Appendix
Bayesian Spatial Modeling of Extreme Precipitation Return Levels

This appendix is for the purpose of further explaining some of the issues that arose during the study of extreme precipitation for Colorado's Front Range. All of the issues below are mentioned in the manuscript, however, because of article length constraints, the discussion in the manuscript is quite brief.

1 Prior elicitation

One may ask how "empirical information" is used to assign the priors for the spatial parameters $\beta_{\cdot,0}$ (which forms the sill of the variogram), and $\beta_{\cdot,1}$ (which is inversely proportional to the range of the variogram).

The parameters in question control the spatial behavior of the data layer parameters to the latent process that drives the extreme precipitation events. For example $\beta_{\phi,0}$ and $\beta_{\phi,1}$ control the spatial behavior of the log-transformed scale parameter of the GPD distribution. Unlike the data, the values of ϕ are not directly observable which makes elicitation about how this parameter varies over space very difficult. Further complicating matters is the fact that we work in a space whose coordinates are determined by climatological factors. And, unlike the other parameters (such as $\alpha_{\cdot}$), we are unable to use improper and uninformative priors for these parameters and obtain a proper posterior (Berger et al., 2001).

In order to set a proper prior for $\beta_{\cdot,1}$ we use knowledge of the space in which the latent process is modeled. The geographic area we model is roughly 100 x 275 miles (160 x 440 km). We set the prior for $\beta_{\cdot,1}$ to be $Unif[0.075, 0.6]$ which yields maximum range of the exponential variogram to be about 40 miles and the minimum range to be about 5 miles. Similarly when modeling in the climate space which has a domain of roughly [1,4] x [2,6] (see Figure 3 in the manuscript), we set the prior for $\beta_{\cdot,1}$ to be $Unif[6/7, 12]$ which yields a range roughly between .25 and 3.5 units.

It is in setting the priors for the $\beta_{\cdot,0}$ parameter where we utilize empirical information. The method we employ is to take MLE estimates of the GPD parameters (e.g. ϕ) independently at each of the station locations, and then fit an empirical variogram to these estimates. We then choose priors that create an envelope around the empirical variogram (Figure 1). Recognizing that using the data to set the priors is a non-traditional Bayesian approach, the authors choose an especially wide envelope. The prior chosen for $\beta_{\cdot,0}$ was $Unif[.005, .09]$.

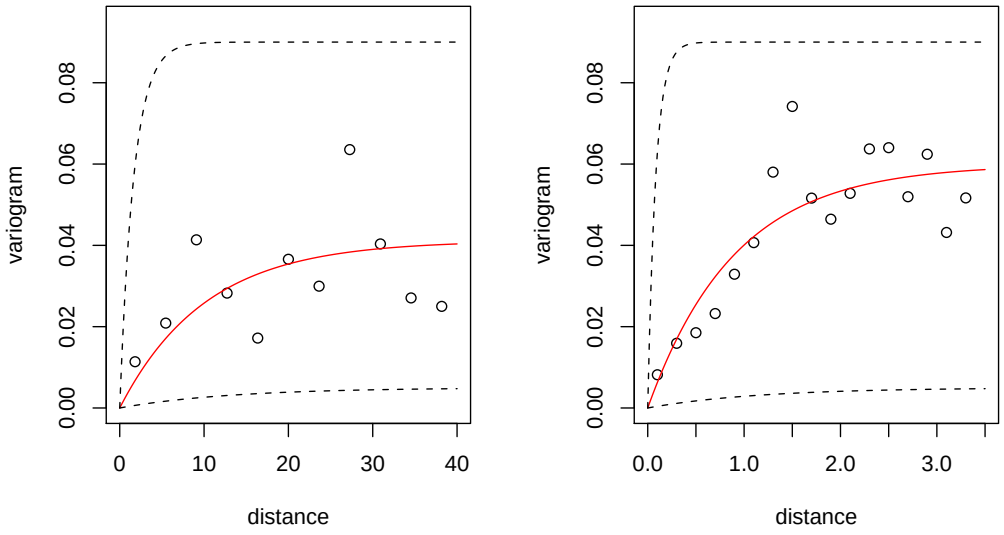

Figure 1: Shows the empirical variogram estimates in the traditional space (left) and the climate space (right). Binned variogram estimates (points) and the SSE-minimizing variogram (solid line) are plotted for the MLE-estimated ϕ parameters. The dashed lines denote the envelope of possible variograms given the priors for $\beta_{\phi,0}$ (sill) and $\beta_{\phi,1}$ (1/range).

