

# Exploration and inference in spatial extremes using empirical basis functions

Samuel A Morris<sup>1</sup>, Brian J Reich<sup>1</sup>, and Emeric Thibaud<sup>2</sup>

July 31, 2016

## Abstract

words...

**Key words:** Bayesian data analysis; principal components analysis; max-stable process; spectral representation.

---

<sup>1</sup>North Carolina State University

<sup>2</sup>Colorado State University

# 1 Introduction

The spatial Extreme Value Analysis (EVA) literature is expanding rapidly (Davison et al., 2012) to meet the demands of researchers to improve estimates of rare-event probabilities by borrowing information across space and to estimate the probability of extreme events occurring simultaneously at multiple locations. Environmental datasets commonly include observations from hundreds or thousands of locations, and advanced tools are required to explore and analyze these data. For Gaussian data, Principle Components Analysis (Everitt and Hothorn, 2008, PCA), also known as Empirically Orthogonal Functions (Hannachi et al., 2007, EOF), has proven to be a powerful tool to study correlation between spatial locations; understand the most important large-scale spatial features; and reduce the dimension of the problem to allow for simple computation even for massive datasets. Computation and exploration are arguably more difficult for EVA than Gaussian data, yet to our knowledge no tool analogous to spatial PCA has been developed for EVA.

In EVA, extremes are separated from the bulk of the distribution by either analyzing only points above a threshold or block maximums (Coles, 2001), e.g., the annual maximum of the daily precipitation. A natural spatial model for block maximum at several spatial locations is the max-stable process, which, under certain conditions, arises as the limit of the location-wise maximum of infinitely-many spatial processes (de Haan and Ferreira, 2006). De Haan (1984) showed that any max-stable process can be represented in terms of a countable number of spatial processes (e.g., stationary log Gaussian processes), and a finite truncation of this representation has been used for conditional simulation (Wang and Stoev, 2011). The proposed empirical basis function (EBF) approach also uses a finite truncation of the spectral representation, and develops a method-of-moments estimator for the underlying spatial processes.

In addition to exploratory analysis, we show that the EBFs can be used for Bayesian inference on the marginal parameters at each location and to test for covariate effects. Fully-Bayesian analysis using max-stable processes is cumbersome for large data sets (Wadsworth and Tawn, 2014; Thibaud and Opitz, 2015).

One option is to use non-max-stable models that retain extremal dependence such as the skew-t process in Morris et al [put on arxiv](#). Alternatively, Reich and Shaby (2012) propose a low-rank method based on spatial kernel functions, and others have used pairwise (Padoan et al., 2010; Huser and Davison, 2014) and trivariate (Genton et al., 2011) likelihood methods for parameter estimation.

In this paper we develop methodology to use the spectral representation of a max-stable process to identify a small set of EBFs that capture the most important spatial features of the data. Unlike PCA/EOFs, the EBFs are not orthogonal, nonetheless these spatial functions can be plotted for exploratory analysis to reveal important spatial trends. The EBFs can also be used in a second-stage statistical analysis. By basing the spatial dependence on EBFs, the resulting spatial analysis does not require dubious assumptions such as stationarity. In addition, a Bayesian analysis for either block-maximum or point above a threshold is computationally feasible for large datasets because the entire spatial process is represented by a small number of basis functions.

The paper proceeds as follows. In Section 2 we present the low-rank model. [More here about the sections](#)

## 2 Model

Let  $Y_t(\mathbf{s})$  be the observation at spatial location  $\mathbf{s}$  and time  $t$ . We temporarily drop the subscript  $t$  and describe the model for the process  $Y(\mathbf{s})$  for a single time point, but return to the spatiotemporal setting in Section 3. To focus attention on the extreme values, we emphasize the statistical model for exceedances above a location-specific threshold  $T(\mathbf{s})$ . We begin by specifying a spatial model for the complete data  $Y(\mathbf{s})$  and then use the censored likelihood defined by  $T(\mathbf{s})$  for inference as described in Section 4. Although the model presented implements a censored likelihood, the model also can fit uncensored data (such as block-maxima) by setting  $T(\mathbf{s}) = -\infty$ .

53 Spatial dependence is captured by modeling  $Y(\mathbf{s})$  as a max-stable process (de Haan and Ferreira, 2006).  
 54 Max-stable processes have generalized extreme value (GEV; see A.1) marginal distribution. The GEV has  
 55 three parameters: location  $\mu(\mathbf{s})$ ; scale  $\sigma(\mathbf{s})$ ; and shape  $\xi(\mathbf{s})$ . Spatial dependence is present both in the GEV  
 56 parameters but also the standardized residual process

$$Z(\mathbf{s}) = \left\{ 1 + \frac{\xi(\mathbf{s})}{\sigma(\mathbf{s})} [Y(\mathbf{s}) - \mu(\mathbf{s})] \right\}^{1/\xi(\mathbf{s})}, \quad (1)$$

57 which has unit Fréchet (i.e., GEV with location, scale, and shape all equal one) marginal distribution for all  
 58  $\mathbf{s}$ .

59 Our objective is to identify a low-rank model for the spatial dependence of  $Z(\mathbf{s})$ . (de Haan, 1984,  
 60 Chapter 9) show that any max-stable process can be written as

$$Z(\mathbf{s}) = \sup_l B(\mathbf{s}, \mathbf{t}_l) A_l \quad (2)$$

61 where the function  $B$  satisfies  $B(\mathbf{s}, \mathbf{t}) > 0$  for all  $(\mathbf{s}, \mathbf{t})$  and  $\int B(\mathbf{s}, \mathbf{t}) d\mathbf{t} = 1$  for all  $\mathbf{s}$ , and  $(\mathbf{t}_l, A_l)$  for  
 62  $l = 1, \dots, \infty$  are a Poisson process with intensity measure  $dA d\mathbf{t}/A^2$ .

