Bayesian Spatial Modeling of Extreme Precipitation Return Levels

Daniel COOLEY, Douglas NYCHKA, and Philippe NAVEAU

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 it to a region in Colorado. 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 a Markov chain Monte Carlo algorithm and spatial interpolation method, which provide a natural method for estimating uncertainty.

KEY WORDS: Colorado; Extreme value theory; Generalized Pareto distribution; Hierarchical model; Latent process.

1. INTRODUCTION

On July 28, 1997, a rainstorm in Fort Collins, Colorado, produced a flood that killed five people and caused $250 million in damage. Similar large precipitation events in Colorado caused the 1976 Big Thompson flood near Loveland, which killed 145 people, and the 1965 South Platte flood, which caused $600 million in damages around Denver (National Weather Service 2005). 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 article we address the problem through the use of a Bayesian hierarchical model that leverages statistical extreme value theory. This approach has the advantage that one is able to quantify the uncertainty of these maps due to the limited amount of data and their sparse spatial representation. Here we report the results of a pilot study of 24-hour precipitation extremes for the Front Range region of Colorado. Although this statistical research focuses on hydrometeorological extremes, we note that our spatial models are not limited to this context, and 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 National Weather Service (NWS) maintains precipitation atlases and a companion digital database (see *hdsc.nws.noaa.gov/hdsc/pfds/pfds_maps.html*) 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 United States (Bonnin et al. 2004a) and the mid-Atlantic states (Bonnin et al. 2004b). Neither of these regions encompasses Colorado, 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 that has probability $1/r$ of being exceeded in a particular year. 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 2–500 years. In this article, 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.

We illustrate our methods by applying them to the Front Range of Colorado. The most recent NWS precipitation atlas for Colorado was produced in 1973 (Miller, Frederick, and Tracey 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 that 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.

Daniel Cooley is Postdoctoral Researcher, Department of Statistics, Colorado State University, Fort Collins, CO 80523, and Postdoctoral Researcher, Geophysical Statistics Project, National Center for Atmospheric Research, Boulder, CO 80307 (E-mail: *cooleyd@ucar.edu*). Douglas Nychka is Director, Institute for Mathematics in Geosciences, National Center for Atmospheric Research, Boulder, CO 80307. Philippe Naveau is Research Scientist, Laboratoire des Sciences du Climat et de l'Environnement, IPSL-CNRS, Gif-sur-Yvette, France, and Adjunct Professor, Department of Applied Mathematics, University of Colorado, Boulder, CO 80309. The authors thank two anonymous reviewers for their comments, which greatly improved the article. The authors also thank Rick Katz, Mary Downton, Rebecca Morss, and Claudia Tebaldi from the Institute for the Study of Society and the Environment at the National Center for Atmospheric Research, Boulder, CO. Dan Cooley's work was funded by NSF VIGRE grant DMS-98-10751 at the University of Colorado, by the Geophysical Statistics Project (NCAR) DMS-03-55474, by PRIMES (CSU) DGE-0221595, and by the Weather and Climate Impacts Initiative (NCAR). Part of Philippe Naveau's work was supported by a E2-C2 European grant.

Journal of the American Statistical Association
September 2007, Vol. 102, No. 479, Applications and Case Studies
DOI 10.1198/016214506000000780

1.2 Extreme Value Statistics

Extreme value theory (EVT) provides statistical models for the tail of a probability 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 that use only each year's annual maximum disregard other extreme data that 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 likewise has an asymptotic justification (Pickands 1975). Exceedances (the amounts that observations exceed a threshold u) should approximately follow a GPD as u becomes large and sample size increases. In this case, the tail of the distribution is characterized by the equation

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

where $a_+ = a$ if $a \geq 0$ and $a_+ = 0$ if $a < 0$. The scale parameter σ_u must be greater than 0, 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. For more background on EVT, see Embrechts, Klüppelberg, and Mikosch (1997); the book by 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 (1),

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

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:

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

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

1.3 Spatial Dependence for Climatological Extremes

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. Much of the work in multivariate and spatial extremes has centered around describing the dependence of extreme observations (de Haan 1985; Coles, Heffernan, and Tawn 1999; Schlather and Tawn 2003; Heffernan and Tawn 2004; Cooley, Naveau, and Poncet 2006). Such work could be used to model the multivariate structure of extreme weather events.

In contrast to weather that describes the state of the atmosphere at a given time, the climate for a particular location is the distribution over a long period of time. Return levels are climatological quantities, and their spatial dependence must be modeled outside of the framework provided by the multivariate analyses cited previously. In this application, our focus is on how the *distribution* of precipitation varies over space, not the multivariate structure of particular precipitation events.

Let $Z(\mathbf{x})$ denote the total precipitation for a given period of time (e.g., 24 hours) and at location $\mathbf{x}$. From the preceding discussion, our goal is to provide inference for the probability $P(Z(\mathbf{x}) > z + u)$ for all locations, $\mathbf{x}$, 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 a location $\mathbf{x}$ in the study area. We assume that the values of $\sigma_u(\mathbf{x})$, $\xi(\mathbf{x})$, and $\zeta_u(\mathbf{x})$ result from a latent spatial process that characterizes the extreme precipitation and arises from climatological and orographic effects. The dependence of the parameters characterizes the similarity of climate at different locations.

Notationally, $\mathbf{x}$ denotes a location in a generic sense, and it takes on two meanings in this article. Traditionally, $\mathbf{x}$ 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 (also denoted by $\mathbf{x}$). The coordinates of $\mathbf{x}$ in the climate space are given by orographic and climatological measures. Our reason for working in the climate space is explained in Section 3.1, and it will be clear by the context what spatial coordinates are being used.

To understand the latent spatial process that drives the climatological dependence of precipitation extremes, we implement a Bayesian approach that 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 the considered return periods. Although there have been several studies using Bayesian methods in extremes (e.g., Smith and Naylor 1987; Coles and Tawn 1996a; Bottolo, Consonni, Dellaportas, and Lijoi 2003; Stephenson and Tawn 2005) only a few have built models that borrowed strength across different spatial locations. Coles (2001a) and Jagger and Elsner (2006) proposed information-sharing models that pooled data from various locations to better estimate the parameters for EVT distributions. Neither work, however, attempted to model any spatial nature of the extremes. Cooley, Naveau, Jomelli, Rabatel, and Grancher (2006) built a Bayesian hierarchical GEV model that pooled lichenometry data from different locations, but the model was not fully spatial as it did not utilize location information. 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. To our knowledge, this is the first extremes study that employs Bayesian hierarchical models to study both the intensity and the frequency of extremes, models a two-dimensional latent spatial process, and spatially interpolates the results.

1.4 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 multiple-step procedure that first defines homogeneous spatial regions, normalizes the data from each region using an index flood measure (e.g., the mean of annual maxima), and fits a regional probability distribution to the pooled data. RFA then scales the regional distribution by each station's index flood measure to produce local distributions. L moments (a method-of-moments type of 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, Parlange, and Naveau 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. Fowler and Kilsby (2003) and Fowler, Ekstrom, Kilsby, and Jones (2005) studied regional precipitation in the United Kingdom via RFA. For its current precipitation atlas update project, the NWS uses RFA to obtain probability distributions at the station locations and then produces their maps by using the parameter-elevation regression on independent slopes model (PRISM) method (Daly, Neilson, and Phillips 1994; Daly, Gibson, Taylor, Johnson, and Pasteris 2002) to spatially interpolate their index flood measure. Several studies followed the aforementioned Fort Collins flood, and one used RFA methods to estimate the return period associated with the event (Sveinsson, Salas, and Boes 2002). Because Fort Collins is situated at the foothills of the mountains, the authors found it difficult to define a region that satisfactorily explained precipitation in the city.

The method we present in Section 3 differs in several ways from the RFA-based studies discussed previously. 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 stations' data. Our method does not require one to define homogeneous regions at the outset of the analysis but does allow one to define different subregions as part of the model-fitting process. 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 obtain measures of uncertainty from the estimated posterior distributions. Finally, because RFA is a multiple-step algorithm, it is difficult to assess the error propagation through the steps. In comparison, uncertainty measures result naturally from the Bayesian framework of our approach, and uncertainty due to both parameter estimation and interpolation are taken into account when producing the maps.

1.5 Outline

Section 2 describes the precipitation data sources and issues such as threshold selection and declustering. In Section 3 we describe the models that produce the return level map. We discuss the GPD-based hierarchical model for threshold exceedances in Section 3.1, present the method for modeling the threshold exceedances rate in Section 3.2, briefly describe our Markov chain Monte Carlo (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

Our study region in Colorado 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 (Fig. 1). The mountainous area has peaks in excess of 4,270 m and meets the plains region along what is called the Front Range. The Front Range includes the cities of Fort Collins, Boulder, Denver, and Colorado Springs. The plains region extends more than 100 km east of the Front Range, where the elevation is less than 1,400 m, and is divided into two major drainages separated by the Palmer Divide, an area of increased elevation between the cities of Colorado Springs and Denver. This study of extreme precipitation began as 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. We study only precipitation that occurs between April 1 and October 31 because most of the precipitation falls as rain during this period, which has different flood characteristics than snowfall. The region is semiarid; the city of Denver receives approximately 29 cm 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.