2 Prior sensitivity

Estimating both the sill and range parameters in variogram models is problematic. Zhang (2004) shows that these two parameters cannot be estimated consistently for a fixed domain. Our purpose for estimating these parameters is to be able to interpolate the data-layer parameters (ϕ, ξ and ζ) over the study area. For interpolation purposes, the individual parameters $\beta_{\cdot,0}$ and $\beta_{\cdot,1}$ are less important than the value of their product $\beta_{\cdot,0}\beta_{\cdot,1}$ (Zhang, 2004; Stein, 1999), which Zhang demonstrates can be estimated consistently.

From a preliminary sensitivity analysis and the posterior plots, it appears that the model is most sensitive to the lower bound of the prior for $\beta_{\phi,1}$. To test this specifically, we run Model 7 with an alternate prior of $Unif[0.214, 6]$ 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] x [2,6] this prior would seem to yield non-sensible values. However, much of the mass for $\beta_{\phi,1}$ is again near the new lower bound (Figure 2) indicating a very long range. 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 even though its prior was not changed. The resulting posterior of the product $\beta_{\phi,0}$ and $\beta_{\phi,1}$ appears to be less affected by the change in prior than the individual parameters. Although there is some difference because of the differences in priors' lower bounds, the posteriors of the product are similar (Figure 2, right). Figure 3 shows a scattergram of posterior realizations for three sets of priors. The scattergrams clearly show that the realizations, though affected by the prior, fall in an area defined by the functions $xy = .01$ and $xy = .08$. Return levels maps were produced for both the original and alternate priors for both the pointwise mean and pointwise .025 quantile (Figures 4 and 5). Maps for both sets of priors were nearly indistinguishable to the eye. The sensitivity of the individual range and sill parameters to their priors has little influence on the marginal distribution of posterior return levels.

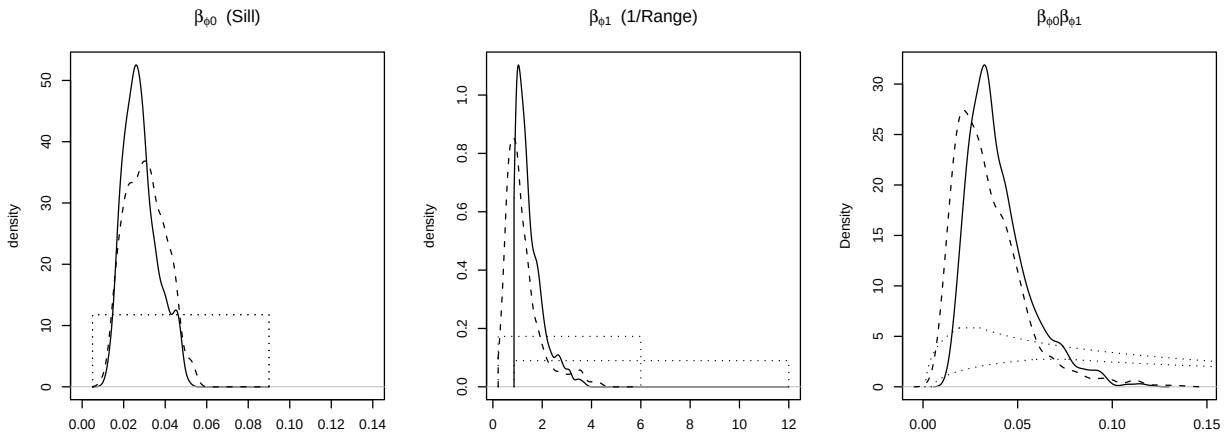

Figure 2: Shows the posterior distributions for $\beta_{\phi,0}$ (left), $\beta_{\phi,1}$ (center), and $\beta_{\phi,0}\beta_{\phi,1}$ (right). Solid line shows the posterior given the original priors, and dashed line shows the posterior given the alternative priors. Dotted lines show the prior distributions. Notice that despite the differences in the posteriors for the individual parameters, the posterior of the product is similar. Notice the range of differences is

