x} + \boldsymbol{h}) \leq z_2) \quad &= \quad \exp\left\{-\frac{1}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right)\right. \\ &\qquad \left.\left(1 + \sqrt{1 - 2\left(\rho(h) + 1\right)\frac{z_1 z_2}{(z_1 + z_2)^2}}\right)\right\}\end{aligned}$$

where $\rho(x)$ is the covariance function of the underlying Gaussian process. This yields

$$\theta(h) = 1 + \sqrt{1 - \frac{1}{2}\left(\rho(h) + 1\right)}$$

In Figure 4.1, the covariance function $\rho(h)$ for the center plot is chosen to be equal to $\rho(h) = \exp(-h/40)$ and in the right plot $\rho(h) = \exp(-(h/\sqrt{40})^2)$. Interestingly, for Schlather's model, as $h \to \infty$, $\theta(h) \to 1 + \sqrt{1/2}$, which means that observations do not approach independence as the distance between them increases.

The main advantage of the extremal coefficient is that it is easily understandable via (4.13). It is important to note that the value of θ is independent of the value of x in relation (4.13). Another important property is

that, although we have defined it in the case of unit Fréchet margins, the extremal coefficient is invariant to transformation of the margins to $GEV(\mu, \sigma, \xi)$ (Smith, 1990; Schlather and Tawn, 2002). However, while providing a measure of the amount of dependence, the extremal coefficient does not completely characterize the dependence structure. Two bivariate random variables (X, Y) and (X', Y') can share the same value of θ and yet have different dependence structures. In fact, θ provides information about the values of the joint distribution $G(x, y)$ only along the line $x = y$.

4.2.2 Pickands dependence function

A second dependence measure for extremes arises from an alternative representation of the bivariate EVD and gives a more complete picture of the extremes than the extremal coefficient. A bivariate EVD can be written as

$$G(x, y) = \exp\left\{-\left(\frac{1}{x} + \frac{1}{y}\right) A\left(\frac{y}{x+y}\right)\right\}, \tag{4.17}$$

where the dependence function $A(w)$ on $[0, 1]$ is convex, and $\max(w, 1-w) \leq A(w) \leq 1$ (Pickands, 1981; Fougères, 2004) (Figure 4.2). This representation of a bivariate EVD shows that the argument of the exponential in the distribution function can be separated into the product of the radial part and the angular part. A is a function of only the angular $y/(x+y)$ and completely characterizes the dependence. The Pickands dependence function is related to the extremal coefficient through the equation $\theta = 2A(1/2)$. For the bivariate logistic random variable with dependence measure α, $A_\alpha(w) = \{w^{(1/\alpha)} + (1-w)^{(1/\alpha)}\}^\alpha$, and for the asymmetric logistic $A_{\alpha,t_1,t_2}(w) = (1-t_1)w + (1-t_2)(1-w) + [(t_1 w)^{(1/\alpha)} + (t_2(1-w))^{(1/\alpha)}]^\alpha$. Also, if the bivariate EVD has a symmetric dependence structure (e.g. $G(x, y) = G(y, x)$), then $A(w) = A(1-w)$.

Unlike θ, the dependence function A gives complete information regarding the dependence structure. That is, given the function A, one can immediately determine the value of the joint distribution function $G(x, y)$ for all (x, y). However, the function takes values in $[1/2, 1]$ and these values do not have an interpretation like the value of θ.

To our knowledge, the dependence function A has not been extended to the spatial context, but it can easily be done. As an alternative to (4.9), we can define the bivariate distribution of a max-stable process as

$$\mathbb{P}\left(Z(\boldsymbol{x}) \leq u_1, Z(\boldsymbol{x}+\boldsymbol{h}) \leq u_2\right) = \exp\left\{-\left(\frac{1}{u_1} + \frac{1}{u_2}\right) A\left(\frac{u_2}{u_1+u_2}, h\right)\right\}.$$

Hence, the dependence function $A(w, h)$ is measures the dependence between points who lie a distance of h units apart, and has the same properties as the bivariate dependence function.

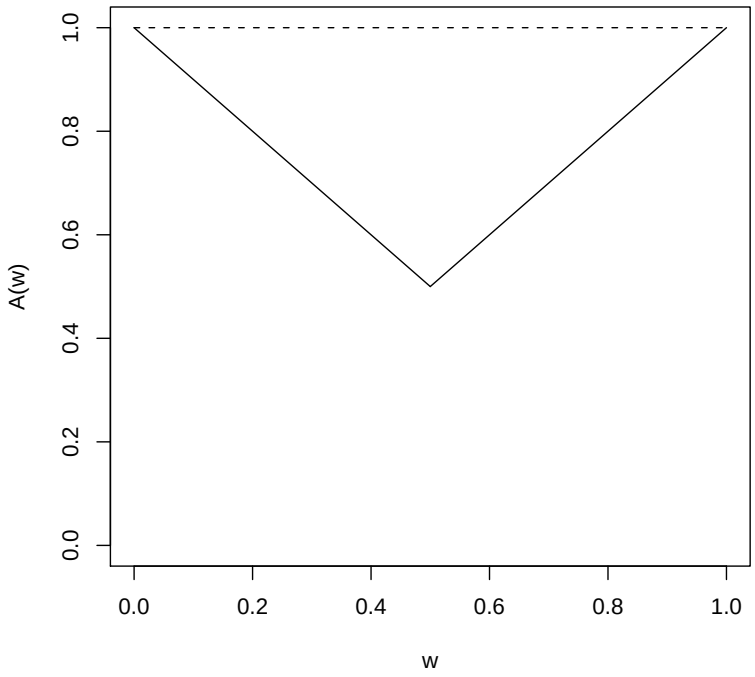

Figure 4.2: The range of values for the Pickands dependence function. The dashed line represents independence and the solid line represents $X = Y$ in probability.

4.2.3 Estimation

A few nonparametric estimators of A have been proposed, and since $\theta = 2A(1/2)$, they can also be viewed as estimators for the extremal coefficient. Pickands (1981) includes an estimator in his initial work:

$$A_n(w) = n \left\{ \sum_{i=1}^{n} \min \left(\frac{x_i}{w}, \frac{y_i}{1-w} \right) \right\}^{-1},$$

where $(x_1, y_1), \ldots, (x_n, y_n)$ are observations from a bivariate EVD. It can be shown that $1/A_n(w)$ is an unbiased estimator of $1/A(w)$, but it does not share all of the properties of A such as convexity and the values at the endpoints (Deheuvels, 1991). Since $A(0) = A(1) = 1$, Deheuvels (1991) proposed an estimator which corrected the Pickands estimator's values at $w = 0$ and $w = 1$:

$$A_n(w) = n \left\{ \sum_{i=1}^{n} \min \left(\frac{x_i}{w}, \frac{y_i}{1-w} \right) \right\}^{-1} - (1-u) \sum_{i=1}^{n} y_i - u \sum_{i=1}^{n} x_i + n$$

Hall and Tajvidi (2000) proposed a similar estimator which has the additional property that it is convex:

$$A_n(w) = n \left\{ \sum_{i=1}^{n} \min \left(\frac{x_i}{\bar{x}w}, \frac{y_i}{\bar{y}(1-w)} \right) \right\}^{-1},$$

where $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$ and similarly for $\bar{y}$. Capéràa et al. (1997) proposed a very different estimator which relied on integrating a function of the empirical distribution of H. Capéràa et al. define

$$A_n^0(w) = \exp\left\{\int_0^w \frac{H_n(z) - z}{z(1-z)}dz\right\}, A_n^1(w) = \exp\left\{-\int_w^1 \frac{H_n(z) - z}{z(1-z)}dz\right\},$$

where $H_n(z)$ is the empirical distribution function for H, and suggest the estimator

$$\log A_n(w) = p(w)\log A_n^0 + (1 - p(w))\log A_n^1,$$

where $p(w)$ is a weighting function. The authors claim that this estimator outperforms the estimators proposed by Pickands and Deheuvels. These estimators are investigated along with our proposed estimator in the next section.

4.3 Madogram

Section 4.1 discussed the variogram, which assesses spatial dependence. A lesser-known dependence measure from geostatistics is the first-order version of the variogram or madogram:

$$\nu(h) = \frac{1}{2}\mathbb{E}\left|Z(\boldsymbol{x}) - Z(\boldsymbol{x} + \boldsymbol{h})\right|. \tag{4.18}$$

This measure has been studied by Matheron (1987), although its properties have not been completely worked out and has not been previously employed with extremes. We suggest the madogram as a dependence measure for extremes for two reasons. Heavy-tailed distributions with infinite second moments are not infrequent in the study of extremes, and for such distributions, the variogram is not finite. However, this explanation is not completely satisfactory as distributions with infinite first moments are also studied in extremes, and the madogram as defined in (4.18) is not finite for such distributions. The second and more compelling argument for applying the madogram to extreme data is that the extremal coefficient $\theta(h)$ can be expressed as a function $\nu(h)$ as shown in the following section. Hence, just as the variogram is an especially useful tool for measuring dependence of Gaussian data because of its relationship with covariance, the madogram is a useful tool for measuring dependence in extremal data because of its relationship with the extremal coefficient.

4.3.1 Madogram for bivariate extremes

For the moment, assume that (X, Y) comes from a bivariate EVD with extremal coefficient θ and common $GEV(\mu, \sigma, \xi)$ marginals such that $\xi < .5$ (hence, we are *not* in the Fréchet marginal case). The bivariate madogram is defined to be

$$\nu = \frac{1}{2}\mathbb{E}\left|X - Y\right|. \tag{4.19}$$

We employ the relation

$$\frac{1}{2}\left|X - Y\right| = \max\left(X, Y\right) - \frac{1}{2}\left(X + Y\right) \tag{4.20}$$

and take expectations of both sides to obtain $\nu = \theta\mathbb{E}[XF^{\theta-1}(X)] - \mathbb{E}[X]$. The first term of the expression for ν is recognizable as a probability weighted moment (PWM), which are primarily employed as method-of-moments type estimators for parameters. PWMs have been studied in the context of extremes especially in the field of hydrology, and their values for the GEV distribution are known (Hosking and Wallis, 1987; 1985). Taking advantage of these known relationships, we obtain

$$\nu = \begin{cases} \frac{\sigma}{\xi}(\theta^{\xi} - 1)\Gamma(1-\xi), & \text{for } \xi \neq 0 \\ \sigma \log \theta, & \text{for } \xi = 0. \end{cases} \tag{4.21}$$

However, the primary interest is in the case when the marginal distributions of the bivariate EVD are standard Fréchet, and unfortunately the madogram as defined in (4.19) is not finite for this case. Several definitions of Fréchet madograms were previously examined in (Naveau et al., 2005c) which rely on the concept of regularly varying tails (Embrechts et al., 1997). One such Fréchet madogram was defined as

$$\nu_t = \frac{1}{2}\frac{\mathbb{E}|X_t - Y_t|}{\mathbb{E}[X_t]}; \text{ where } X_t = X\mathbf{1}(X \leq t) \text{ and } Y_t = X\mathbf{1}(Y \leq t),$$

and $\mathbf{1}(A)$ is the indicator function that is equal to 1 if A is true and equal to 0 otherwise. This madogram ν_t is finite for all t and $\lim_{t\to\infty} \nu_t = \theta - 1$. However, from simulation studies, it was found that the estimators for ν_t and the other definitions of Fréchet madograms were very slow to converge.

Instead, a different approach is taken for defining a madogram which is finite for the standard Fréchet marginal case. Let (X, Y) come from a bivariate EVD.Then the normalized madogram is defined by

$$\nu_F = \frac{1}{2}\mathbb{E}\left|F(X) - F(Y)\right|. \tag{4.22}$$

Since $F(X) \sim Unif[0, 1]$, the normalized madogram is defined for the standard Fréchet case, but the normalized madogram can be applied to a bivariate EVD with any common marginal distribution $F = GEV(\mu, \sigma, \xi)$. If X and Y are independent, then $\nu_F = 1/6$ and if $X = Y$ in probability, then $\nu_F = 0$.

Similar to the previous madogram, the normalized madogram has a convenient relationship with the extremal coefficient.

Proposition 1 Let (X, Y) be a bivariate extreme value random variable with $GEV(\mu, \sigma, \xi)$ marginals and an extremal coefficient of θ. Then

$$\nu_F = \frac{\theta}{1+\theta} - \frac{1}{2}, \tag{4.23}$$

or inversely,

$$\theta = \frac{\nu_F + 1/2}{1/2 - \nu_F}. \tag{4.24}$$

The proof of this proposition is a special case of proposition 4 which can be found in Appendix .2.

The definition of normalized madogram lends itself to a natural estimator. Assuming the marginal distribution F is known, then

$$\hat{\nu}_{F,n} = \frac{1}{2n} \sum_{i=1}^{n} |F(X_i) - F(Y_i)| \tag{4.25}$$

If F is unknown, the estimator is done in two steps: the marginal distribution is estimated (numerical maximum likelihood methods are used in this work), and F in the above estimator is replaced with the estimated distribution function $\hat{F}$.

A few properties of the estimator $\hat{\nu}_{F,n}$ are also easily derived. It can be shown that the estimator is unbiased, and converges in distribution to a Gaussian distribution with mean ν_F and variance $n^{-1}\left[1/2\gamma_F - \nu_F\right]$ where $\gamma_F = 1/2\mathbb{E}\left(F(X_i) - F(Y_i)\right)^2$. The proof of this is a special case of the first part of Proposition 5 found in Appendix .2.

By plugging $\hat{\nu}_{F,n}$ into (4.24), one can generate a new estimator for θ. The estimate, $\hat{\theta}_n$ can be shown to converge in distribution to a Gaussian with mean θ and and variance $n^{-1}(1-(\nu_F+1/2))^{-2}(1/2\gamma_F - \nu_F)$. The proof of this is a special case of the the second part of Proposition 5. One disadvantage of using the relation (4.24) as an estimator for θ is if $\hat{\nu}_{F,n}$ is near 1/2, then the estimate $\hat{\theta}_n$ could potentially have a large variance. In simulation studies of bivariate random variables, this did not seem to be an issue. Extending the idea further, using the relation $A(1/2) = 1/2\theta$, we can obtain an estimator for the Pickands measure $\hat{A}_n(1/2)$.

A simulation experiment was performed to compare various estimators of extremal dependence. Three different bivariate dependence measures which yield information about the joint distribution function $G(x, y)$ along the line $x = y$: θ, $A(1/2)$, and ν_F have been presented. Although they quantify the same thing, the dependence measures have different values. To compare the different estimators, one must convert to a common measure, and θ was chosen. The estimators suggested by Capéraà et al. (1997), Pickands (1981), Deheuvels (1991), Hall (2000), and the normalized madogram are compared in the experiment. For each experiment, one hundred bivariate EV random variables with unit Fréchet margins were simulated. The marginal distributions of each component were estimated under the assumption of independence, then each estimator was calculated and converted to a value of θ. Each experiment was repeated 100 times and the mean-absolute-error was calculated. Three different bivariate distributions were simulated: a logistic with weak dependence ($\alpha = .7$), a logistic with strong dependence ($\alpha = .3$), and an asymmetric logistic with $\alpha = .2, t_1 = .8,$ and $t_2 = .5$.

The results from the experiments are shown in Table 4.1 and Figure 4.3. For all three distributions, the estimator for θ based on the normalized madogram has a smaller mean-absolute-error than any of the other estimators.

	Simulation 1	Simulation 2	Simulation 3
Capéraà	0.042 (0.058)	0.033 (0.057)	0.038 (0.037)
Pickands	0.068 (0.080)	0.023 (0.027)	0.049 (0.052)
Deheuvels	0.056 (0.067)	0.021 (0.024)	0.040 (0.041)
Hall	0.052 (0.057)	0.022 (0.024)	0.036 (0.034)
Madogram	0.030 (0.022)	0.013 (0.010)	0.028 (0.025)

Table 4.1: The mean-absolute-error (standard errors of mae) for the different estimators of θ. The normalized madogram appears to outperform the other estimators.

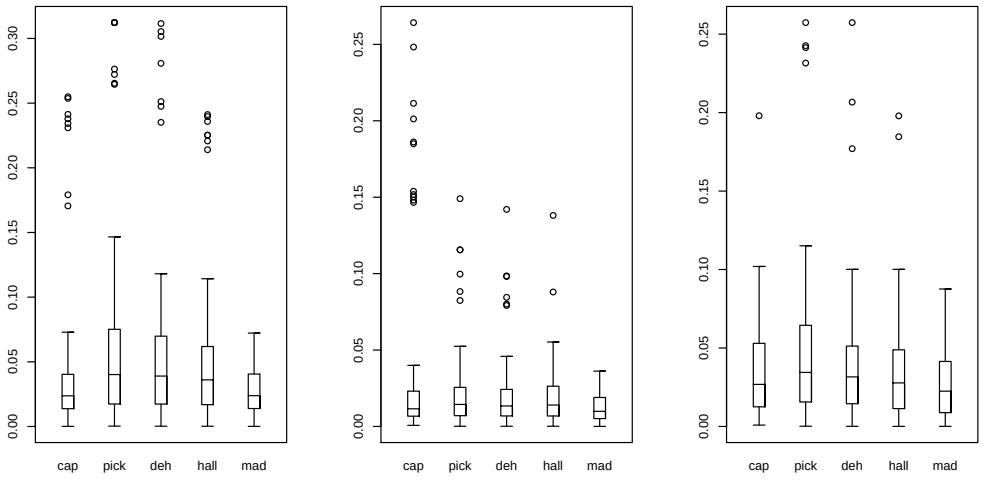

Figure 4.3: Box plots of the absolute errors for 100 experiments. The three figures are for simulation 1, simulation 2, and simulation 3. Box plots within each figure are from left-to-right are for the Capéraà, Pickands, Deheuvels, Hall, and Madogram estimators.

4.3.2 Madograms for spatial extremes

The original focus of this work is to produce a spatial dependence measure. In the introductory paragraph of this chapter, the spatial madogram (4.18) was introduced. Of course, (4.18) is not finite for the Fréchet marginal case, and as before the normalized spatial madogram is defined to remedy this problem:

$$\nu_F(h) = \frac{1}{2}\mathbb{E}|F(Z(\boldsymbol{x})) - F(Z(\boldsymbol{x}+\boldsymbol{h}))|. \tag{4.26}$$

As this is just an extension of the bivariate normalized madogram, it has all the same properties.

Given a stationary, isotropic max-stable field, the madogram is a function of distance. As is typically done with the variogram, we propose an estimator which bins distances. It is straightforward to define such an estimator, we let

$$\hat{\nu}_F(h) = \frac{1}{2|\mathcal{N}_h|}\sum_{(x_i,x_j)\in\mathcal{N}_h}|F(Z(x_i)) - F(Z(x_j))|, \tag{4.27}$$

where $\mathcal{N}_h$ is the set of all sample pairs for which $||x_i - x_j|| \in [h-\delta, h+\delta)$.

To test the estimator, max-stable random fields of size 40x40 units^2 are simulated using the models proposed by Smith (1990) and Schlather(2002). A set of 105 points is used to simulate observation stations in the field. A field with a known spatial dependence structure was estimated and (4.27) was used to estimate the madogram's values. The experiment was repeated for 300 fields. Figure 4.4 summarizes the results. On average, the madogram does a very good job of tracking the theoretical spatial dependence. As is expected, the range of values of the madogram for the individual fields varies quite a bit as shown by the range of the box plots. This is also true for variogram estimation (Cressie, 1993), and suggests that employing parametric models of the madogram would be advantageous in practice. Such parametric models could take advantage of the fact that $\nu(h) \leq 1/6$, which the estimator (4.27) does not.

As in the bivariate case, the madogram can also be used as an estimator for $\theta(h)$ or $A(1/2, h)$. Using the madogram results from the earlier experiments, estimates for θ were also obtained and the results are shown the lower plots of in Figure 4.4.

An advantage of the madogram's relationship with $\theta(h)$ is that is makes proving properties of the dependence function measure much easier. Schlather and Tawn's (2003; 2002) proofs of 4.15 and 4.16 are rather complicated, but similar results can be obtained rather easily using the triangle inequality and the madogram (see also Poncet, Cooley and Naveau, 2005).