63 To arrive at a low-rank model, we assume there are a finite and known number of spatial basis functions  
 64  $B_1(\mathbf{s}), \dots, B_L(\mathbf{s})$  that explain the important spatial variation in the process. As in de Haan's expansion, the  
 65 basis functions are restricted so that  $B_l(\mathbf{s}) > 0$  and  $\sum_{l=1}^L B_l(\mathbf{s}) = 1$  for all  $\mathbf{s}$ . Because it is unrealistic to assume  
 66 that realizations of  $Z$  are exactly functions of  $L$  basis functions, we include independent error variables  $\epsilon(\mathbf{s})$   
 67 to capture variation not explained by the  $B_l(\mathbf{s})$ . We follow Reich and Shaby (2012) and decompose  $Z(\mathbf{s})$  as  
 68  $Z(\mathbf{s}) = \theta(\mathbf{s})\epsilon(\mathbf{s})$  where  $\theta(\mathbf{s})$  is a spatial process and  $\epsilon(\mathbf{s}) \stackrel{iid}{\sim} \text{GEV}(1, \alpha, \alpha)$  is independent error. The spatial

69 component is

$$\theta(\mathbf{s}) = \left( \sum_{l=1}^L B_l(\mathbf{s})^{1/\alpha} A_l \right)^\alpha. \quad (3)$$

70 If  $B_l(\mathbf{s}) > 0$ ,  $\sum_{l=1}^L B_l(\mathbf{s}) = 1$  for all  $\mathbf{s}$ , and the  $A_l$  have positive stable (PS; A.2) distribution  $A_l \stackrel{iid}{\sim} \text{PS}(\alpha)$ ,  
 71 then  $Z(\mathbf{s})$  is max-stable and has unit Fréchet marginal distributions.

72 Extremal spatial dependence for max-stable processes can be summarized by the extremal coefficient  
 73 (Schlather and Tawn, 2003, EC)  $\vartheta(\mathbf{s}, \mathbf{t}) \in [1, 2]$ , where

$$\text{Prob}[Z(\mathbf{s}) < c, Z(\mathbf{t}) < c] = \text{Prob}[Z(\mathbf{s}) < c]^{\vartheta(\mathbf{s}, \mathbf{t})}. \quad (4)$$

74 For the PS random effects model the EC has the form

$$\vartheta(\mathbf{s}, \mathbf{t}) = \sum_{l=1}^L \left[ B_l(\mathbf{s})^{1/\alpha} + B_l(\mathbf{t})^{1/\alpha} \right]^\alpha. \quad (5)$$

75 In particular,  $\vartheta(\mathbf{s}, \mathbf{s}) = 2^\alpha$  for all  $\mathbf{s}$ .

### 76 **3 Estimating the spatial dependence function**

77 To estimate the extremal coefficient function, we consider the process at  $n_s$  spatial locations  $\mathbf{s}_1, \dots, \mathbf{s}_{n_s}$   
 78 and  $n_t$  times  $t = 1, \dots, n_t$ . The basis functions are fixed over time, but the random effects and errors are  
 79 independent over time. That is  $Z_t(\mathbf{s}) = \theta_t(\mathbf{s})\epsilon_t(\mathbf{s})$  where  $\theta_t(\mathbf{s}) = \left( \sum_{l=1}^L B_l(\mathbf{s})^{1/\alpha} A_{lt} \right)^\alpha$ ,  $A_{lt} \stackrel{iid}{\sim} \text{PS}(\alpha)$ , and  
 80  $\epsilon_t(\mathbf{s}) \stackrel{iid}{\sim} \text{GEV}(1, \alpha, \alpha)$ . Denote  $Y_t(\mathbf{s}_i) = Y_{it}$ ,  $B_l(\mathbf{s}_i) = B_{il}$ ,  $T(\mathbf{s}_i) = T_i$ , and  $\vartheta(\mathbf{s}_i, \mathbf{s}_j) = \vartheta_{ij}$ .

81 In this section we develop an algorithm to estimate the spatial dependence parameter  $\alpha$  and the  $n_s \times L$   
 82 matrix  $\mathbf{B} = \{B_{il}\}$ . Our algorithm has the following steps:

- 83 (1) Obtain an initial estimate of the extremal coefficient for each pair of locations,  $\hat{\vartheta}_{ij}$ .
- 84 (2) Spatially smooth these initial estimates  $\hat{\vartheta}_{ij}$  using kernel smoothing to obtain  $\tilde{\vartheta}_{ij}$ .
- 85 (3) Estimate the spatial dependence parameters by minimizing the difference between model-based coefficients,  $\vartheta_{ij}$ , and smoothed coefficients,  $\tilde{\vartheta}_{ij}$ .

87 The first-stage estimates are obtained using an empirical estimate as follows. To estimate the spatial  
 88 dependence we first remove variation in the marginal distribution. Let  $U_{it} = \sum_{k=1}^{n_t} I[Y_{ik} < Y_{it}]/n_t$ , so that  
 89 the  $U_{it}$  are approximately uniform at each location. Then for some extreme probability  $q \in (0, 1)$ , solving  
 90 (4) suggests the estimate

$$\hat{\vartheta}_{ij}(q) = \frac{\log[Q_{ij}(q)]}{\log(q)}, \quad (6)$$

91 where  $Q_{ij}(q) = \sum_{t=1}^{n_t} I[U_{it} < q, U_{jt} < q]/n_t$  is the sample proportion of the time points at which both  
 92 sites are less than  $q$ . Since all large  $q$  give valid estimates, we average over a grid of  $q$  with  $q_1 < \dots < q_{n_q}$

$$\hat{\vartheta}_{ij} = \frac{1}{n_q} \sum_{j=1}^{n_q} \hat{\vartheta}_{ij}(q_j). \quad (7)$$