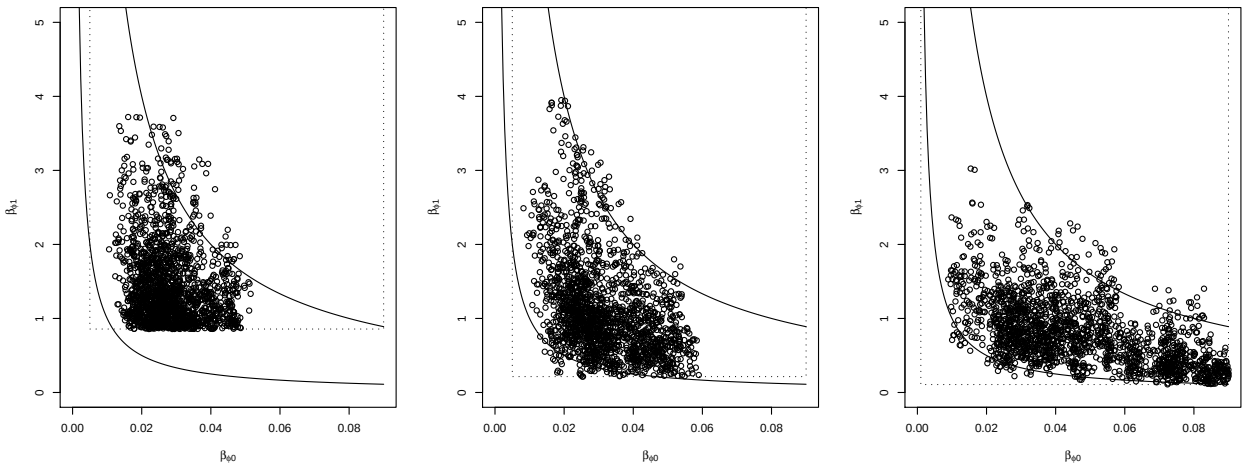

Figure 3: Scatterplot of the realizations of $\beta_{\phi,0}$ and $\beta_{\phi,1}$ for three different sets of priors (denoted by dotted lines). Notice that despite differences in the individual values, the points primarily lie in the region enclosed by the lines $xy = .01$ and $xy = .08$ (solid lines).

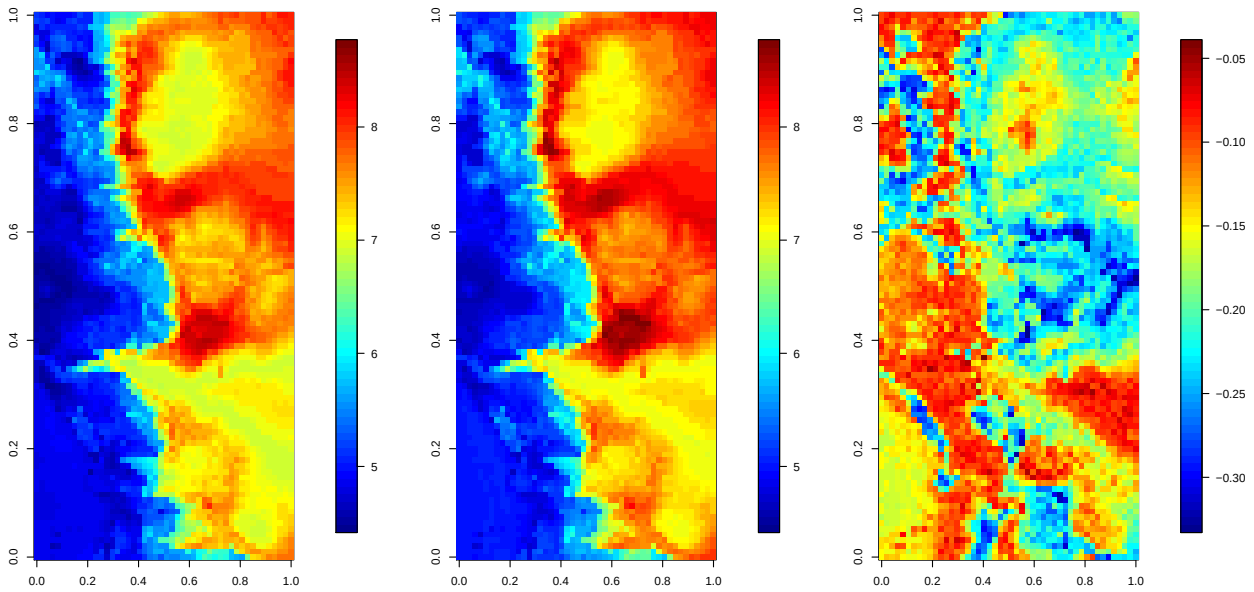

Figure 4: Shows the 25-year return level point estimate (pointwise posterior mean) produced using the original priors (left), the corresponding map using the alternate priors (center), and the difference of two (right).