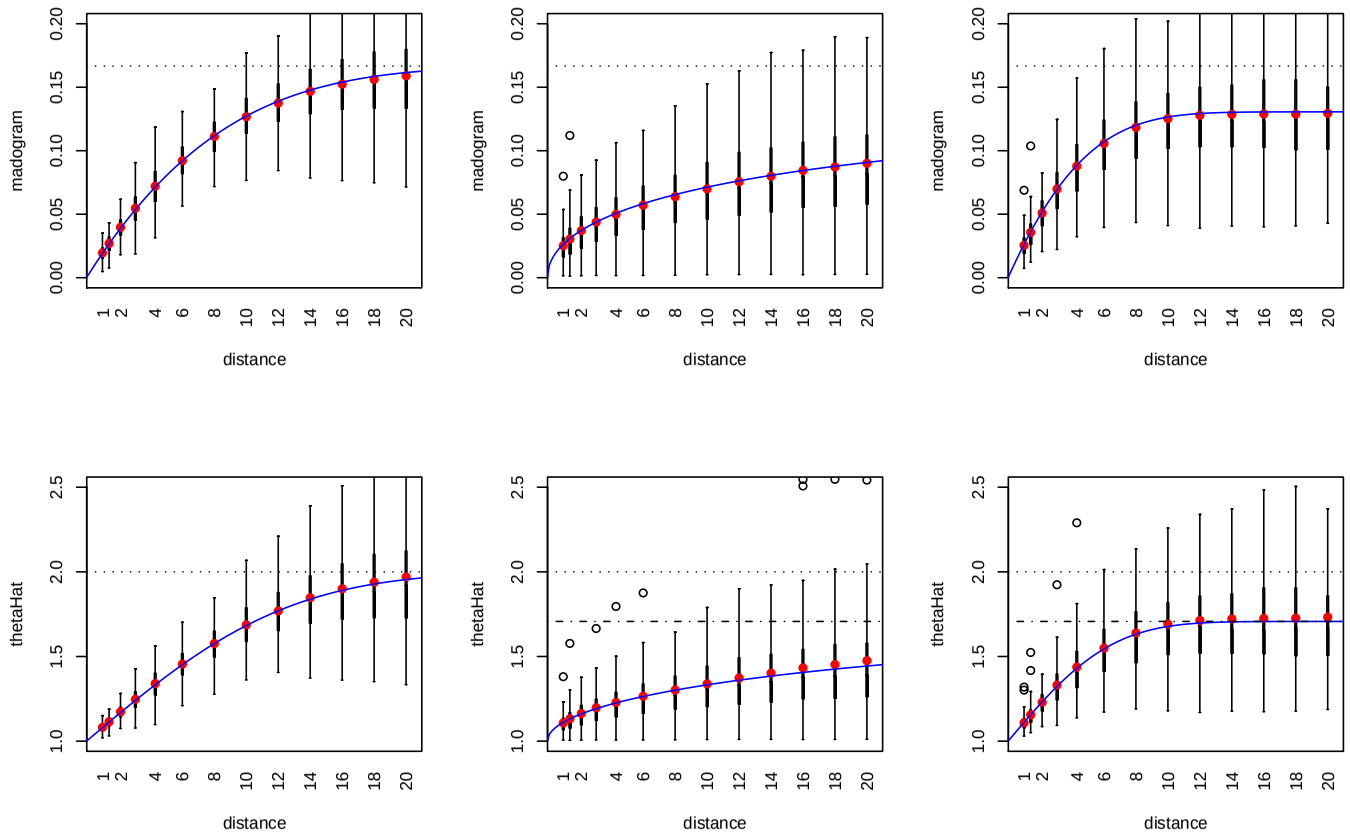

Figure 4.4: Top Row: Madogram estimates for the Smith model (left), Schlather model with exponential variogram (center), and Schlather model with Gaussian variogram (right). Circles correspond to the mean madogram value of the 300 experiments, and box plots summarize the range of estimates. Dashed horizontal line at 1/6 corresponds to independence. Bottom Row: Estimates θ for the three models as before. Dashed lines at height of 2 corresponds to independence, and at height of $1 + \sqrt{1/2}$ corresponds to the limit behavior of the Schlather model.

Proposition 2 Any extremal coefficient function $\theta(h)$ is such that

$$(1) \qquad 2 - \theta(h) \text{ is positive semi-definite,} \tag{4.28}$$

$$(2) \qquad \theta(h+k) \leq \theta(h)\theta(k), \tag{4.29}$$

$$(3) \qquad \theta^\tau(h+k) \leq \theta^\tau(h) + \theta^\tau(k) - 1 \text{ for } 0 \leq \tau \leq 1, \text{ and} \tag{4.30}$$

$$(4) \qquad \theta^\tau(h+k) \geq \theta^\tau(h) + \theta^\tau(k) - 1 \text{ for } \tau \leq 0\ . \tag{4.31}$$

The proof of this proposition is given in Appendix .2. Obviously, (4.28) corresponds with (4.15) and if $\tau = 1$ (4.30) corresponds with (4.16).

4.3.3 Application: France precipitation data

To illustrate the normalized madogram, it is applied to precipitation from the Dijon region of France. The data consists of the maximum daily precipitation measurement recorded over a 30-year period at 146 locations. As the measurements are maxima over a long period of time, it is reasonable to assume that they come from a distribution which is approximately max-stable. The data were obtained from Metéo France, who standardized the measurements so that the field could be assumed to be stationary. The locations of the stations are given by Lambert coordinates which transforms the space to have Euclidean distances between the stations. The locations and measurements are shown in Figure 4.5. The marginal distribution of the data was estimated by assuming independence and fitting a GEV to the measurements by numerical maximum-likelihood estimates.

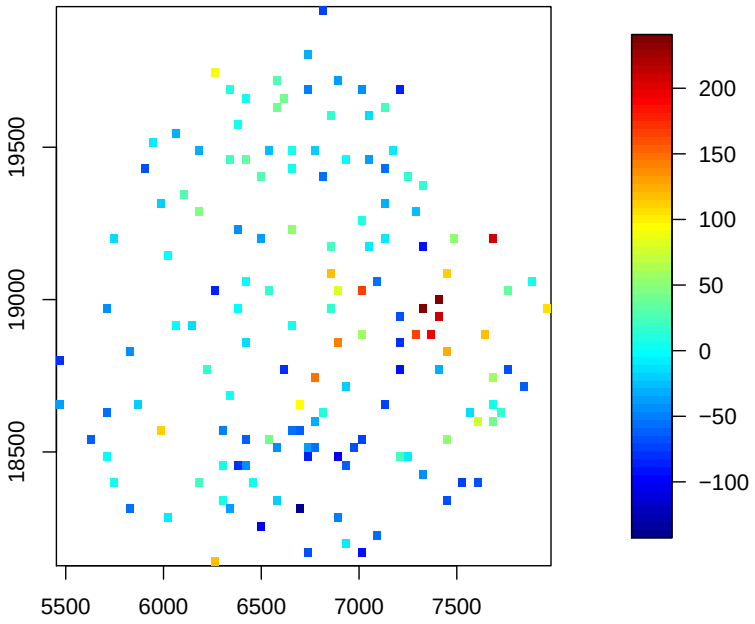

Figure 4.5: The locations and normalized 30-year maximum precipitation measurements for the Dijon region of France.

The normalized madogram was applied to the data, using the estimated marginal distribution as F. Figure 4.6 shows the binned madogram estimates along with the asymptotic 95% confidence interval. Also plotted is a parametric estimate of the madogram. Employing the method of least-squares, an exponential madogram model $\nu_F(h) = c_0 + c(1 - \exp(-h/a))$ was fit to the 10585 pairwise madogram estimates, with the estimates being $c_0 = 0.057, c = 0.110,$ and $a = 120.06$. A more sophisticated method of fitting the parametric madogram model could be devised which utilizes the information that $\max(v(h)) = 1/6$, however, our unrestricted method

obtains a value $\nu(h)$ very close to 1/6 for large values of h.

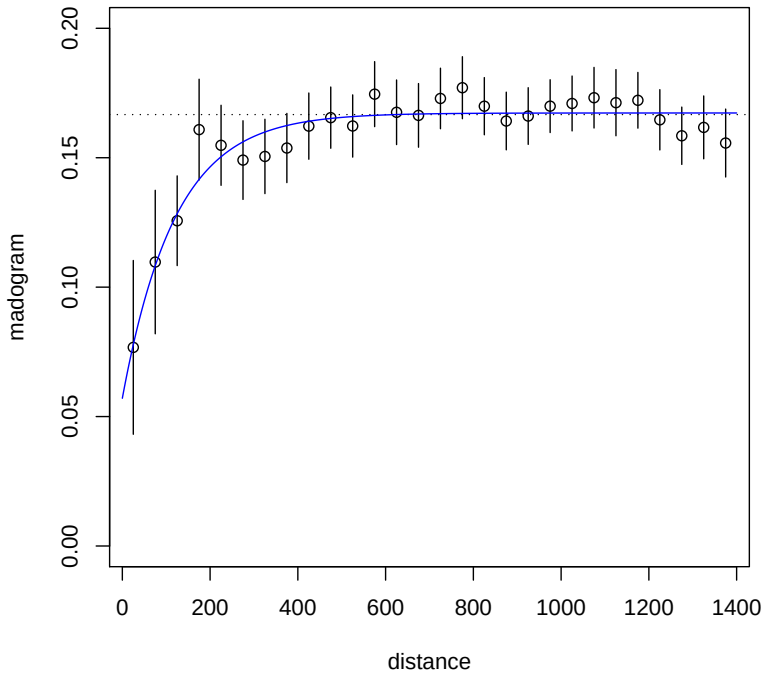

Figure 4.6: Non-parametric and parametric madogram estimates for the France precipitation data.

4.4 Extending the extremal coefficient and madogram

Let F be the distribution function of a unit Fréchet random variable. Given a bivariate EVD with joint distribution function G, unit Fréchet marginals, and extremal coefficient θ, we know that $G(x, x) = F^{\theta}(x)$. But what of $G(x, y)$ where $y \neq x$? Because the extremal coefficient gives incomplete dependence information, it is not possible to write $G(x, y)$ in terms of θ, as it is defined. The notion of the extremal coefficient (and later the madogram) can be extended to give a complete description of dependence.

4.4.1 Bivariate Case

Let $(x, y); x, y > 0$ be a point given by Cartesian coordinates. The point can be alternatively defined by letting $\lambda = x/y$ and describing the same location with coordinates (x, λ). The parameter $\lambda \in [0, \infty)$ is an angular measure and any point (x, λ) can be translated to Cartesian coordinates $(x, x/\lambda)$.

Given a unit Fréchet bivariate EVD, we define the parameterized extremal coefficient $\theta(\lambda)$ as

$$\theta(\lambda) = V\left(1, \frac{1}{\lambda}\right) = \int_0^1 \max\left(w, \lambda(1-w)\right) dH(w). \tag{4.32}$$

Given this definition, then

$$\begin{aligned} G(x,y) &= \mathbb{P}\left(X \leq x, Y \leq y\right) \\ &= \mathbb{P}\left(X \leq x, Y \leq x/\lambda\right), \text{ where } \lambda = x/y \\ &= \exp\left\{-V(x, x/\lambda)\right\} \\ &= \exp\left\{-\frac{1}{x}\theta(\lambda)\right\} \\ &= F^{\theta(\lambda)}(x) \end{aligned} \tag{4.33}$$

Hence, $\theta(\lambda), \lambda \in [0, \infty)$ has a similar relation as the previously-defined extremal coefficient, but now gives the complete dependence structure. $\theta(1)$ corresponds to the unparameterized extremal coefficient.

There are several properties that can be easily derived for $\theta(\lambda)$.

Proposition 3 Let (X, Y) be a random variable from a bivariate EVD with Fréchet margins and let $\theta(\lambda)$ be defined as in (4.32). Then the following properties hold:

(1) If $\lambda_1 < \lambda_2$, then $\theta(\lambda_1) < \theta(\lambda_2)$.

(2) $\max(1, \lambda) \leq \theta(\lambda) \leq 1 + \lambda$

(3) If the dependence structure is symmetric, e.g $G(x, y) = G(y, x)$, then $\theta(1/\lambda) = \frac{1}{\lambda}\theta(\lambda)$.

These properties are proved in Appendix .2. As a direct consequence of the second item, as $\lambda \to \infty$, $\theta(\lambda) \to \infty$, and $\theta(0) = 1$. One difference from the unparameterized measure is that only if $\lambda = 1$ is $\theta(\lambda)$ invariant to marginal transformations. The Pickands dependence function can be related to $\theta(\lambda)$ through the relation

$$A(w) = \frac{1}{\lambda + 1}\theta(\lambda), \tag{4.34}$$

where $w = (1 + \lambda)^{-1}$.

The measure $\theta(\lambda)$ is a slightly awkward dependence measure. Although the relation (4.33) is similar to the unparameterized case, $\theta(\lambda)$ can no longer be interpreted as the number of independent random variables from which the maximum is drawn. The measure $\theta(\lambda)$ takes values in $[1, \infty)$, and the value of the measure strongly depends on the value of the parameter λ, as seen in the second item in Proposition 3. Hence, it is difficult to interpret the value of $\theta(\lambda)$. The parametrization is also a bit strange, as $\lambda \in [0, \infty)$ and symmetry (in the sense of Proposition 3 part (3)) is via reciprocal values about $\lambda = 1$.

The definition of the madogram can likewise be extended to yield a complete measure of dependence. If (X, Y) is a random variable with a bivariate EVD distribution with Fréchet margins, then the λ-madogram is defined as:

$$\nu(\lambda) = \frac{1}{2}\mathbb{E}\left|F(X) - F(\lambda Y)\right|. \tag{4.35}$$

It is easily shown that $\nu(0) = 1/4$ and as $\lambda \to \infty$, $\nu(\lambda) \to 1/4$. Furthermore, if X and Y are independent, $\nu(\lambda) = \frac{1+\lambda}{2+\lambda} - \frac{3\lambda+1}{4(\lambda+1)}$, and if $X = Y$ in probability, $\nu(\lambda) = \left|\frac{1}{4} - \frac{\lambda}{2(\lambda+1)}\right|$. The values which $\nu(\lambda)$ can achieve are plotted in Figure 4.7. Interestingly, $\nu(\lambda)$ is not reciprocal-symmetric about $\lambda = 1$, and in fact

$$\nu(1/\lambda) = \frac{1 + 3\lambda + 4(1+\lambda)\nu(\lambda)}{1 + 4\nu(\lambda) + \lambda(6 + \lambda - 4\lambda\nu(\lambda))} - \frac{\lambda + 3}{4(\lambda + 1)}$$

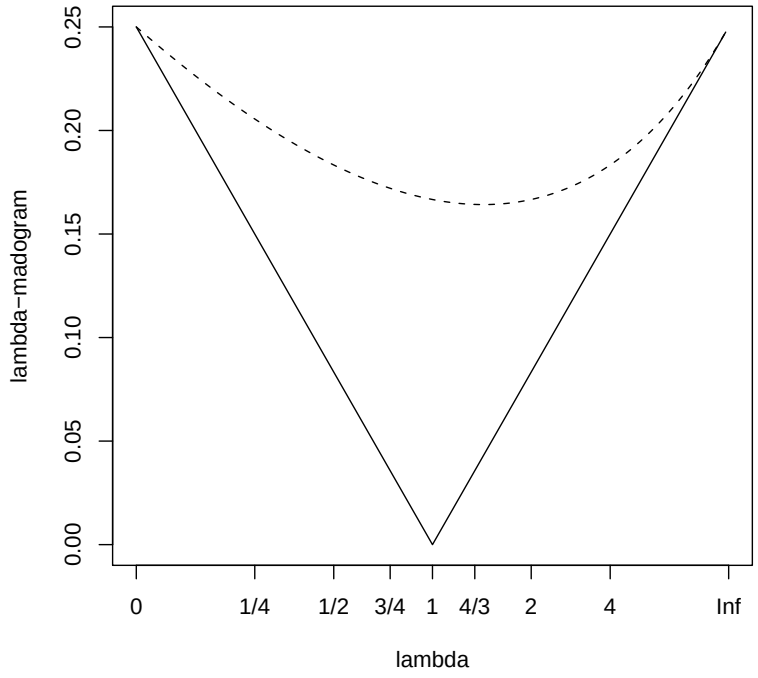

Figure 4.7: The range of values of the λ-madogram. The dashed line represents the case where X and Y are independent, and the solid line represents $X = Y$ in probability. Interestingly, the λ-madogram is not reciprocal symmetric. Compare with Figure 4.2

Just as we related the normalized madogram to the extremal coefficient, the λ-madogram can be related to $\theta(\lambda)$.

Proposition 4 Let (X, Y) be a bivariate extreme value random variable with unit Fréchet marginals and an extremal coefficient of θ. Then

$$\nu(\lambda) = \frac{\theta(\lambda)}{1 + \theta(\lambda)} - \frac{1 + 3\lambda}{4(1 + \lambda)}, \tag{4.36}$$

or inversely,

$$\theta(\lambda) = \frac{\nu(\lambda) + c_\lambda}{1 - (\nu(\lambda) + c_\lambda)}, \text{ where } c_\lambda = \frac{1 + 3\lambda}{4(1 + \lambda)}. \tag{4.37}$$

The proof can be found in Appendix .2.

4.4.1.1 Estimators for $\theta(\lambda)$ and $\nu(\lambda)$

The definition of $\nu(\lambda)$ lends itself to the natural estimator

$$\hat{\nu}_n(\lambda) = \frac{1}{2n}\sum_{i=1}^{n} |F(X) - F(\lambda Y)| \,. \tag{4.38}$$

It is obvious that $\hat{\nu}_n(\lambda)$ is unbiased for $\nu(\lambda)$. However, $\hat{\nu}_n(\lambda)$ does not have all the desirable properties that an estimator might. For instance, $\hat{\nu}_n(0)$ and $\hat{\nu}_n(\infty)$ (taken as the limit), do not necessarily equal 1/4. In simulation studies, $\hat{\nu}_n(\lambda)$ proved to be an unreliable estimator of $\nu(\lambda)$ for values of λ near zero or very large.

Similarly to what Deheuvels (1991) did with the Pickands (1981) estimator, a correction can be made such that the ends have the correct values. Let

$$\begin{aligned}\hat{\hat{\nu}}_n(\lambda) &= \frac{1}{2n}\sum_{i=1}^{n} |F(X_i) - F(\lambda Y_i)| - (1-w(\lambda))\frac{1}{2n}\sum_{i=1}^{n}(1-F(X_i)) \\ &\quad -w(\lambda)\frac{1}{2n}\sum_{i=1}^{n} F(X_i) + \frac{1}{4},\end{aligned} \tag{4.39}$$

where $w(\lambda)$ is a function such that $w(0) = 1$, as $\lambda \to \infty$, $w \to 1$, and $w(1) = 1/2$ ($w(\lambda) = \frac{\lambda}{1+\lambda}$ within this work). Note that $\hat{\hat{\nu}}_n(\lambda)$ is still unbiased for $\nu(\lambda)$, but now $\hat{\hat{\nu}}_n(0) = 1/4$ and $\hat{\hat{\nu}}_n(\infty) = 1/4$.

Similarly to before, if an estimator for $\theta(\lambda)$ is desired, we propose that $\hat{\hat{\nu}}_n(\lambda)$ is plugged into relation (4.37) to obtain $\hat{\theta}_n(\lambda)$. Furthermore, $\hat{\theta}_n(\lambda)$ can be plugged into relation (4.34) to obtain $\hat{A}_n(w)$. The estimators $\hat{\hat{\nu}}_n(\lambda), \hat{\theta}_n(\lambda)$, and $\hat{A}(w)$ can be shown to consistent.

Proposition 5 Let $(x_i, y_i), i = 1, \ldots, n$ be observations from a bivariate EVD with Fréchet margins. The estimators converge in distribution as follows

(1) $\hat{\hat{\nu}}_n(\lambda) \to N\left(\nu(\lambda), \frac{1}{n}h(\lambda)\right)$.

(2) $\hat{\theta}_n(\lambda) \to N\left(\theta(\lambda), \frac{1}{n}\frac{1}{1-(\hat{\hat{\nu}}_n(\lambda)+c_\lambda)}h(\lambda)\right)$.

(3) $\hat{A}_n(w) \to N\left(A(w), \frac{1}{n}\frac{1}{\lambda+1}\frac{1}{1-(\hat{\hat{\nu}}_n(\lambda)+c_\lambda)}h(\lambda)\right)$.

where $h(\lambda) = \frac{1}{2}\gamma(\lambda) - \frac{3}{4}\nu(\lambda) - \frac{1}{2}\eta(\lambda) + \frac{1}{48}(1 - w(\lambda) - (w(\lambda))^2)$, $\gamma(\lambda) = \mathbb{E}[(F(X) - F(\lambda Y))^2]$, $\eta(\lambda) = \mathbb{E}[F(X)|F(X) - F(\lambda Y)|]$ and $w = 1/(\lambda + 1)$. The proof is in Appendix .2.

	$\lambda = 99$	$\lambda = 7/3$	$\lambda = 1$	$\lambda = 3/7$	$\lambda = 1/99$
Capéràa	0.003 (0.002)	0.032 (0.034)	0.038 (0.037)	0.030 (0.026)	0.004 (0.001)
Hall	0.002 (0.001)	0.024 (0.021)	0.036 (0.034)	0.040 (0.030)	0.004 (0.001)
Madogram	0.156 (0.132)	0.021 (0.017)	0.028 (0.025)	0.027 (0.022)	0.011 (0.008)

Table 4.2: The mean absolute error (standard error of mae) from 100 simulated asymmetric logisitic random variables.

The naive and end-corrected estimators for $\theta(\lambda)$ based on (4.39) are compared to the Deheuvels and Capéraà estimators for $A(t)$. Bivariate random variables with known dependence structures were simulated and the various estimators were applied. Figure 4.8 shows an illustration of each type of simulation, and Figure 4.9 and Table 4.2 give results from multiple simulations. The λ-madogram based estimator seems to outperform the Deheuvels estimator and also seems to be a slight improvement of the Capéràa estimator, at least for moderate values of the parameter λ. Also, as Table 4.2 shows, the standard errors for the madogram are less than other estimators for moderate values of λ, but is less stable for very high or very low values of λ.