93 Assuming the true EC is smooth over space, the initial estimates  $\hat{\vartheta}_{ij}$  can be improved by smoothing. Let

$$\tilde{\vartheta}_{ij} = \frac{\sum_{u=1}^{n_s} \sum_{v=1}^{n_s} w_{iu} w_{jv} \hat{\vartheta}_{uv}}{\sum_{u=1}^{n_s} \sum_{v=1}^{n_s} w_{iu} w_{jv}}, \quad (8)$$

94 where  $w_{iu} = \exp[-(|\mathbf{s}_i - \mathbf{s}'_u|/\phi)^2]$  is the Gaussian kernel function with bandwidth  $\phi$ . The elements  $\hat{\vartheta}_{ii}$   
 95 do not contribute any information as  $\hat{\vartheta}_{ii} = 1$  for all  $i$  by construction. To eliminate the influence of these  
 96 estimates we set  $w_{ii} = 0$ . However, this approach does give imputed values  $\tilde{\vartheta}_{ii}$ , which provide information  
 97 about small-scale spatial variability.

98 The dependence parameters  $B_{lt}$  and  $\alpha$  are estimated by comparing estimates  $\tilde{\vartheta}_{ij}$  with the model-based  
 99 values  $\vartheta_{ij}$ . For all  $i$ ,  $\vartheta_{ii} = 2^\alpha$ , and therefore we set  $\alpha$  to  $\hat{\alpha} = \log_2(\sum_{i=1}^{n_s} \tilde{\vartheta}_{ii}/n_s)$ . Given  $\alpha = \hat{\alpha}$ , it remains  
 100 to estimate  $\mathbf{B}$ . Similarly to Smith (1990) for a stationary max-stable process, we use squared-error loss, so  
 101 the estimate  $\hat{\mathbf{B}}$  is the minimizer of

$$\sum_{i < j} \left( \tilde{\vartheta}_{ij} - \vartheta_{ij} \right)^2 = \sum_{i < j} \left( \tilde{\vartheta}_{ji} - \sum_{l=1}^L [B_{il}^{1/\hat{\alpha}} + B_{jl}^{1/\hat{\alpha}}]^{\hat{\alpha}} \right)^2 \quad (9)$$

102 under the restrictions that  $B_{il} \geq 0$  for all  $i$  and  $l$  and  $\sum_{l=1}^L B_{il} = 1$  for all  $i$ . Since the minimizer of  
 103 (9) does not have a closed form, we use block coordinate descent to obtain  $\hat{\mathbf{B}}$ . We cycle through spatial  
 104 locations and update the vectors  $(B_{i1}, \dots, B_{iL})$  conditioned on the values for the other location and repeat  
 105 until convergence. At each step, we use the restricted optimization routine in the R function `optim`. This  
 106 algorithm gives estimates of the  $B_{il}$  at the  $n_s$  data locations, but is easily extended to all  $\mathbf{s}$  for spatial  
 107 prediction. The kernel smoothing step ensures that the estimates for  $\hat{B}_{il}$  are spatially smooth, and thus  
 108 interpolation of the  $\hat{B}_{il}$  gives spatial functions  $\hat{B}_l(\mathbf{s})$ .

109 These functions provide useful exploratory data analysis techniques. Maps of  $\hat{B}_l(\mathbf{s})$  show important  
 110 spatial features in the extremal dependence. Furthermore, they allow for a non-stationary spatial dependence  
 111 structure. The relative contribution of each term can be measured by

$$v_l = \frac{1}{n_s} \sum_{i=1}^{n_s} \hat{B}_{il}. \quad (10)$$

112 Since  $\sum_{l=1}^L \hat{B}_{il} = 1$  for all  $i$ , we have  $\sum_{l=1}^L v_l = 1$ . Therefore, terms with large  $v_l$  are the most important.  
 113 The order of the terms is arbitrary, and so we reorder the terms so that  $v_1 \geq \dots \geq v_L$ .

## 4 Bayesian implementation details

For our data analysis in Section 5 we allow the GEV location and scale parameters, denoted  $\mu_t(\mathbf{s})$  and scale  $\sigma_t(\mathbf{s})$  respectively, to vary with space and time. The model we choose is as follows

$$\mu_t(\mathbf{s}) = \beta_{1,\text{int}}(\mathbf{s}) + \beta_{1,\text{time}}(\mathbf{s})t \quad (11)$$

$$\log[\sigma_t(\mathbf{s})] = \beta_{2,\text{int}}(\mathbf{s}) + \beta_{2,\text{time}}(\mathbf{s})t \quad (12)$$

where

$$\begin{aligned} \beta_{1,\text{int}}(\mathbf{s}) &\sim \text{N}(\mu_{1,\text{int}}, \sigma_{1,\text{int}}^2 \Sigma) & \beta_{1,\text{time}}(\mathbf{s}) &\sim \text{N}(\mu_{1,\text{time}}, \sigma_{1,\text{time}}^2 \Sigma) \\ \beta_{2,\text{int}}(\mathbf{s}) &\sim \text{N}(\mu_{2,\text{int}}, \sigma_{2,\text{int}}^2 \Sigma) & \beta_{2,\text{time}}(\mathbf{s}) &\sim \text{N}(\mu_{2,\text{time}}, \sigma_{2,\text{time}}^2 \Sigma) \end{aligned} \quad (13)$$

are Gaussian process priors and  $\Sigma$  is an exponential spatial correlation matrix obtained from  $\rho(h) = \exp\left\{-\frac{h}{\phi}\right\}$  where  $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$  is the Euclidean distance between sites  $\mathbf{s}_1$  and  $\mathbf{s}_2$ . The GEV shape parameter  $\xi$  is held constant over space and time because this parameter is challenging to estimate. Collectively, let the marginal GEV parameters at location  $i$  and time  $t$  be  $\Theta_{it} = \{\mu_{it}, \sigma_{it}, \xi\}$  where  $\mu_{it} = \mu_t(\mathbf{s}_i)$  and  $\sigma_{it} = \sigma_t(\mathbf{s}_i)$ .