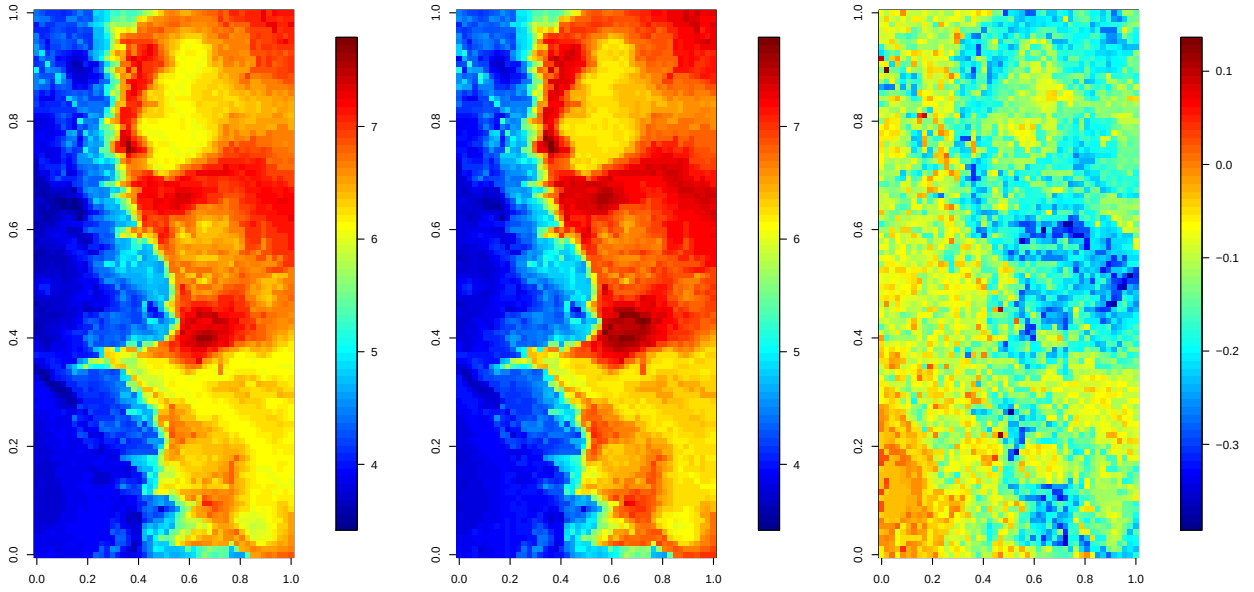

Figure 5: Shows the 25-year return level lower bound (pointwise .025 quantile of the posterior) produced using the original priors (left), the corresponding map using the alternate priors (center), and the difference of two (right).

3 Data precision

Much of the data is recorded to a low level of precision (.1 inches). This causes a bias in the parameter estimates depending on where the threshold is chosen within the precision interval. In the original manuscript, we attempted to address this problem by selecting the threshold in the center of the precision interval which appears to minimize the bias in the parameter estimates. The reviewer suggests that the data should be treated as interval censored.

The authors believe that this data precision issue is a large one. Although the true precipitation distribution is continuous, having records at such a low level of precision results in data that are basically discrete. All of EVT is based on the assumption of a continuous distribution, and how best to reconcile low precision data with EVT is a question beyond the scope of this paper.

As explained in the manuscript, a simulation experiment was performed to find the threshold at which the bias in the two parameters was minimized. GPD data of sample size 200 were simulated and then rounded, and then a GPD was fit to the rounded data at various thresholds. Figures 6 and 7 illustrate the results which indicate that the lowest bias occurs in the center of the precision interval.

A second experiment compared the results of an interval analysis with those of simply choosing the threshold in the center of the interval. A simple interval analysis can be performed by writing the likelihood based on interval observations rather than as exact observations. Rather than defining the likelihood in terms of the GPD density, we defined the likelihood to be $l(\vec{x};\theta) = \prod_{i=1}^{n} \frac{1}{d}[G(x_i + d/2;\sigma_u,\xi) - G(x_i - d/2;\sigma_u,\xi)]$, where $\vec{x} = [x_1, x_2, \ldots, x_n]$ represents the truncated data, G represents the GPD distribution function, and d represents the length of the interval (0.1 inches in our application). Again, we used maximum-likelihood methods in the experiments.