Our data come from 56 weather stations scattered throughout the study region (Fig. 1). These stations record hourly precipitation amounts, but for this study only the daily totals are used. We apply our method to data recorded during the years 1948–2001. Included in the 56 are stations that discontinued operation during the period of study and others that 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

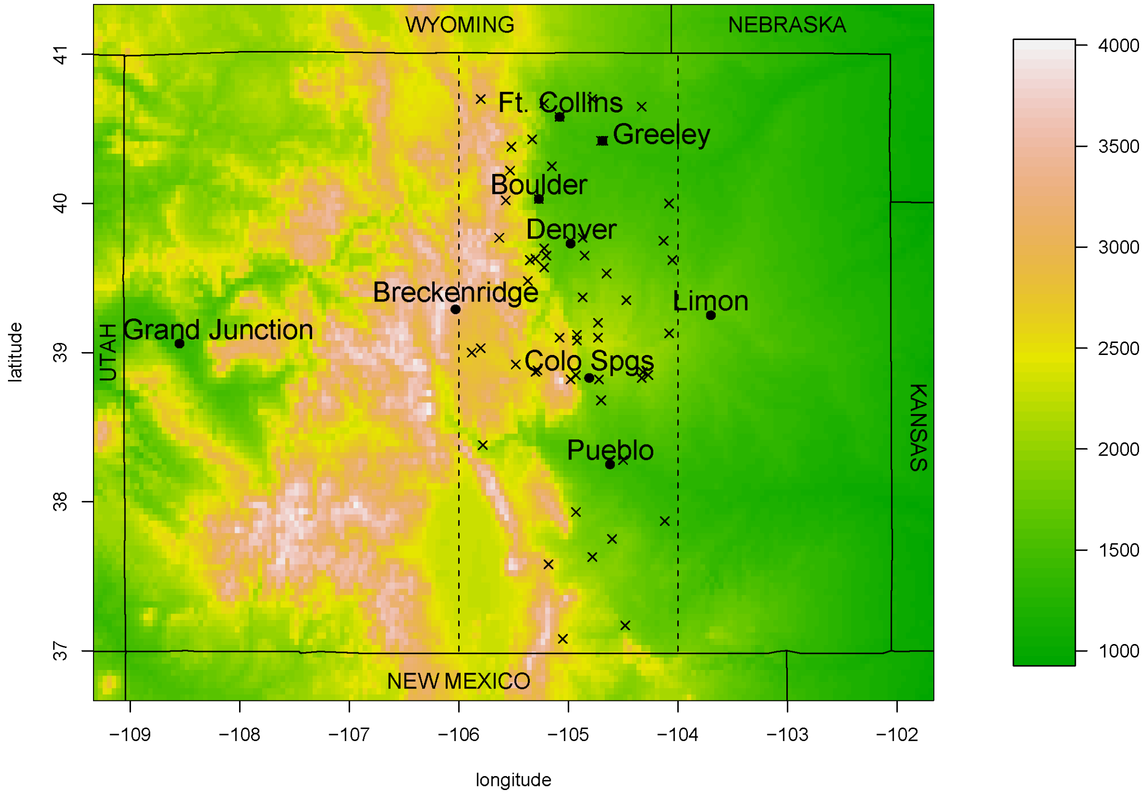

Figure 1. Map of Colorado showing study location within dashed lines and elevation (in meters). Station locations are marked with a "×."

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 dataset in England, Coles and Tawn (1996b) found that mean precipitation was a stronger covariate for extreme precipitation than elevation. Because the mean precipitation data are highly correlated with the elevation data and take into account other factors such as aspect, slope, and meteorology, we expect at the outset that it may be the stronger of the two covariates.

2.2 Data Precision and Threshold Selection

Figure 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 $1/100$th of an inch (.25 mm), and after 1971 to the nearest $1/10$th of an inch (2.5 mm). All but three stations similarly switched their level of precision around 1970. Although the actual precipitation follows a continuous distribution, daily precipitation measurements recorded to the nearest $1/10$th of an inch have relatively low precision, making the data practically discrete.

The low-precision data introduce a problem with threshold selection. Figures 2(b) and 2(c) show the maximum likelihood estimates for the threshold-independent scale ($\sigma^* = \sigma_u - \xi u$) and shape parameters of the GPD as a function of threshold. Above a high enough threshold, these parameters should be constant; however, a sawtooth pattern is clearly evident for both as one moves the threshold through the precision interval.

EVT is based on the assumption of a continuous distribution, and how to handle truncated data such as we have is unclear. One method would be to build data truncation into the model. Jagger and Elsner (2006) built data truncation into their hurricane wind speed model using a Bayesian approach. They assumed that the true wind speed was uniformly distributed around the observed wind speed. However, it is not clear what effect assuming a uniform (or other) distribution might have on modeling the tail behavior. A second approach would be to adjust the likelihood to account for the data truncation. Rather than defining the likelihood in terms of the GPD density, one could define the likelihood to be $l(\mathbf{x};\theta) = \prod_{i=1}^{n} \frac{1}{d} G(x_i + d/2; \sigma_u, \xi) - G(x_i - d/2; \sigma_u, \xi)$, where $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ represents the truncated data, G represents the generalized Pareto distribution function, and d represents the length of the interval. This likelihood could be integrated into either a maximum likelihood or a Bayesian model. A third approach would be to simply find the threshold location in the precision interval at which the parameters show the least bias.

A full investigation of the precision issue is beyond the scope of this article, but two simple simulation experiments were performed to decide how best to proceed. The first experiment found the threshold at which the parameter bias was minimized. Random samples were drawn from a GPD with similar parameters to the precipitation data and then rounded to the nearest $1/10$th. Parameters were estimated at various thresholds for both the rounded and the original data, and the bias was least

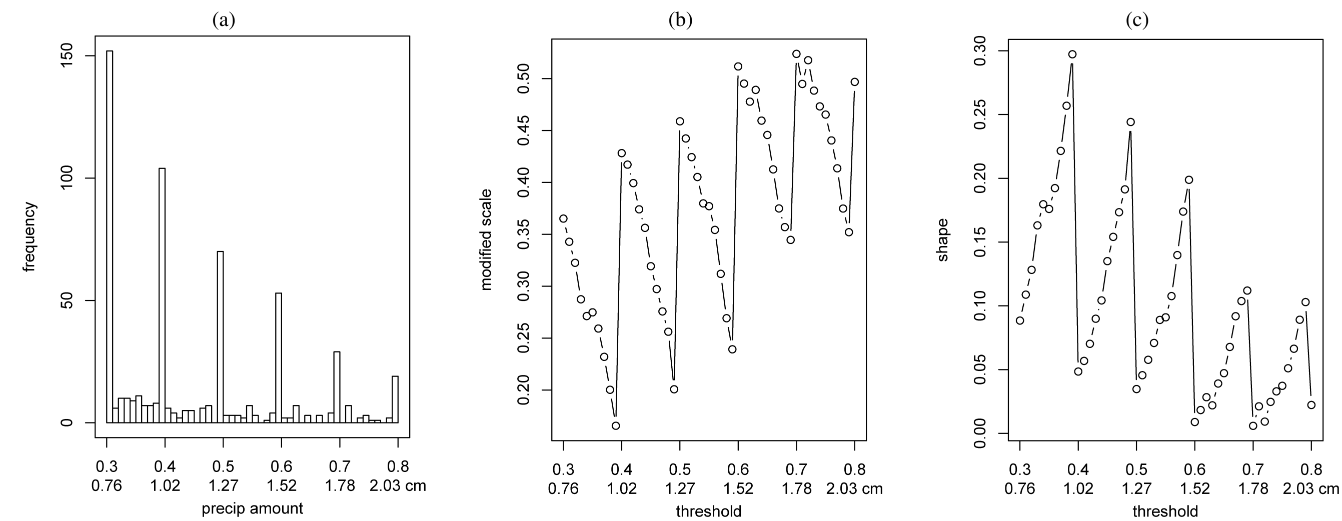

Figure 2. Bias due to lack of precision in recorded data. Part (a) shows a partial histogram of the Boulder station data, which illustrates that most of the data are recorded to the nearest $1/10$th of an inch. Part (b) shows the MLE parameter estimate for the modified scale parameter, and part (c) shows the estimate for ξ. Parts (b) and (c) 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 (in centimeters below).

when the threshold was set at the center of the precision interval, that is, at .35, .45, or .55 inches (.89, 1.14, or 1.40 cm). A second experiment compared the interval likelihood method to the center-threshold analysis. Once again, GPD samples were simulated and rounded, and then a GPD was fit to the rounded data. Both methods yield very similar parameter estimates and the mean squared error of the 25-year return level for the interval method was slightly larger than that of simply choosing the middle of the threshold interval.

Based on the results of the simulations, we proceed by choosing a threshold in the center of a precision interval; however, we must still choose one at an appropriate level to model the exceedances with the GPD. Choosing a threshold is a delicate procedure as there is a classic bias versus precision trade-off. Usually threshold selection is done using diagnostic plots that show how quantities such as the mean residual life or shape parameter vary as the threshold changes (Coles 2001b). For most of the stations in our study, the diagnostic plots seem to indicate that a threshold of around .45 inches (1.14 cm) would be adequate. However, a threshold sensitivity analysis of model runs (Sec. 4.2) indicates that the shape parameter is more consistently estimated above a threshold of .55 inches (1.40 cm). Setting the threshold at this level, we had 7,789 exceedances (after declustering), which represents 2.0% of the original data or 9.0% of the nonzero observations.

2.3 Residual Dependence in Observations and Declustering

The EVT distributions are based on iid data, and 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. In our case, conditional independence may not 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. It has been shown that maxima of stationary series still converge in distribution to a GEV, provided that dependence is short range and that extremes do not occur in clusters (Embrechts et al. 1997). In practice, short-range dependence can often be assumed, and thus temporal dependence is dealt with via declustering (Coles 2001b). If a station had consecutive days that exceeded the threshold, we declustered the data by keeping only the highest measurement, though declustering the data did not noticeably change the results. Temporal declustering may be more important for shorter duration periods (e.g., 2-hour precipitation amounts).

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) claimed 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 (Cooley et al. 2006) and found that there was a low level of dependence between stations within 24 km (15 miles) of one another and no detectable dependence beyond this distance. Because there are very few stations within this distance that 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 just annual maximum data, and, after modeling spatial dependence within the parameter space, how much dependence remains in these data is unanswered. Our precipitation application makes the assumption of conditional spatial independence less problematic, especially given that heavy precipitation is characterized by localized thunderstorms in this region. In other applications such as stream flow measurements, where downstream flows are strongly dependent on those upstream, such an assumption would be unjustifiable.

We have not accounted for any seasonal effects in our data. Restricting our analysis to the nonwinter 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 (Karl and Knight 1998; Trenberth 1999). However, the purpose of this study is not to study climate change. Both climate change and seasonal effects would be interesting extensions of this study.

