Bayesian Spatial Modeling of Extreme Precipitation Return Levels

Daniel Cooley[1,2]

Doug Nychka[2]

Philippe Naveau[1,3]

[1]Department of Applied Mathematics

University of Colorado at Boulder, USA

[2]Geophysical Statistics Project

National Center for Atmospheric Research, Boulder, CO

[3]Laboratiore des Sciences du Climat et de l'Environnement

IPSL-CNRS, Gif-sur-Yvette, France

cooleyd@colorado.edu

phone: 303-492-4668

fax: 303-492-4066

July 22, 2005

Abstract

Quantification of precipitation extremes is important for flood planning purposes, and a common measure of extreme events is the r-year return-level. We present a method for producing maps of precipitation return levels and uncertainty measures and apply to a Colorado region. Separate hierarchical models are constructed for the intensity and the frequency of extreme precipitation events. For intensity, we model daily precipitation above a high threshold at 56 weather stations with the Generalized Pareto Distribution. For frequency, we model the number of exceedances at the stations as binomial random variables. Both models assume that the regional extreme precipitation is driven by a latent spatial process characterized by geographical and climatological covariates. Effects not fully described by the covariates are captured by spatial structure in the hierarchies. Spatial methods were improved by working in a space with climatological coordinates. Inference is provided by an MCMC algorithm and spatial interpolation method, which provide a natural method for estimating uncertainty.

Keywords: Colorado, Extreme Value Theory, Generalized Pareto Distribution, Hierarchical Model, Latent Process

1 Introduction

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. 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. 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.

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 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 paper 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. More details for this extension will be given in the discussion section.

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

We illustrate our methods by applying them to the Front Range of Colorado. 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.

1.2 Extreme value statistics

Extreme value theory (EVT) provides statistical models for the tail of a probability distribution and is a complement to modeling the mean or central part of a distribution. EVT for univariate data is well understood and is based on the asymptotic arguments that lead to the generalized extreme value (GEV) distribution. Given iid continuous data $Z_1, Z_2, \ldots, Z_n$, and letting $M_n = \max(Z_1, Z_2, \ldots, Z_n)$, it is known that if the normalized distribution of M_n converges as $n \to \infty$, then it converges to a GEV (Fisher and Tippett, 1928; Gnedenko, 1943; Von Mises, 1954). Because of its asymptotic justification, the GEV is used to model maxima of finite-sized blocks such as annual maxima. However, if daily observations are available, models which use only each year's annual maximum disregard other extreme data which could provide additional information. Another distribution from EVT, the Generalized Pareto Distribution (GPD) is based on the exceedances above a threshold rather than just the annual maxima, and it also has an asymptotic justification (Pickands, 1975). Exceedances (the amounts which observations exceed a threshold u) should approximately follow a GPD as u gets large and sample size increases. In this case, the tail of the distribution is characterized by the equation

$$\mathbb{P}(Z > z + u | Z > u) = \left(1 + \xi \frac{z}{\sigma_u}\right)_+^{-1/\xi}, \tag{1}$$

where $a_+ = a$ if $a \geq 0$ and $a_+ = 0$ if $a < 0$. The scale parameter σ_u must be greater than zero, and the shape parameter ξ controls whether the tail is bounded ($\xi < 0$), light ($\xi \to 0$), or heavy ($\xi > 0$). In practice, a threshold is chosen at a level where the data above it approximately follow a GPD and the shape and scale parameters are estimated. In our precipitation study, the GPD provides the foundation for our modeling procedure. For more background on EVT, Embrechts (1997) is an excellent source, and Coles (2001b) gives a good introduction of its statistical applications.

EVT provides the link between data recorded on a daily (or hourly) time frame and quantities of longer time scales such as return levels. From basic probability rules and Equation (1)

$$\mathbb{P}(Z > z + u) = \zeta_u \left(1 + \xi \frac{z}{\sigma_u}\right)^{-1/\xi} \text{ with } \zeta_u = \mathbb{P}(Z > u). \tag{2}$$

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 $\mathbb{P}(Z > z_r) = \frac{1}{rn_y}$ for z_r:

$$z_r = u + \frac{\sigma_u}{\xi}\left[(rn_y\zeta_u)^{\xi} - 1\right]. \tag{3}$$

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

1.3 Extremes, climate and weather

Although the tools for statistically modeling univariate extremes are well-developed, extending these tools to model spatial extreme data is a very active area of research. For this application, our focus is on how the *distribution* of precipitation varies over space, not the multivariate structure of particular precipitation events. This is a distinction between climate and weather. The climate for a particular location and specific meteorological variables is the distribution over a long period of time. Weather is the realization of the process and the climatological distribution is found by aggregating over many weather events.