As shown in Reich and Shaby (2012), the uncensored responses  $Y_t(\mathbf{s})$  are conditionally independent given the spatial random effects, with conditional distribution

$$Y_{it} | \theta_{it}, \Theta_{it} \overset{\text{indep}}{\sim} \text{GEV}(\mu_{it}^*, \sigma_{it}^*, \xi^*), \quad (14)$$

where  $\mu_{it}^* = \mu_{it} + \frac{\sigma_{it}}{\xi}(\theta_{it}^\xi - 1)$ ,  $\sigma_{it}^* = \alpha \sigma_{it} \theta_{it}^\xi$ , and  $\xi^* = \alpha \xi$ . Therefore, the conditional likelihood conveniently factors across observations; marginalizing over the random effect  $\theta_{it}$  induces extremal spatial



dependence. To focus on the extreme values above the local threshold  $T_i$ , we use the censored likelihood

$$d(y; \theta_{it}, \Theta_{it}, T_i) = \begin{cases} F(y; \mu_{it}^*, \sigma_{it}^*, \xi^*) & y \leq T_i \\ f(y; \mu_{it}^*, \sigma_{it}^*, \xi^*) & y > T_i, \end{cases} \quad (15)$$

where  $F$  and  $f$  are the GEV distribution and density functions, respectively, defined in ??.

In summary, given the estimates of  $\alpha$  and  $\mathbf{B}$ , the hierarchical model is

$$\begin{aligned} Y_{it} | \theta_{ij} &\overset{indep}{\sim} d(y; \theta_{it}, \Theta_{it}, T_i) \\ \theta_{it} &= \left( \sum_{l=1}^L \hat{B}_{il}^{1/\hat{\alpha}} A_{lt} \right)^{\hat{\alpha}} \quad \text{where} \quad A_{lt} \overset{iid}{\sim} PS(\hat{\alpha}) \\ \mu_{it} &= \beta_{1,int}(\mathbf{s}_i) + \beta_{1,time}(\mathbf{s}_i)t \\ \log(\sigma_{it}) &= \beta_{2,int}(\mathbf{s}) + \beta_{2,time}(\mathbf{s})t. \end{aligned} \quad (16)$$

We estimate parameters  $\Theta = \{A_{lt}, \beta_1, \beta_2, \xi\}$  using Markov chain Monte Carlo methods. We use a Metropolis-Hastings algorithm to update the model parameters with random walk candidate distributions for all parameters. The PS density is challenging to evaluate as it does not have a closed form. One technique to avoid this complication is to incorporate auxiliary random variables (Stephenson, 2009), but we opt for a numerical approximation to the integral as described in A.2. The hyperparameters  $\mu_{1,int}, \mu_{1,time}, \mu_{2,int}, \mu_{2,time}$  and  $\sigma_{1,int}^2, \sigma_{1,time}^2, \sigma_{2,int}^2, \sigma_{2,time}^2$  are updated using Gibbs sampling since their prior distributions are conjugate.

The first-stage estimate of the extremal coefficients has three tuning parameters: the quantile thresholds  $q_1, \dots, q_{n_q}$ , the kernel bandwidth  $\phi$ , and the number of terms  $L$ . In Section 5 we explore a few possibilities for  $L$  and discuss sensitivity to this choice. The second-stage Bayesian analysis requires selecting thresholds  $T_i, \dots, T_{n_s}$ . For this we use spatially smoothed sample quantiles. That is, we set  $T_i$  to the 0.95 quantile of

140 the  $Y_{it}$  and its five nearest neighbors.

## 141 **5 Data analysis**

142 In this section, we illustrate our method with two data analyses. In Section 5.2, we present a points above  
143 a threshold analysis using annual acreage burned due to wildfires in Georgia from 1965 – 2014. This is  
144 followed in Section 5.4 by an analysis of block maxima precipitation data in the eastern U.S. We compare  
145 our method with another method that uses standardized Gaussian kernels for the spatial basis functions.

### 146 **5.1 Gaussian kernel basis functions**

147 To provide a comparison of our model with another approach, we also fit a model that uses standardized  
148 Gaussian kernels for the spatial basis functions (Reich and Shaby, 2012). In this method, Reich and Shaby  
149 introduce a set of  $\mathbf{k}_1, \dots, \mathbf{k}_L$  spatial knots and use standardized Gaussian kernel functions (GSK) instead of  
150 using EBFs for the  $\hat{B}_l(\mathbf{s})$ . For the comparison between EBF and GSK methods, we use the same number of  
151 basis functions. We obtain estimates the spatial bandwidth  $\hat{\rho}$  and spatial dependence  $\hat{\alpha}$ , using the same block  
152 coordinate descent as with the EBF method, and treat these as fixed in the MCMC. For more information  
153 about the basis functions, see A.3.

### 154 **5.2 Analysis of extreme Georgia fires**

155 The dataset used for our application is composed of yearly acreage burned due to wildfires for each county in  
156 Georgia from 1965 – 2014 (<http://weather.gfc.state.ga.us/FireData/>). Figure 1 shows  
157 the time series of  $\log(\text{acres burned})$  for 25 randomly selected counties. Based on this plot and other ex-  
158 ploratory analysis, we see no evidence of non-linear trends and proceed with linear time trends for the GEV  
159 location and scale parameters.