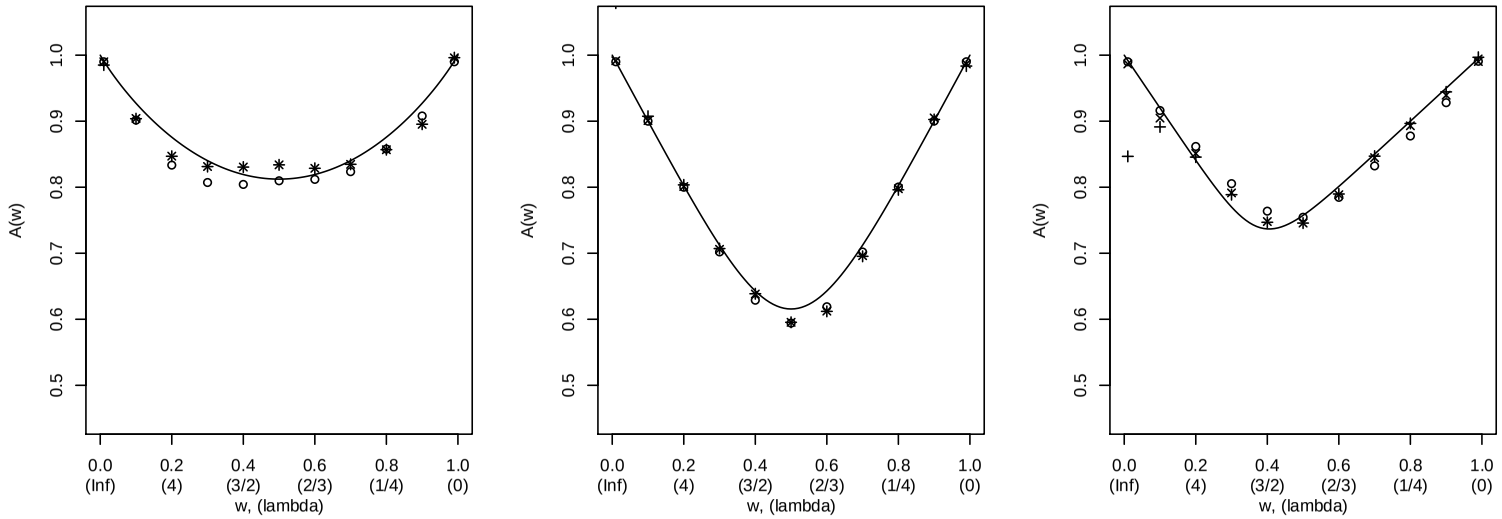

Figure 4.8: Three examples of the simulations which compare the dependence estimators. Circles denote the estimator proposed by Caperaa, plusses denote the madogram estimator without the end correction, and X's denote the madogram estimator with the end correction. For the left and center plots, bivariate logistic random variables were simulated with dependence measures $\alpha = .7$, and $\alpha = .3$ respectively. For the right plot, asymmetric logistic bivariate random variables with $\alpha = .2, t_1 = .8$, and $t_2 = .5$ were simulated. The Caperaa estimate and the corrected madogram estimate seem to give similar results. The uncorrected madogram estimate can sometimes be very bad for large λ (small w), such as in the right plot.

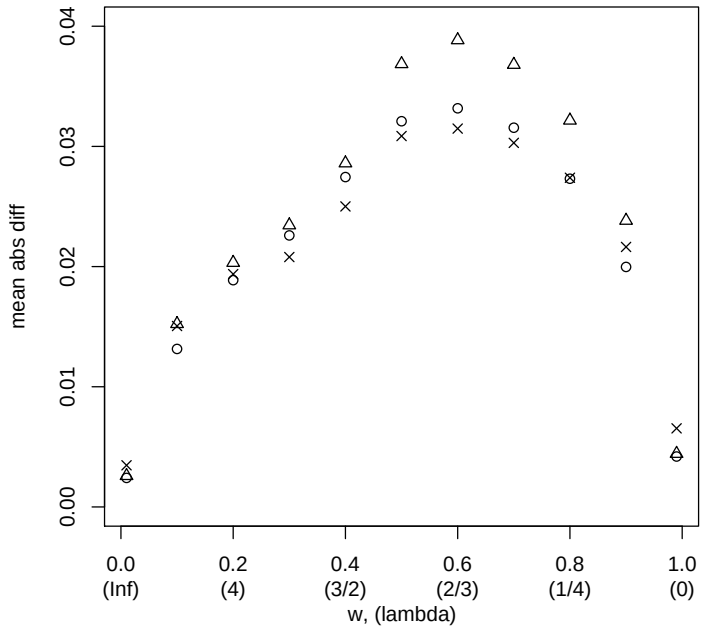

Figure 4.9: The mean absolute difference of $A(w)$ and its estimates. One-hundred pairs of asymmetric logistic bivariate random variables with $\alpha = .2, t_1 = .8$, and $t_2 = .5$ were simulated, and the estimators calculated. Experiment was repeated 100 times, and the mean absolute difference from the theoretical value was calculated. Madogram estimators with end-correction are X's, Caperaa estimators are circles, and Deheuvels estimators are trianges. The madogram appears to outperform the Caperaa estimator for moderate values of λ (w), and the Caperaa estimator appears superior for more extreme λ values. Both estimators appear to give relatively similar estimates and are superior to Deheuvels.

4.4.2 Spatial Case

The definitions of the parametrized extremal coefficient and λ-madogram can be extended in a straightforward manner to the spatial case. As before, let $Z(\boldsymbol{x})$ be a max-stable spatial process. The parametrized pairwise extremal coefficient can be defined as

$$\theta(\lambda, h) = \int \max\{g(s, x), \lambda g(s, x + h)\}\delta(ds). \tag{4.40}$$

As before, by letting $\lambda = z_1/z_2$, knowledge of $\theta(\lambda, h)$ completely determines the pairwise dependence.

$$G_{Z(\boldsymbol{x}),Z(\boldsymbol{x}+\boldsymbol{h})}(z_1, z_2) = G_{Z(\boldsymbol{x}),Z(\boldsymbol{x}+\boldsymbol{h})}(z_1, z_1/\lambda) = \exp(-\theta(\lambda, h)/z_1).$$

The λ-madogram is also defined for the spatial case as

$$\nu(\lambda, h) = \mathbb{E}\frac{1}{2}|F(Z(\boldsymbol{x}) - F(\lambda Z(\boldsymbol{x} + \boldsymbol{h}))|. \tag{4.41}$$

To get an idea of the behavior of the λ-madogram, the measure for the Schlather max-stable field model with a Gaussian variogram structure is plotted in Figure 4.10. One can see that at $h = 0$, the madogram shows the V-shaped behavior of complete dependence (reference Figure 4.7) and as h increases, the dependence structure approaches the more (unsymmetrical) U-shaped behavior of independence. Although it is hard to see in the figure, the spatial λ madogram is not reciprocal-symmetric about the value $\lambda = 1$.

Similarly to the the bivariate case, an estimator can be constructed for the λ-madogram which has the correct behavior for $\lambda = 0$ and $\lambda \to \infty$.

$$\begin{aligned}\hat{\nu}(\lambda, h) &= \frac{1}{2|\mathcal{N}_h|}\sum_{(x_i,x_j)\in\mathcal{N}_h}|F(Z(x_i)) - F(\lambda Z(x_j))| \qquad (4.42)\\ &\quad -(1 - w(\lambda))\frac{1}{2|\mathcal{N}_h|}\sum_{(x_i,x_j)\in\mathcal{N}_h}(1 - F(X_i))\\ &\quad -w(\lambda)\frac{1}{2|\mathcal{N}_h|}\sum_{(x_i,x_j)\in\mathcal{N}_h}F(X_i) + \frac{1}{4} \qquad (4.43)\end{aligned}$$

where $\mathcal{N}_h$ is the set of all sample pairs for which $||x_i - x_j|| \in [h - \delta, h + \delta)$.

We tested the estimator on the max-stable random fields models of Schlather and Smith. Similarly to before, $\hat{n}u(\lambda, h)$ was applied to 300 simulated fields and the average of the estimates for the Schlather exponential model plotted in Figure 4.11. As in the unparameterized case, the λ-madogram does an excellent job on average of tracking the theoretical dependence structure for all values of λ.

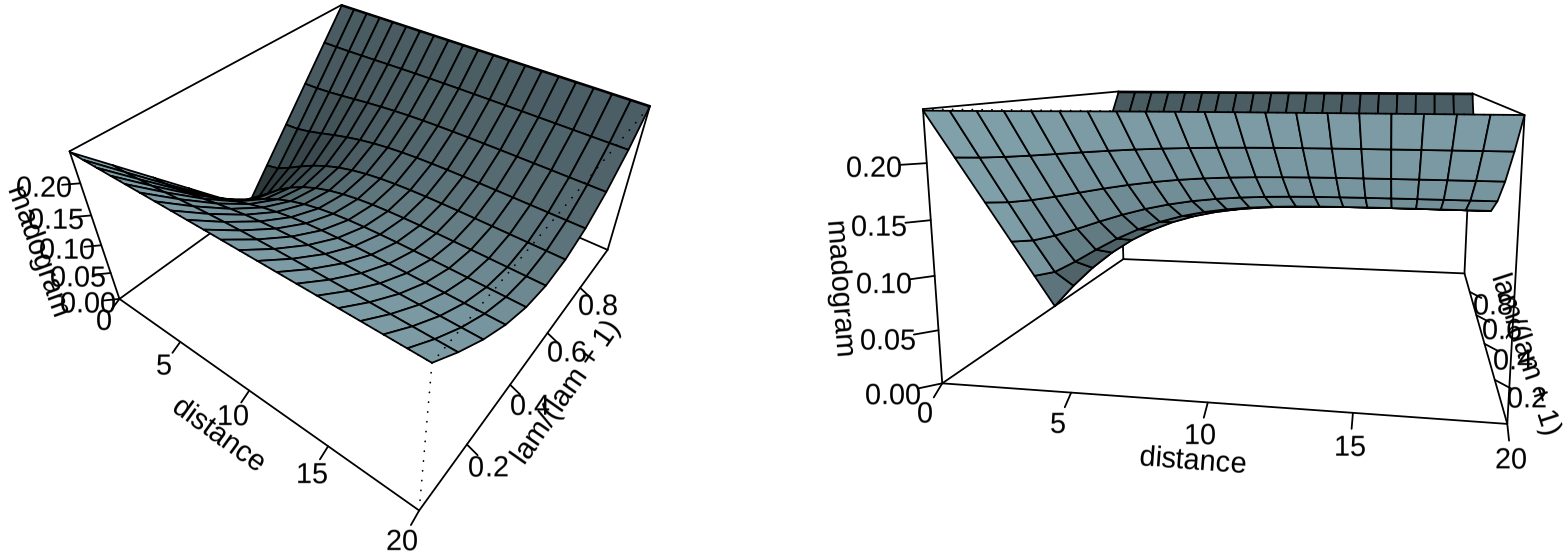

Figure 4.10: The behavior of the spatial λ-madogram. At small h the λ-madogram displays the "V" shape indicating complete dependence, and as h grows, the madogram becomes an unsymmetric "U" shape indicating less and less dependence. The scale for the λ's has been changed to $\lambda/(\lambda+1)$ for ease of graphing.

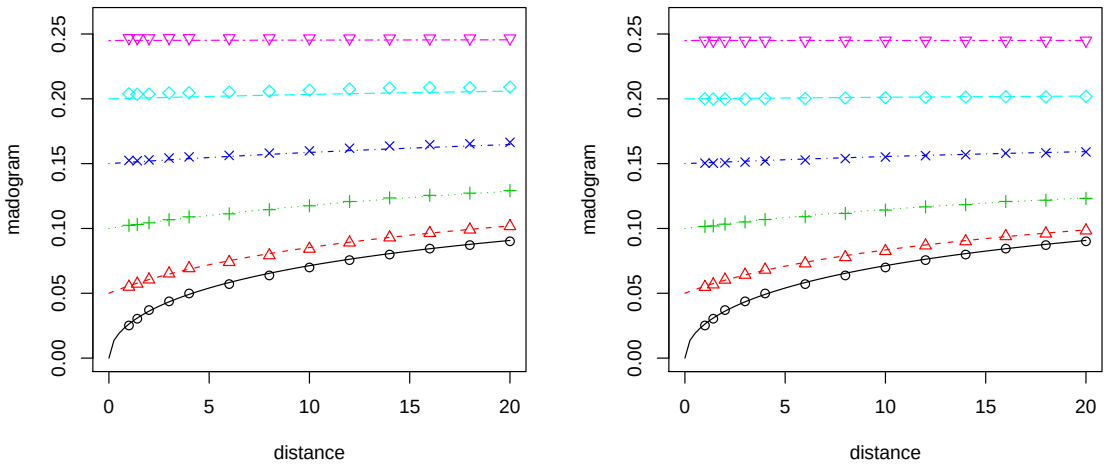

Figure 4.11: Points are the mean of the madogram estimates for 300 simulated fields. Lines indicate the theoretical value. Left plot shows the results for (from bottom) $\lambda = 1$, $\lambda = 2/3$, $\lambda = 3/7$, $\lambda = 1/4$, $\lambda = 1/9$, and $\lambda = 1/99$. Right plot shows the results for (from bottom) $\lambda = 1$, $\lambda = 3/2$, $\lambda = 7/3$, $\lambda = 4$, $\lambda = 9$, and $\lambda = 99$.

4.5 Conclusions

The madogram, ν is a dependence measure new to extremes In addition to introducing the madogram to the study of max-stable distributions, the normalized madogram was developed to allow for the Fréchet marginal case to be measured, and the definition of the extremal coefficient and the initial definition of the madogram were extended to provide measures of the complete dependence structure $\theta(\lambda)$ and $\nu(\lambda)$.

The madogram was compared to two earlier dependence measures: the extremal coefficient and the Pickands dependence function. Both measures have strengths and weaknesses. The extremal coefficient's value is the easiest to interpret, but it loses this advantage when extended to give complete dependence information. $\theta(\lambda)$ is a bit awkward because it takes on values in $[1, \infty)$ and its angular parameter λ takes on values in $[0, \infty)$. The Pickands dependence measure takes on more sensible values, $A(w) \in [1/2, 1]$ and its angular parameter w takes on values in $[0, 1]$. Furthermore, for bivariate EV random variables with symmetric dependence structure, $A(w) = A(1 - w)$. These properties make the dependence measure A the easiest to represent graphically. However, the definitions of θ, $\theta(\lambda)$, and A do not lend themselves to straightforward estimators.

The madogram and λ-madogram offer two advantages over the other measures. First, the definitions of the madogram and λ-madogram lead to straightforward estimators. The estimators for both the madogram and λ-madogram were tested in the bivariate and spatial cases with simulated data. These estimators seem to perform at least as well as existing estimators for $A(w)$. Second, the madogram is of a form familiar to spatial statisticians. The definition of the madogram and the madogram plots (such as in Figure 4.6) are similar to those of the variogram. The λ-madogram shares with $\theta(\lambda)$ the disadvantage of being parameterized by $\lambda \in [0, \infty)$. Another disadvantage is that the λ-madogram does not show any apparent symmetry in its parameter.

Fortunately, one does not have to choose just one measure of extremal dependence. It is easy to covert from one measure to another, and all differently measure the same thing. Therefore the practitioner can utilize the strengths of each measure to his or her advantage.

Chapter 5

An extreme precipitation return-levels map for Colorado's Front Range

5.1 Background

On July 31st 1976, west of Loveland, Colorado, a large precipitation event over a short period of time produced the Big Thompson Flood killing 145 people (Henz et al., 1976). On July 28th 1997, a storm which produced over 23 cm (9 inches) of rain in some areas caused extensive flood damage to the campus of Colorado State University in Fort Collins. The Fort Collins flood resulted in 5 deaths, 62 injuries and over $250 million in property damage (Weaver et al., 2000). Although these extreme precipitation events are rare, understanding their frequency and intensity is important for public safety and long term planning. Estimating the probability of extreme meteorological events is difficult because of limited temporal records and the need to extrapolate the distributions to locations where observations are not available. In this work we address the problem through the use of a Bayesian hierarchical model that leverages statistical extreme value theory. This approach can complement current methods used by the National Weather Service (NWS) to construct maps of precipitation extremes, but also has the advantage that one is able to quantify the uncertainty of these maps due to the limited amount of data and its sparse spatial representation. Here we report the results of a pilot study for extremes in 24 hour precipitation for the Front Range region of Colorado which encompasses the areas affected by the aforementioned events.

This statistical research benefits from a focus on hydrometeorological extremes, however we note that our spatial models are not limited to this context. Indeed, we expect spatial dependence of extremes to be important in many other areas this methodology can be adapted to other disciplines.

5.1.1 A precipitation atlas for Colorado's Front Range

An estimate of potential flooding is necessary for city and development planning, engineering, and risk assessment. To support this requirement, the NWS maintains precipitation atlases and a companion digital database [1] that are used as a primary resource for inferring the probability of an extreme at a particular

[1] See `hdsc.nws.noaa.gov/hdsc/pfds/pfds_maps.html`

location. Currently, the NWS is updating these precipitation maps for regions of the United States. At this time, atlases have been produced for the arid southwestern US (Bonnin et al., 2004a) and the mid-Atlantic states (Bonnin et al., 2004b). Neither of these regions encompasses Colorado, however, and part of our motivation was to consider a climatic region that has not yet been revised by the NWS's latest effort.

A common and relatively easy-to-understand measure of extreme events is the return level, and this is the measure furnished by the precipitation atlases. The r-year return level is the quantile which has probability $1/r$ of being exceeded in a particular year, and the most familiar example of a return level is hydrology's so-called "100-year flood". Precipitation return levels must be given in the context of the duration of the precipitation event; for example, the r-year return level of a d-hour (e.g. 6 or 24-hour) duration interval is reported. The standard levels for the NWS's most recent data products are quite extensive with duration intervals ranging from 5 minutes to 60 days and with return levels for 2 to 500 years. In this study we focus on providing return level estimates for daily precipitation (24 hours) but the methodology could be implemented to determine return levels for any duration period.

EVT provides the link between data recorded on a daily (or hourly) time frame and quantities of longer time scales such as return levels. A return level can be calculated as a function of the parameters of an extreme value distribution. Assume that the tail of the distribution is described by a $GPD(\tilde{\sigma}, \xi)$. From basic probability rules and Equation (1.4)

$$\mathbb{P}(X > x + u) = \zeta_u \left(1 + \xi \frac{x}{\tilde{\sigma}_u}\right)^{-1/\xi} \text{ with } \zeta_u = \mathbb{P}(X > u).$$

Letting n_y represent the number of observations taken in a year, one obtains the r-year return level z_r by solving the equation $\mathbb{P}(X > z_r) = \frac{1}{rn_y}$ for x_r:

$$z_r = u + \frac{\tilde{\sigma}_u}{\xi} \left[(rn_y\zeta_u)^{\xi} - 1\right]. \tag{5.1}$$

Although the return level has a closed form, it is a nonlinear function of the GPD parameters and the probability of exceedance.

A precipitation atlas was produced for the entire state of Colorado in 1973 by the NWS (Miller et al., 1973). Still in use today, the atlas provides point estimates of 2, 5, 10, 25, 50, and 100 year return levels for duration intervals of 6 and 24 hours. One shortcoming of this atlas is that it does not provide uncertainty measures of its point estimates even though one might expect significantly different levels of reliability between say 2 and 100 year return levels. Our method aims to produce a similar atlas which also provides measures of uncertainty. Additionally, we make use of statistical and computational techniques that have been developed since the previous atlas was produced, and we benefit from 30 years of additional data.

5.1.2 Modeling distributional dependence

Since we ultimately want to model return levels which are a climatological quantity, our focus is on how the *distribution* of precipitation varies over space, not the multivariate structure of particular precipitation events. Let $Z(\boldsymbol{x})$ denote the total precipitation for a given period of time (e.g. 24 hours) and at location $\boldsymbol{x}$. We want to model return levels, which are simply high quantiles for all locations $\boldsymbol{x}$. Because the distributions explicitly vary over space, the quantities derived from the distribution will also have a spatial dependence. Conceptually our approach is simple. Given the GPD approximation to the tail of a distribution, we add a spatial component by considering $\tilde{\sigma}_u$, ξ and ζ_u to be functions of space. We assume that the values of the parameters result from a latent spatial process which characterizes the extreme precipitation and which arises from climatological and orographic effects. The dependence of the parameters characterizes the similarity of climate at different locations.

Notationally, $\boldsymbol{x}$ denotes a location in a generic sense, and it takes on two meanings in this chapter. Traditionally, $\boldsymbol{x}$ represents a location in a space whose coordinates are given by longitude and latitude. In addition to working with space in the usual sense, we alternatively define a station's location in a "climate" space (which is also denoted by $\boldsymbol{x}$). The coordinates of $\boldsymbol{x}$ in the climate space are given by orographic and climatological measures. Each location in the longitude/latitude space corresponds to a location in the climate space, and there is an invertible transformation which takes points from one space to the other. Our reason for working in this alternative climate space is explained in Section 5.3.1, and it will be clear by the context what spatial coordinates are being used.

In a Bayesian setting, the functions $\tilde{\sigma}_u(\boldsymbol{x})$ $\xi(\boldsymbol{x})$ and $\zeta_u(\boldsymbol{x})$ are considered a priori as random processes and one of the main advantages is that a posteriori inferences for these surfaces provide natural measures of uncertainty. In order to understand the latent spatial process that drives the climatological dependence of precipitation extremes, we implement a Bayesian approach which integrates all the stations' data into one model. This pooling of data is especially important when studying extremes as these events are rare and the data record is relatively short for long return periods. Although there have been several studies using Bayesian methods in extremes (Coles and Tawn (1996a), Smith and Naylor (1987)), only a few have built models which borrowed strength across different spatial locations. Coles (2001a) proposed an information-sharing model which pooled wind speed data from different locations to better estimate the parameters for GEV distributions. This work, however, did not attempt to model any spatial nature of the extremes. Casson and Coles (1999) built a spatial model for the point process representation of exceedances of a threshold. The spirit of the model is quite similar to the one we propose, but the model was applied to a simulated, one dimensional process and no spatial interpolation was performed. Cooley et al. (2005a) built Bayesian a hierarchical GEV model which pooled lichenometry data from different locations, but the model was not fully spatial as it did not utilize location

information. To our knowledge, this is the first extremes study which employs Bayesian hierarchical models to study both the intensity and frequency of extremes, which models a two-dimensional latent spatial process, and which spatially interpolates the results.