Much of the work in multivariate and spatial extremes has centered around describing the dependence of extreme observations (de Haan (1985); Schlather and Tawn (2003); Naveau et al. (2005); Coles et al. (1999); Heffernan and Tawn (2004)). In the case of precipitation, one would expect that the spatial dependence of the extreme observations, i.e. weather, would have a short range. Particularly with localized thunderstorms that are associated with convective precipitation, measurements can vary largely over a region. In contrast, an r-year precipitation return level is a climatological quantity rather than a quantity associated with the weather observations of a particular day or year. The rth-year return level of two locations tens or even hundreds of kilometers apart might be dependent, provided the locations have similar climates.

1.4 Spatial dependence of extremal distributions

Let $Z(\boldsymbol{x})$ denote the total precipitation for a given period of time (e.g. 24 hours) and at location $\boldsymbol{x}$. From the discussion above, our goal is to provide inference for the probability $P(Z(\boldsymbol{x}) > z + u)$ for all locations, $\boldsymbol{x}$, in a particular domain and for u large. Given this survival function, one can then compute return levels and other summaries for the tail probabilities. Because the distributions now 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 σ_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 paper. Traditionally, $\boldsymbol{x}$ represents a location in a space whose coordinates are given by longitude and latitude. Alternatively, in addition to working with space in the usual sense, we 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 3.1, and it will be clear by the context what spatial coordinates are being used.

In a Bayesian setting, the functions $\sigma_u(\boldsymbol{x})$ $\xi(\boldsymbol{x})$ 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. (2005) 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.

1.5 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 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 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.

1.6 Outline

The next section describes the precipitation data sources and some preliminary analysis of the choice of thresholds. In Section 3 we describe the models which produce the return level map. We discuss the GPD-based hierarchical model for threshold exceedances in Section 3.1, discuss the method for modeling the threshold exceedances rate in Section 3.2, briefly describe our MCMC method for model inference in Section 3.3, and discuss how our model was interpolated on the region in Section 3.4. We then present our method for model selection and results in Section 4. We conclude with a discussion in Section 5.

2 Data

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 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 most of the precipitation falls as rain during this period, which has different flood characteristics than snowfall. The region is semi-arid; the city of Denver receives approximately 29cm (11.5 inches) of precipitation during the months of April-October. Most of the 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 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 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.

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 3 shows a 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 3) 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 .64cm (.25 inches) might be adequate. However, we choose to set the threshold at (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.

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. 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. We tested for spatial dependence in the annual maximum residuals of the stations using a first-order variogram (Naveau et al. (2005)) and found that there was a low level of dependence between stations within 24km (15 miles) of one another and no detectable dependence beyond this distance. 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.

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.

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 3.1 and 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 (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 $\sigma_u(\boldsymbol{x})$, $\xi(\boldsymbol{x})$, 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 $\boldsymbol{\sigma}_u = [\sigma_u(x_1), \ldots, \sigma_u(x_s)]^T$ and we similarly define $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}_u$. The two hierarchies model $\boldsymbol{\sigma}_u, \boldsymbol{\xi}, \boldsymbol{\zeta}_u$ and parameters which relate these processes to orographic and climatological information. The process $\sigma(\boldsymbol{x})$ given $\boldsymbol{\sigma}_u$ is 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:

$$p(\boldsymbol{\theta}|\boldsymbol{Z}(\vec{\boldsymbol{x}})) \propto p(\boldsymbol{Z}(\vec{\boldsymbol{x}})|\boldsymbol{\theta})\ p(\boldsymbol{\theta}), \tag{4}$$

where p denotes a probability density. Based on the conditional distributions of our hierarchical model Equation (4) becomes

$$p(\boldsymbol{\theta}|\boldsymbol{Z}(\vec{\boldsymbol{x}})) \propto p_1(\boldsymbol{Z}(\vec{\boldsymbol{x}})|\boldsymbol{\theta}_1)\ p_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)\ p_3(\boldsymbol{\theta}_2). \tag{5}$$

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

3.1 Hierarchical Model for Threshold Exceedances

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 $\phi = \log \sigma_u$, 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:

$$\mathbb{P}\{Z_k(\boldsymbol{x}_i) - u > z | Z_k(\boldsymbol{x}_i) > u\} = \left(1 + \frac{\xi(\boldsymbol{x}_i) z}{\exp \phi(\boldsymbol{x}_i)}\right)^{-1/\xi(\boldsymbol{x}_i)}. \tag{6}$$

Differentiating the distribution function associated with (6) to obtain a probability density we get the first piece of Equation (5):

$$p_1(\boldsymbol{Z}(\vec{\boldsymbol{x}}) | \boldsymbol{\theta}_1) = \prod_{i=1}^{s} \prod_{k=1}^{n_i} \frac{1}{\exp \phi(\boldsymbol{x}_i)} \left(1 + \frac{\xi(\boldsymbol{x}_i) z}{\exp \phi(\boldsymbol{x}_i)}\right)^{-1/\xi(\boldsymbol{x}_i) - 1}. \tag{7}$$

where $\boldsymbol{\theta}_1 = [\boldsymbol{\phi}, \boldsymbol{\xi}]^T$.