3. MODELS

To produce the return level map, both the exceedances and their rate of occurrence must be modeled, and we construct separate hierarchical models for each. Hierarchical models allow one to statistically model a complex process and its relationship to observations in several simple components. For an introduction to such models, see Gelman, Carlin, Stern, and Rubin (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 that drives the climatological extreme precipitation for the region. The third layer consists of the prior distributions of the parameters that control the latent process.

As the return level $z_r(\mathbf{x})$ is a function of $\sigma_u(\mathbf{x})$, $\xi(\mathbf{x})$, and $\zeta_u(\mathbf{x})$, we capture the latent process through these parameters. At the s station locations $x_1, \dots, x_s$, we denote the values of the GPD scale parameter by $\boldsymbol{\sigma}_u = [\sigma_u(x_1), \dots, \sigma_u(x_s)]^T$ and we similarly define $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}_u$. The two hierarchies model $\boldsymbol{\sigma}_u$, $\boldsymbol{\xi}$, and $\boldsymbol{\zeta}_u$ and parameters that relate these processes to orographic and climatological information. The process $\sigma(\mathbf{x})$ given $\boldsymbol{\sigma}_u$ is conditionally independent of the data, and likewise for $\xi(\mathbf{x})$ and $\zeta_u(\mathbf{x})$.

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

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

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

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

where p_j is the density associated with level j of the hierarchical model and depends on parameters $\boldsymbol{\theta}_j$. The model for threshold exceedances is described in Section 3.1, and the parallel model for exceedance rates is more briefly described in Section 3.2.

3.1 Hierarchical Model for Threshold Exceedances

3.1.1 Data Layer. A GPD 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, 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). Let $Z_k(\mathbf{x}_i)$ be the kth recorded precipitation amount at location $\mathbf{x}_i$, with $i = 1, \dots, s$. Given that $Z_k(\mathbf{x}_i)$ exceeds the threshold u, we assume that it is described by a GPD whose parameters are dependent on the station's location. Letting $\phi(\mathbf{x}_i)$ and $\xi(\mathbf{x}_i)$ represent the parameters at the location $\mathbf{x}_i$ yields

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

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

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

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 process by constructing a structure that 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) that 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, and this leads us to work in the alternative climate space.

Each location in the longitude/latitude space corresponds to a location in the climate space, and there is an invertible transformation that takes points between the spaces. The coordinates of each point in the climate space are given by its elevation and its MSP. We transform these coordinates by treating the points as observations from a bivariate normal and transforming so that they have a covariance matrix of identity, yielding coordinates on roughly the same scale.

In the climate space, stations that 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 (Fig. 3). As a result, the process is smoother in the climate space, and as shown in Section 4.1, our models performed better in this space.

Let $\phi(\mathbf{x})$ and $\xi(\mathbf{x})$ be the log-transformed scale and shape parameter processes for the region. Being in a Bayesian framework, we treat the parameters $\phi(\mathbf{x})$ and $\xi(\mathbf{x})$ as random variables and must choose a prior distribution. Stephenson and Tawn (2005) suggested using a dependent prior structure as a negative dependence between the GPD's scale and shape parameters is expected. They constructed a joint prior using elicited information and applied it to a univariate time series. How to build dependence into the prior distribution for the spatial processes $\phi(\mathbf{x})$ and $\xi(\mathbf{x})$ is unclear, and we do not have prior information as had Stephenson and Tawn (2005). We instead focus on the primary task of constructing priors that will best model the spatial processes and put independent priors on $\phi(\mathbf{x})$ and $\xi(\mathbf{x})$. Although the priors are constructed independently, the posterior distributions for $\phi(\mathbf{x})$ and $\xi(\mathbf{x})$ do show the expected negative dependence.

We anticipate that the log-transformed scale parameter $\phi(\mathbf{x})$ will be sensitive to regional climate effects and build a model

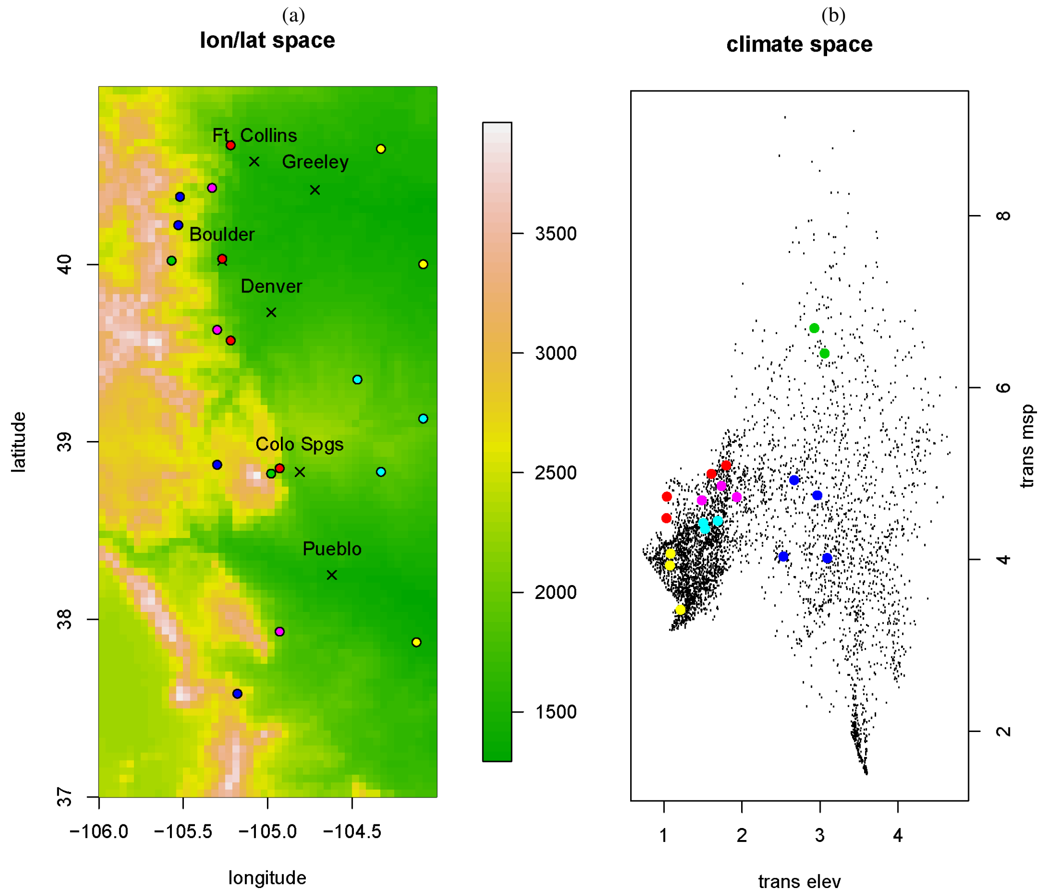

Figure 3. Translation of points in longitude/latitude space (a) (elevation in meters) to points in climate space (b). Letters correspond to selected station locations and represent foothill cities (C), plains (P), Palmer Divide (D), Front Range (F), mountain valley (V), and high elevation (H).

that describes its relationship with the latent spatial process. Drawing on standard geostatistical methods, we model $\phi(\mathbf{x})$ as a Gaussian process with $\mathbb{E}[\phi(\mathbf{x})] = \mu_\phi(\mathbf{x})$ and $\text{cov}(\phi(\mathbf{x}), \phi(\mathbf{x}')) = k_\phi(\mathbf{x}, \mathbf{x}')$. The mean $\mu_\phi(x)$ is a function of parameters $\boldsymbol{\alpha}_\phi$ and the covariates:

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

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

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

which corresponds to an exponential variogram model. The parameters $\beta_{\phi,0}$ and $1/\beta_{\phi,1}$ are sometimes called the "sill" and "range" in the geostatistics literature. Our covariance model assumes the process is isotropic and stationary; we found it impossible to detect any nonstationarity or anisotropy with only 56 stations. We choose to work with exponential models because of their simplicity and because they assume no level of smoothness in the latent process.

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 $\xi(\mathbf{x})$ as a single value and increasingly add complexity until we have a reasonable fit. We model the shape parameter in three ways: (a) as a single value for the entire study region with a $\text{Unif}(-\infty, \infty)$ prior; (b) as two values, one for the mountain stations and one for the plains stations each with $\text{Unif}(-\infty, \infty)$ priors; and (c) as a Gaussian process with structure similar to that of the prior for $\phi(\mathbf{x})$.

Modeling the GPD parameters $\phi(\mathbf{x})$ and $\xi(\mathbf{x})$ as before, data at the station locations provide information about the latent spatial process that characterizes these parameters. Hence, the second piece in (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 (8) evaluated at the covariates of the locations $\mathbf{x}_i$, Σ is the covariance matrix generated by (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 hierarchical layer, we assign priors to the parameters $\boldsymbol{\alpha}_\phi$, $\boldsymbol{\beta}_\phi$, and $\boldsymbol{\theta}_\xi$ that characterize the latent process. We assume each parameter in this layer is independent of the others.

We have no prior information on how the GPD parameter ϕ is related to the covariates, and thus we choose uninformative priors for the regression parameters $\boldsymbol{\alpha}_\phi$. Because a proper posterior is obtained even with an improper prior (Banerjee, Carlin, and Gelfand 2004), we set $\alpha_{\phi,i} \sim \text{Unif}(-\infty, \infty)$ for all models.

Setting priors for $\boldsymbol{\beta}_\phi$ is more difficult. Berger, DeOiveira, and Sanso (2001) and Banerjee et al. (2004) cautioned that improper priors for the sill and range parameters generally result in improper posteriors. Banerjee et al. (2004) suggested that the safest strategy is to choose informative priors. In our application, the parameters $\boldsymbol{\beta}_\phi$ describe the spatial structure of the log-transformed scale parameter of the GPD, a quantity for which it is difficult to elicit prior information.