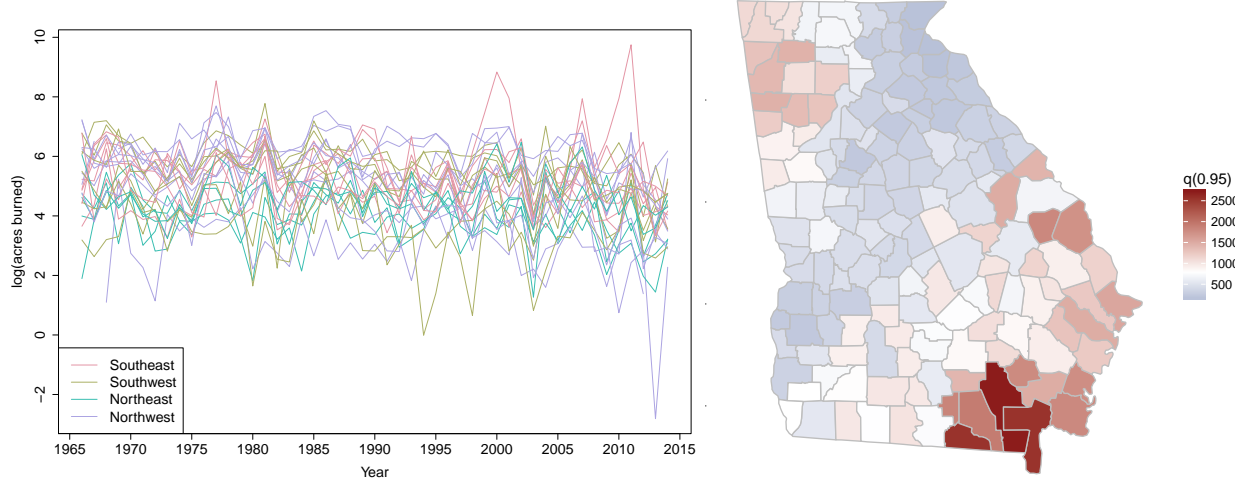


Figure 1: Time series of log acres burned for 25 randomly selected counties with colors coding the county's quadrant (left), and spatially smoothed threshold values,  $T_i$  for each county (right).

We estimate the extremal coefficient function  $\hat{\theta}_{ij}$  by setting  $q_1 = 0.90$  and using  $n_q = 100$ . With more data, it would be possible to increase  $q_1$ , but we set  $q_1 = 0.90$  to increase the stability when estimating  $\hat{\theta}_{ij}$ . Because these data are not block-maxima, we select a site-specific threshold  $T_i$  to use in the analysis with the following algorithm. Without some adjustment to the data, it is challenging to borrow information across sites to inform the threshold selection. We first standardize the data, separated by county, by subtracting the site's median and dividing by the site's interquartile range. Denote the standardized data by  $\tilde{Y}_i$ . Then we combine all sites together and plot a mean residual plot for  $\tilde{Y}_{it}, i = 1, \dots, n_s$  and  $t = 1, \dots, n_t$ . The mean residual plot is given in Figure 2. Based upon the mean residual plot, we select the 95th percentile for the threshold. To calculate  $T_i$  for each county, we use the 95th percentile for the combined data for county  $i$  and its five closest counties (see Figure 1).

The empirical basis functions for the analysis can be used to help explore spatial dependence in the extremes. See Figure ?? for the first six (of 25) EBFs for the Georgia fire data. We also plot the cumulative sum of the contributions for each basis function in 4. As a comparison, we provide the first six principal components of the fire data along with the cumulative sum of the first 25 eigenvalues in Figure 19.

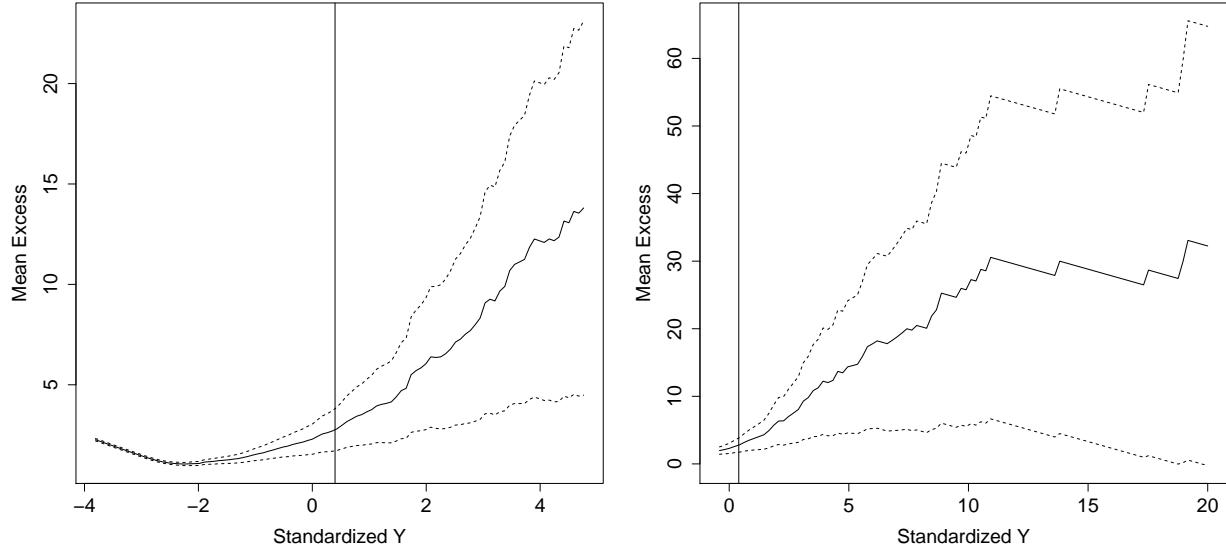


Figure 2: Mean residual plot for the data pooled across counties after standardizing using the county's median and interquartile range. The two panels show different ranges on the x-axis and include a vertical line at the sample 95th percentile.

### 5.3 Results for fire analysis

We use 10-fold cross-validation to assess the predictive performance of a model. For each method, we randomly select 90% of the observations across counties and years to be used as a training set to fit the model. The remaining 10% of sites and years are withheld for testing model predictions. To assess the predictions for the test set, we use quantile scores and Brier scores (Gneiting and Raftery, 2007). The quantile score is given by [give formula](#). The Brier score is given by [give formula](#). For both of these methods, a lower score indicates a better fit. The Brier and quantile scores for the fire analysis are given in Table 3.