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 2). In effect the climate space reduces the number of regions in the model producing a much stronger spatial signal. As shown in Section 4, our models performed better in the climate space.

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:

$$\mu_\phi(\boldsymbol{x}) = f_\phi(\boldsymbol{\alpha}_\phi, covariates). \tag{8}$$

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 between stations and parameters $\boldsymbol{\beta}_\phi$ and it is given by:

$$k_\phi(\boldsymbol{x}, \boldsymbol{x}') = \beta_{\phi,0} \times \exp(-\beta_{\phi,1} \times ||\boldsymbol{x} - \boldsymbol{x}'||) \tag{9}$$

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) is

$$p_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) = \frac{1}{\sqrt{(2\pi)^s|\Sigma|}} \exp\left[-\frac{1}{2}(\boldsymbol{\phi}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{\phi}-\boldsymbol{\mu})\right] \times p_\xi(\boldsymbol{\xi}|\boldsymbol{\theta}_\xi) \tag{10}$$

where $\boldsymbol{\mu}$ is a vector defined by Equation (8) evaluated at the covariates of the locations $\boldsymbol{x}_i$, Σ is the covariance matrix generated by Equation (9) 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 $\boldsymbol{\theta}_2 = [\boldsymbol{\alpha}_\phi, \boldsymbol{\beta}_\phi, \boldsymbol{\theta}_\xi]^T$.

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 3.3), we are able to choose improper and uninformative priors for the regression parameters $\boldsymbol{\alpha}_\phi$. For all the models tested, $\alpha_{\phi,i} \sim Unif(-\infty, \infty)$.

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 $Unif(1, 6)$.

With the priors set as above, the third piece of Equation 5 is:

$$p_3(\boldsymbol{\theta}_2) = p_{\alpha_\phi}(\boldsymbol{\alpha}_\phi) \times p_{\beta_\phi}(\boldsymbol{\beta}_\phi) \times p_{\alpha_\xi}(\boldsymbol{\alpha}_\xi) \times p_{\beta_\xi}(\boldsymbol{\beta}_\xi) \propto 1 \times p_{\beta_\phi}(\boldsymbol{\beta}_\phi) \times 1 \; p_{\beta_\xi}(\boldsymbol{\beta}_\xi), \tag{11}$$

and the model for threshold exceedances is completely specified.

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 (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 4.

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 (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 defined as in equations (8) and (9). 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 4.

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 (3) $z_r(\boldsymbol{x})$ 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 6, 7). 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 8, 9).

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 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 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) 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 6). 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 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 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 7) 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 two models; Figure 8 shows the point estimate of the 25 year daily precipitation return level. 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 9).

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 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 for this paper 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 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 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 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. An important extension would be to make a comprehensive model for all duration periods.

Typically, data from different duration periods are modeled separately, and not being coupled, it is possible that the return level estimate of a 12-hour duration period could be higher than that of a 24-hour duration period. We propose a Bayesian model where the data from all duration periods are pooled and the GPD parameters are not only functions of location, but also duration period.

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.

Acknowledgments: The authors would like to thank Rick Katz, Mary Downton, and Rebecca Morss from the Institute for the Study of Society and the Environment at the National Center for Atmospheric Research, Boulder, CO. This project was funded by the Weather and Climate Impacts Initiative (NCAR) and by NSF VIGRE grant #DMS-9810751 (CU). Part of Philippe Naveau's work was supported by the E2-C2 European grant.

References

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.

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

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

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., Jomelli, V., Rababtel, A., and Grancher, D. (2005). A bayesian hierarchical extreme value model for lichenometry. *Environmetrics (in press)*.

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.

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

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

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.

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.

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.

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

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

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.

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.

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.

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

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., Cooley, D., and Poncet, P. (2005). Variograms for spatial max-stable random fields. *Submitted*.

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

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

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

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

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

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.*, pages 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.

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.

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.