5.1.3 Regional frequency analysis

Our Bayesian method is an alternative to regional frequency analysis (RFA). RFA originates from the index flood procedure of Dalrymple (1960) and has been extensively studied by Hosking and Wallis (1997). RFA is a three-step procedure to identify regional distributions for spatial extremes. First, RFA prescribes homogeneous spatial regions. Second, an index flood measure (often the mean or median of the annual maxima) is calculated for each station in the region and the annual maximum data are normalized by this measure. Third, a probability distribution (e.g. GEV, log-normal, or other) is chosen and fit to the pooled, dimensionless data and is then scaled by the stations' index flood measures to obtain distributions at the station locations. L-moments (a method-of-moments type estimator based on order statistics) are used for parameter estimation and criteria based on L-moments are suggested for both selecting homogeneous regions and choosing a probability distribution. Uncertainty measures for parameters estimated using L-moments are usually obtained via bootstrap methods. A drawback of L-moments is that they cannot readily incorporate covariates into the parameters (Katz et al., 2002).

RFA based on L-moments has been implemented in several precipitation studies. Schaefer (1990) used mean annual precipitation to construct homogeneous regions while performing an RFA analysis of Washington State. Zwiers and Kharin (1998) used L-moments to study precipitation data produced by global climate models to compare current climate to modeled climate under carbon-dioxide doubling. Two papers by Fowler (2003, 2005) studied regional precipitation in the UK via RFA. For its current precipitation atlas update project, the NWS employs RFA and relies on the PRISM method (Daly et al. (1994, 2002)) to spatially interpolate the mean annual maximum precipitation which is employed as the index flood measure. The Ft. Collins flood precipitated several studies, one of which used RFA methods to estimate the return period associated with the event (Sveinsson et al., 2002). Because Ft. Collins is situated at the foothills of the mountains, the authors found it difficult to prescribe a region which satisfactorily explained precipitation in the city.

The method we present in Section 5.3 differs in several ways from RFA-based studies above. Our general model has the flexibility to accommodate different covariate relationships, and models can be evaluated in a manner similar to that in a regression analysis. An explicit spatial model does not require the data to be normalized and instead allows changes in the parameters to account for differences in the data. Also, this approach avoids discontinuities due to regional boundaries that can arise in RFA. In contrast to the NWS team's spatial interpolation method, we adapt techniques from geostatistics. Instead of relying on L-moments

to obtain parameter estimates, we utilize Bayesian techniques and we obtain measures of uncertainty from the estimated posterior distributions. Finally, RFA is a three-step algorithm and it is difficult to assess the error propagation through the steps. In comparison, the sources of uncertainty in our approach are more easily tracked and all are taken into account when producing uncertainty maps.

5.1.4 Outline

The next section describes the precipitation data sources and some preliminary analysis of the choice of thresholds. In Section 5.3 we describe the models which produce the return level map. We discuss the GPD-based hierarchical model for threshold exceedances in Section 5.3.1, discuss the method for modeling the threshold exceedances rate in Section 5.3.2, briefly describe our MCMC method for model inference in Section 5.3.3, and discuss how our model was interpolated on the region in Section 5.3.4. We then present our method for model selection and results in Section 5.4. We conclude with a discussion in Section 5.5.

5.2 Data

5.2.1 Study region, weather stations, and covariates

Colorado's Front Range is a diverse geographic region where the Great Plains of North America meet the Rocky Mountains. The region we are interested in mapping lies between 104 and 106 W longitude and 37 and 41 N latitude (Figure 5.1). This area has peaks in excess of 4270m and extends more than 100km east of the mountains where the elevation is less than 1400m. We study only precipitation which occurs between April 1st and October 31st because virtually all of the floods for the Front Range region have occurred in these months. The region is semi-arid; the city of Denver receives approximately 29cm (11.5 inches) of precipitation during the months of April-October. Many of the extreme precipitation events during these months are localized convective cells. More than 75% of Colorado's population lives in the area we study, and the region is experiencing much growth and development. This study of extreme precipitation is part of a larger study of potential flooding for this region being done by the Institute for the Study of Society and the Environment at the National Center for Atmospheric Research.

Our data come from 56 weather stations scattered throughout the study region (Figure 5.1). These stations record hourly precipitation amounts, but for this study only the daily totals are utilized. We apply our method to data recorded during the years 1948-2001. Included in the 56 are stations which discontinued operation during the period of study and others which have come into existence. Twenty-one stations have over 50 years of data, fourteen have less than 20 years of data, and all stations have some missing values.

For the Front Range region, we anticipate that covariates could bear important information in describing the latent spatial process of the extremes. To interpolate over the study area and produce a map, we must have

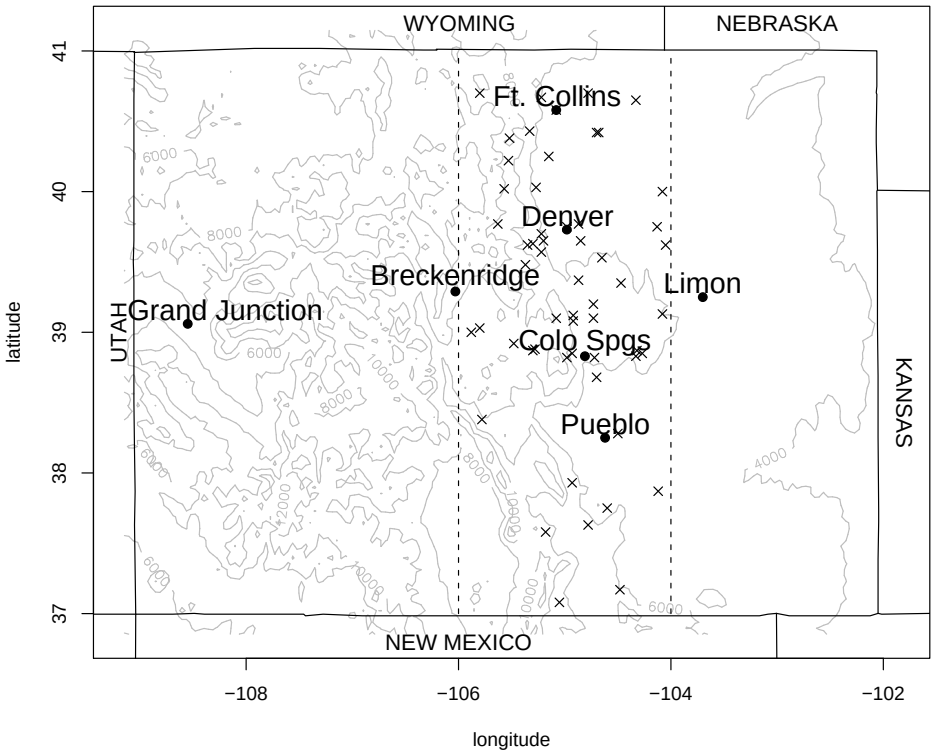

Figure 5.1: Map of Colorado shows the study location within the dashed lines. Contours show elevation which is given in meters. Station locations are marked with an "X".

covariate information for the entire region and not just at the weather station locations. We focus on two readily available covariates: elevation (Spatial Climate Analysis Service (1995)) and mean precipitation for the months April-October (MSP), which is itself an interpolated data product from the Spatial Climate Analysis Service (2004). For an area with both mountain and plains geographies, it is likely that elevation will have a significant influence on the climatological behavior of extreme precipitation. It is also likely that mean precipitation will be a strong covariate. For a spatial data set in England, Coles and Tawn (1996b) found that mean precipitation was a stronger covariate for extreme precipitation than elevation. Certainly the elevation varies more dramatically in the Front Range region than in the one studied in England, but since our mean precipitation data is highly correlated with the elevation data and takes into account other factors such as aspect, slope and meteorology, we expect that it may be the stronger of the two covariates.

5.2.2 Threshold selection

When working with a GPD model, one must select a threshold above which the data approximately follow a GPD distribution, and it can be chosen at either a particular level or quantile. Figure 5.2 shows a partial histogram of the data from the Boulder station. Its odd distribution is due to the fact that, prior to 1971, precipitation was recorded to the nearest 1/100th of an inch (.25 mm), and after 1971 to the nearest 1/10th of an inch (2.5 mm). Only three stations have complete records recorded to the nearest 1/100th of an inch. For daily precipitation, measurements recorded to the nearest 1/10th of an inch have relatively low precision. This low precision introduces a bias into the maximum likelihood estimates of the parameters (center and right plots of Figure 5.2) and affects how we can choose the threshold.

To understand how low precision affects the parameter estimates, a simple simulation experiment was performed. Random samples were drawn from a GPD distribution with similar parameters to the precipitation data and then rounded to the nearest 1/10th. Parameters were estimated for both the rounded and the original data, and the bias appeared to be the least when the threshold was set at the middle of the precision interval (not shown). Therefore, to minimize bias, we choose a threshold in the middle of a precision interval for all stations rather than setting the threshold at a certain quantile.

There are many diagnostics designed to determine a threshold above which the data are approximately GPD. These include mean residual life plots which should behave linearly for GPD data, and plots of the shape and rescaled scale parameter which should be constant as the threshold varies for GPD data (Coles, 2001b). Choosing a threshold is a delicate procedure as there is always a bias verses precision trade-off. As the GPD is derived asymptotically, the higher the threshold is set, the closer the data is to being GPD-distributed and consequently there is less bias in the parameter estimates. However with a high threshold, there is less data which results in less precise parameter estimates. For most of the stations in our study, the diagnostic plots (not shown) seem to indicate that a threshold of around u = .64cm (.25 inches) might be adequate. However, we choose to set the threshold at u = 1.1 cm (.45 inches) for all stations, preferring to reduce the bias and reasoning that since our model pools the data across stations, we should still obtain reasonably precise parameter estimates. With this threshold we had 10582 exceedances (after declustering), which represents 2.7% of the original data.

5.2.3 Declustering

Inherent in our model is an assumption of spatial and temporal conditional independence of the precipitation observations once the spatial dependence in the stations' parameters has been accounted for. This may not always be true, as the occurrence of an extreme event one day may influence the probability of an extreme occurrence the next day, and the spatial extent of an extreme event may not be limited to one station.

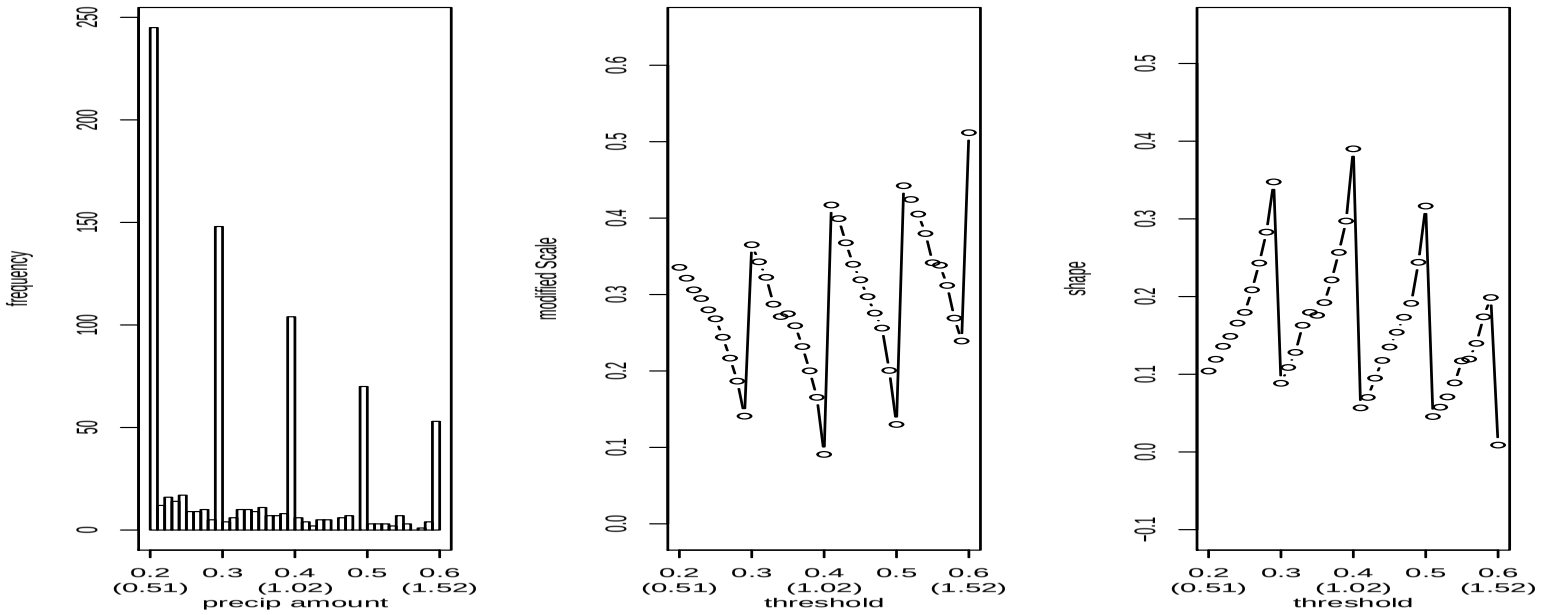

Figure 5.2: Shows the bias due to the lack of precision in the recorded data. The left plot shows a partial histogram of the Boulder station data which illustrates that most of the data is recorded to the nearest 1/10th of an inch. The center plot shows the mle parameter estimate for the modified scale parameter and the right shows the estimate for ξ. Both plots should be approximately constant as the threshold changes, but both show a bias depending on where the threshold is chosen within the precision interval. Precipitation amounts and thresholds are given in inches (cm).

Temporal dependence is a common issue with univariate extremes studies and it is dealt with via declustering. If a station had consecutive days which exceeded the threshold, we declustered the data by keeping only the highest measurement.

Spatial dependence of extremes is not as well understood as temporal dependence, and there is no declustering method for spatially dependent data in the extremes literature. Hosking and Wallis (1988) claim that any effects of spatial dependence between the observations are outweighed by the advantages of a regional (RFA) analysis. For now, we assume that there is no spatial dependence between the observations at different stations, conditional on the the stations' parameters (which are themselves spatially dependent). However, we test this assumption of spatial independence using the madogram in Section 5.4.

We have not accounted for any seasonal effects in our data. Restricting our analysis to the non-winter months reduces seasonality and inspecting the data from several sites showed no obvious seasonal effect. Likewise, we have not accounted for any temporal trends in the data. It has been suggested that anthropogenic climate change could cause regional precipitation extremes to become more severe (Trenberth (1999), Karl and Knight (1998)). However, our series are relatively short and it may be difficult to discern any trend in extreme precipitation over the last 50 years. Furthermore, the purpose of this study is not to focus on climate change. Both climate change and seasonal effects would be interesting extensions of this study.

5.3 Models

To produce the return levels map, we must model both the exceedances and their rate of occurrence. We construct separate hierarchical models for each and these are explained in the sections 5.3.1 and 5.3.2. By building a hierarchy, one can statistically model a complex process and its relationship to observations in several simple components. For an introduction to such models see Gelman et al. (2003).

There are three layers in both of our hierarchical models. The base layer models the data (either exceedance amounts or number of exceedances) at each station. The second layer models the latent process which drives the climatological extreme precipitation for the region. The third layer consists of the prior distributions of the parameters which control the latent process.

The latent process is our primary interest since return levels are climatological quantities. As the return level $z_r(\boldsymbol{x})$ is a function of $\tilde{\sigma}_u(\boldsymbol{x})$, $\xi(\boldsymbol{x})$, and $\zeta_u(\boldsymbol{x})$, we capture the latent process through these parameters. At the s station locations $x_1, \ldots, x_s$, we denote the values of the GPD scale parameter with $\tilde{\boldsymbol{\sigma}}_u = [\tilde{\sigma}_u(x_1), \ldots, \tilde{\sigma}_u(x_s)]^T$ and we similarly define $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}_u$. The two hierarchies model $\tilde{\boldsymbol{\sigma}}_u, \boldsymbol{\xi}, \boldsymbol{\zeta}_u$ and parameters which relate these processes to orographic and climatological information. The process $\tilde{\sigma}(\boldsymbol{x})$ given $\tilde{\boldsymbol{\sigma}}_u$ is assumed conditionally independent of the data, and likewise for $\xi(\boldsymbol{x})$ and $\zeta_u(\boldsymbol{x})$.

The inference for the parameters in our models $\boldsymbol{\theta}$ given the stations' data $\boldsymbol{Z}(\vec{\boldsymbol{x}})$ simply comes from Bayes'

rule:

$$p(\boldsymbol{\theta}|\boldsymbol{Z}(\vec{\boldsymbol{x}})) \propto p(\boldsymbol{Z}(\vec{\boldsymbol{x}})|\boldsymbol{\theta})\, p(\boldsymbol{\theta}), \tag{5.2}$$

where p denotes a probability density. Based on the conditional distributions of our hierarchical model Equation (5.2) becomes

$$p(\boldsymbol{\theta}|\boldsymbol{Z}(\vec{\boldsymbol{x}})) \propto p_1(\boldsymbol{Z}(\vec{\boldsymbol{x}})|\boldsymbol{\theta}_1)\, p_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)\, p_3(\boldsymbol{\theta}_2). \tag{5.3}$$

where p_j is the density associated with level j of the hierarchical model and depends on parameters $\boldsymbol{\theta}_j$.

5.3.1 Hierarchical model for threshold exceedances

5.3.1.1 Data layer

A GPD distribution forms the base level of our hierarchical model, as this takes advantage of the fact that we have daily data rather than just annual maxima. We reparameterize the GPD distribution, letting $\phi = \log \tilde{\sigma}_u$, which allows the parameter ϕ to take on both positive and negative values (to simplify notation, we drop the subscript u in the definition of ϕ). Let $Z_k(\boldsymbol{x}_i)$ be the kth recorded precipitation amount at location $\boldsymbol{x}_i$, with $i = 1, \ldots, s$. Given that $Z_k(\boldsymbol{x}_i)$ exceeds the threshold u, we assume that it is described by a GPD distribution whose parameters are dependent on the station's location. Letting $\phi(\boldsymbol{x}_i)$ and $\xi(\boldsymbol{x}_i)$ represent the parameters at the location $\boldsymbol{x}_i$ yields:

$$\mathbb{P}\{Z_k(\boldsymbol{x}_i) - u > z | Z_k(\boldsymbol{x}_i) > u\} = \left(1 + \frac{\xi(\boldsymbol{x}_i) z}{\exp \phi(\boldsymbol{x}_i)}\right)^{-1/\xi(\boldsymbol{x}_i)}. \tag{5.4}$$

Differentiating the distribution function associated with (5.4) to obtain a probability density we get the first piece of Equation (5.3):

$$p_1(\boldsymbol{Z}(\vec{\boldsymbol{x}})|\boldsymbol{\theta}_1) = \prod_{i=1}^{s} \prod_{k=1}^{n_i} \frac{1}{\exp \phi(\boldsymbol{x}_i)} \left(1 + \frac{\xi(\boldsymbol{x}_i) z}{\exp \phi(\boldsymbol{x}_i)}\right)^{-1/\xi(\boldsymbol{x}_i)-1}. \tag{5.5}$$

where $\boldsymbol{\theta}_1 = [\boldsymbol{\phi}, \boldsymbol{\xi}]^T$ and n_i is the number of exceedances for station located at $\boldsymbol{x}_i$.

5.3.1.2 Process layer

In the second layer of our hierarchy, we characterize the spatial latent processby constructing a structure which relates the parameters of the data layer to the orography and climatology of the region. The study area has many different subregions (e.g. plains, foothills, mountain valleys, mountain ranges) which are not necessarily contiguous. We expect the subregions to exhibit different extreme precipitation characteristics that might not be fully explained by simple functions of the covariates and employ spatial methods to capture these effects. However, we find that with only 56 stations in such a geographically diverse region, it is difficult to discern much of a spatial signal in the traditional longitude/latitude space which leads us to work in the alternative

climate space. The coordinates of each station in the two-dimensional climate space are given by its elevation and its MSP. Both dimensions are transformed so that the climate space is roughly isotropic. In the climate space, stations which have similar climate characteristics (e.g. stations along the foothills) are naturally grouped together even though their locations may be fairly distant in the traditional sense (Figure 5.3). As shown in Section 5.4, our models performed better in the climate space.

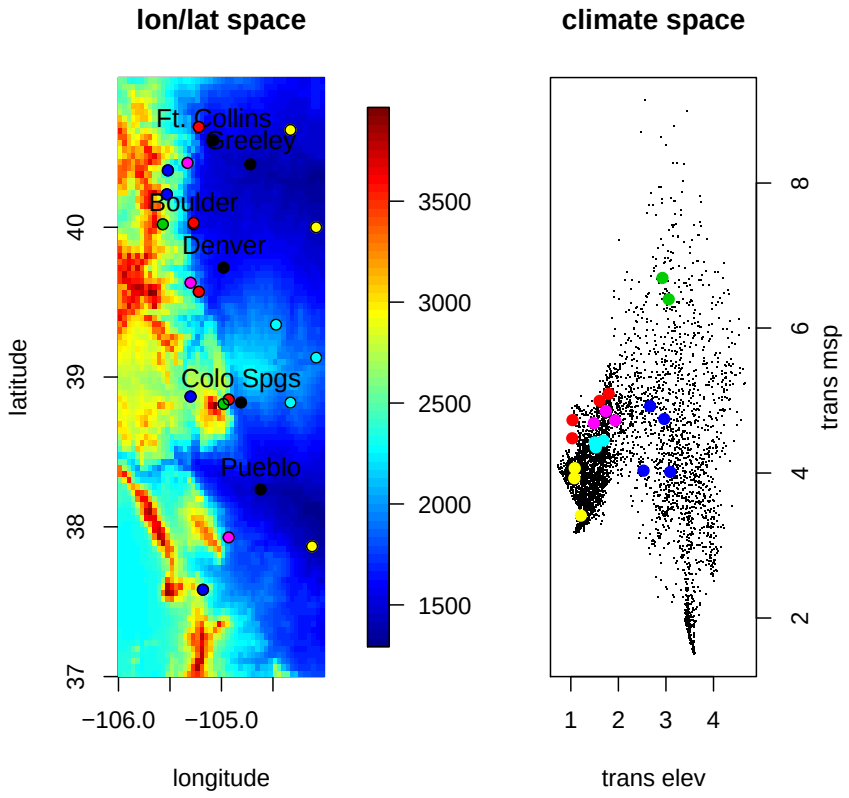

Figure 5.3: Translates points in longitude/latitude space to points in the climate space. Colored dots all correspond to selected station locations and represent foothills (red), plains (yellow), Palmer Divide (cyan), Front Range (magenta), mountain valley (blue), and high elevation (green).