We rely on knowledge of the space in which we model to set priors for $\beta_{\phi,1}$. When modeling in the latitude/longitude space, we use Unif[.075, .6] as our prior for $\beta_{\phi,1}$, which sets the maximum range of the exponential variogram model to be approximately 40 miles (64 km) and the minimum range to be approximately 5 miles. When modeling in the climate space whose domain is approximately $[1,4] \times [2,6]$ units (Fig. 3), we set the prior for $\beta_{\phi,1}$ to be Unif[6/7, 12], which yields an effective range of approximately .25 to 3.5 units.

Because the values of ϕ are not observed, we have little other than our dataset on which to base our prior for $\beta_{\phi,0}$, which controls the sill of the variogram model. Using maximum likelihood (ML), we fit a GPD to each station independently and then fit an empirical variogram to the $\hat{\phi}$'s. Based on the variogram, we choose a prior for $\beta_{\phi,0}$, which gives a wide envelope around the variogram of the ML estimates. Because we recognize that this is a nontraditional Bayesian methodology, we choose Unif[.005, .09] as the prior, which gives a very wide envelope and which can be used for all models tested (Fig. 4).

Even though we use proper priors for the spatial parameters $\boldsymbol{\beta}_\phi$, Berger et al. (2001) cautioned that simple truncation of the parameter space (as we have done) still leads to difficulties with posteriors. They recommended a careful sensitivity study with respect to the parameter bounds. Such a study is reported in Section 4.2.

For the shape parameter $\xi(\mathbf{x})$, only when modeled as a Gaussian process are there parameters that must be assigned priors in level 3. In this case, $\xi(\mathbf{x})$ has regression coefficients $\boldsymbol{\alpha}_\xi$ and spatial parameters $\boldsymbol{\beta}_\xi$. As with $\boldsymbol{\phi}$, we use $\text{Unif}(-\infty, \infty)$ 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 before, the third piece of (5) is

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

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 exceeds 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.

Because we chose the threshold to be 1.4 cm (.55 inches) for all stations, the exceedance rate parameter $\zeta_{.55}$ (henceforth ζ) will differ at each station and must be modeled spatially. We let $\zeta(\mathbf{x})$ be the exceedance rate parameter for the location $\mathbf{x}$. As with the GPD parameters, we assume there is a latent spatial process that drives the exceedance probability.

Our model to obtain inference about $\zeta(\mathbf{x})$ 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 $\zeta(\mathbf{x}_i)$ of being a cluster maximum.

The process layer of our hierarchy is quite similar to that of the GPD parameter $\phi(\mathbf{x})$. We follow the methodology of Diggle, Tawn, and Moyeed (1998) and transform $\zeta(\mathbf{x})$ using a logit transformation and then model the transformed

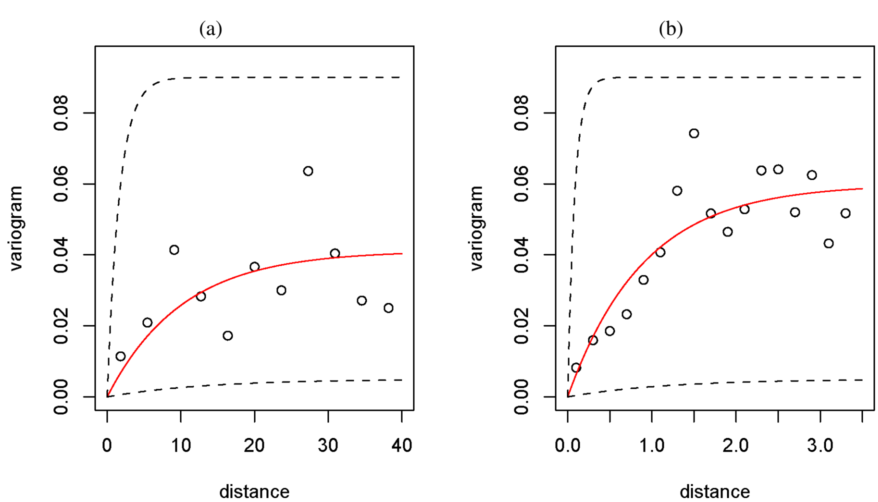

Figure 4. Empirical variogram estimates in traditional space (a) and climate space (b). Binned variogram estimates (○) and the SSE-minimizing variogram (——) are plotted for the MLE-estimated ϕ parameters. The dashed lines denote the envelope of possible variograms given the priors for $\beta_{\phi,0}$ (sill) and $\beta_{\phi,1}$ (1/range).

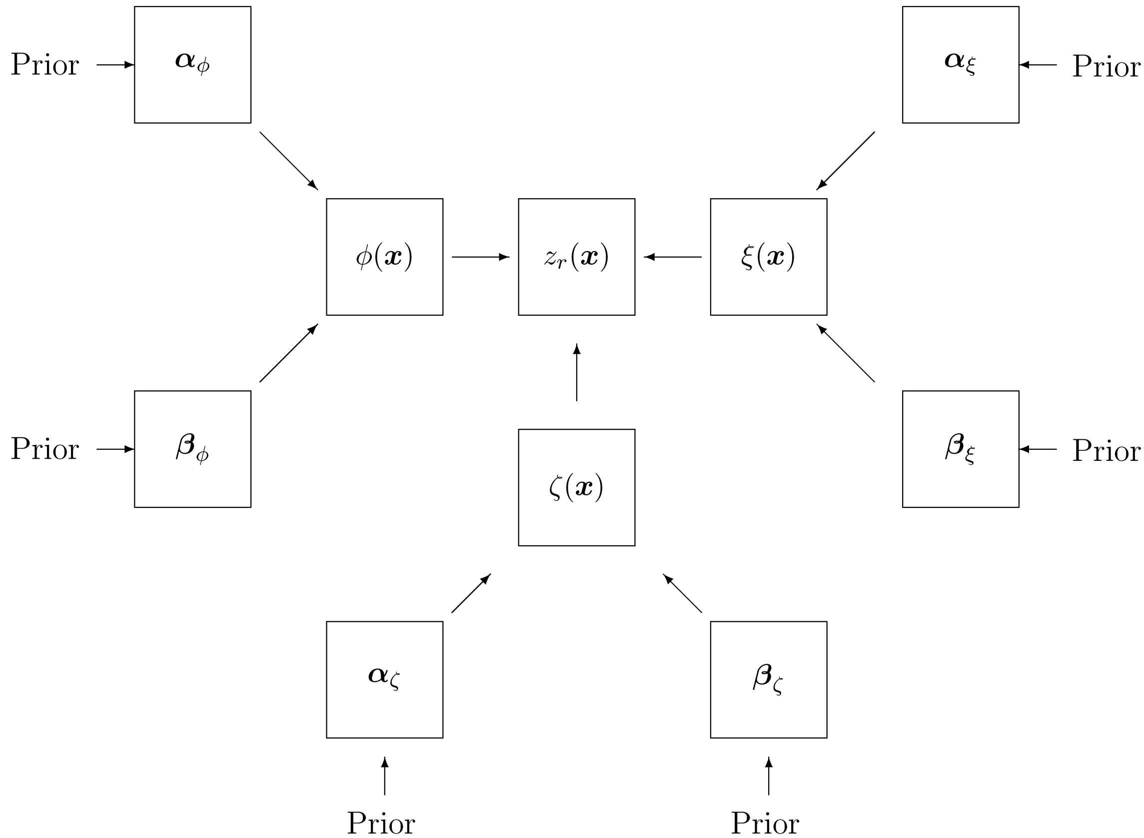

Figure 5. Schematic of model used to estimate return level $z_r(\mathbf{x})$. The return level is a function of the GPD parameters $\phi(x)$ and $\xi(x)$ and of the exceedance rate parameter $\zeta(x)$. All of these can be modeled spatially by a Gaussian process where the parameters α describe the mean structure and β describe the covariance structure.

parameters as a Gaussian process with $\mathbb{E}[\zeta(\mathbf{x})] = \mu_\zeta(\mathbf{x})$ and $\operatorname{cov}(\zeta(\mathbf{x}), \zeta(\mathbf{x}')) = k_\zeta(\mathbf{x}, \mathbf{x}')$. As with $\phi(\mathbf{x})$, the mean $\mu_\zeta(\mathbf{x})$ 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 noninformative $\mathrm{Unif}(-\infty, \infty)$ 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 $\mathrm{Unif}(1, 6)$. The exceedance rate model is taken to be independent of the model for the GPD parameters, and the overall structure to derive return levels is illustrated in Figure 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 exceedance model and the exceedance rate model, we employ Metropolis–Hastings (MH) steps within a Gibbs sampler to update each parameter of the model.

We illustrate our method within the context of the exceedance 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 (MLEs) for the GPD parameter $\boldsymbol{\phi}$. From the asymptotic properties of MLEs, $\hat{\boldsymbol{\phi}} \approx \boldsymbol{\phi} + \epsilon$, where $\epsilon \sim \mathrm{MVN}(0, \mathcal{I}^{-1})$ and $\mathcal{I}$ is the Fisher information matrix. Assuming that ϵ is independent of $\boldsymbol{\phi}$ and given the prior distribution $\boldsymbol{\phi} \sim \mathrm{MVN}(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu}$ and Σ are defined as in (8) and (9), we can then write the joint distribution of $\hat{\boldsymbol{\phi}}$ and $\boldsymbol{\phi}$:

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

We use the conditional distribution

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

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.

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 $\boldsymbol{\xi}$ as a random walk when modeled as a single or pair of values and in the same manner as $\boldsymbol{\phi}$ when modeled spatially. The exceedance rate model is handled analogously.

We ran three parallel chains for each model. Each simulation consisted of 20,000 iterations, the first 2,000 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 assumed that values below the suggested critical value of 1.1 imply convergence. For all parameters of all models, the value of $\hat{R}$ is below 1.05 unless otherwise noted in Section 4.1.

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 (3), $z_r(\mathbf{x})$ is a function of the GPD parameters $\phi(\mathbf{x})$, $\xi(\mathbf{x})$, and the (independent) exceedance rate parameter $\zeta(\mathbf{x})$; 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 $z_r(\mathbf{x})$.