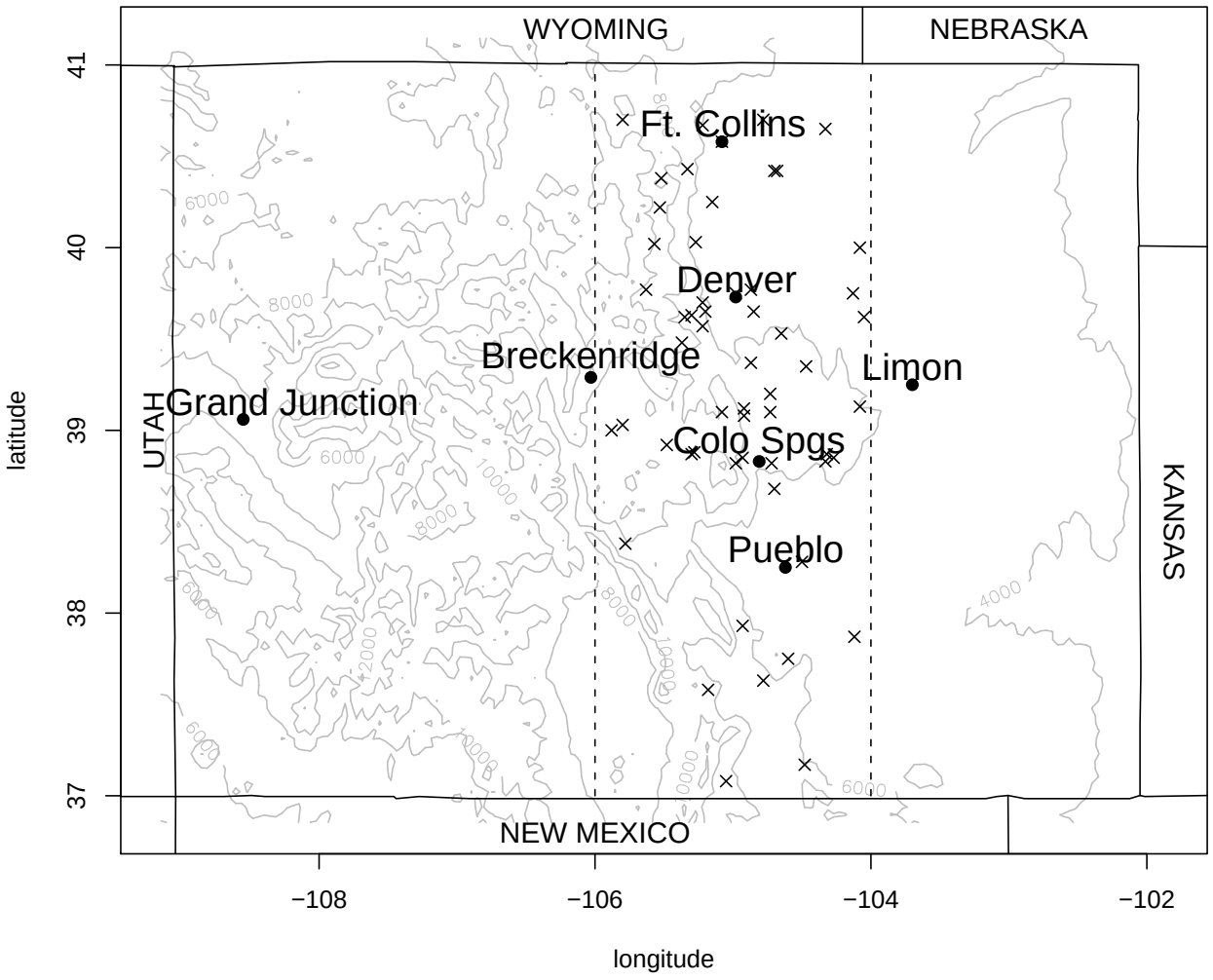

Figure 1: Map of Colorado shows the study location within the dashed lines. Contours show elevation which is given in meters. Station locations are marked with an "X".