As before, GPD data of sample size 200 was simulated and rounded and the GPD parameters were estimated using both the center-threshold and interval likelihood methods. Individual experiments were run and the parameter estimates for the interval-likelihood and the center-threshold method The two methods' parameter estimates were practically identical for individual simulations. One hundred such experiments were run and the mean-squared error for the two methods were compared. The mean-squared error for the scale parameter ϕ was slightly lower for the interval analysis, but the mse for the shape parameter ξ, was lower for the center-threshold method, which resulting in a lower mean-squared error for the estimated return levels. The difference between the estimates was small compared to the variance of the estimates as:

$$\frac{\frac{1}{n}\sum_{i=1}^{n} |\hat{\phi}_{intvl} - \hat{\phi}_{center}|}{\sqrt{Var[\hat{\phi}_{intvl}]}} = .085 \quad \text{and} \quad \frac{\frac{1}{n}\sum_{i=1}^{n} |\hat{\xi}_{intvl} - \hat{\xi}_{center}|}{\sqrt{Var[\hat{\xi}_{intvl}]}} = .069.$$

Table 1 compares the maximum-likelihood parameter estimates for both the interval and center-threshold methods for the precipitation data at several stations. The estimates are practically identical.

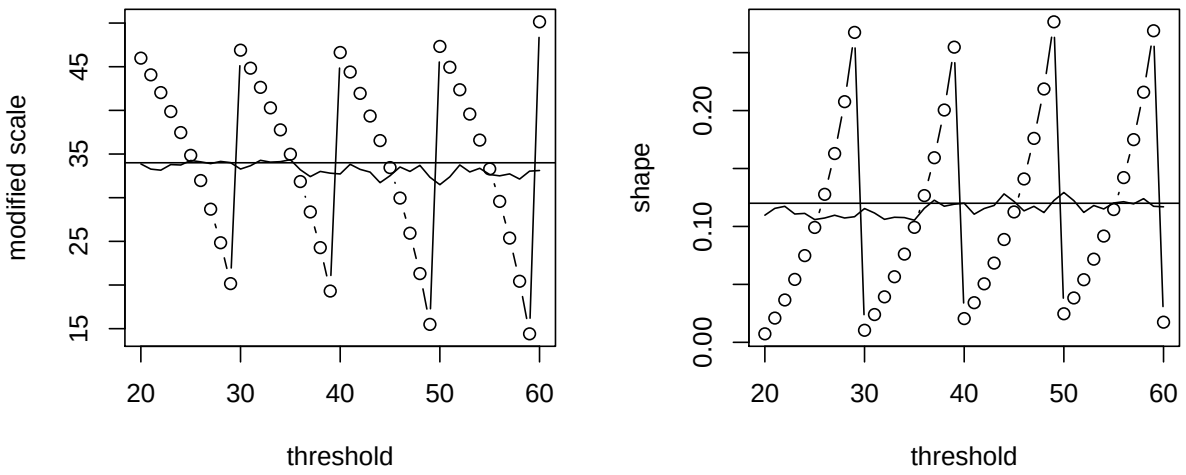

Figure 6: Shows one example of the simulation to find the best threshold value. GPD data were simulated and rounded and ML methods were used to fit both the original and rounded data. Dots show the parameter estimates for the rounded data, the solid line shows estimates for the original data, and the horizontal line shows the true parameter value.

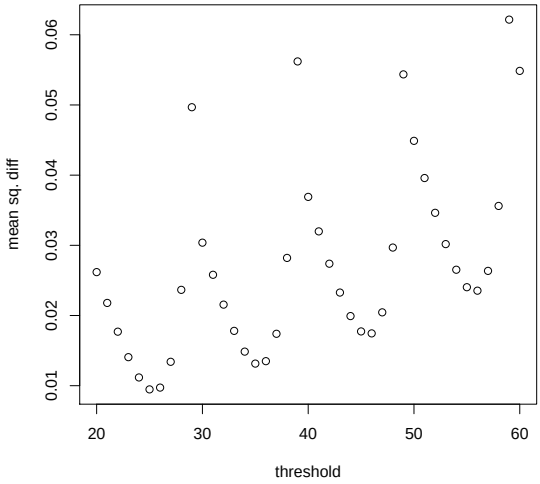

Figure 7: Shows the mean-squared error of the shape parameter for 100 experiments of sample size 200 each. Notice that the error appears to be minimized if the threshold is chosen in the center of the precision interval at .25 .35 or .45 inches.