To illustrate our interpolation method, consider the log-transformed GPD scale parameter of the exceedance model. We begin with values for $\boldsymbol{\phi}$, $\boldsymbol{\alpha}_\phi$, and $\boldsymbol{\beta}_\phi$ from which we need to interpolate the value of $\phi(\mathbf{x})$. 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 $\phi(\mathbf{x})$. Using the values of $\boldsymbol{\alpha}_\phi$ and $\boldsymbol{\beta}_\phi$, we are able to draw from the conditional distribution for $\phi(\mathbf{x})$ given the current values of $\boldsymbol{\phi}$. Doing this for each iteration of the MCMC algorithm provides draws from the posterior distribution of $\phi(\mathbf{x})$.

We do the same for the exceedance rate parameter $\zeta(\mathbf{x})$ and for the GPD shape parameter $\xi(\mathbf{x})$ if it is modeled spatially. Pointwise means are used as point estimates for each of the parameters (Fig. 7). The entire collection of draws from the posterior distributions of $\phi(\mathbf{x})$, $\xi(\mathbf{x})$, and $\zeta(\mathbf{x})$ 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 (Figs. 8 and 9).

4. RESULTS

4.1 Model Selection and Map 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, Best, Carlin, and van der Linde 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 (lower is better). As the DIC scores result from the realizations of an MCMC run, there is some randomness in them, and, therefore, nested models do not always have improved fits. 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 $\phi(\mathbf{x})$ is modeled as in Section 3.1 and $\xi(\mathbf{x})$ 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, and 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 $\phi(\mathbf{x})$ show relatively high values around the cities of Boulder and Fort Collins but do not show similar values for the stationless region between the cities despite that it has a similar climate and geography.

Table 1. Several of the different GPD hierarchical models tested and their corresponding DIC scores

Baseline model		$\bar{D}$	p_D	DIC
Model 0:	$\phi = \phi$ $\xi = \xi$	73,595.5	2.0	73,597.2
Models in latitude/longitude space		$\bar{D}$	p_D	DIC
Model 1:	$\phi = \alpha_0 + \epsilon_\phi$ $\xi = \xi$	73,442.0	40.9	73,482.9
Model 2:	$\phi = \alpha_0 + \alpha_1(\text{msp}) + \epsilon_\phi$ $\xi = \xi$	73,441.6	40.8	73,482.4
Model 3:	$\phi = \alpha_0 + \alpha_1(\text{elev}) + \epsilon_\phi$ $\xi = \xi$	73,443.0	35.5	73,478.5
Model 4:	$\phi = \alpha_0 + \alpha_1(\text{elev}) + \alpha_2(\text{msp}) + \epsilon_\phi$ $\xi = \xi$	73,443.7	35.0	73,478.6
Models in climate space		$\bar{D}$	p_D	DIC
Model 5:	$\phi = \alpha_0 + \epsilon_\phi$ $\xi = \xi$	73,437.1	30.4	73,467.5
Model 6:	$\phi = \alpha_0 + \alpha_1(\text{elev}) + \epsilon_\phi$ $\xi = \xi$	73,438.8	28.3	73,467.1
Model 7:	$\phi = \alpha_0 + \epsilon_\phi$ $\xi = \xi_{\text{mtn}}, \xi_{\text{plains}}$	**73,437.5**	**28.8**	**73,466.3**
Model 8:	$\phi = \alpha_0 + \alpha_1(\text{elev}) + \epsilon_\phi$ $\xi = \xi_{\text{mtn}}, \xi_{\text{plains}}$	73,436.7	30.3	73,467.0
Model 9:	$\phi = \alpha_0 + \epsilon_\phi$ $\xi = \xi + \epsilon_\xi$	73,433.9	54.6	73,488.5

NOTE: Models in the climate space had better scores than models in the longitude/latitude space. $\epsilon_\cdot \sim \text{MVN}(0, \Sigma)$, where $[\sigma]_{i,j} = \beta_{\cdot,0} \exp(-\beta_{\cdot,1} \|\mathbf{x}_i - \mathbf{x}_j\|)$.

When we perform the analysis for the climate space, we obtain better results. Both the model fit score and the effective number of parameters are lower in the climate space, yielding lower DIC scores for corresponding models (e.g., Models 1 and 5 or Models 3 and 6). However, in the climate space, adding elevation (or MSP) as a covariate does not seem to improve the model as these covariates are already integrated into the analysis as the locations' coordinates. Most important, when the points are translated back to the original space, we obtain parameter estimate maps that seem to better agree with the geography.

We then begin to add complexity to the shape parameter $\xi(\mathbf{x})$. Allowing the mountain stations and plains stations to have separate shape parameter values slightly improves model fit (Model 7), but a fully spatial model for $\xi(\mathbf{x})$ does not improve

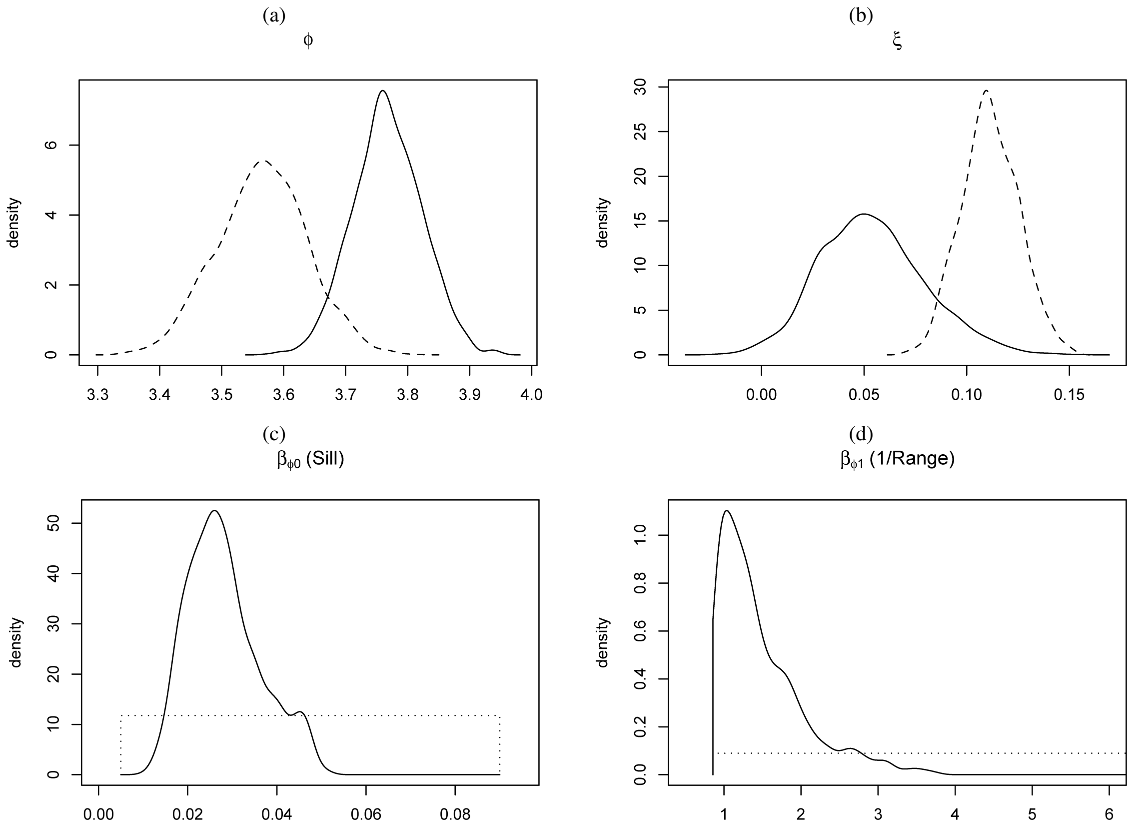

Figure 6. Posterior densities from Model 7. Plot (a) shows ϕ parameters associated with the Boulder station (——) and the Greeley/UNC station (- - - - -). Plot (b) shows the parameters ξ for the mountain stations (——) and the plains stations (- - - - -). Plot (c) shows $\beta_{\phi,0}$, which corresponds to the sill of the variogram (——) and its prior distribution (⋯⋯). Plot (d) shows $\beta_{\phi,1}$, which corresponds to the inverse range parameter in the climate space (lower parameter value indicates longer range).

model fit enough to warrant the added complexity (Model 9). Model 7 is chosen as the most appropriate model tested based not only on its DIC score but also on the posteriors for the parameters ξ_{mtn} and ξ_{plains}. Selected posterior densities from Model 7 are plotted in Figure 6. The left plot in Figure 7 shows Model 7's point estimate (pointwise posterior mean) for $\phi(\mathbf{x})$, which is strongly influenced by the geography of the region.

Our analysis for the exceedance rate model follows a similar methodology, and results are given in Table 2. We begin in the latitude/longitude space and proceed by adding covariates. Interestingly, adding elevation and MSP only slightly improves the model fit scores (Models A–D). Contrary to the exceedance model, the rate models run in the traditional space have lower DIC scores than those in the climate space. However, the maps produced by traditional space models were judged to be inferior to those produced in the climate space. Climate space maps aligned with the geography of the region, whereas traditional space maps were dominated by the mean structure and effects at the stations were represented by small circular areas of disagreement. Model H was selected as most appropriate. The exceedance rate point estimate map [Fig. 7(b)] 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.

Field practitioners are interested in return level estimates rather than those of individual parameters. Return level ensembles are produced by combining the MCMC results of the exceedance and exceedance rate models. Figure 8 shows maps of the point estimate (pointwise posterior mean) for the 25-year daily precipitation return level produced by modeling in both the climate and the traditional space. The map produced by the climate space analysis shows interesting geographic effects, which the traditional space analysis cannot detect. Stations in Front Range cities such as Boulder and Fort Collins show an increased intensity of extreme precipitation compared to those farther east. The traditional space analysis is unable to adequately pool these stations and interpolate to stationless regions between the cities, resulting in circular areas of increased intensity around the cities. The climate space analysis better represents this apparent Front Range effect by clearly showing an area of increased intensity all along the Front Range. The climate space model also shows a region of increased intensity to the north of the Palmer Divide region, contrasting with a region of decreased intensity around Greeley and to the northeast of Denver.