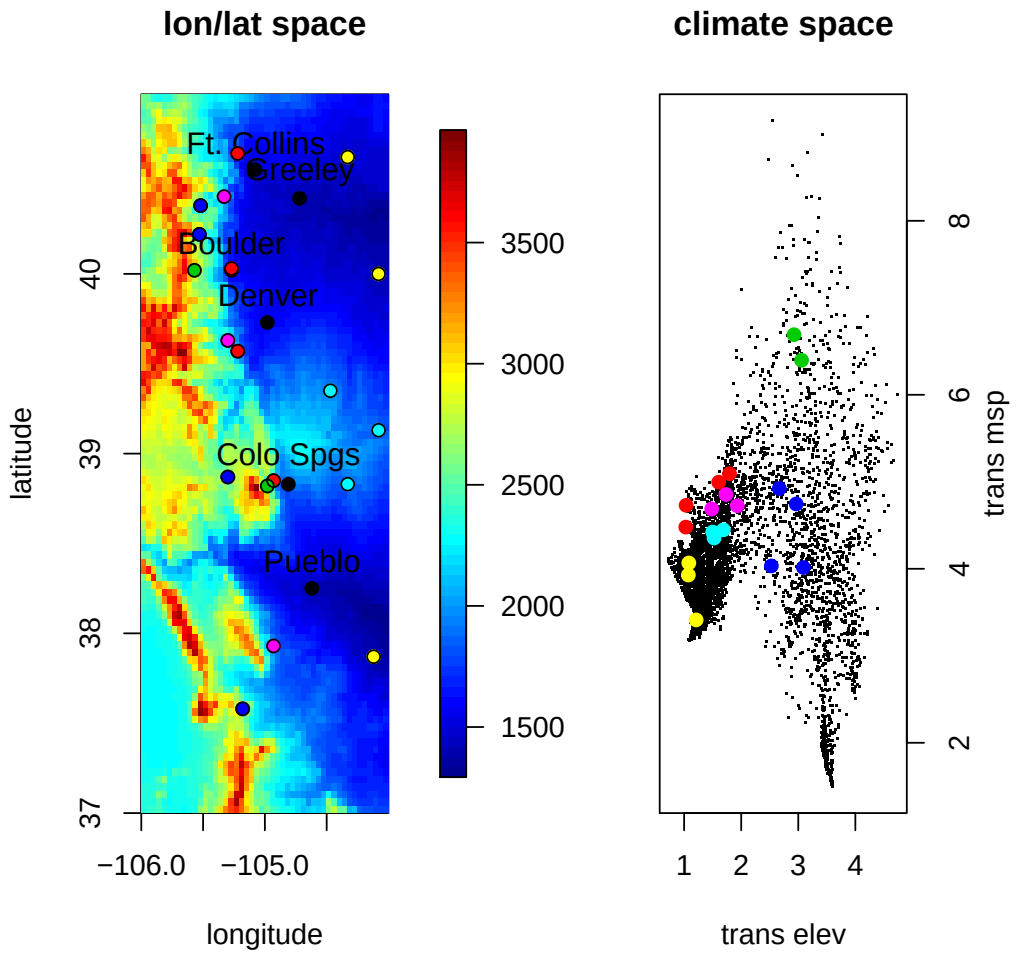

Figure 2: Translates points in longitude/latitude space to points in the climate space. Colored dots all correspond to selected station locations and represent foothills (red), plains (yellow), Palmer Divide (cyan), Front Range (magenta), mountain valley (blue), and high elevation (green).

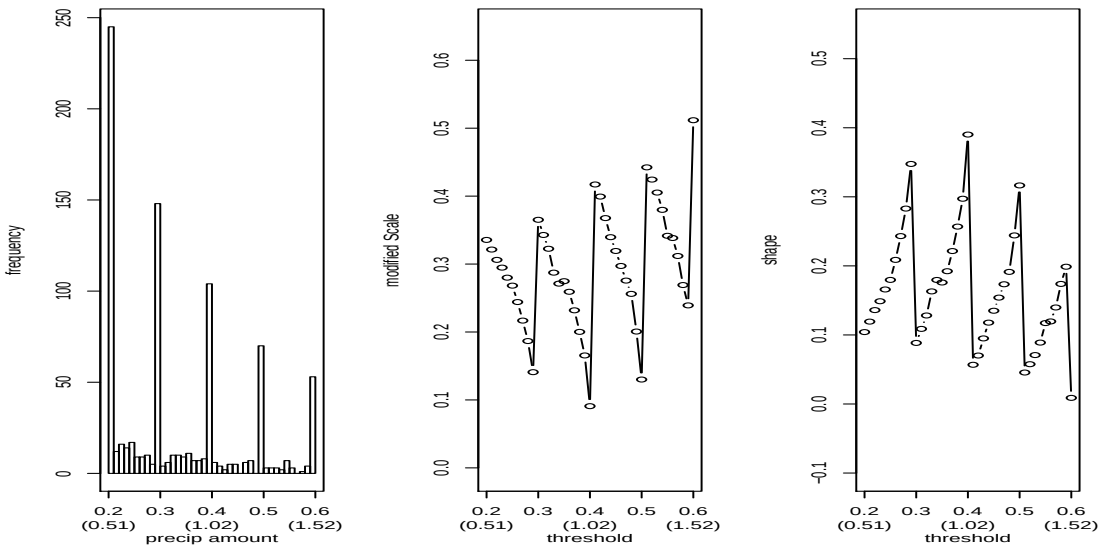

Figure 3: Shows the bias due to the lack of precision in the recorded data. The left plot shows a partial histogram of the Boulder station data which illustrates that most of the data is recorded to the nearest 1/10th of an inch. The center plot shows the mle parameter estimate for the modified scale 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).