[Figure here with BS and QS for the fire data](#)

Based on the cross-validation results, we run a full analysis using all of the data with  $L = 25$ . Figure 13 and Figure 14 give the county by county posterior means of  $\beta_{1,time}$  and  $\beta_{2,time}$  for EBF and GSK respectively. Figure 15 and Figure 16 give the county by county posterior  $P(\beta_{1,time} > 0)$  and  $P(\beta_{2,time} > 0)$  for EBF and GSK respectively.

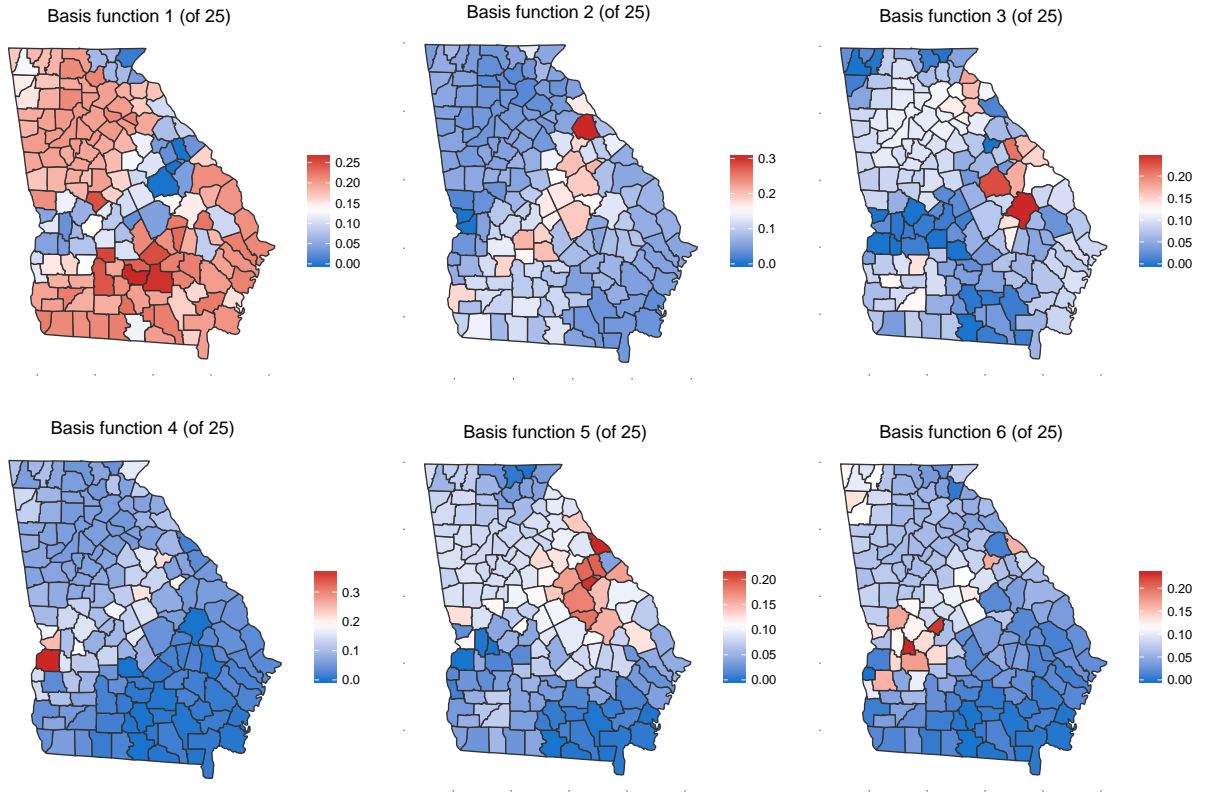


Figure 3: First six EBFs for the Georgia fire data.

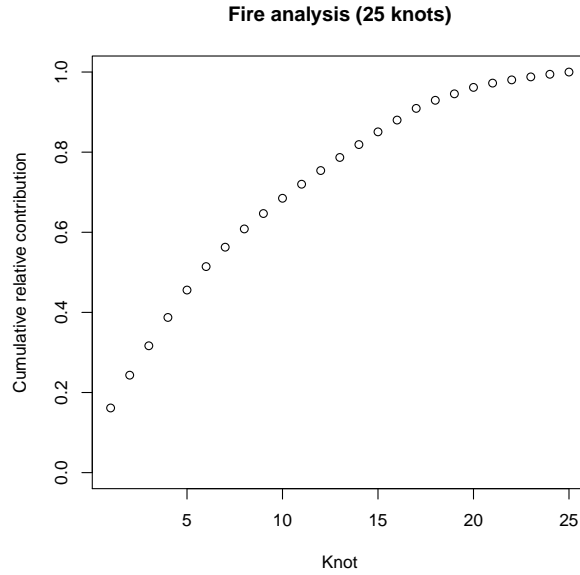


Figure 4: Cumulative sum of contributions  $v_1, \dots, v_{25}$ .

Table 1: Average Brier scores ( $\times 100$ ) for selected thresholds, quantile scores for selected quantiles, and timing for fire analysis.