Let $\phi(\boldsymbol{x})$ and $\xi(\boldsymbol{x})$ be the log-transformed scale and shape parameter processes for the region. Since we are in a Bayesian framework, we treat the parameters $\phi(\boldsymbol{x})$ and $\xi(\boldsymbol{x})$ as random variables and choose a prior distribution which allows us to model the latent spatial process. We put independent priors on $\phi(\boldsymbol{x})$ and $\xi(\boldsymbol{x})$.

We anticipate that the log-transformed scale parameter $\phi(\boldsymbol{x})$ will be sensitive to regional climate effects and build a model which describes its relationship with the latent spatial process. Drawing on standard geostatistical methods, we model $\phi(\boldsymbol{x})$ as a Gaussian process with $\mathbb{E}[\phi(\boldsymbol{x})] = \mu_\phi(\boldsymbol{x})$ and $Cov\left(\phi(\boldsymbol{x}), \phi(\boldsymbol{x}')\right) = k_\phi(\boldsymbol{x}, \boldsymbol{x}')$.

The mean $\mu_\phi(x)$ is a function of parameters $\boldsymbol{\alpha}_\phi$ and the covariates:

$$\mu_\phi(\boldsymbol{x}) = f_\phi(\boldsymbol{\alpha}_\phi, covariates). \tag{5.6}$$

The function f is changed easily to allow different relationships with the covariates and an example of one of the models tested is: $\boldsymbol{\mu}_\phi(x) = \alpha_{\phi,0} + \alpha_{\phi,1} \times (elevation)$. Covariance is a function of the distance (in the space, traditional or climate, in which the analysis is performed) between stations and parameters $\boldsymbol{\beta}_\phi$ and it is given by:

$$k_\phi(\boldsymbol{x}, \boldsymbol{x}') = \beta_{\phi,0} \times \exp(-\beta_{\phi,1} \times ||\boldsymbol{x} - \boldsymbol{x}'||) \tag{5.7}$$

This corresponds to exponential variogram models with no nugget effect. We choose to work with exponential models because of their simplicity, but they could be easily exchanged for other covariance models. We also tested an exponential model with a nugget effect, but with our data, it was difficult to discern any nugget effect and the fits of the two models were comparable.

In contrast to the transformed scale parameter, we are less certain of the shape parameter's sensitivity to regional variables. Because the shape parameter is more difficult to estimate than the scale parameter, we start to model $\xi(\boldsymbol{x})$ as a single value and increasingly add complexity until we have a reasonable fit. We model the shape parameter in three ways: (1) as a single value for the entire study region with a $Unif(-\infty, \infty)$ prior, (2) as two values, one for the mountain stations and one for the plains stations each with $Unif(-\infty, \infty)$ priors, and (3) as a Gaussian process with structure similar to that of the prior for $\phi(\boldsymbol{x})$.

Modeling the GPD parameters $\phi(\boldsymbol{x})$ and $\xi(\boldsymbol{x})$ as above, data at the station locations provide information about the latent spatial process that characterizes these parameters. Hence, the second piece in Equation (5.3) is

$$p_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) = \frac{1}{\sqrt{(2\pi)^s|\Sigma|}} \exp\left[-\frac{1}{2}(\boldsymbol{\phi} - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{\phi} - \boldsymbol{\mu})\right] \times p_\xi(\boldsymbol{\xi}|\boldsymbol{\theta}_\xi) \tag{5.8}$$

where $\boldsymbol{\mu}$ is a vector defined by Equation (5.6) evaluated at the covariates of the locations $\boldsymbol{x}_i$, Σ is the covariance matrix generated by Equation (5.7) at the station locations (in either the traditional or climate space), the density function p_ξ comes from the prior distribution we choose for the shape parameter $\boldsymbol{\xi}$ with parameters $\boldsymbol{\theta}_\xi$, and $\boldsymbol{\theta}_2 = [\boldsymbol{\alpha}_\phi, \boldsymbol{\beta}_\phi, \boldsymbol{\theta}_\xi]^T$.

5.3.1.3 Priors

In the third layer of our hierarchy, we assign priors to the parameters $\boldsymbol{\alpha}_\phi$, $\boldsymbol{\beta}_\phi$, and $\boldsymbol{\theta}_\xi$ which characterize the latent process. We assume each parameter in this layer is independent of the others. Because of the MCMC method we use to obtain model inference (Section 5.3.3), we are able to choose improper and uninformative priors for the regression parameters $\boldsymbol{\alpha}_\phi$. For all the models tested, $\alpha_{\phi,i} \sim Unif(-\infty, \infty)$.

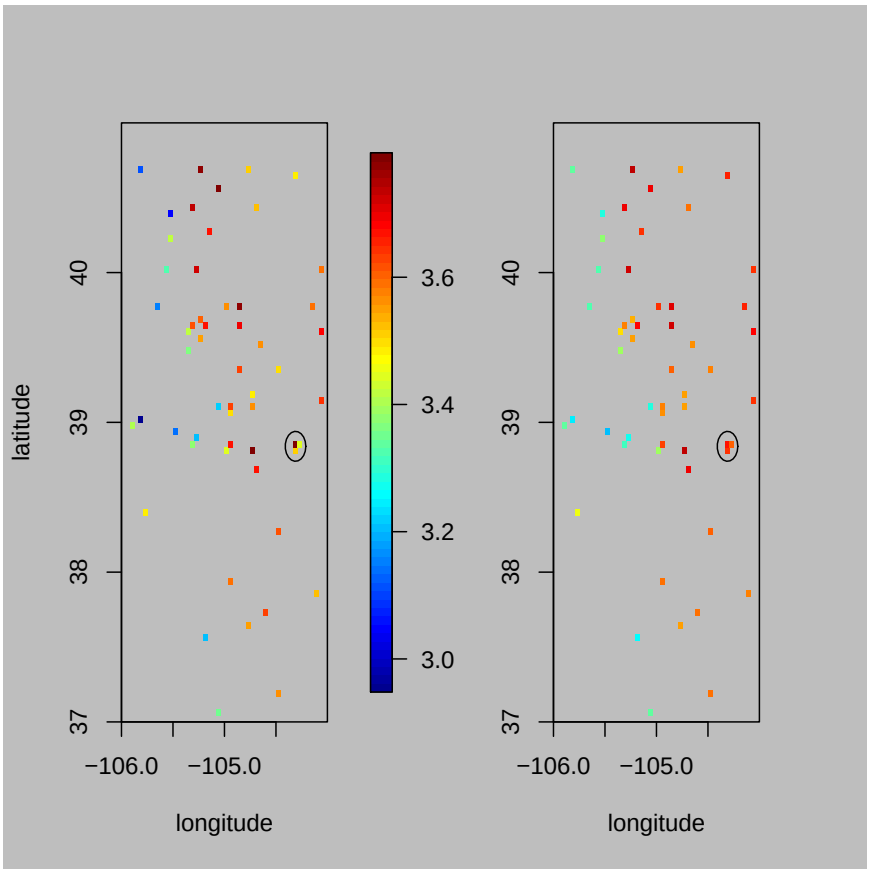

Figure 5.4: Illustrates the effect of pooling the data in the Bayesian model. On the left are the maximum likelihood estimates of ϕ when fit independently (with a common shape parameter ξ). On the right are the posterior means from the Bayesian model. The circle denotes an area where the station has been moved twice during the study period. The mle estimates treat these stations independently and obtains disparate estimates because of the short data record of each station. The Bayesian model pools the data and obtains a common result for the region.

The parameters $\boldsymbol{\beta}_\phi$ describe the spatial structure of the transformed scale parameter of the GPD distribution, a quantity for which it is difficult to elicit prior information. We therefore used empirical information to construct priors for these spatial parameters. When we modeled in the traditional longitude/latitude space, we were forced to use priors that were quite informative to obtain convergence. However, when we modeled in the climate space, we were able to relax the information in the priors using uniform priors on a range deemed reasonable based on empirical information. We choose a prior of $Unif(0.005, 0.03)$ for $\beta_{\phi,0}$ and a prior of $Unif(1, 6)$ for $\beta_{\phi,1}$.

For the shape parameter $\xi(\boldsymbol{x})$, only when modeled as a multivariate normal are there parameters which must be assigned priors in level three. In this case, $\xi(\boldsymbol{x})$ has regression coefficients $\boldsymbol{\alpha}_\xi$ and spatial parameters $\boldsymbol{\beta}_\xi$. As with $\boldsymbol{\phi}$, we use $Unif(-\infty, \infty)$ for the priors on $\boldsymbol{\alpha}_\xi$ and use empirical information to determine appropriate priors for $\boldsymbol{\beta}_\xi$. In the climate space, the prior for $\beta_{\xi,0}$ was $Unif(.001, .020)$ and the prior for $\beta_{\xi,1}$ was $Unif(1, 6)$.

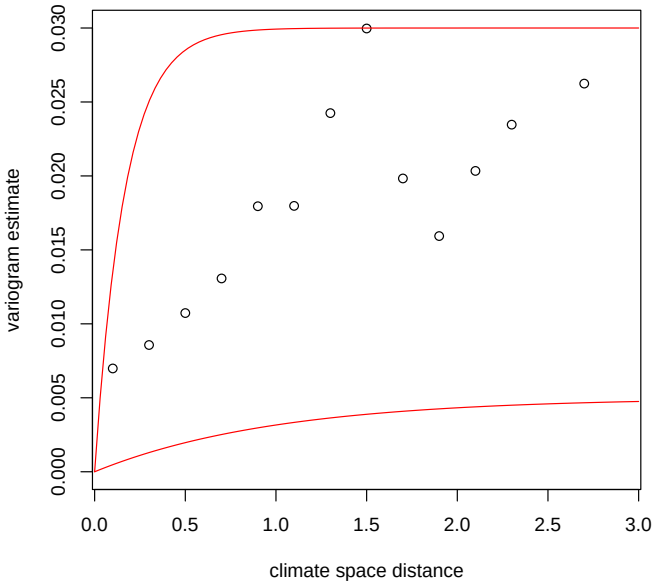

Figure 5.5: The variogram of the mle estimates for the ϕ parameter (points). This plot was used to define reasonable priors for $\vec{\beta}\phi$. Also plotted are the bounds on the spatial structure imposed by the priors (lines).

With the priors set as above, the third piece of Equation 5.3 is:

$$p_3(\boldsymbol{\theta}_2) = p_{\alpha_\phi}(\boldsymbol{\alpha}_\phi) \times p_{\beta_\phi}(\boldsymbol{\beta}_\phi) \times p_{\alpha_\xi}(\boldsymbol{\alpha}_\xi) \times p_{\beta_\xi}(\boldsymbol{\beta}_\xi) \propto 1 \times p_{\beta_\phi}(\boldsymbol{\beta}_\phi) \times 1 \; p_{\beta_\xi}(\boldsymbol{\beta}_\xi), \tag{5.9}$$

and the model for threshold exceedances is completely specified.

5.3.2 Exceedance rate model

To estimate return levels, we not only need to estimate the GPD parameters, but also must estimate ζ_u the rate at which a cluster of observations exceed the threshold u. Because we have temporally declustered our data, rather than being the probability that an observation exceeds the threshold, ζ_u is actually the probability that an observation is a cluster maximum. However, we will continue to refer to ζ_u as the exceedance rate parameter.

Since we chose the threshold to be 1.1 cm (.45 inches) for all stations, the exceedance rate parameter $\zeta_{.45}$ (henceforth ζ) will differ at each station, and must be modeled spatially. Had we been able to set our threshold at a certain quantile for all the stations, then modeling exceedance rates would be quite easy, as it would be the same for all stations. We let $\zeta(\boldsymbol{x})$ be the exceedance rate parameter for the location $\boldsymbol{x}$. As with the GPD parameters, we assume there is a latent spatial process which drives the exceedance probability.

Our model to obtain inference about $\zeta(\boldsymbol{x})$ is a hierarchical model, again with data, process and prior layers.

At the data layer of this model, we assume that each station's number of declustered threshold exceedances N_i is a binomial random variable with m_i (total number of observations) trials each with a probability of $\zeta(\boldsymbol{x}_i)$ of being a cluster maximum.

The process layer of our hierarchy is quite similar to that of the GPD parameter $\phi(\boldsymbol{x})$. We follow Diggle's et al. (1998) methodology and transform $\zeta(\boldsymbol{x})$ using a logit transformation and then model the transformed parameters as a Gaussian process with $\mathbb{E}[\zeta(\boldsymbol{x})] = \mu_\zeta(\boldsymbol{x})$ and $Cov\left(\zeta(\boldsymbol{x}), \zeta(\boldsymbol{x}')\right) = k_\zeta(\boldsymbol{x}, \boldsymbol{x}')$. As with $\phi(\boldsymbol{x})$, the mean $\mu_\zeta(\boldsymbol{x})$ is a function of parameters $\boldsymbol{\alpha}_\zeta$ and the covariate information, and the covariance function is based on an exponential variogram model with parameters $\boldsymbol{\beta}_\zeta$.

The prior layer of the hierarchy consists of the priors for these parameters, and as before we put non-informative $Unif(-\infty, \infty)$ priors on the regression coefficients $\boldsymbol{\alpha}_\zeta$ and use empirical information to choose priors on the spatial parameters $\boldsymbol{\beta}_\zeta$. In the climate space, the prior for $\beta_{\zeta,0}$ is $Unif(.005, .02)$ and the prior for $\beta_{\zeta,1}$ is $Unif(1, 6)$.

The model for $\zeta(\boldsymbol{x})$ is taken to be independent of the model for the GPD parameters, but, of course, these two independent models must be combined to obtain return level estimates. The overall structure to derive return levels is illustrated in Figure 5.6.

5.3.3 MCMC structure

As is often the case with complicated Bayesian models, we obtain approximate draws from the posterior distribution via an MCMC algorithm. For background on MCMC methods see Robert and Casella (1999) and for a reference on Bayesian inference via MCMC see Gelman et al. (2003). For both the exceedances model and the exceedance rate model, we employ a Metropolis Hastings (MH) steps within a Gibbs sampler to update each parameter of the model.

We illustrate our method within the context of the exceedances model. When applying the Gibbs sampler, we partition the sampling for $\boldsymbol{\phi}$ and $\boldsymbol{\xi}$. The update of $\boldsymbol{\phi}$ uses an MH step, drawing a value from a candidate density and then accepting or rejecting it with the appropriate rate. To speed up convergence, when updating $\boldsymbol{\phi}$, we use information from the maximum likelihood estimates of the GPD parameters at each of the stations to obtain a suitable candidate density. Let $\hat{\boldsymbol{\phi}}$ represent the maximum likelihood estimates (mle's) for the GPD parameter $\boldsymbol{\phi}$. From the asymptotic properties of mle's,

$$\hat{\boldsymbol{\phi}} \approx \boldsymbol{\phi} + MVN(0, \mathcal{I}^{-1}),$$

where $\mathcal{I}$ is the Fisher information matrix. Given the prior distribution $\boldsymbol{\phi} \sim MVN(\mu, \Sigma)$ where μ and Σ are

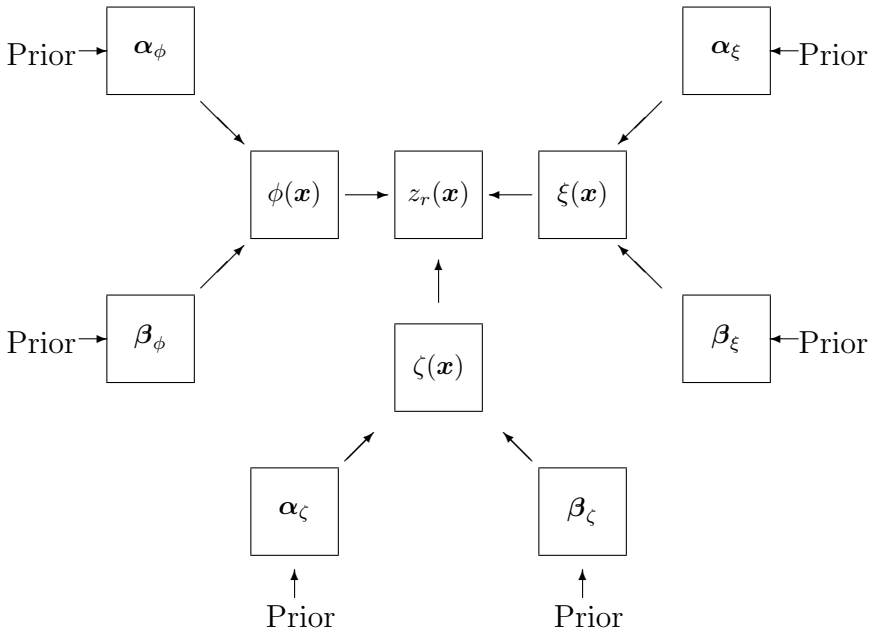

Figure 5.6: Schematic of the model used to estimate the return level $z_r(\boldsymbol{x})$. The return level is a function of the GPD parameters $\phi(x)$ and $\xi(x)$, and of the exceedance rate parameter $\zeta(x)$. All of these can be modeled spatially by a Gaussian process where the parameters $\alpha.$ describes the mean structure, and $\beta.$ describes the covariance structure.

defined as in equations (5.6) and (5.7). We can then write the joint distribution of $\hat{\boldsymbol{\phi}}$ and $\boldsymbol{\phi}$:

$$\begin{pmatrix} \hat{\boldsymbol{\phi}} \\ \boldsymbol{\phi} \end{pmatrix} = MVN\left(\begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \end{pmatrix}, \begin{bmatrix} \Sigma + \mathcal{I}^{-1} & \Sigma \\ \Sigma & \Sigma \end{bmatrix}\right).$$

We use the conditional distribution

$$\boldsymbol{\phi}|\hat{\boldsymbol{\phi}} \sim MVN\left(\boldsymbol{\mu} + \Sigma(\mathcal{I}^{-1} + \Sigma)^{-1}(\hat{\boldsymbol{\phi}} - \boldsymbol{\mu}), \Sigma - \Sigma(\mathcal{I}^{-1} + \Sigma)^{-1}\Sigma\right)$$

as the candidate density in our MH step. The intuition behind this choice is that for large samples, the mle inference will be close to the Bayesian posterior. Hence, the sampling distribution for the mle should provide a good candidate distribution for this part of the posterior. The mle approach to generate good candidate distributions substantially improved the acceptance rate of the MH steps.

After the parameter $\boldsymbol{\phi}$, we then update the mean and covariance parameters $\boldsymbol{\alpha}_\phi$ and $\boldsymbol{\beta}_\phi$. The MH candidate densities of $\boldsymbol{\alpha}_\phi$ and $\boldsymbol{\beta}_\phi$ are implemented as random walks.

For the shape parameter, we repeat the process, updating $\boldsymbol{\xi}$ and any second-level parameters which correspond to it. The exceedance rate model is handled analogously.

We ran several parallel chains for each model. Each simulation consisted of 20000 iterations, the first 2000 iterations were considered to be burn-in time. Of the remaining iterations, every 10th iteration was kept to reduce dependence. We used the criterion $\hat{R}$ as suggested by Gelman (1996) to test for convergence and assume that values below the suggested critical value of 1.1 implies convergence. For all parameters of all models the value of $\hat{R}$ is below 1.05 unless otherwise noted in Section 5.4.

5.3.4 Spatial interpolation and inference

Our goal is to estimate the posterior distribution for the return level for every location in the study region. From Equation (5.1) $z_r(\boldsymbol{x})$ is a function of $\phi(\boldsymbol{x})$, $\xi(\boldsymbol{x})$, and $\zeta(\boldsymbol{x})$, thus it is sufficient to estimate the posteriors of these processes. Our method allows us to draw samples from these distributions which in turn can be used to produce draws from $z_r(\boldsymbol{x})$.

To illustrate our interpolation method, consider the log-transformed GPD scale parameter of the exceedances model. We begin with values for $\boldsymbol{\phi}, \boldsymbol{\alpha}_\phi,$ and $\boldsymbol{\beta}_\phi$ from which we need to interpolate the value of $\phi(x)$. We have assumed that the parameters $\boldsymbol{\alpha}_\phi$ and $\boldsymbol{\beta}_\phi$ respectively determine the mean and covariance structure of the Gaussian process for $\phi(\boldsymbol{x})$. Using the values of $\boldsymbol{\alpha}_\phi$ and $\boldsymbol{\beta}_\phi$, we are able to draw from the conditional distribution for $\phi(\boldsymbol{x})$ given the current values of $\boldsymbol{\phi}$ in each iteration. Doing this for each iteration of the MCMC algorithm provides draws from the posterior distribution of $\phi(\boldsymbol{x})$.

We do the same for the exceedance rate parameter $\zeta(\boldsymbol{x})$, and for the GPD shape parameter $\xi(\boldsymbol{x})$ if it is modeled spatially. Pointwise means are used as point estimates for each of the parameters (figures 5.8, 5.9). The entire collection of draws from the posterior distributions of $\phi(\boldsymbol{x})$, $\xi(\boldsymbol{x})$, and $\zeta(\boldsymbol{x})$ are used to produce draws from the return-level posterior distribution. The pointwise quantiles and pointwise means of the posterior draws are used for the return-level maps (figures 5.11, 5.10).