The return level map also shows a dramatic difference between plains and mountain locations. This difference results from the estimates for both $\phi(\mathbf{x})$ (Fig. 7) and $\xi(\mathbf{x})$. Because $\xi(\mathbf{x})$ is modeled as different values for the mountain and plains locations, the difference between the mountain and foothill locations may be less discontinuous than the map indicates. The posterior plot for ξ (Fig. 6) shows that precipitation at both the mountain and plains stations has a heavy-tailed distribution with plains stations having a heavier tail. The analysis that

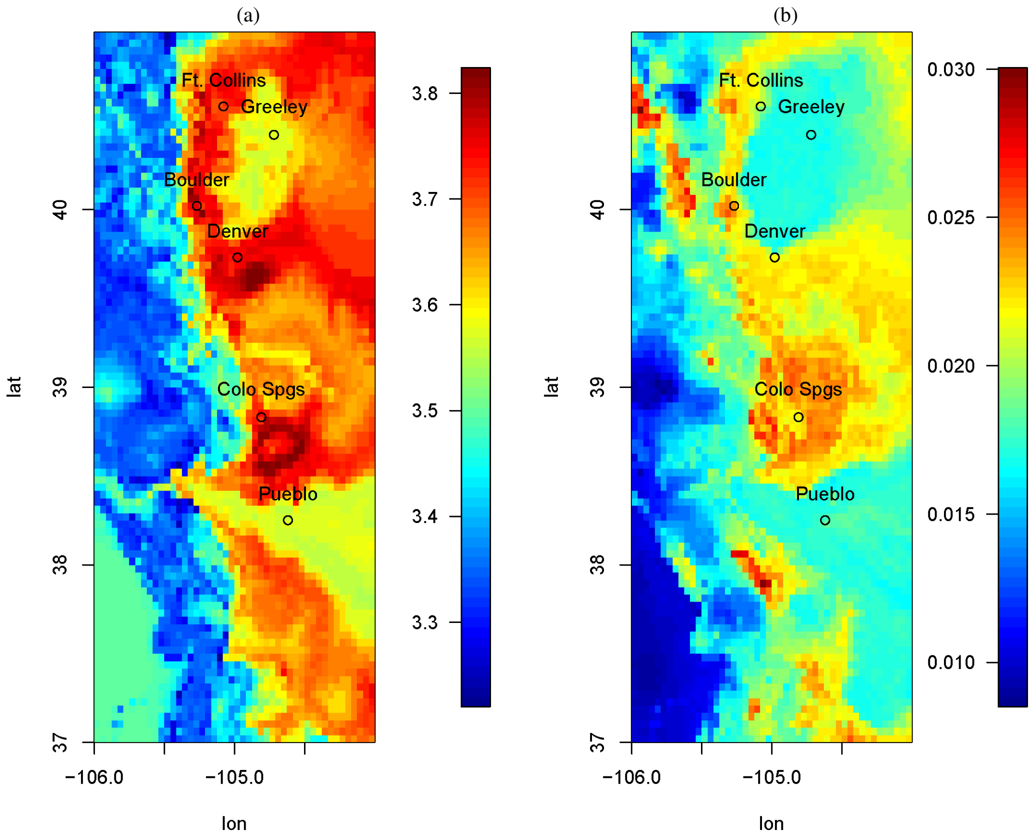

Figure 7. Point estimate for log-transformed GPD scale parameter ϕ for Model 7 (a) and point estimate for exceedance rate parameter ζ from Model H (b).

the mountainous areas have a lighter tail and, therefore, less extreme precipitation agrees with the studies of Jarrett (1990, 1993), who claimed that the hydrologic and paleohydrologic evidence shows that intense rainfall does not occur at higher elevations.

Part of the reason for adopting a Bayesian methodology was to obtain natural uncertainty estimates for the return levels. Figure 9 shows lower and upper bound estimates of the 25-year return level, which were calculated by taking the pointwise .025 and .975 empirical quantiles from the return level draws. Figure 9 also shows a map of the uncertainty range, which is simply the difference between the upper and lower bounds. When viewing the uncertainty map, one must recognize that points located somewhat distant from the nearest station may not be that distant in the climate space, resulting in less uncertainty than one might expect. The levels of uncertainty are greatest 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.

Table 2. Exceedance rate hierarchical models tested and their DIC scores

Baseline model		$\bar{D}$	p_D	DIC
Model O:	$\zeta = \zeta$	733.3	1	734.3
Models in latitude/longitude space		$\bar{D}$	p_D	DIC
Model A:	$\zeta = \alpha_0 + \epsilon_\zeta$	445.1	55.1	500.2
Model B:	$\zeta = \alpha_0 + \alpha_1(\text{msp}) + \epsilon_\zeta$	445.5	55.0	500.5
Model C:	$\zeta = \alpha_0 + \alpha_1(\text{elev}) + \epsilon_\zeta$	446.9	56.4	503.2
Model D:	$\zeta = \alpha_0 + \alpha_1(\text{msp}) + \alpha_2(\text{elev}) + \epsilon_\zeta$	443.3	53.9	497.2
Models in climate space		$\bar{D}$	p_D	DIC
Model E:	$\zeta = \alpha_0 + \epsilon_\zeta$	460.7	61.2	521.9
Model F:	$\zeta = \alpha_0 + \alpha_1(\text{msp}) + \epsilon_\zeta$	450.2	56.1	506.3
Model G:	$\zeta = \alpha_0 + \alpha_1(\text{elev}) + \epsilon_\zeta$	449.9	58.1	508.0
Model H:	$\zeta = \alpha_0 + \alpha_1(\text{msp}) + \alpha_2(\text{elev}) + \epsilon_\zeta$	**446.5**	**53.2**	**499.7**

NOTE: Models in the longitude/latitude space generally had lower DIC scores than models in the climate space. However, maps produced by models in the climate space were judged to be superior to those produced by model in the traditional space. $\epsilon_\cdot \sim \text{MVN}(0, \Sigma)$, where $[\sigma]_{i,j} = \beta_{\cdot,0} \exp(-\beta_{\cdot,1} \|\mathbf{x}_i - \mathbf{x}_j\|)$.

Given the motivation for this article based on the Fort Collins flood, the reader may wonder how our analysis compares with the estimates of other studies. Our 100-year return level estimate for 24-hour precipitation has a 95% credible interval of (9.01 cm, 12.12 cm). Interestingly, our estimates for this particular location do not differ dramatically from those in the 1973 NWS atlas, which has a point estimate of 11.32 cm. Because our model pools data from all sites, the return levels for Fort Collins are not heavily influenced by the 1997 event. In comparison, Sveinsson et al.'s (2002) GEV-based RFA analysis gave point estimates for the 100-year return level of 15.4, 12.9, and 12.4 cm, for analyses of one, three, and thirteen stations, respectively. A maximum likelihood GPD analysis [threshold = .45 inches (1.14 cm)] of the Fort Collins station data yields a 95% profile likelihood (Coles 2001b) confidence interval of (9.53 cm, 19.86 cm), showing that the model's data pooling

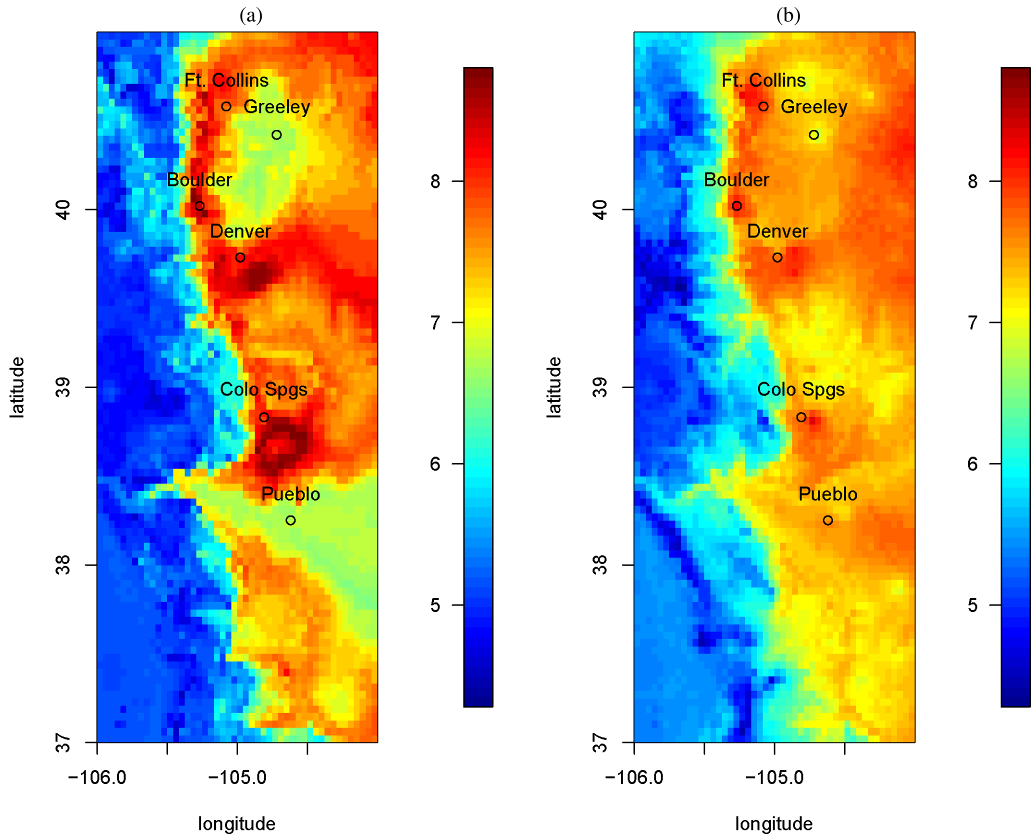

Figure 8. Comparison of point estimates for 25-year return level for daily precipitation. On part (a) is the estimate working in the climate space; on part (b) the estimate working in the longitude/latitude space.