	ϕ		ξ	
Station	center	interval	center	interval
Boulder	3.75	3.74	0.091	0.098
White Rock	3.59	3.58	0.027	0.029
Allenspark	3.52	3.51	0.037	0.042
Hoyt	3.67	3.66	0.063	0.066
Manitou Spgs	3.53	3.51	0.205	0.217

Table 1: Comparison of parameter estimates using the center-threshold method verses the interval-likelihood method (threshold = .55 inches).

4 Threshold sensitivity

Threshold sensitivity was tested by running corresponding models with thresholds set at .35, .45, .55, .65, and .75 inches. Sensitivity was assessed by viewing the posterior distributions of the individual parameters and the 25-year return level. Figure 8 shows the 25-year return level posterior for a station in the mountains, on the Front Range, and on the plains for various thresholds. Figure 9 shows the posterior for the GPD shape parameter ξ. The return level plots indicate only a little sensitivity to the choice of threshold. Both the plains location and the Front Range location display slightly increased values at the .35 threshold, perhaps indicating that this threshold choice is too low for such locations. More interesting are the plots of ξ. This parameter is hard to estimate, and it is not surprising to see some variability in the posterior due to the choice of threshold. However, differences in the posterior ξ_{plains} (and to a lesser extent $\xi_{mountains}$) seem to indicate that this parameter can be estimated reliably for thresholds above .55 inches. This sensitivity analysis persuaded us to change our threshold from .45 inches to .55 inches. Notice that the return level estimates for Fort Collins and Hoyt (both of which use this parameter) do not have much difference between the .45 and .55 thresholds, indicating that the GPD scale parameter ϕ has also changed at these locations to yield similar return levels despite different values for ξ.

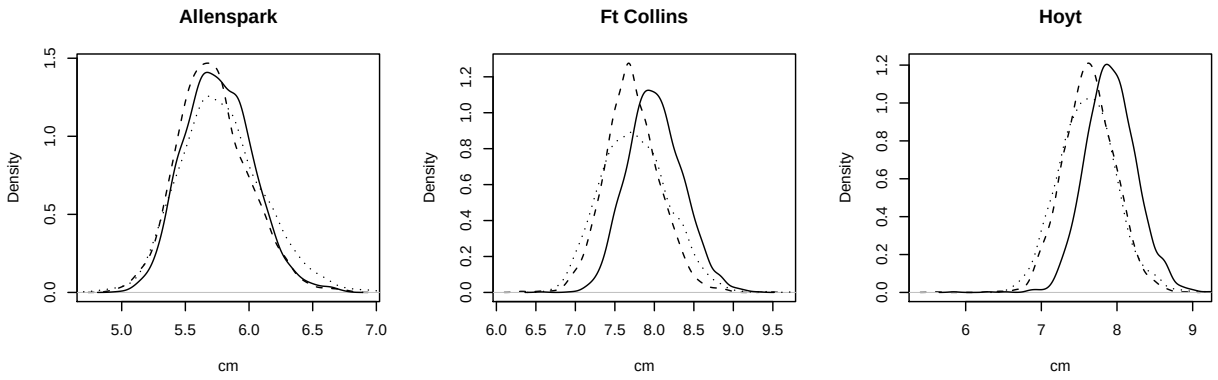

Figure 8: Shows the posterior distribution of the 25-year return level at three representative station locations at different thresholds. The solid line is at .35 inches, the dashed at .45 inches, and the dotted at .55 inches. Allenspark is a mountain station, Fort Collins is on the Front Range, and Hoyt is on the plains.