5.4 Model selection and results

As in a regression study, we test both the threshold exceedance and the exceedance rate models with different covariates. To assess model quality, we use the deviance information criterion (DIC) (Spiegelhalter et al., 2002) as a guide. The DIC produces a measure of model fit $\bar{D}$ and a measure of model complexity p_D and sums them to get an overall score. A lower score implies a better model. However, we do not solely rely on the DIC to choose the most appropriate model. Because our project is product-oriented (i.e. we want to produce a map), we also considered the statistical and climatological characteristics of each model's map, as well as their uncertainty measures.

We first discuss the model for threshold exceedances. Table 5.1 shows the models tested and their corresponding DIC scores. We begin developing models in the traditional latitude/longitude space and start with simple models where $\phi(\boldsymbol{x})$ is modeled as in Section 5.3.1 and $\xi(\boldsymbol{x})$ is modeled as a single value throughout the region. We allow the mean of the scale parameter to be a linear function of elevation and/or MSP (Models 2,3,4). To our surprise, we find that elevation outperforms MSP as a covariate and in fact adding MSP does not improve the model over including elevation alone. Unfortunately, the maps produced by these simple models in the traditional space seem to inadequately describe the extreme precipitation. For example, the point estimate maps for $\phi(\boldsymbol{x})$ show relatively high values around the Northern Front Range cities of Boulder and Ft. Collins but do not show similar values for the stationless region between the cities despite that it has a similar climate and geography. Furthermore, our posterior estimates for our spatial parameters $\boldsymbol{\beta}$ are highly influenced by the prior we choose and the $\hat{R}$ criterion is consistently above 1.1 for our range parameter β_2.

However when we turn to the climate space, we obtain much better results. Both the model fit score and the effective number of parameters are lower in the climate space yielding lower DIC scores for comparable models. Our estimates for $\boldsymbol{\beta}_\phi$ are now clearly picking up a spatial signal in the climate space (Figure 5.7) and the values of $\hat{R}$ for the range parameter β_2 are now below the critical value 1.1 (about 1.08 for all climate-space models). Most importantly, when the points are translated back to the original space, we obtain parameter estimate maps which seem to agree with the geography. We then begin to add complexity to the shape parameter $\xi(\boldsymbol{x})$. Allowing the mountain stations and plains stations to have separate values of the shape parameter slightly improves model fit (Model 7), but a full spatial model for $\xi(\boldsymbol{x})$ does not improve model fit enough to warrant the added complexity (Model 8). As before, adding MSP does not significantly improve the model (Model 9). Model 7 is chosen as the most appropriate, but its differences from models 5 and 6 are slight.

Although the point estimate map for $\phi(\boldsymbol{x})$ uses elevation as its only covariate, there are several features that are not explained by elevation alone (Figure 5.8). There is a Front Range effect which leads to relatively high values just east of the mountains between the cities of Fort Collins and Boulder and west of the cities of Denver and Colorado Springs. There is an east-west area of higher $\phi(\boldsymbol{x})$ values north (and to a lesser extent south) of the Palmer Divide, an area of higher elevation between the cities of Colorado Springs and Denver.

Our analysis for the exceedance rate model follows a similar methodology where we start with simple models and add layers of complexity. One difference the exceedance rate model has from the threshold exceedance model is that the strongest covariate is now MSP rather than elevation (Table 5.2). It is perhaps not surprising that a higher mean precipitation corresponds to a higher probability an observation exceeds the threshold. Another difference we find is that the model fit as measured by the DIC is not very different between the traditional space and the climate space. Unlike the GPD model, the $\hat{R}$ value for the range parameter is below the critical value for models in both spaces (around 1.07 for both), and in fact, the model with the lowest

Baseline Model			$\bar{D}$	p_D	DIC
Model 0:	$\phi =$	ϕ	112264.2	2.0	112266.2
	$\xi =$	ξ			
Models in Latitude/Longitude Space			$\bar{D}$	p_D	DIC
Model 1:	$\phi =$	$\phi + \epsilon_\phi$	98533.2	33.8	98567.0
	$\xi =$	ξ			
Model 2:	$\phi =$	$\alpha_0 + \alpha_1(\text{msp}) + \epsilon_\phi$	98532.3	33.8	98566.1
	$\xi =$	ξ			
Model 3:	$\phi =$	$\alpha_0 + \alpha_1(\text{elev}) + \epsilon_\phi$	98528.8	30.4	98559.2
	$\xi =$	ξ			
Model 4:	$\phi =$	$\alpha_0 + \alpha_1(\text{elev}) + \alpha_2(\text{msp}) + \epsilon_\phi$	98529.7	29.6	98559.6
	$\xi =$	ξ			
Models in Climate Space			$\bar{D}$	p_D	DIC
Model 5:	$\phi =$	$\phi + \epsilon_\phi$	98524.3	27.3	98551.6
	$\xi =$	ξ			
Model 6:	$\phi =$	$\alpha_0 + \alpha_1(\text{elev}) + \epsilon_\phi$	98526.0	25.8	98551.8
	$\xi =$	ξ			
Model 7:	$\phi =$	$\alpha_0 + \alpha_1(\text{elev}) + \epsilon_\phi$	**98524.0**	**26.0**	**98550.0**
	$\xi =$	ξ_{mtn}, ξ_{plains}			
Model 8:	$\phi =$	$\alpha_0 + \alpha_1(\text{elev}) + \epsilon_\phi$	98518.5	79.9	98598.4
	$\xi =$	$\xi + \epsilon_\xi$			
Model 9:	$\phi =$	$\alpha_0 + \alpha_1(\text{elev}) + \alpha_2(\text{msp}) + \epsilon_\phi$	98523.6	27.3	98550.9
	$\xi =$	ξ_{mtn}, ξ_{plains}			

$\epsilon_\cdot \sim MVN(0, \Sigma)$ where $[\sigma]_{i,j} = \beta_{\cdot,0} \exp(-\beta_{\cdot,1} ||\boldsymbol{x}_i - \boldsymbol{x}_j||)$

Table 5.1: Shows several of the different GPD hierarchical models tested and corresponding DIC scores. Corresponding models in the climate space had better scores than in the longitude/latitude space (e.g. Models 3 and 6). The scores for Models 5, 6, and 7 are very close, and Model 7 was chosen as best. Model 8 shows that modeling ξ spatially improves the apparent fit, but makes the model much more complex resulting in a higher DIC score.

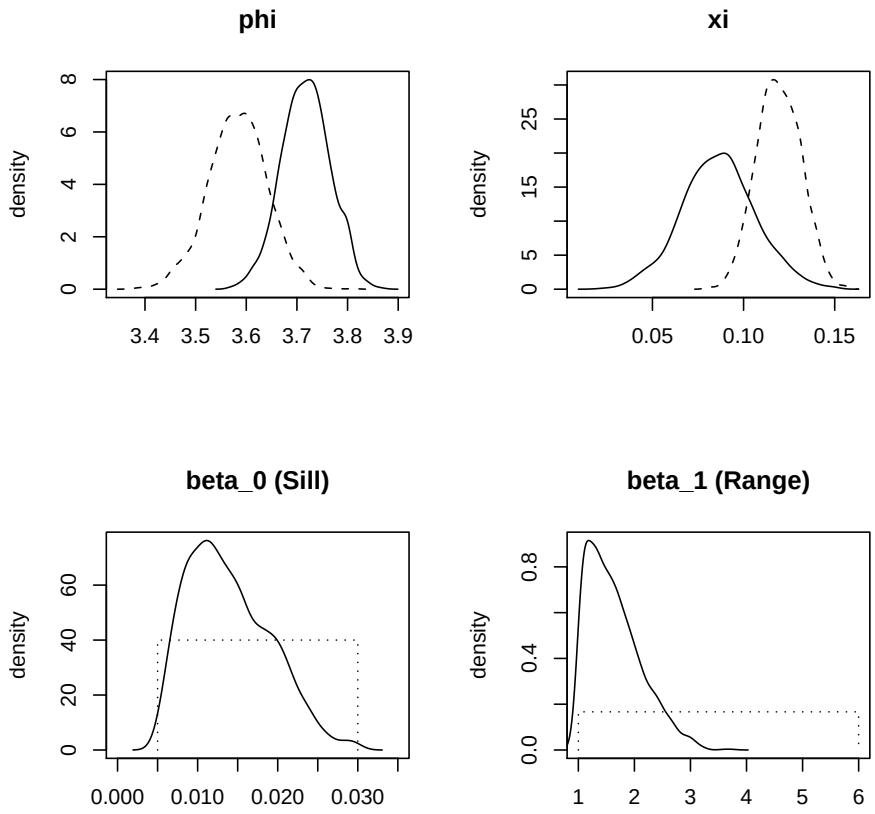

Figure 5.7: Posterior densities from Model 7. The upper left shows ϕ parameters associated with the Boulder station which has data for 54 years (solid) compared to the wider density of the Greeley/UNC station which has data for 16 years (dashed). Upper right shows the parameters ξ for the mountain stations (solid) and the plains stations (dashed), indating heavier tail behavior on the plains. Lower left shows β_0 which corresponds to the sill of the variogram (solid) and its prior distribution (dashed). Lower right shows β_1 which corresponds to the inverse range parameter in the climate space (lower parameter value indicates longer range).

DIC score uses the longitude/latitude space. However, to be consistent with the threshold exceedances model and since the differences in the DIC scores are small we select Model C (in the climate space) to produce our return levels map. The exceedance rate point estimate map (Figure 5.9) shows low exceedance rates around Greeley and east of Pueblo, and high exceedance rates at areas of very high elevation where there are no stations and the model is forced to extrapolate.

Of course, field practitioners are not interested in parameter estimate maps, as they want to know what is the expected return level. Return level maps are produced by combining the results of the exceedances and exceedance rates model. It is interesting to compare the maps produced by defining locations in the traditional space and in the climate space (Figure 5.11). The map from the traditional space shows higher return levels around the cities of Fort Collins and Boulder, but lower return levels for the region between the cities which has

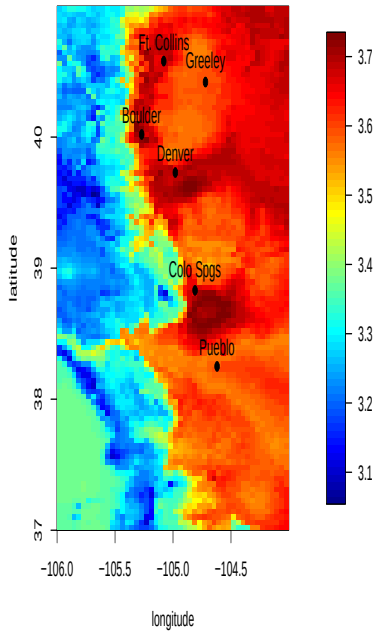

Figure 5.8: Point estimate for the log-transformed GPD scale parameter ϕ for Model 7.

similar climate and orographic characteristics. Because weather stations are in the cities, the model characterizes the extreme precipitation well at these locations, however it regresses to the mean away from locations that it has data. On the other hand, the climate space recognizes that the area between the two cities is close (climatologically) to these cities and shows a more realistic representation of the return levels. Other orographic features (such as the Palmer Divide area between Colorado Springs and Denver) also appear to have an effect on potential extreme precipitation.

Our lower and upper bound estimates were calculated by taking the pointwise .025 and .975 empirical quantiles from the return level draws, and the uncertainty range map is simply the difference of the two (Figure 5.10).

Recall that our GPD model assumes that the data are temporally independent and spatially independent conditional on the stations' GPD parameters. We accounted for the temporal dependence in the data by declustering; however, we did not account for any residual spatial dependence in the observations. We use the madogram to test how much spatial dependence remains in the data. Using the GPD model parameter point estimates, each station's data can be transformed to have a standard Frećhet distribution. Since the madogram has so far only been studied in the case of max-stable random fields, we analyze only the spatial dependence in the annual maxima which should be approximately max-stable. The madogram is applied to the normalized

Models in Latitude/Longitude Space			D	p_D	DIC
Model A:	ζ	$\alpha_0 + \alpha_1(\text{msp}) + \epsilon_\zeta$	458.3	54.0	512.3
Model B:	ζ	$\alpha_0 + \alpha_1(\text{msp}) + \alpha_2(\text{elev}) + \epsilon_\zeta$	456.7	53.1	509.8
Models in Climate Space			$\bar{D}$	p_D	DIC
Model C:	ζ	$\alpha_0 + \alpha_1(\text{msp}) + \epsilon_\zeta$	**460.4**	**51.4**	**511.8**
Model D:	ζ	$\alpha_0 + \alpha_1(\text{msp}) + \alpha_2(\text{elev}) + \epsilon_\zeta$	460.2	51.4	511.6

$\epsilon_\cdot \sim MVN(0, \Sigma)$ where $\sigma_{i,j} = \beta_{\cdot,0} \exp(-\beta_{\cdot,1}||\boldsymbol{x}_i - x_j||)$

Table 5.2: Shows exceedance rate hierarchical models tested and their DIC scores. Models in the longitude/latitude space had very similar DIC scores to models in the climate space. However, differences were slight and model C in the climate space was chosen for consistency with the exceedance model.

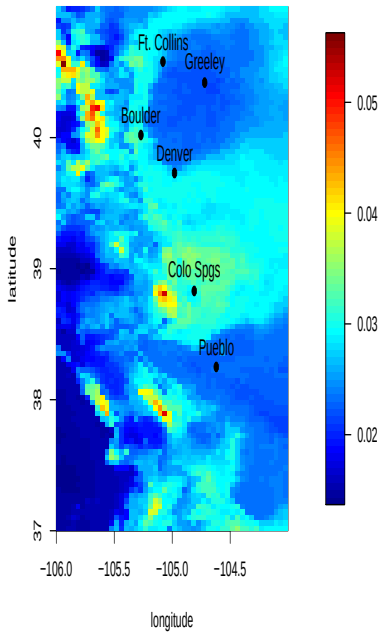

Figure 5.9: Point estimate for the exceedance rate parameter ζ for Model C.

annual max data (Figure 5.12), and shows that beyond a distance of 24 km (15 miles), the annual maxima are effectively independent. Since there are very few stations within this distance which record data for the same time period, it would seem to indicate that any spatial dependence in the observations not accounted for in the latent process is of little consequence. However, our model analyzes threshold exceedance data, not annual maximum data, and, after modeling spatial dependence within the parameter space, how much dependence remains in this data is unanswered.

The point estimate and uncertainty interval maps for the return levels exhibit very interesting information. The point estimate map shows a significant difference between the plains and mountains which arises from both

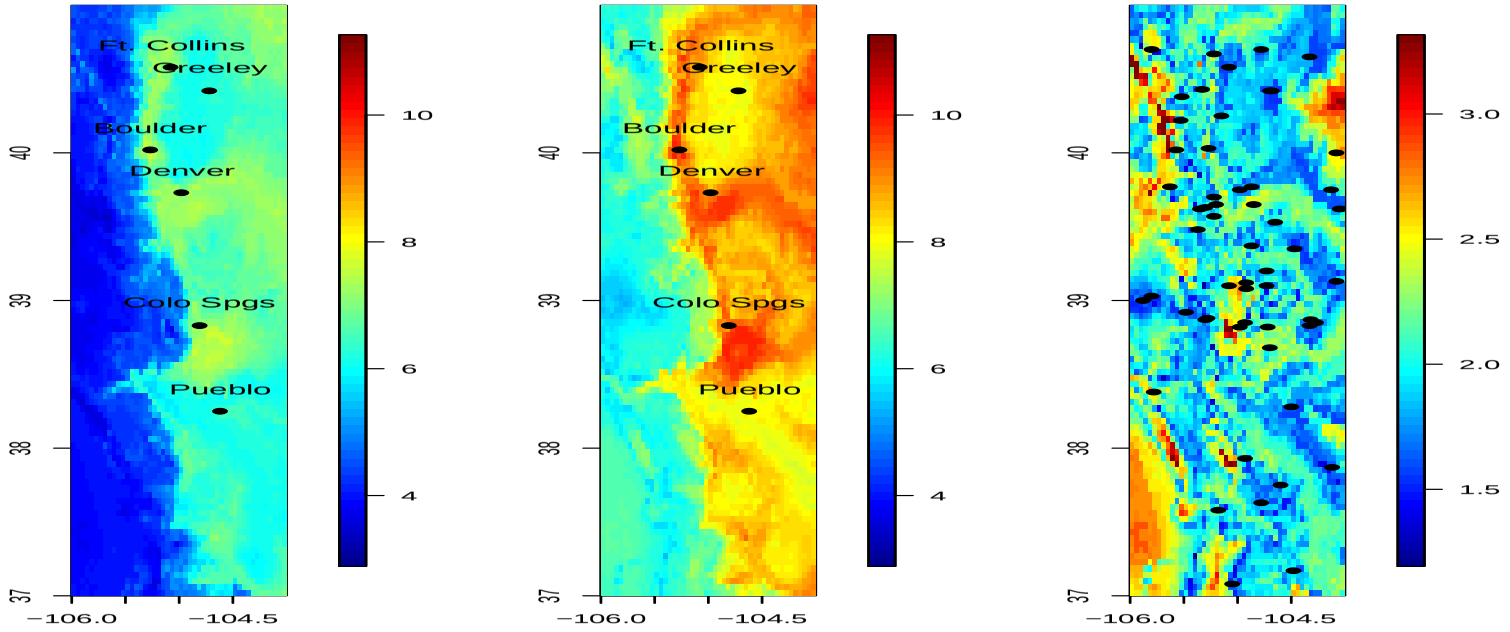

Figure 5.10: Uncertainty bounds. Left plot is the .025 quantile of the daily precipitation 25 year return level. Center plot is the .975 quantile, and right plot is the difference of the two, or an estimate of the range of 95% credible interval (with station locations plotted) Not surprisingly, the uncertainty is greatest in areas where there are no stations.

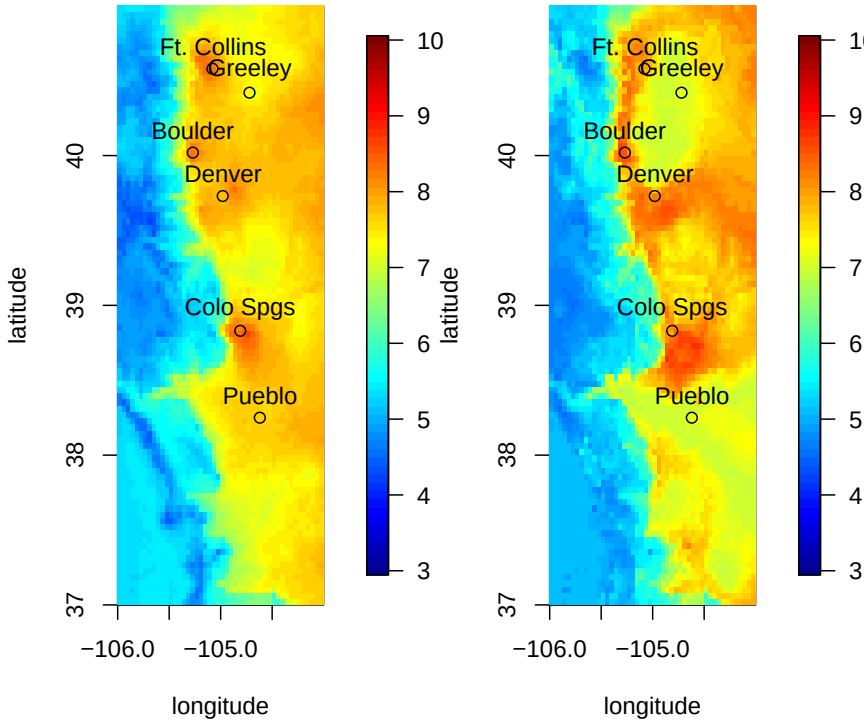

Figure 5.11: Point estimate for the daily precipitation 25 year return level as described by corresponding models in the traditional space (left) and the climate space (right). Scale is in centimeters.

the estimates for $\phi(\boldsymbol{x})$ and $\xi(\boldsymbol{x})$, with mountain areas having a significantly lower return level. The analysis that the mountainous areas have a lighter tail and therefore less extreme precipitation agrees with the studies of Jarrett (1990; 1993) who claims that the hydrologic and paleohydrologic evidence shows that intense rainfall does not occur at higher elevations. The point estimate map also clearly shows a Front Range effect where areas that lie at the foothills of the mountains have higher return levels. From the point estimate for ϕ, the model predicts that areas north of the Palmer Divide have more extreme precipitation than areas northeast of Denver toward Greeley. Not surprisingly, the uncertainty interval map shows the highest levels of uncertainty in the San Luis Valley (extreme southwest part of the study region) and east of Greeley where no stations are located, and in areas of very high elevation where the model extrapolates trends.