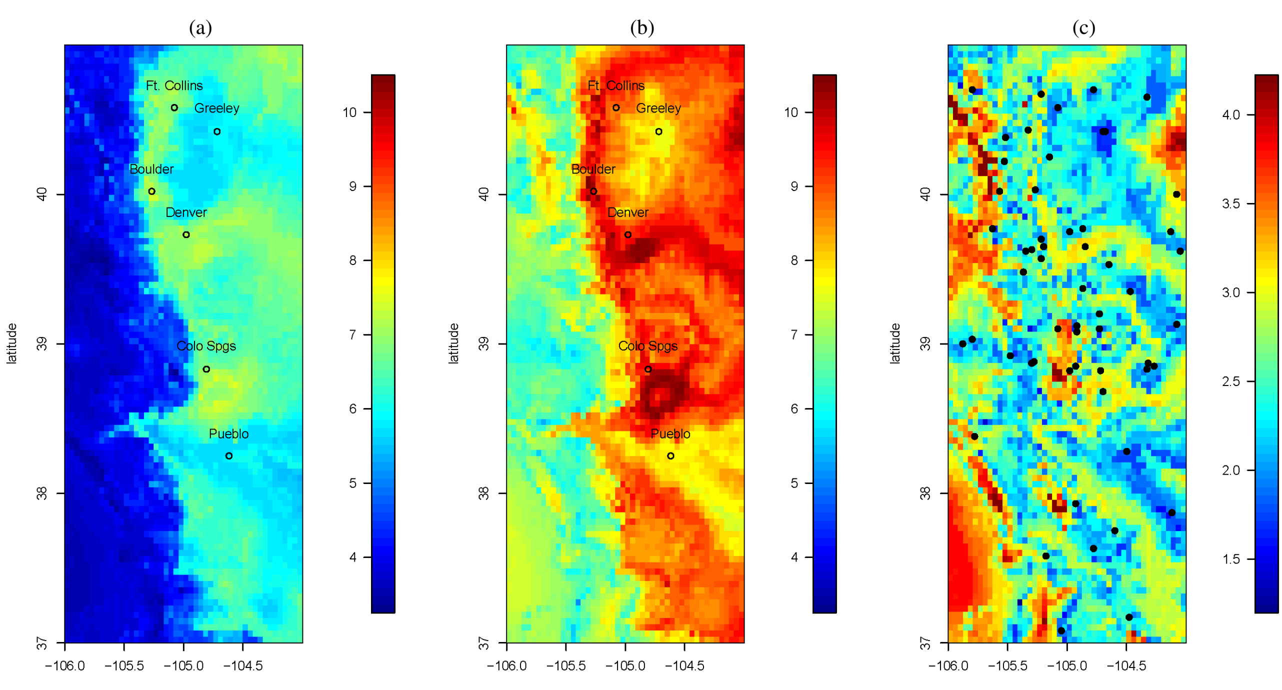

Figure 9. Estimates of the lower and upper bounds of the 25-year return level. Plot (a) is the .025 quantile of the daily precipitation 25-year return level. Plot (b) is the .975 quantile, and plot (c) is the difference of the two, or an estimate of the range of 95% credible interval (with station locations plotted).

also significantly decreases the uncertainty. The 24-hour precipitation recorded at the Fort Collins station on the day of the 1997 flood was 15.7 cm (6.2 inches) and was the second-largest recorded precipitation amount in our dataset. This amount corresponds roughly to a 500-year event, for which our model gives a 95% credible interval of (11.84 cm, 16.68 cm). When one considers that over 1,900 (nonindependent) station-years of data were used in our model, the likelihood of observing an event of this magnitude seems reasonable.

Although the model and the 1973 NOAA atlas do not differ significantly in Fort Collins, there are areas where there are large differences. Estes Park is a mountain town northwest of Boulder and its weather station has a short data record from 1978 to 1994. Our model gives the location a 95% credible interval of (5.50 cm, 8.36 cm) for the 100-year return level, whereas the 1973 NOAA atlas (which had no Estes Park data available) gives a point estimate of 11.07 cm. Longmont, a city 15 miles northeast of Boulder, has a nearly complete data record from 1948 to 2001, but our model's credible interval for this location is (8.64 cm, 11.95 cm), whereas the NOAA estimate is 12.92 cm for the 100-year return level.

4.2 Sensitivity Analysis

In a Bayesian analysis, it is natural to ask how sensitive the results are to the choice of prior, and in this study, the priors of most interest are those for $\boldsymbol{\beta}_\phi$. Estimation of the sill and range parameters of a variogram model is difficult. Zhang (2004) showed that these parameters cannot be estimated consistently for a fixed domain. However, for interpolation purposes, the individual parameters $\beta_{\phi,0}$ and $\beta_{\phi,1}$ are less important than the value of their product $\beta_{\phi,0}\beta_{\phi,1}$ (Stein 1999; Zhang 2004), which Zhang demonstrated can be estimated consistently.

A preliminary sensitivity analysis indicated that the model was most sensitive to the lower bound on $\beta_{\phi,1}$, which controls the upper bound of the variogram's range. To specifically test the sensitivity of this bound, we ran Model 7 with an alternative prior of Unif[.214, 6] for $\beta_{\phi,1}$ while keeping the prior for $\beta_{\phi,0}$ the same. This alters the possible range of the variogram from approximately [.25, 3.5] units in the climate space to approximately [.5, 14] units. As the domain of the climate space is $[1,4]\times[2,6]$, this prior would seem to yield nonsensible values. However, much of the mass for $\beta_{\phi,1}$ is again near the new lower bound (Fig. 10), indicating a range beyond the domain of the climate space. Clearly, the posteriors for $\beta_{\phi,0}$ are quite sensitive to the prior.

However, the mass of $\beta_{\phi,0}$ has also shifted slightly upward though its prior was not changed. Interestingly, although there is a difference at the lower bound, the posterior of the product $\beta_{\phi,0}\beta_{\phi,1}$ appears to be less sensitive than the individual parameters to the change of prior (Fig. 10(c), Fig. 11). This agrees with Zhang's (2004) results that the product of these parameters is more easily estimated than the individual values. Because it is the product that is important for interpolation, the maps produced by the two sets of priors are nearly identical.

In addition to prior sensitivity, it is natural to ask how sensitive the results are to the choice of threshold. We ran the model with thresholds of .35, .45, .55, .65, and .75 inches (.89, 1.14, 1.40, 1.65, and 1.91 cm) and compared the posterior densities for the 25-year return levels at mountain, Front Range, and plains stations. The posterior densities for the return levels were similar for thresholds above .45 inches (1.14 cm), with the densities becoming wider as the threshold was increased. However, we found that the shape parameter ξ was more consistently estimated for thresholds above .55 inches (1.40 cm). The fact that the return level plots were similar for thresholds of .45 and .55 inches (1.14 and 1.40 cm) despite the difference in the values for ξ was because estimates for ϕ had adjusted for the value of ξ. Because correct estimation of ξ is vital for estimating return levels of long periods (e.g., 100 or 500 years), we decided that a threshold of .55 inches (1.40 cm) was best.

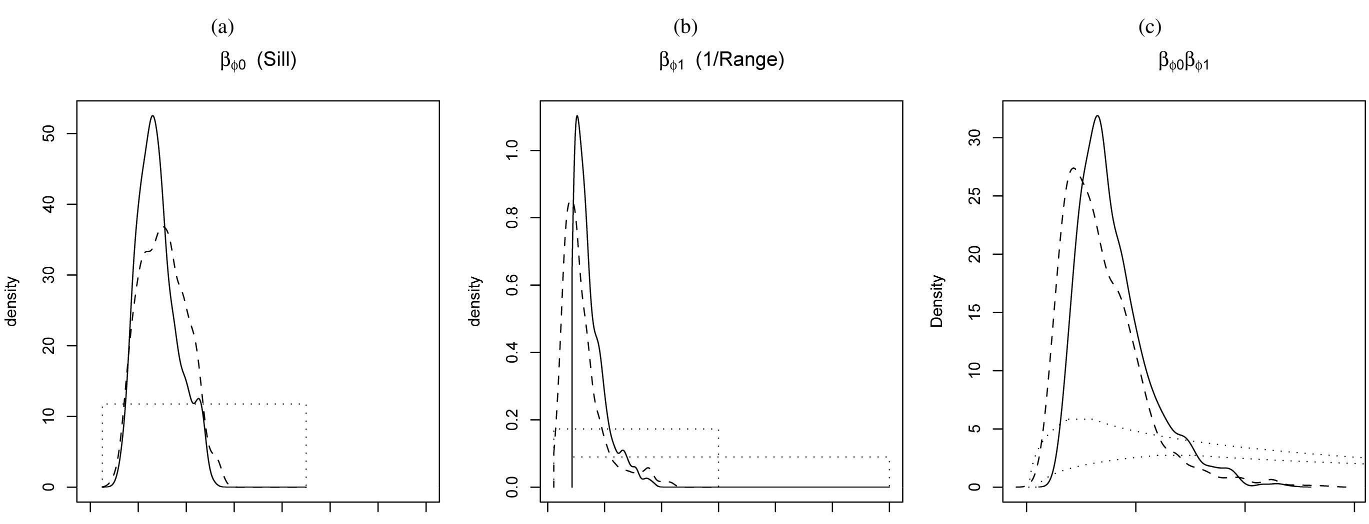

Figure 10. Posterior distributions for $\beta_{\phi,0}$ (a), $\beta_{\phi,1}$ (b), and $\beta_{\phi,0}\beta_{\phi,1}$ (c). The solid line shows the posterior given the original priors, and the dashed line shows the posterior given the alternative priors. Dotted lines show the prior distributions. Notice that, despite the differences in the posteriors for the individual parameters, the posterior of the product is similar for both sets of parameters, with a small difference due to the new lower bound of the support.