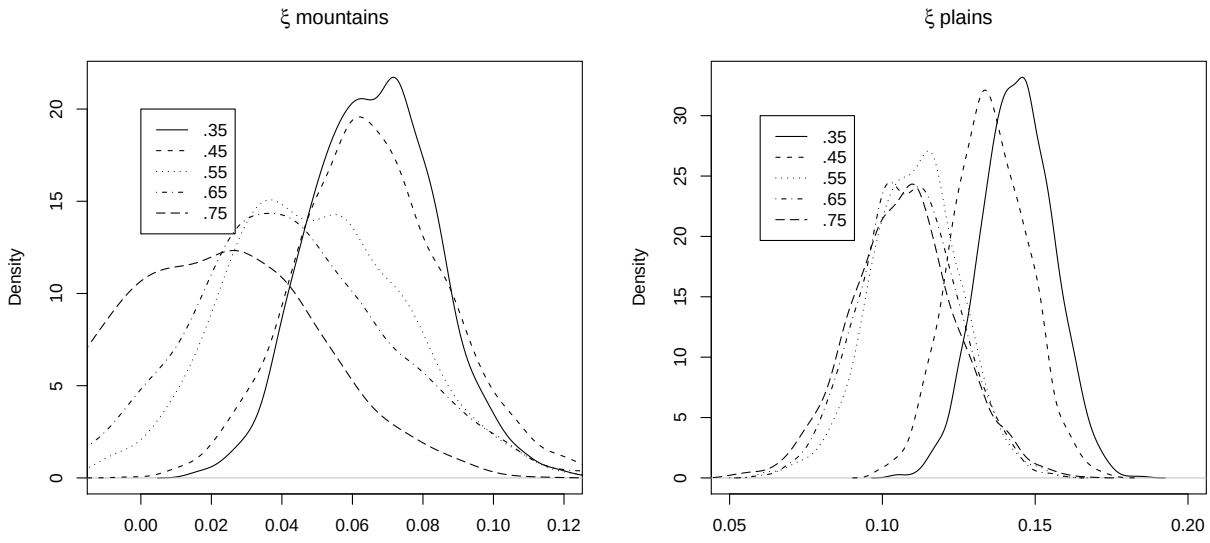

Figure 9: Shows the posterior distribution of the GPD shape parameter ξ for the mountain locations (left) and the plains locations (right).

5 Residual dependence in the observations

The main point of Section 2.3 is to clarify the assumption we are making about spatial dependence, i.e. that our model assumes the data are spatially conditionally independent given the stations parameters. As stated in the paper, this is not really true as one storm could yield high measurements at more than one station location. We do feel that any remaining spatial dependence in the data has little effect on the results. We also tried to assess how much residual spatial dependence is in the data using the first-order variogram (Figure 10), which showed the residual dependence was of very limited range. As a caution to general application, we note in the paper that this assumption of conditional spatial independence could not be made in other applications such as stream flow measurements.

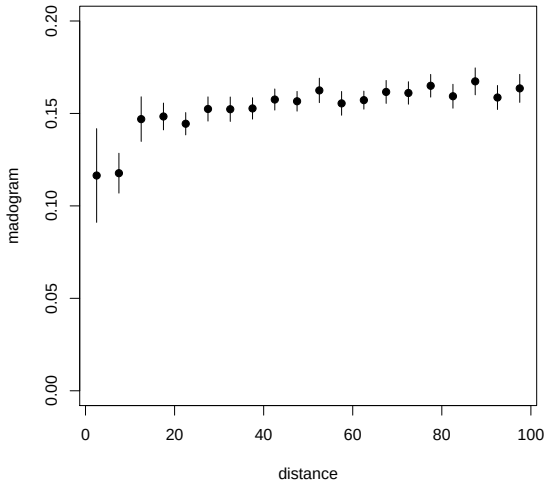

Figure 10: Shows the first-order variogram Cooley et al. (2006) used to assess residual dependence in the data. A value of 1/6th indicates independence. Very little dependence is shown beyond a distance of about 15 miles.

6 Independence of priors for $\phi(\boldsymbol{x})$ and $\xi(\boldsymbol{x})$

We assume independent priors for the GPD shape and scale parameters to allow us to model each parameter's relationship with the latent process individually. A priori, we know that there should be a negative dependence with these parameters, but how to build this into the prior is unclear. Our primary objective is to determine the spatial behavior of these parameters and this is most expeditiously done by modeling them independently. The posterior distribution of these parameters, however, does show a weak but significant negative dependence (Figure 11).

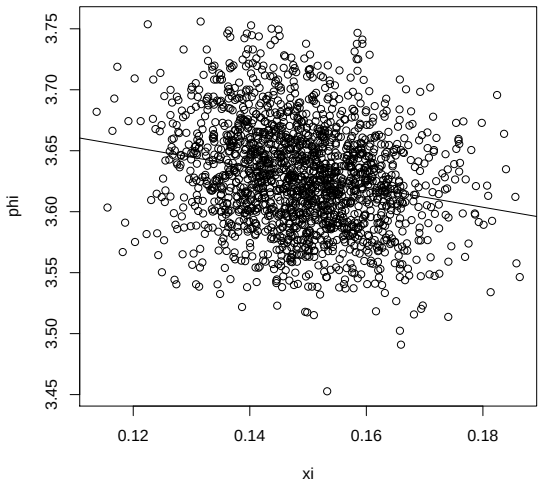

Figure 11: Shows a scatterplot of the posterior realizations of ϕ for a plains location and ξ_{plains}.