	Process	Brier Scores ( $\times 100$ )		Quantile Scores		Time (in hours)
		$q(0.95)$	$q(0.99)$	$q(0.95)$	$q(0.99)$	
L = 5	EBF GSK					
L = 10	EBF GSK					
L = 15	EBF GSK					
L = 20	EBF GSK					
L = 25	EBF GSK					
L = 30	EBF GSK					
L = 35	EBF GSK					
L = 40	EBF GSK					

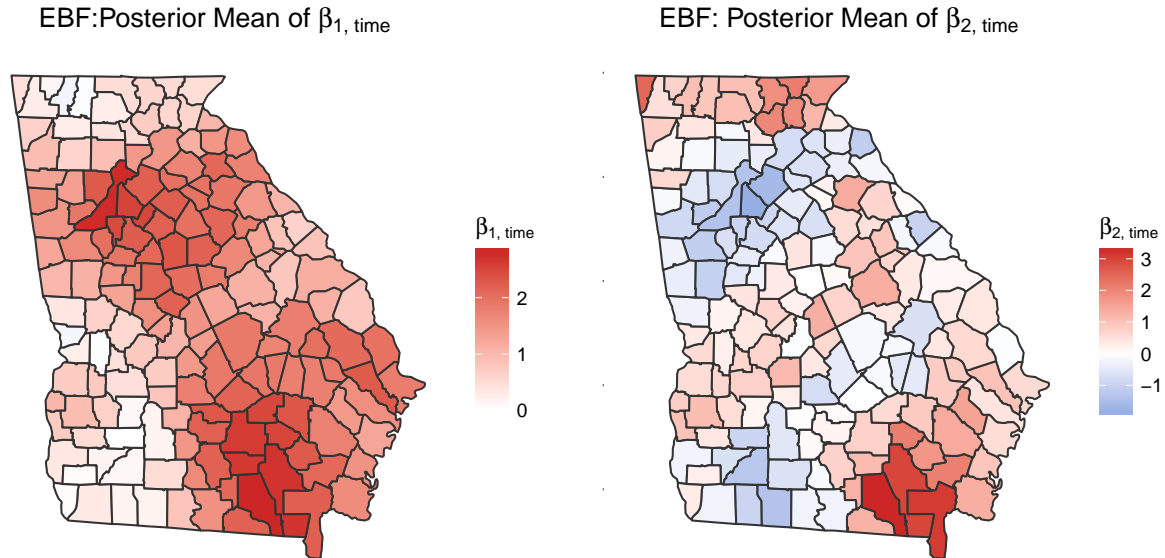
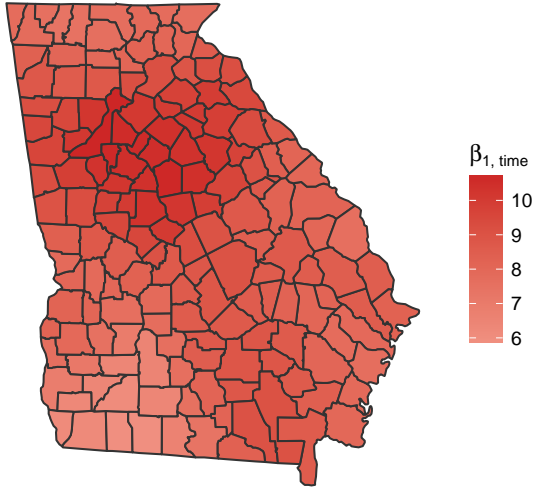


Figure 5: Posterior mean of  $\beta_{1, \text{time}}$  (left) and  $\beta_{2, \text{time}}$  (right) for fire data using EBF.

GSK:Posterior Mean of  $\beta_{1, \text{time}}$



GSK: Posterior Mean of  $\beta_{2, \text{time}}$

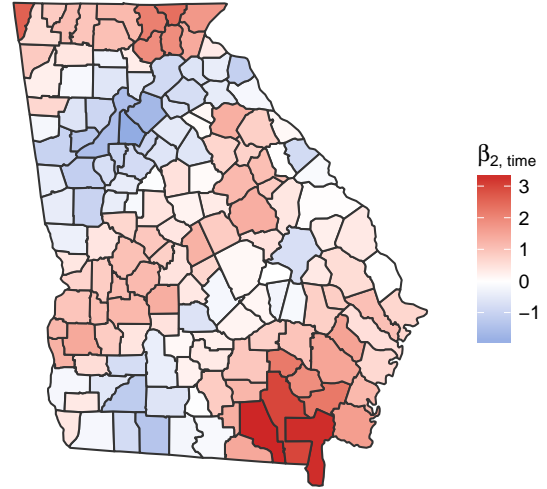
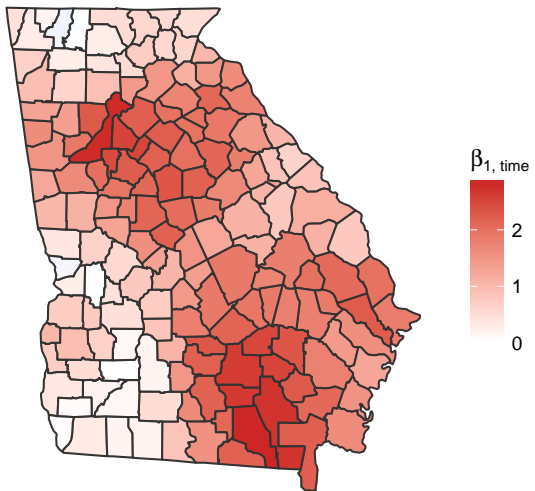


Figure 6: Posterior mean of  $\beta_{1, \text{time}}$  (left) and  $\beta_{2, \text{time}}$  (right) for fire data using GSK.