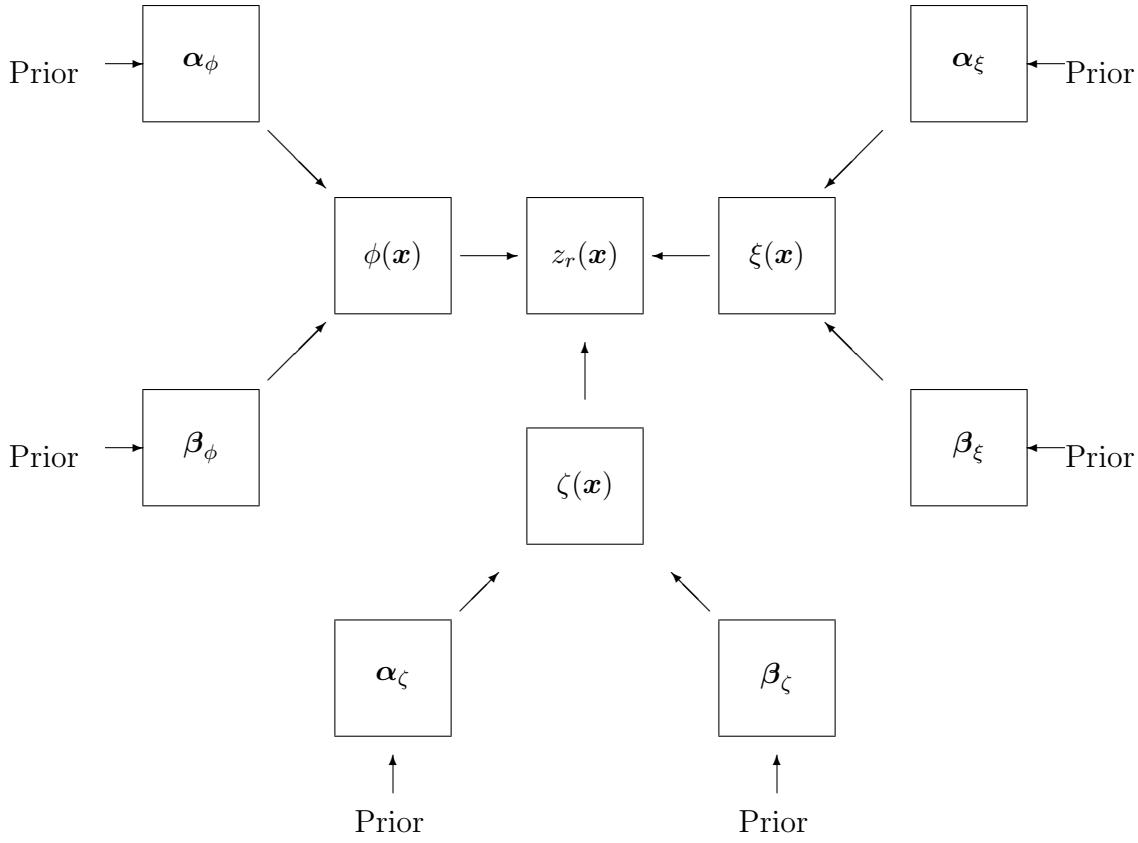

Figure 4: Schematic of the model used to estimate the return level $z_r(\boldsymbol{x})$. The return level is a function of the GPD parameters $\phi(x)$ and $\xi(x)$, and of the exceedance rate parameter $\zeta(x)$. All of these can be modeled spatially by a Gaussian process where the parameters $\alpha_\cdot$ describes the mean structure, and $\beta_\cdot$ describes the covariance structure.

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

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

Table 1: Shows several of the different GPD hierarchical models tested and corresponding DIC scores. Models in the climate space had better scores than in the longitude/latitude space.

Models in Latitude/Longitude Space			$\bar{D}$	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 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.