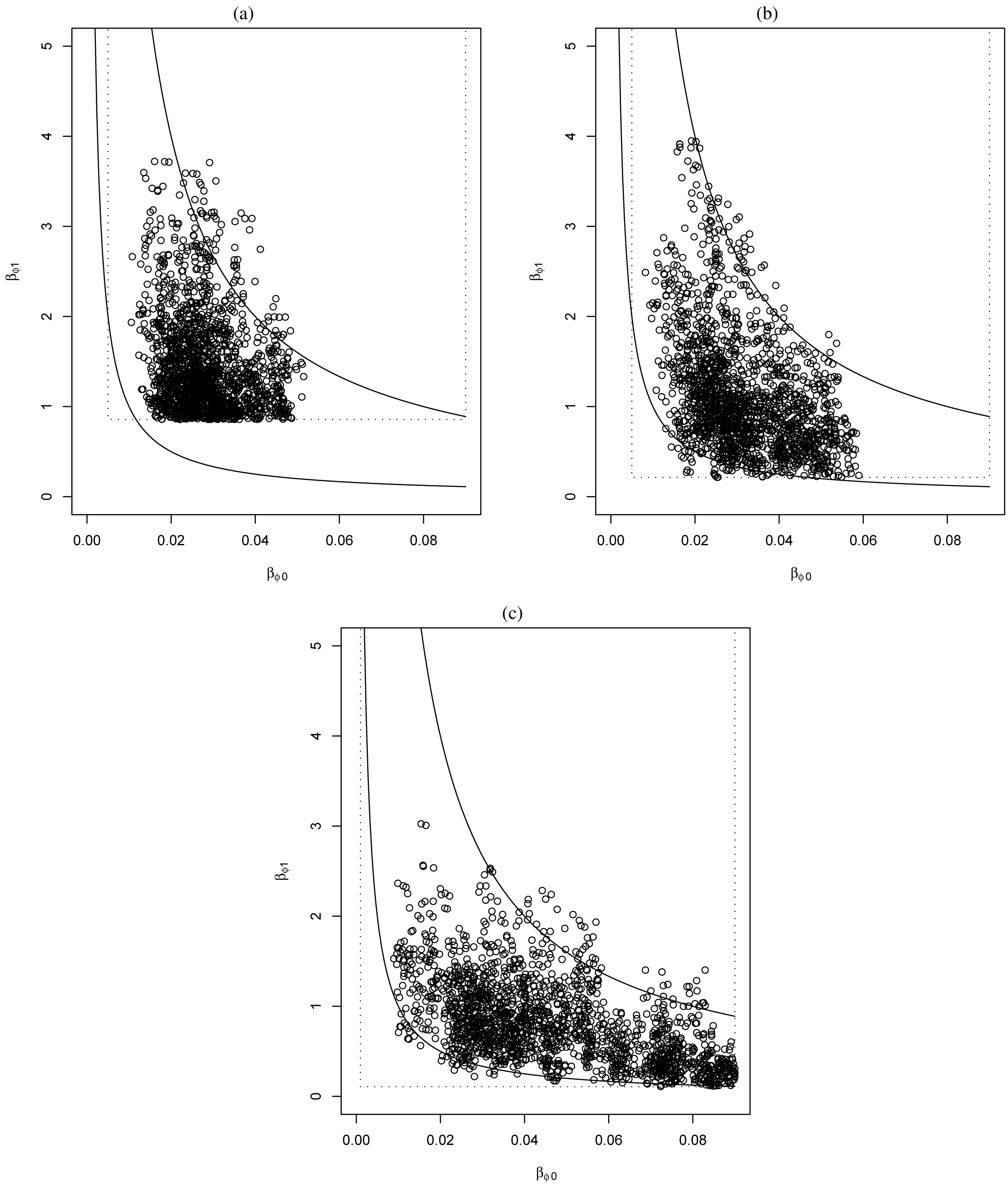

Figure 11. Scatterplot of realizations of $\beta_{\phi,0}$ and $\beta_{\phi,1}$ for three different sets of priors ($\cdots\cdots$). Notice that, despite differences in the individual values, the points primarily lie in the region enclosed by the lines $xy = .01$ and $xy = .08$ (——), indicating that the product of the two parameters is not sensitive to the prior.

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 a common threshold was chosen for the entire region. To our knowledge, this study is the first to spatially model the exceedance rate parameter in the context of extremes. By performing the spatial analysis on locations defined by climatological coordinates, we are able to better model regional differences for this geographically diverse study area.

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 predetermining distinct regions in which to pool normalized data. The multistep 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 this 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 to 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 because they are not coupled, it is possible to obtain nonsensical return level estimates (e.g., the return level 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 of duration period.

Finally, we were able to produce practical maps of return levels. Our map has several features not well shown by the 1973 NOAA atlas such as an east–west region of higher return levels north of the Palmer Divide and a region of lower return levels around Greeley. 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.

[Received July 2005. Revised February 2006.]

REFERENCES

Banerjee, S., Carlin, B., and Gelfand, A. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL: Chapman & Hall/CRC Press.

Berger, J., DeOiveira, V., and Sanso, B. (2001), "Objective Bayesian Analysis of Spatially Correlated Data," *Journal of the American Statistical Association*, 96, 1361–1374.

Bonnin, G., Todd, D., Lin, B., Parzybok, T., Yekta, M., and Riley, D. (2004a), *NOAA Atlas 14, Precipitation Frequency Atlas of the United States*, Vol. 1, Silver Spring, MD: U.S. Department of Commerce.

——— (2004b), *NOAA Atlas 14, Precipitation Frequency Atlas of the United States*, Vol. 2, Silver Spring, MD: U.S. Department of Commerce.

Bottolo, L., Consonni, G., Dellaportas, P., and Lijoi, A. (2003), "Bayesian Analysis of Extreme Values by Mixture Modeling," *Extremes*, 6, 25–47.

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.

——— (1996b), "Modeling Extremes of the Areal Rainfall Process," *Journal of the Royal Statistical Society*, Ser. 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*, 11, 23–34.

——— (2001b), *An Introduction to Statistical Modeling of Extreme Values*, London: Springer-Verlag.

Cooley, D., Naveau, P., Jomelli, V., Rabatel, A., and Grancher, D. (2006), "A Bayesian Hierarchical Extreme Value Model for Lichenometry," *Environmetrics*, 17, 555–574.

Cooley, D., Naveau, P., and Poncet, P. (2006), "Variograms for Spatial Max-Stable Random Fields," in *Dependence in Probability and Statistics*, Springer Lecture Notes in Statistics, New York: Springer, pp. 373–390.

Dalrymple, T. (1960), "Flood Frequency Analyses," Water Supply Paper 1543-a, U.S. Geological Survey.

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 Phillips, D. L. (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," in *Proceedings of the 45th Session of the International Statistical Institute*.

Diggle, P., Tawn, J., and Moyeed, R. (1998), "Model-Based Geostatistics," *The Applied Statistics*, 47, 299–350.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*, Berlin: Springer-Verlag.

Fisher, R. A., and Tippett, L. H. C. (1928), "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample," *Proceedings of the Cambridge Philosophical Society*, 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, Part 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. (1996), "Inference and Monitoring Convergence," in *Markov Chain Monte Carlo in Practice*, eds. W. R. Gilks, S. Richarson, and D. J. Spiegelhalter, London: Chapman & Hall.

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

Gnedenko, R. (1943), "Sur la Distribution Limite du Terme Maximum d'une Série Aléatoire," *The Annals of Mathematics*, 44, 423–453.

Heffernan, J. E., and Tawn, J. A. (2004), "A Conditional Approach for Multivariate Extreme Values," *Journal of the Royal Statistical Society*, Ser. B, 66, 497–546.

Hosking, J., and Wallis, J. (1988), "The Effect on Intersite Dependence on Regional Frequency Analysis," *Water Resources Research*, 24, 588–600.

——— (1997), *Regional Frequency Analysis: An Approach Based on L-Moments*, Cambridge, U.K.: Cambridge University Press.

Jagger, T., and Elsner, J. (2006), "Climatology Models for Extreme Hurricane Winds in the United States," *Journal of Climate*, 19, 3220–3236.

Jarrett, R. D. (1990), "Paleohydrologic Techniques Used to Define the Spatial Occurrence of Floods," *Geomorphology*, 3, 181–195.

——— (1993), "Flood Elevation Limits in the Rocky Mountains," in *Engineering Hydrology: Proceedings of the Symposium Sponsored by the Hydraulics Division of the American Society of Civil Engineers*, ed. C. Kuo, San Francisco, CA: American Society of Civil Engineers, pp. 180–185.

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*, Vol. 3, Silver Spring, MD: U.S. Department of Commerce.

National Weather Service (2005), "Noteworthy Colorado Floods" [on-line], available at *http://www.crh.noaa.gov/bou/awebphp/floods.php*.

Pickands, J. (1975), "Statistical Inference Using Extreme Order Statistics," *The Annals of Statistics*, 3, 119–131.

Robert, C. P., and Casella, G. (1999), *Monte Carlo Statistical Methods*, New York: Springer-Verlag.

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," *The Applied Statistics*, 36, 358–369.

Spatial Climate Analysis Service (1995), "2.5 Minute Digital Elevation Model for the Coterminous United States" [on-line], available at *http://www.ocs.orst.edu/prism/docs/meta/dem_25m.htm*.

——— (2004), "May–October Precipitation Normals 1971–2000" [on-line], available at *http://www.ocs.oregonstate.edu/prism*.

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*, Ser. B, 64, 583–639.

Stein, M. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer-Verlag.

Stephenson, A., and Tawn, J. (2005), "Bayesian Inference for Extremes: Accounting for the Three Extremal Types," *Extremes*, 7, 291–307.

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," in *Selected Papers II*, American Mathematical Society, pp. 271–294.

Zhang, H. (2004), "Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics," *Journal of the American Statistical Association*, 99, 250–261.

Zwiers, F., and Kharin, V. (1998), "Changes in the Extremes of the Climate Simulated by CCC GCM2 Under CO_2 Doubling," *Journal of Climate*, 11, 2200–2222.