Given the motivation based on the Fort Collins flood, the reader may wonder how our analysis compares with the estimates of other studies. Interestingly, our return-level estimates for this particular location do not differ dramatically from those in the 1973 NWS atlas and are slightly less than the estimates from the RFA analysis of Sveinsson et al. (2002). For the 100-year return level, our 95% credible interval for 24-hour

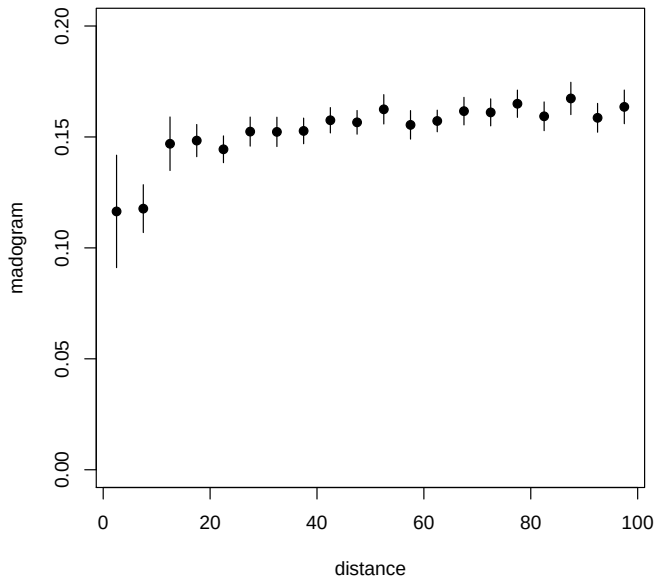

Figure 5.12: The madogram of the normalized annual maximum data at each station. A madogram value of 1/6th implies independence. The range of dependence of the annual maxima appears to be about 15 miles.

precipitation is (9.33, 11.79 cm), whereas the point estimate from NWS is 11.32 cm and the Sveinsson estimate is 12.4 or 12.9 cm, depending on how many stations are included in their analysis. The 24-hour precipitation recorded at the Fort Collins station on the day of the 1997 flood was 15.7cm (6.2 inches) and was the second-largest recorded precipitation amount in our data set. This amount corresponds roughly to a 500-year event, for which our model gives a 95% credible interval of (12.45, 16.33 cm). When one considers that over 1900 station-years of data were used in our model, the likelihood of observing an event of this magnitude seems quite reasonable.

5.5 Conclusions and discussion

The statistical contribution of this work lies in developing and applying a Bayesian analysis for spatial extremes. There are few examples of hierarchical models in the extremes literature, and to our knowledge this is the first use of such a model to produce a map characterizing extreme behavior across a geographic region. In addition to a model for exceedances, a separate spatial model for the threshold exceedance rate proved to be necessary because the level of precision in the recorded data did not allow us to choose thresholds based on common empirical quantiles at each station. To our knowledge, this study is the first to spatially model the exceedance rate parameter in the context of extremes. Obtaining convergence of the spatial range parameter is

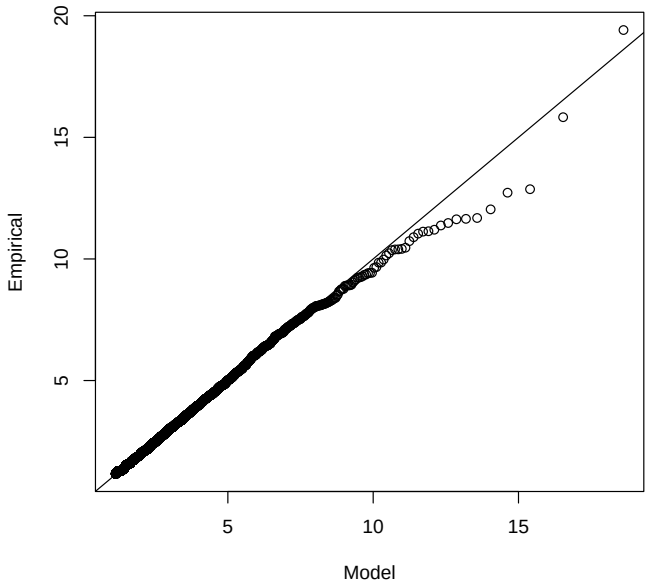

Figure 5.13: Shows a qq plot of all the threshold exceedances. Each station's data was normalized using the posterior mean of ϕ and ξ. The data was then converted to have a common GPD $\phi = 3.7$ $\xi = 0.12$ distribution. Axes labels are in centimeters

often difficult in spatial hierarchical models and we identified a useful transformation of the spatial coordinates to aid convergence and which produced a map which agreed with the regional geography. Additionally we used maximum likelihood theory to increase MCMC efficiency.

Because we are studying a relatively small area with a small number of stations, our best model treats the GPD shape parameter in a simple manner fitting one value of ξ to all the mountain stations and another to all the plains stations. For a larger study area, it would most likely be advantageous to allow the shape parameter to vary more over the region. However, to do so, it would be necessary to have more stations to spatially model this difficult-to-estimate parameter.

The hydrological contribution is a methodology to study extreme precipitation of a region. Our process differs substantially from the commonly-used RFA. We implement a Bayesian spatial model to combine all the information from different stations rather than determining distinct regions in which to pool normalized data. The three-step RFA algorithm makes it difficult to account for all the sources of uncertainty, whereas uncertainty that arises from all the parameter estimates as well as from the interpolation procedure is accounted for in our method. Using the methodology we produced a 25-year daily precipitation return level map for Colorado's Front Range along with measures of uncertainty. The presented model could be employed to produce maps for other return levels or duration periods for this region, and the methodology could be adapted for other regions.

Finally, we were able to produce practical maps of return levels. The maps have several features not predicted by the 1973 NOAA atlas such as the east-west region of higher return levels north of the Palmer Divide. Unlike the older study, we are also able to produce region-wide uncertainty measures. We hope that a study of potential flooding or any other application would not only employ point estimates of precipitation return levels but also consider their uncertainty estimates.

Chapter 6

Conclusion and Future Work

In this thesis, three projects in extreme value theory which were motivated by issues in climate and weather studies have been presented. Although all three of the projects incorporated a spatial component to their analysis, the details of the spatial analysis in the three projects were very different.

The first project developed two extreme-value models for lichenometry. The goal of this project, to determine the values of hidden ages of the moraines, is unique to extreme value theory. Two approaches were taken: the first was to treat the ages and lichen measurements as a bivariate couple, and the second was to treat the ages as hidden covariates of the measurements. The first approach was more simple, but made several more assumptions than the second model which relied on a Bayesian framework.

The second of the two models incorporated data from different spatial locations (specifically, different glaciers on the same mountain), and the model was able to test the hypothesis that there was no spatial effect in the data. More importantly, it was able to pool data from corresponding moraines of different glaciers to more accurately determine the age of the climatic event which formed the moraine.

At the time the second model was constructed, it was believed that there would be an effort to gather lichen data from a large region of the Andes. With these measurements would be recorded precise location information. It was hoped to build a more robust spatial model for this data within the existing framework of the second lichenometry model. At this time, it is unclear whether the project to gather the data will go forward. However, if it is collected, or if other large-scale lichenometry projects are undertaken, the model has the flexibility to incorporate any location information which is recorded with the data.

Both models provide a foundation that is superior to the analysis typically done for lichenometry studies. The disadvantage of the models is that they are more complex and require more sophisticated statistical knowledge. Our collaborators in France hope to provide a "black-box" program for geomorphologists without much statistical background to use to analyze their data. Such a program has been developed for the bivariate lichenometry model and has been used to analyze the Charquini data with different growth curves than the linear one presented in this thesis. It has also been applied to other data sets, including a set from the Alps.

Developing a black-box program for the Bayesian model would be much more difficult, and it is unclear whether our collaborators will undertake this project. Despite the complexity of the two models, the analysis time and resources still pale in comparison to the time required to gather the data.

A study has been done comparing the effectiveness both lichenometry models with the types of analysis which has been typically done in the past. (Jomelli et al. (2005)). Data was collected with known dates, the dates were hidden, and the analysis performed. The authors concluded that the Bayesian model seemed to outperform all other analysis methods.

The madogram project developed a dependence measure for observations which are assumed to come from a bivariate max-stable distribution or a spatial max-stable process. The madogram and λ-madogram were compared to existing dependence measures the extremal coefficient θ and the Pickands dependence function $A(w)$. These three dependence measures are inter-related, and each has strengths and weaknesses. Describing dependence for extremes is not straightforward, and there is no one simple measure promoted in the literature. We believe that the madogram should be among the tools used by practitioners.

One advantage of the madogram is that it has its roots in geostatistics. Statisticians familiar with the variogram will immediately recognize many of the properties of the madogram. Because of its geostatistical roots, the madogram is a natural estimator for spatial dependence.

Another advantage of the madogram and λ-madogram is that they have straightforward estimators. Unlike the existing dependence measures' estimators, the madogram's estimators do not require taking maxima, minima, or estimating the empirical distribution function of the angular component. Hence, they are closer to the estimators which statisticians are most familiar. The madogram's estimators seem to perform at least as well as existing estimators.

At this time it is unclear how applicable max-stable dependence measures such as the madogram are to real-world questions. To assume that the underlying distribution is max-stable, one probably must be analyzing data which are componentwise block maxima. Although there are a some instances where dependence between multivariate or spatial annual-max observations might be studied, there may be more interest in describing the dependence between the extreme observations that correspond to a particular time. It is hoped that the madogram can be extended to this case where we only assume that the data are in the domain of attraction of a max-stable distribution.

Like the madogram project, spatial dependence was the central focus of the Colorado precipitation project. However, rather than trying to describe the dependence between the extreme observations, the project's goal was to describe the dependence between the distributions of the extremes at the locations. To model the dependence in the distributions, a Bayesian hierarchical model was created which built a spatial structure for the parameters of the GPD. Although hierarchical models are in wide use for problems which model related

data with many covariates, there are very few examples in the extremes literature. By performing the spatial analysis in a space whose coordinates were elevation and mean seasonal precipitation rather than latitude and longitude, the resulting map showed that the return levels were strongly dependent on the orography of the region.

The climatological extreme precipitation model makes only one assumption with which we are less than comfortable: that the data observations are spatially independent conditional on the stations' spatially dependent GPD parameters. Certainly this is not the case, as even conditional on the climate, precipitation observations are not spatially independent. However, we believe that the return-levels maps are not largely influenced by this simplifying assumption. How to account for spatial dependence in the observations, and how to separate it from the climatological spatial dependence which was of interest in this particular application remain open questions.

The general framework of the model can be extended to other applications. In collaboration with Eric Gilleland of NCAR, the framework has already applied to ozone pollution data for the eastern United States. The pollution data appears to have a bounded tail, whereas the precipitation data had a heavy tail, and this difference forced a considerable change to the MCMC algorithm. As the pollution data was spatially correlated to population centers, the spatial analysis was performed in the traditional latitude/longitude space. Another future application is to apply the model to precipitation data from a regional climate model (RCM) for the western United States. As the nature of precipitation varies considerably over this larger region (e.g. Arizona's and Washington's precipitation) than over Colorado's front range, we anticipate needing to vary both the scale and shape parameters of the GPD. For both the ozone and RCM applications, one might consider some alternative space.

An important extension of the precipitation model may be pursued. Using the methodology presented in Chapter 5, one can create return level maps for the 1-, 2-, 6-, 12-, and 24-hour precipitation levels. Each of these would be done individually, using only the data corresponding to the duration period under study. However, as the data are not coupled, it would be possible that the return level estimate for say a 12-hour duration period could be higher than for the 24-hour period. It seems that one could borrow strength by developing a model which would produce measures for all duration periods which would also not allow inconsistencies in the return-level estimates.

Another question which warrants investigation is the study of extreme observations on different scales or support. At the beginning of Chapter 3, it was mentioned that weather and climate data may come from not only point locations but also may correspond to grid cells. Grid cell data is most often interpreted as an average value over the cell's area. It is unknown how to scale extreme data from one level of support to another, and in fact, this change-of-support problem is still being studied for non-extreme data (Wackernagel, Hans,

2003; Wackernagel H, Lajaunie C, Blond N, Roth C and Vautard R, 2004). Originally, this question was to be investigated as part of this thesis, and may be investigated in the future.

The statistical study of extremes is a growing field, and its applications to atmospheric sciences are only beginning. The diversity of the projects studied in this thesis indicates that the field is wide-open and there are many questions and applications that are not yet answered. All of the projects presented here have many more extensions, and there are other projects left untouched. It will be interesting to continue to study in the field of extremes and spatial extremes as one expects there to be many new developments in the near future.

Bibliography

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

Benedict J.B. (1967). Recent glacial history of an alpine area in the Colorado Front Range, USA. 1. Establishing a lichen growth curve. Journal of Glaciology, 6:817–832.

Benedict J.B. (1990). Lichen mortality due to late-lying snow, results of a transplant study. Arctic and Alpine Research, 22:81–99.

Benedict J.B. (1991). Experiments on lichen growth ii. effects of a seasonal snow cover. Arctic and Alpine Research, 23:189–199.

Beschel R.E. (1961). Dating rock surfaces by lichen growth and its application to glaciology and physiography (lichenometry). In Raasch G.O. (ed.), Geology of the Arctic. Toronto; University of Toronto Press, 2:1044–1062.

Bonnin, G., Todd, D., Lin, B., Parzybok, T., Yekta, M., and Riley, D. (2004a). NOAA Atlas 14, Precipitation Frequency Atlas of the United States, Volume 1. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Silver Spring, Maryland.

Bonnin, G., Todd, D., Lin, B., Parzybok, T., Yekta, M., and Riley, D. (2004b). NOAA Atlas 14, Precipitation Frequency Atlas of the United States, Volume 2. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Silver Spring, Maryland.

Brockwell, P. and Davis, R. (1991). Time Series: Theory and Methods. Springer Verlag.

Bull W.B. and Brandon M.T. (1998). Lichen dating of earthquake-generated regional rockfall events, Southern Alps, New Zealand. GSA Bulletin 1, 10 1:60–84.

Capéraà, P. and Fougéres, A.-L. (2000). Estimation of a bivariate extreme value distribution. Extremes, 3:311–329.

Caperaa, P., Fougeres, A.-L., and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. Biometrika, 84:567–577.

Carlin, B. and Louis, T. (2000). Bayes and Empirical Bayes Methods for Data Analysis. Chapman & Hall/CRC, New York.

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

Casson, E. and Coles, S. (1999). Spatial regression models for extremes. Extremes, 1:449–468.

Chiles, J. and Delphiner, P. (1999). Geostatistics: Modeling Spatial Uncertainty. Wiley, New York.

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

Coles, S. and Tawn, J. (1991). Modeling multivariate extreme events. Journal of the Royal Statistical Society, Series B, 53:377–92.

Coles, S. and Tawn, J. (1996a). A Bayesian analysis of extreme rainfall data. Applied Statistics, 45:463–478.

Coles, S. and Tawn, J. (1996b). Modeling extremes of the areal rainfall process. Journal of the Royal Statistical Society, Series B, 58:329–347.

Coles, S. G. (2001a). Improving the analysis of extreme wind speeds with information-sharing models. Notes de l'Institut Pierre Simon Laplace, ISSN 1626-8334, 11:23–34.

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

Cooley, D., Naveau, P., and Jomelli, V. (2004). Spatio-temporal analysis of extreme values from lichenometric studies and their relationships to climate. In Proceedings of the 2004 Conference. The International Environmetrics Society.

Cooley, D., Naveau, P., Jomelli, V., Rababtel, A., and Grancher, D. (2005a). A Bayesian hierarchical extreme value model for lichenometry. Environmetrics (in press).

Cooley, D., Nychka, D., and Naveau, P. (2005b). Bayesian spatial modeling of extreme return levels. Submitted.

Cressie, N. (1993). Statistics for Spatial Data. Wiley, New York.

Dalrymple, T. (1960). Flood frequency analyses. Water supply paper 1543-a, U.S. Geological Survey, Reston, VA.

Daly, C., Gibson, W., Taylor, G., Johnson, G., and Pasteris, P. (2002). A knowledge-based approach to the statistical mapping of climate. Climate Research, 23:99–113.

Daly, C., Neilson, R., and D.L.Phillips (1994). A statistical topographic model for mapping climatological precipitation in mountainous terrain. Journal of Applied Meteorology, 33:140–158.

Davis, R. and Resnick, S. (1993). Prediction of stationary max-stable processes. Ann. of Applied Prob, 3:497–525.

Davison, A. and Smith, R. (1990). Models for exceedances over high thresholds. J. R. Statist. Soc. B, 52:393–442.

de Haan, L. (1984). A spectral representation for max-stable processes. Annals of Probability, 12:1194–2004.

de Haan, L. (1985). Extremes in higher dimensions: the model and some statistics. Proceedings 45th Session of the International Statistical Institute.

Deheuvels, P. (1991). On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. Statistics and Probability Letters, 12:429–439.

Dempster A.P., Laird N.M., Rubin D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, 39:1–21.

Diggle, P., Tawn, J., and Moyeed, R. (1998). Model-based geostatistics. Applied Statistics, 47:299–350.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance, volume 33 of Applications of Mathematics. Springer-Verlag, Berlin.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proc. Cambridge. Philos. Soc., 24:180–190.

Fougères, A. (2004). Multivariate extremes. In B. Finkenstadt and H. Rootzen, editor, Extreme Values in Finance, Telecommunications and the Environment, pages 373–388. Chapman and Hall CRC Press, London.

Fowler, H., Ekstrom, M., Kilsby, C., and Jones, P. (2005). New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. 1. assessment of control climate. Journal of Hydrology, 300:212–233.

Fowler, H. and Kilsby, C. (2003). A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. International Journal of Climatology, 23:1313–1334.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2003). Bayesian Data Analysis, 2nd ed. Chapman & Hall, Boca Raton, FL.

Gelman, A. (1996). Inference and monitoring convergence. In W.R. Gilks, S. Richarson, and D.J. Spiegelhalter, editor, Markov Chain Monte Carlo in Practice. Chapman and Hall, London.

Gnedenko, R. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. Ann. of Math., 44:423–453.

Gumbel, E. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. Publ. Inst. Statis. Univ. Paris, 9:171–173.

Hall, P. and Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. Statistical Science, 15:153–167.

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

Henz, J., Scheetz, V., and Doehring, D. (1976). The Big Thompson Flood of 1976 in Colorado. Weatherwise, 29:278–285.

Horn, R. (1990). The hadamard product. In Johnson, C., editor, Matrix Theory and Applications. American Mathematical Society, Providence, RI.

Hosking, J. and Wallis, J. (1987). Parameter and quantile estimation for the generalized Pareto distribution. Technometrics, 29:339–349.

Hosking, J. and Wallis, J. (1988). The effect on intersite dependence on regional frequency analysis. Water Resources Research, 24:588–600.

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

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

Innes J. L. (1985). Lichenometry. Progress in Physical Geography, 9:187 – 254.

IPCC (2001a). Annex b. glossary of terms (based on third assessment report). http://www.ipcc.ch/pub/syrgloss.pdf.

IPCC (2001b). Climate change 2001: the scientific basis. In Houghton, J. T., Ding, Y., Griggs, D., Noguer, M., van der Linden, P. J., and Xiaosu, D., editors, Third Assessment Report. Cambridge University Press, Cambridge.

Jarrett, R. D. (1990). Paleohydrologic techniques used to define the spatial occurrence of floods. Geomorphology, 3:181–195.

Jarrett, R. D. (1993). Flood elevation limits in the Rocky Mountains. In Kuo, C., editor, Engineering Hydrology: Proceedings of the Symposium Sponsorded by the Hydraulics Division of the American Society of Civil Engineers, pages 180–185. American Society of Civil Engineers.

Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. Statistics and Probability Letters, 9:75–81.

Jomelli, V., Grancher, D., Naveau, P., and Cooley, D. (2005). Assessment study of lichenometric methods for dating surfaces. Manuscript.

Karl, T. and Knight, R. (1998). Secular trends of precipitation amount, frequency, and intensity in the United States. Bulletin of the American Meteorological Society, 79:231–241.

Karlen W. and Black J. (2002). Estimates of lichen growth-rate in northern sweden. Geografisca Annaler, 84:A:225–232.

Katz, R., Parlange, M., and Naveau, P. (2002). Extremes in hydrology. Advances in Water Resources, 25:1287–1304.

Lamperti, J. (1996). Probability: A Survey of the Mathematical Theory. Wiley, New York.

Leadbetter, M., Lindgren, G., and Rootzén, H. (1983). Extremes and related properties of random sequences and processes. Springer Verlag, New York.

Matheron, G. (1987). Suffit-il, pour une covariance, d'être de type positif? Sciences de la Terre, série informatique géologique, 26:51–66.

McLachlan G.J., Thriyambakam, K. (1997). The EM Algorithm and Extensions. Wiley, New York.

Miller, J., Frederick, R., and Tracey, R. (1973). Precipitation-Frequency Atlas of the Western United States, Volume III-Colorado. U.S Department of Commerce, Silver Spring, MD.

Naveau, P., Jomelli, V., Cooley, D., and Rabatel, A. (2005a). Modeling uncertainties in lichenometry studies. Submitted.

Naveau, P., Nogaj, M., Ammann, C., Yiou, P., Cooley, D., and Jomelli, V. (2005b). Statistical analysis of climate extremes. Comptes rendus Géosciences de l'Académie des Sciences.

Naveau, P., Poncet, P., and Cooley, D. (2005c). First-order variograms for extreme bivariate random vectors. Submitted.

O'Neal M. and Schoenenberger K. R. (2003). A rhizocarpon geographicum growth curve for the Cascade Range of Washington and northern Oregon, usa. Quaternary Research, 60:233–241.

Pentecost A. (1979). Aspect and slope preferences in a saxicolous lichen community. Lichenologist, 11:81–83.

Pickands, J. (1975). Statistical inference using extreme order statistics. Ann. Statist., 3:119–131.

Pickands, J. (1981). Multivariate extreme value distributions. Bulletin of the International Statistical Institution, pages 859–878.

P.Poncet, Cooley, D., and Naveau, P. (2005). Variograms for spatial max-stable random fields. Under revision.

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

Robert, Christian P.; Casella, George (1999). Monte Carlo statistical methods. Springer, New York.

Rodbell D.T. (1992). Lichenometric and radiocarbon dating of holocene glaciation, Cordillera Blanca, Peru. The Holocene, 2:19–29.

Schaefer, M. (1990). Regional analyses of precipitation annual maxima in Washington State. Water Resources Research, 26:119–131.

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

Schlather, M. and Tawn, J. (2002). Inequalities for the extremal coefficients of multivariate extreme value distributions. Extremes, 5(1):87–102.