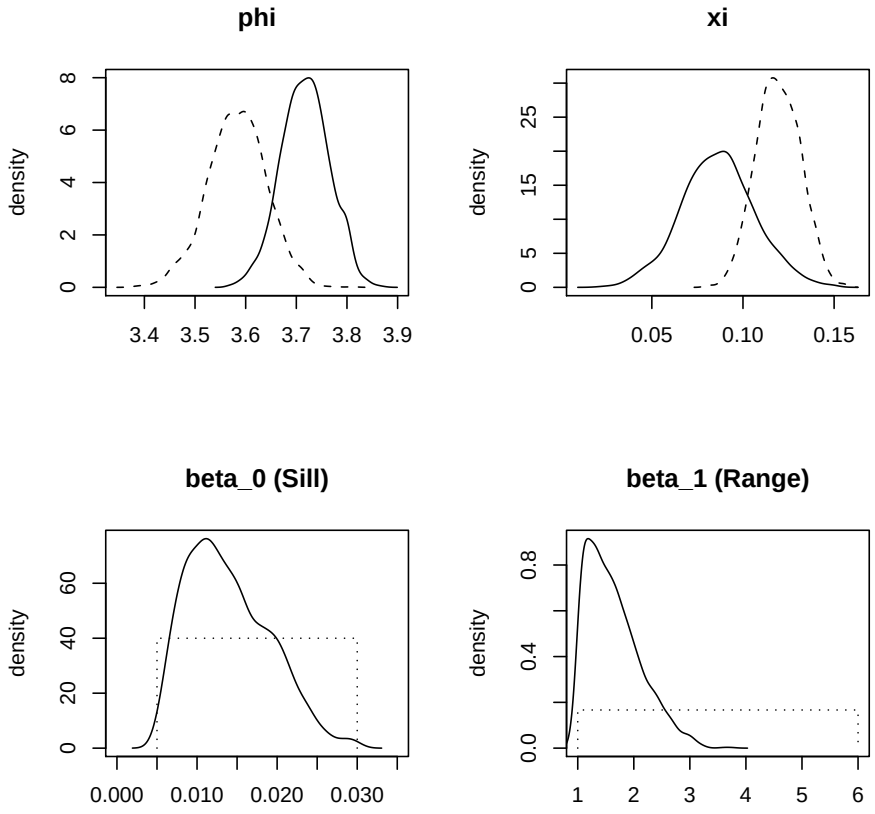

Figure 5: Posterior densities from Model 7. The upper left shows ϕ parameters associated with the Boulder station which has data for 54 years (solid) and the Greeley/UNC station which has data for 16 years (dashed). Upper right shows the parameters ξ for the mountain stations (solid) and the plains stations (dashed). 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).

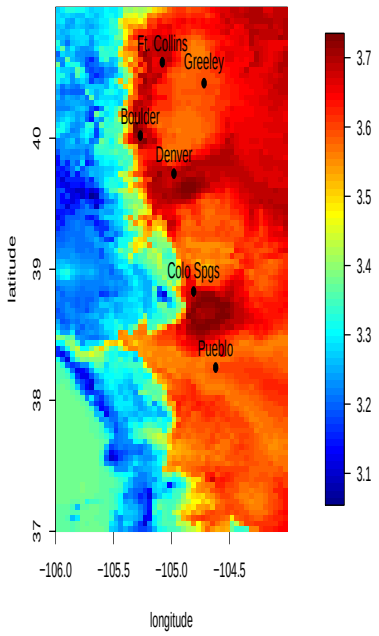

Figure 6: Point estimate for the log-transformed GPD scale parameter ϕ for Model 7.

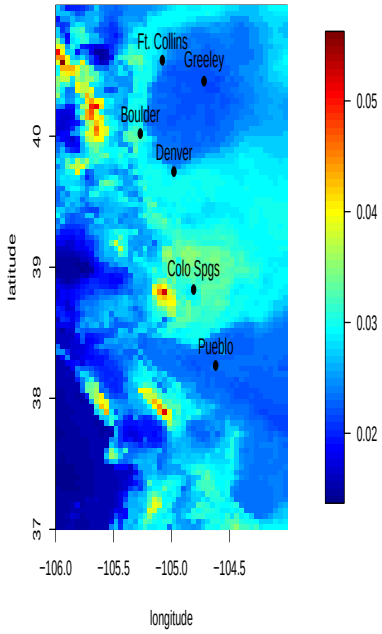

Figure 7: Point estimate for the exceedance rate parameter ζ for Model C.

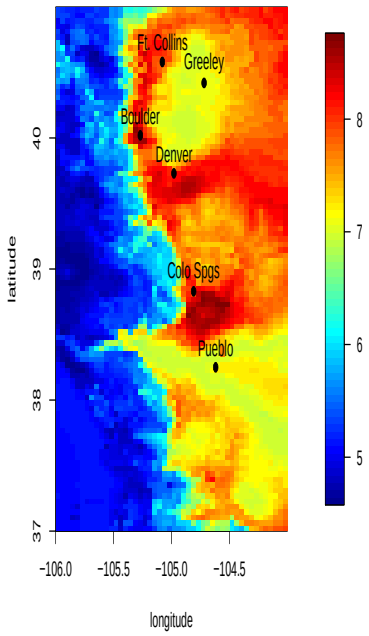

Figure 8: Point estimate for the daily precipitation 25 year return level as described by Models 7 and C. Scale is in centimeters.