7 Assumptions of stationarity and isotropy

The climate space uses elevation and mean seasonal precipitation to define its coordinates. Because these coordinates are measured on different scales, there is no reason to assume a common range parameter.

Each location in the study region corresponds to a location in the climate space. Plotting these locations in the original scales yields a cloud of points with a domain roughly given by [4000,12000 ft] x [20,70 cm] (Figure 12). Certainly there is no reason to believe that a common range parameter should be used in such a space. We transform these coordinates so that a common range is not unreasonable. We treat the points as observations from a bivariate normal, estimate its parameters, and then transform the points such that they have a covariance matrix of identity (Figure 12). Although the transformation yields points whose domain is roughly the same scale, it does not guarantee isotropy or stationarity.

Our assumptions of isotropy and stationarity are basically simplifying assumptions. With only 56 station locations, we believed that it would be impossible to identify non-stationarity or anisotropy. We tested for geometric anisotropy (Cressie, 1993) in the GPD parameter ϕ's spatial process. We took the maximum-likelihood estimates for ϕ and fit an anisotropic variogram model using maximum likelihood methods. The model constructs a new distance metric which requires two range parameters $\beta_{\phi,1a}$ and $\beta_{\phi,1b}$ and an angle (in radians) for the direction of the major axis θ. The point estimate for ($\beta_{\phi,1a}$, $\beta_{\phi,1a}$, θ) was (5.0, 0.8, -0.25), which gives some indication that the data may be anisotropic, especially due to the difference in the range parameters. However, the asymptotic standard errors of these estimates were (17.1, 3.3, 0.87) showing that an assumption of isotropy could not be rejected with any reasonable confidence interval.

Our assumption of an exponential variogram model is also a simplifying assumption. One could argue that the Matérn class of variogram models would be a better model; however, using that class would either require estimating the smoothness parameter (which is confounded with the sill and range parameters) or a level of smoothness would have to be assumed. When modeling values that are not observed exactly, the correct variogram model is less important than when interpolating.

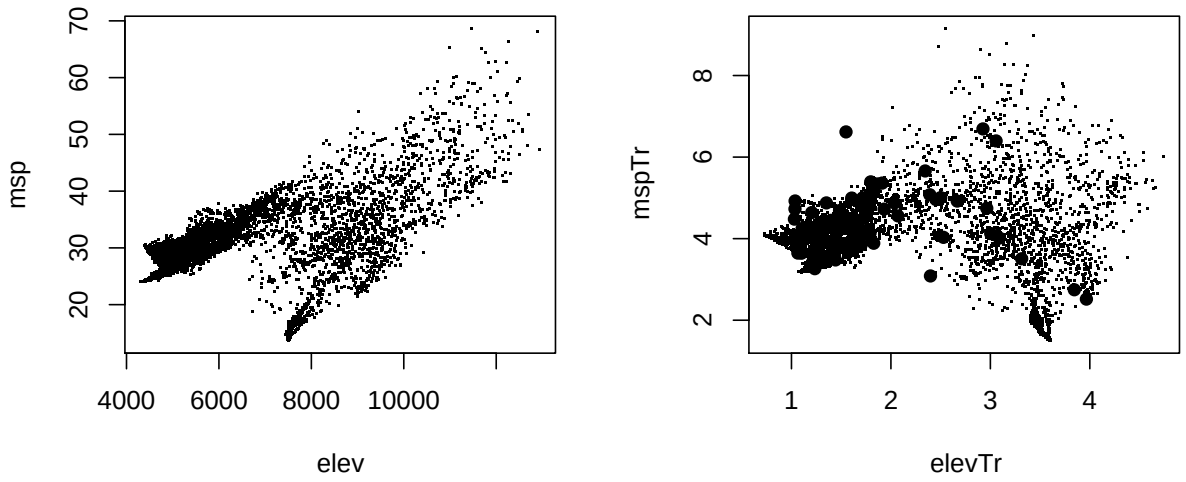

Figure 12: Shows the study area locations in the original and transformed climate space locations. Station locations are shown with large dots in the transformed space.

References

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

Berger, J., DeOiveira, V., and Sanso, B. (2001). Objective bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 96:1361–1374.

Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for spatial max-stable random fields. In *Statistics for Dependendent Data: STATDEP2005*, Springer Lecture Notes in Statistics. Springer, London. In press.

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

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

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.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99:250–261.