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

Smith, R. (1990). Max-stable processes and spatial extremes. Unpublished manusript.

Smith, R. (2004). Statistics of extremes, with applications in environment, insurance and finance. In B. Finkenstadt and H. Rootzen, editor, Extreme Values in Finance, Telecommunications and the Environment, pages 1–78. Chapman and Hall CRC Press, London.

Smith, R. and Naylor, J. (1987). A comparison of maximum likelihood and Bayeisan estimators for the three-parameter Weibull distribution. Applied Statistics, 36:358–369.

Spatial Climate Analysis Service, O. S. U. (created 4 Feb 2004). May-October precipitation normals 1971-2000. http://www.ocs.oregonstate.edu/prism/.

Spatial Climate Analysis Service, O. S. U. (created Apr 1995). 2.5 minute digital elevation model for the coterminous United States. http://www.ocs.orst.edu/prism/docs/meta/dem_25m.htm.

Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B., 64:583–639.

Sveinsson, O., Salas, J., and Boes, D. (2002). Regional frequency analysis of extreme precipitation in Northeastern Colorado and Fort Collins Flood of 1997. Journal of Hydrologic Engineering, 7:49–63.

Tawn, J. (1988). Bivariate extreme value theory: models and estimation. Biometrika, 75:397–415.

Tawn, J. (1990). Modeling multivariate extreme value distributions. Biometrika, 75:245–253.

Trenberth, K. (1999). Conceptual framework for changes of extremes of the hydrological cycle with climate change. Climatic Change, 42:327–339.

Von Mises, R. (1954). La distribution de la plus grande de n valeurs. Selected Papers II, Amer. Math. Soc., pages 271–294.

Wackernagel H, Lajaunie C, Blond N, Roth C and Vautard R (2004). Geostatistical risk mapping with chemical transport model output and ozone station data. Ecological Modelling, 179:177–185.

Wackernagel, Hans (2003). Multivariate Geostatistics, An Introduction with Applications, 3rd ed. Springer-Verlag, Berlin.

Wang, X., Zwiers, F., and Swail, V. (2004). North Atlantic ocean wave climate change scenarios for the 21st century. Journal of Climate, 17:2368–2383.

Weaver, J., Gruntfest, E., and Levy, G. (2000). Two floods in Fort Collins Colorado: Learning from a natural disaster. Bulletin of the American Meteorological Society, 81:2359–2366.

Zwiers, F. and Kharin, V. (1998). Changes in the extremes of the climate simulated by CCC GCM2 under CO2 doubling. Journal of Climate, 11:2200–2222.

.1 Appendix for Lichenometry Chapter

We begin with three statements of fact which can be found in Horn 1990.

Fact 6 A positive-semidefinite matrix is positive-definite if and only if it is non-singular.

Fact 7 If A and B are positive-semidefinite, their Hadamard product $A \circ B$ is positive-semidefinite.

Fact 8 If A is positive-definite, and B is positive-semidefinite with all its main diagonal entries positive, then $A \circ B$ is positive-definite.

Since the temporal matrix U and the spatial matrix V are positive-semidefinite, by Fact 2, $U \circ V$ is positive-semidefinite.

We refer to the block structure of the covariance matrix in Figure 2.4. We denote by A_U the block of U which corresponds to block A of the resulting covariance matrix S, and likewise for the other matrix blocks and the matrix V.

The sub-matrix

A_U	B_U
B_U	C_U

is positive-definite, as it is the covariance matrix of events that do not share the same age. The corresponding sub-matrix of the $U \circ V$ is positive definite by Theorem 3. Therefore the columns of this sub-matrix are linearly independent.

Since $D = \eta C$, it is easy to show that the covariance matrix can be transformed by row operations to

$\mu_{0,1}$ ⋮ μ_{0,n_0}	A	B	B	B
$\mu_{1,1}$ ⋮ $\mu_{1,n}$	a_1B	C	0	0
$\mu_{2,1}$ ⋮ $\mu_{2,n}$	a_2B	0	C	0
$\mu_{3,1}$ ⋮ $\mu_{3,n}$	a_3B	0	0	C

.

Take any column j not in the earlier sub-matrix. Assume it is linearly dependent of the columns $1, \ldots, j-1$. Then the entries of this column found in block C would be a linear combination of the entries in blocks B and C. This contradicts our earlier assertion that the columns of the sub-matrix are linearly independent. Therefore, this matrix is non-singular, and by Theorem 1, the covariance matrix is positive-definite.

.2 Appendix for Madogram Chapter

Proof of Proposition 2

Let $\boldsymbol{Z}(x)$ be a max-stable random field with common marginal distribution F. Without loss of generality (Resnick, 1987) we can assume that F is any $GEV(\mu, \sigma, \xi)$ distribution. Also recall that the extremal coefficient θ is invariant to marginal transformations between GEV distributions. Throughout much of the thesis, we have assumed F is unit-Fréchet, however, for this proof we will simply assume F is $GEV(\mu, \sigma, \xi)$. We will specify the marginal distribution to what is convenient for the proof.

To prove (1), we let $Z(x)$ be a spatial max-stable random field with extremal coefficient $\theta(h)$ and unit Fréchet margins. For all $z \geqslant 0$, it is possible to define the function C_z such that

$$C_z(h) = E[\sqrt{z}\mathbf{1}_{Z(x)>z}\sqrt{z}\mathbf{1}_{Z(x+h)>z}] - E[\sqrt{z}\mathbf{1}_{Z(x)>z}]E[\sqrt{z}\mathbf{1}_{Z(x+h)>z}].$$

By construction $C_z(h)$ is a covariance function and this implies

$$C_z(h) = zF_h(z,z) - zF(z)F(z) = zF(z)^{\theta(h)}(1 - F(z)^{2-\theta(h)}).$$

In the previous equality, we know that $F(z) = \exp(-1/z)$. Taking the limit of $C_z(h)$ as $z \to +\infty$ gives

$$\lim_{z\to+\infty} C_z(h) = \lim_{z\to+\infty} z\left(1 - \exp\left(-\frac{2-\theta(h)}{z}\right)\right) = 2 - \theta(h).$$

Hence, $2 - \theta(h)$, as a limit of covariance functions, is a covariance function.

To prove (2)-(4), we will assume that $F \sim GEV(\mu, \sigma, \xi)$ where $\xi < .5$. We will define the madogram as in (4.18) and rely on the relation given in (4.21).

We first need to show that the madogram satisifies the triangle inequality:

$$\begin{aligned}
\nu(h+k) &= \frac{1}{2}\mathbb{E}|Z(x) - Z(x+h+k)| \\
&= \frac{1}{2}\mathbb{E}|Z(x) - Z(x+h) + Z(x+h) - Z(x+h+k)| \\
&\leq \frac{1}{2}\mathbb{E}|Z(x) - Z(x+h)| + \frac{1}{2}\mathbb{E}|Z(x+h) - Z(x+h+k)| \\
&\leq \nu(h) + \nu(k)
\end{aligned}$$

For the first inequality, we assume that the max-stable field has Gumbel margins. This implies $\nu(h) = \log\theta(h)$. Applying the triangle inequality gives

$\ln\theta(h+k) \leq \ln\theta(h) + \ln\theta(k) \to \theta(h+k) \leq \theta(h)\theta(k)$.

The second and third inequalities are derived in a similar way. We now choose the margins such that the madogram is equal to $\frac{\Gamma(1-\tau)}{\tau}(\theta(h)^\tau - 1)$ with $0 < \tau < 1$. Again, this madogram satisfies the triangular inequality, i.e.

$$\frac{\Gamma(1-\tau)}{\tau}(\theta(h+k)^\tau - 1) \leqslant \frac{\Gamma(1-\tau)}{\tau}(\theta(h)^\tau - 1) + \frac{\Gamma(1-\tau)}{\tau}(\theta(k)^\tau - 1).$$

The required inequalities follow. The special case $\tau = 1$ is obtainted by letting $\tau \uparrow 1$.

Proof of Proposition 3

Let (X, Y) be a random variable from a bivariate EVD with Fréchet margins, let $\theta(\lambda)$ be defined as in (4.32), and let F denote the Fréchet distribution function: $F(x) = \exp(-x^{-1})$

$$\begin{aligned}
\lambda_1 < \lambda_2 &\Rightarrow \mathbb{P}\{X \le x, \lambda_1 Y \le x\} \ge \mathbb{P}\{X \le x, \lambda_2 Y \le x\} \\
&\Rightarrow F^{\theta_{\lambda_1}}(x) \ge F^{\theta_{\lambda_2}}(x) \\
&\Rightarrow \theta_{\lambda_1} \log(F(x)) \ge \theta_{\lambda_2} \log(F(x)) \\
&\Rightarrow \theta\lambda_1 \le \theta\lambda_2
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}\{X \le x, \lambda Y \le x\} &\le \mathbb{P}\left\{Y \le \frac{1}{\lambda}\right\} \\
F^{\theta(\lambda)}(x) &\le \exp(-\lambda x^{-1}) \\
F^{\theta(\lambda)}(x) &\le F^{\lambda}(x) \\
\theta(\lambda) &\ge \lambda
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}\{X \le x, \lambda Y \le x\} &\le \mathbb{P}\{X \le x\} \\
F^{\theta(\lambda)}(x) &\le F(x) \\
\theta(\lambda) &\ge 1
\end{aligned}$$

Because the bivariate EVD only allows for positive dependence:

$$\begin{aligned}
\mathbb{P}\{X \le x, \lambda Y \le x\} &\ge \mathbb{P}\{X \le x\}\,\mathbb{P}\left\{Y \le \frac{x}{\lambda}\right\} \\
F^{\theta(\lambda)}(x) &\ge F(x)F^{\lambda}(x) \\
\theta(\lambda) &\le 1 + \lambda
\end{aligned}$$

Therefore

$$\max(1, \lambda) \le \theta(\lambda) \le 1 + \lambda$$

$$\begin{aligned}
\mathbb{P}\{X \le x, Y \le y\} &= \mathbb{P}\{X \le y, Y \le x\} \\
\mathbb{P}\{X \le x, Y \le x/\lambda\} &= \mathbb{P}\{X \le x/\lambda, Y \le x\} \\
\exp\left\{-\int_0^1 \max\left(\frac{w}{x}, \frac{\lambda(1-w)}{x}\right) dH(w)\right\} &= \exp\left\{-\int_0^1 \max\left(\frac{\lambda w}{x}, \frac{(1-w)}{x}\right) dH(w)\right\} \\
\exp\left\{-\frac{1}{x}\int_0^1 \max\left(w, \lambda(1-w)\right) dH(w)\right\} &= \exp\left\{-\frac{\lambda}{x}\int_0^1 \max\left(w, \frac{1}{\lambda}(1-w)\right) dH(w)\right\} \\
F^{\theta(\lambda)}(x) &= F^{\lambda\theta_{1/\lambda}}(x)
\end{aligned}$$

Proof of Proposition 4

$$\begin{aligned}\nu(\lambda) &= \frac{1}{2}E\left|F(X)-F(\lambda Y)\right| \\ \nu(\lambda) &= \mathbb{E}\left[\max\left(F(X),F(\lambda Y)\right)\right]-\frac{1}{2}\mathbb{E}\left[F(X)+F(\lambda Y)\right]\end{aligned}$$

To evaluate the first term of the above expression, we let $M=\max\left(F(X),F(\lambda Y)\right)$.

$$\begin{aligned}F_M(u) &= \mathbb{P}\left\{\max\left(F(X),F(\lambda Y)\right)<u\right\} \\ &= \mathbb{P}\left\{X<F^{-1}(u),Y<\frac{1}{\lambda}F^{-1}(u)\right\} \\ &= \exp\left\{-V\left(\frac{1}{F^{-1}(u)},\frac{\lambda}{F^{-1}(u)}\right)\right\} \\ &= \exp\left\{-\frac{1}{F^{-1}(u)}\theta(\lambda)\right\} \\ &= u^{\theta}(\lambda),\ \text{for } u\in[0,1],\ \text{because } F^{-1}(y)=-\ln^{-1}y \\ \Rightarrow\ & f_M(u)=\theta(\lambda)u^{\theta(\lambda)-1} \\ \Rightarrow\ & \mathbb{E}\left[\max\left(F(X)+F(\lambda Y)\right)\right]=\frac{\theta(\lambda)}{1+\theta(\lambda)}\end{aligned}$$

For $\mathbb{E}[F(\lambda Y)]$ we write:

$$\begin{aligned}F(\lambda Y) &= \exp\left\{-\frac{1}{\lambda Y}\right\} \\ &= \left(\exp\left\{-\frac{1}{Y}\right\}\right)^{\lambda} \\ &= F^{\lambda}(Y) \\ &= u^{\lambda}\ \text{for } u\in[0,1] \\ \Rightarrow\ & f_{\lambda Y}=\lambda u^{\lambda-1} \\ \Rightarrow\ & \mathbb{E}[F(\lambda Y)]=\frac{\lambda}{1+\lambda}\end{aligned}$$

Putting it all together yields:

$$\nu(\lambda)=\frac{\theta(\lambda)}{1+\theta(\lambda)}-\frac{1+3\lambda}{4(1+\lambda)}, \tag{1}$$

or inversely,

$$\theta(\lambda)=\frac{\nu(\lambda)+c_\lambda}{1-(\nu(\lambda)+c_\lambda)},\ \text{where } c_\lambda=\frac{1+3\lambda}{4(1+\lambda)}. \tag{2}$$

Proof of Proposition 5

Consider $\mathbb{E}[\hat{\nu}_n(\lambda)]$.

$$\begin{aligned}
&= \mathbb{E}\left[\frac{1}{2n}\sum_{i=1}^{n}|F(X_i)-F(\lambda Y_i)| - (1-w(\lambda))\frac{1}{2n}\sum_{i=1}^{n}(1-F(X_i))\right. \\
&\quad \left. -w(\lambda)\frac{1}{2n}\sum_{i=1}^{n}F(X_i)+\frac{1}{4}\right] \\
&= \mathbb{E}\frac{1}{2}|F(X_i)-F(\lambda Y_i)| - (1-w(\lambda))\mathbb{E}\frac{1}{2}(1-F(X_i)) - w(\lambda)\mathbb{E}\frac{1}{2}F(X_i)+\frac{1}{4} \\
&= \nu(\lambda) - (1-w(\lambda))\frac{1}{4} - w(\lambda)\frac{1}{4}+\frac{1}{4} \\
&= \nu(\lambda)
\end{aligned}$$

We define $\gamma(\lambda) = 1/2\mathbb{E}[(F(X)-F(\lambda Y))^2]$ and $\eta(\lambda) = 1/2\mathbb{E}[F(X)|F(X)-F(\lambda Y)|]$. Notice $\gamma(\lambda) \leq 1$, $\nu(\lambda) \leq 1$, and $\eta(\lambda) \leq \nu(\lambda)$.

Consider $Var\left[\frac{1}{2}|F(X_i)-F(\lambda Y_i)|\right]$.

$$\begin{aligned}
&= \mathbb{E}\left[\left(\frac{1}{2}|F(X_i)-F(\lambda Y_i)|\right)^2\right] - \left(\mathbb{E}\left[\frac{1}{2}|F(X_i)-F(\lambda Y_i)|\right]\right)^2 \\
&= \frac{1}{2}\gamma(\lambda) - \nu(\lambda)
\end{aligned}$$

Consider $Cov\left[\frac{1}{2}|F(X_i)-F(\lambda Y_i)|, \frac{1}{2}F(X_i)\right]$.

$$\begin{aligned}
&= \mathbb{E}\left[\frac{1}{4}F(X_i)|F(X_i)-F(\lambda Y_i)|\right] - \mathbb{E}\left[\frac{1}{2}|F(X_i)-F(\lambda Y_i)|\right]\mathbb{E}\left[\frac{1}{2}F(x_i)\right] \\
&= \frac{1}{2}\eta(\lambda) - \frac{1}{4}\nu(\lambda)
\end{aligned}$$

Now, consider $Var[\hat{\hat{\nu}}_n(\lambda)]$.

$$
\begin{aligned}
&= Var\left[\frac{1}{2n}\sum_{i=1}^{n}|F(X_i)-F(\lambda Y_i)|-(1-w(\lambda))\frac{1}{2n}\sum_{i=1}^{n}(1-F(X_i))\right.\\
&\quad \left.-w(\lambda)\frac{1}{2n}\sum_{i=1}^{n}F(X_i)+\frac{1}{4}\right]\\
&= Var\left[\frac{1}{2n}\sum_{i=1}^{n}|F(X_i)-F(\lambda Y_i)|\right]+Var\left[(1-w(\lambda))\frac{1}{2n}\sum_{i=1}^{n}(1-F(X_i))\right]\\
&\quad +Var\left[w(\lambda)\frac{1}{2n}\sum_{i=1}^{n}F(X_i)\right]\\
&\quad -Cov\left[\frac{1}{2n}\sum_{i=1}^{n}|F(X_i)-F(\lambda Y_i)|,(1-w(\lambda))\frac{1}{2n}\sum_{i=1}^{n}(1-F(X_i))\right]\\
&\quad -Cov\left[\frac{1}{2n}\sum_{i=1}^{n}|F(X_i)-F(\lambda Y_i)|,w(\lambda)\frac{1}{2n}\sum_{i=1}^{n}F(X_i)\right]\\
&\quad +Cov\left[(1-w(\lambda))\frac{1}{2n}\sum_{i=1}^{n}(1-F(X_i)),w(\lambda)\frac{1}{2n}\sum_{i=1}^{n}F(X_i)\right]\\
&= \frac{1}{n}Var\left[\frac{1}{2}|F(X_i)-F(\lambda Y_i)|\right]+\frac{(1-w(\lambda))^2}{n}Var\left[\frac{1}{2}(1-F(X_i))\right]\\
&\quad +\frac{(w(\lambda))^2}{n}Var\left[\frac{1}{2}F(X_i)\right]\\
&\quad -\frac{(1-w(\lambda))}{n}Cov\left[\frac{1}{2}|F(X_i)-F(\lambda Y_i)|,\frac{1}{2}(1-F(X_i))\right]\\
&\quad -\frac{(w(\lambda))}{n}Cov\left[\frac{1}{2}|F(X_i)-F(\lambda Y_i)|,\frac{1}{2}F(X_i)\right]\\
&\quad +\frac{w(\lambda)(1-w(\lambda))}{n}Cov\left[\frac{1}{2}(1-F(X_i)),\frac{1}{2}F(X_i)\right]\\
&= \frac{1}{n}\left[\frac{1}{2}\gamma(\lambda)-\nu(\lambda)\right]+\frac{(1-w(\lambda))^2}{n}\left[\frac{1}{48}\right]+\frac{(w(\lambda))^2}{n}\left[\frac{1}{48}\right]-\frac{1-w(\lambda)}{n}\left[\frac{1}{2}\eta(\lambda)-\frac{1}{4}\nu(\lambda)\right]\\
&\quad -\frac{w(\lambda)}{n}\left[\frac{1}{2}\eta(\lambda)-\frac{1}{4}\nu(\lambda)\right]+\frac{w(\lambda)(1-w(\lambda))}{n}\left[\frac{1}{48}\right]\\
&= \frac{1}{n}\left[\frac{1}{2}\gamma(\lambda)-\frac{3}{4}\nu(\lambda)-\frac{1}{2}\eta(\lambda)+\frac{1}{48}(1-w(\lambda)-(w(\lambda))^2)\right]
\end{aligned}
$$

Notice that $\frac{1}{2}\gamma(\lambda)-\frac{3}{4}\nu(\lambda)-\frac{1}{2}\eta(\lambda)+\frac{1}{48}(1-w(\lambda)-(w(\lambda))^2)$ is finite and by the Central Limit Theorem

$$\hat{\hat{\nu}}_n(\lambda)\rightarrow N\left(\nu(\lambda),\frac{1}{n}\left[\frac{1}{2}\gamma(\lambda)-\frac{3}{4}\nu(\lambda)-\frac{1}{2}\eta(\lambda)+\frac{1}{48}(1-w(\lambda)-(w(\lambda))^2)\right]\right).$$

From Brockwell and Davis (1991) we use the following theorem. *If X_n is asymptotically normal with mean μ and variance σ_n^2 where $\sigma_n\rightarrow 0$ and g is differentiable at μ, then $g(X_n)$ is asymptotically normal with mean $g(\mu)$ and variance $g'(\mu)^2\sigma_n^2$.*

We let

$$g(\hat{\hat{\nu}}_n(\lambda))=\frac{\hat{\hat{\nu}}_n(\lambda)+c_\lambda}{1-(\hat{\hat{\nu}}_n(\lambda)+c_\lambda)}=\hat{\theta}_n(\lambda),$$

where $c_\lambda=(1+3\lambda)/(4(1+\lambda))$. Since

$$g'(\hat{\hat{\nu}}_n(\lambda))=\frac{1}{1-(\hat{\hat{\nu}}_n(\lambda)+c_\lambda)},$$

then

$$\hat{\theta}_n(\lambda) \to N\left(\theta(\lambda), \frac{1}{n}\frac{1}{1-(\hat{\hat{\nu}}_n(\lambda)+c_\lambda)}h(\lambda)\right),$$

where $h(\lambda) = \left[\frac{1}{2}\gamma(\lambda) - \frac{3}{4}\nu(\lambda) - \frac{1}{2}\eta(\lambda) + \frac{1}{48}(1 - w(\lambda) - (w(\lambda))^2)\right]$.

Furthermore, since

$$A(w) = \frac{1}{\lambda+1}\theta(\lambda)$$

where $w = 1/(\lambda+1)$, *then*

$$\hat{\theta}_n(\lambda) \to N\left(A(w), \frac{1}{n}\frac{1}{\lambda+1}\frac{1}{1-(\hat{\hat{\nu}}_n(\lambda)+c_\lambda)}h(\lambda)\right).$$