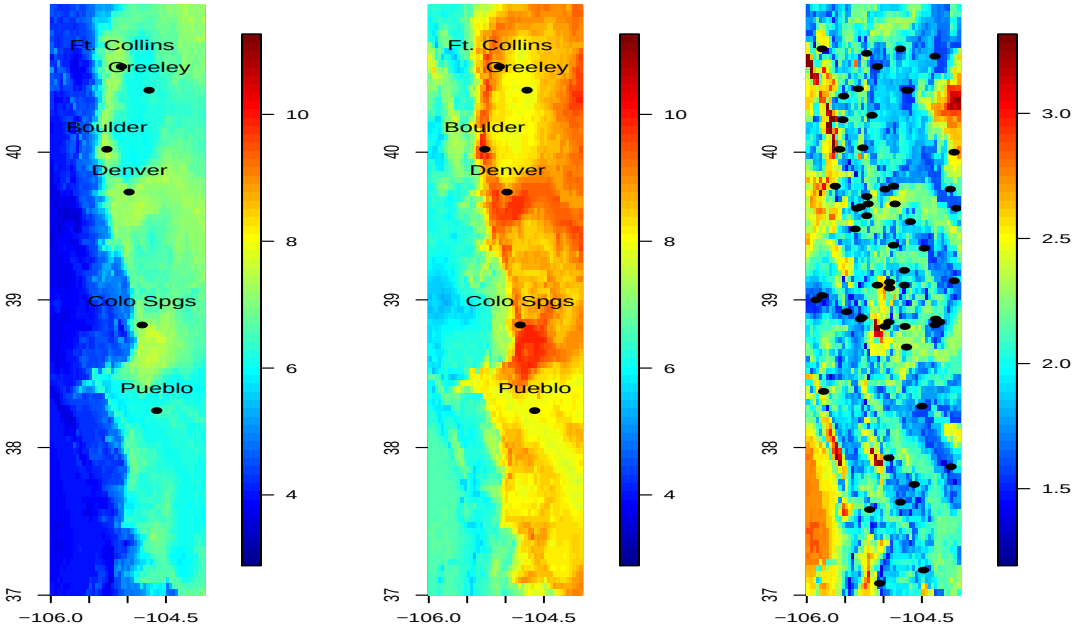

Figure 9: Left plot is the .025 quantile of the daily precipitation 25 year return level. Center plot is the .975 quantile, and right plot is the difference of the two, or an estimate of the range of 95% credible interval (with station locations plotted).

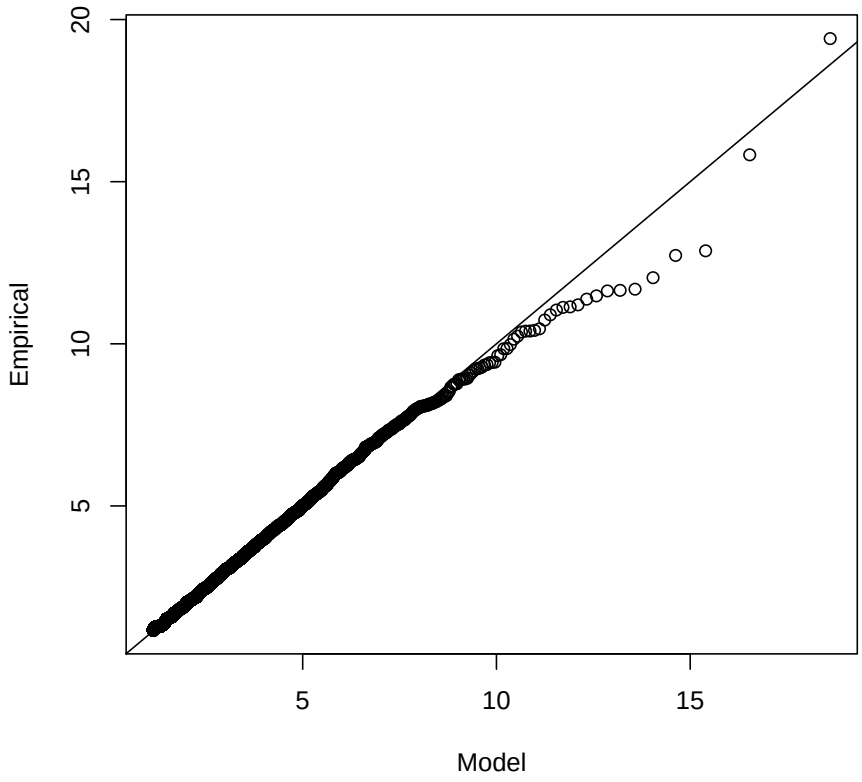

Figure 10: Shows a qq plot of all the threshold exceedances. Each station's data was normalized using the posterior mean of ϕ and ξ. The data was then converted to have a common GPD $\phi = 3.7$ $\xi = 1.2$ distribution. Axes labels are in centimeters