EBF:Posterior Mean of  $\beta_{1, \text{time}}$



EBF: Posterior Mean of  $\beta_{2, \text{time}}$

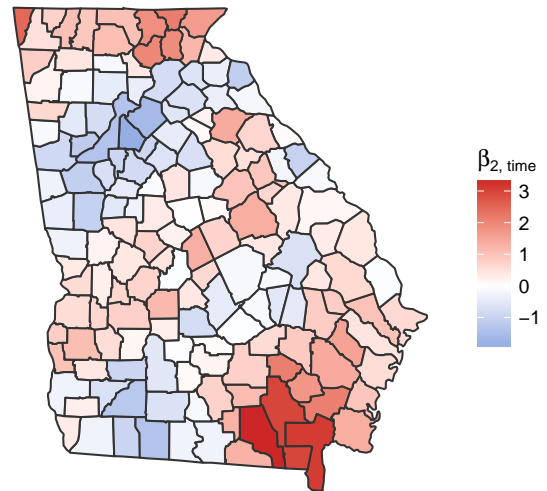
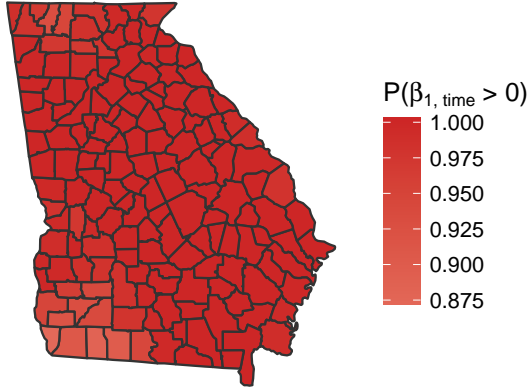


Figure 7: Posterior  $P(\beta_{1, \text{time}} > 0)$  (left) and  $P(\beta_{2, \text{time}} > 0)$  (right) for fire data using EBF.

GSK: Posterior  $P(\beta_{1, \text{time}} > 0)$



GSK: Posterior  $P(\beta_{2, \text{time}} > 0)$

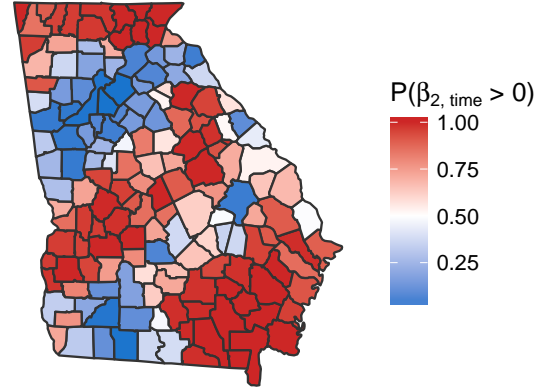


Figure 8: Posterior  $P(\beta_{1, \text{time}} > 0)$  (left) and  $P(\beta_{2, \text{time}} > 0)$  (right) for fire data using GSK.

Is this what you were thinking for the difference in  $q(0.90)$ ? These do not use  $\theta$ .

We also plot the difference in  $\hat{q}(0.90)$ , the estimated 90th quantile, between 2014 and 1965. To obtain the estimate for each county, we first take the posterior mean of  $\beta_{1, \text{int}}$ ,  $\beta_{1, \text{time}}$ ,  $\beta_{2, \text{int}}$ , and  $\beta_{2, \text{time}}$  for each county. Then we find  $\mu(\mathbf{s}, t)$  and  $\sigma(\mathbf{s}, t)$  for each county in 1965 and 2014. Then for the shape parameter we take the posterior mean of  $\xi$ . Finally, we use the inverse distribution function of the GEV to obtain  $\hat{q}(0.90)_{1965}$  and  $\hat{q}(0.90)_{2014}$  and plot  $\hat{q}(0.90)_{\text{diff}} = \hat{q}(0.90)_{2014} - \hat{q}(0.90)_{1965}$  in Figure 9.

#### 5.4 Analysis of annual precipitation

We also conduct an analysis of the precipitation data presented in (Reich and Shaby, 2012). The data are climate model output from the North American Regional Climate Change Assessment Program (NARCCAP). This data consists of  $n_s = 697$  grid cells at a 50km resolution in the eastern US, and includes historical data (1969 – 2000) as well as future conditions (2039 – 2070).

For this dataset, to estimate the EBFs, we use the combined current and future data. The first six EBFs



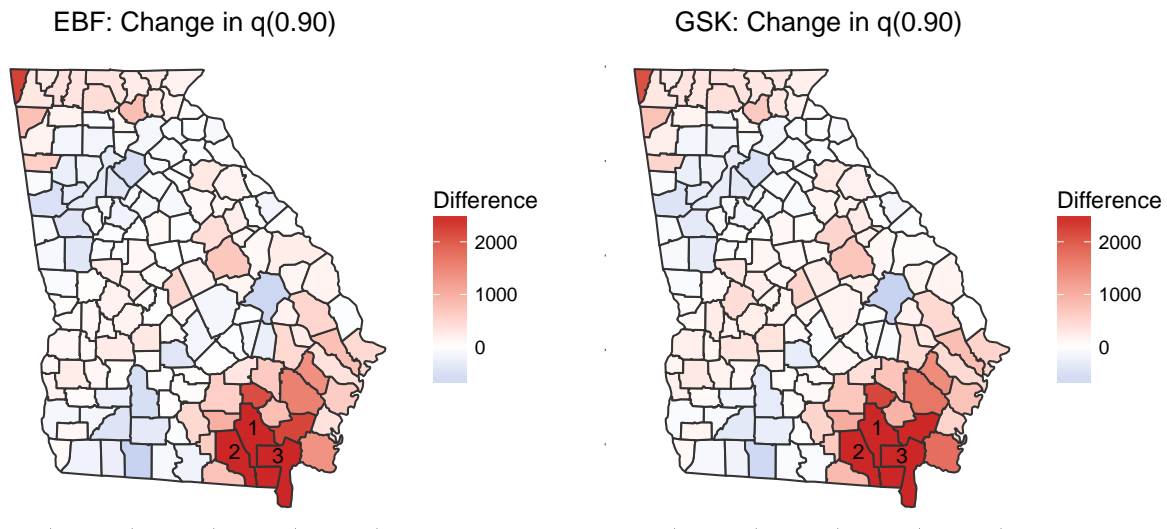


Figure 9: Difference in  $q(0.90)$  for fire data between 2014 and 1965 for EBF (left) and GSK (right). Three counties (labeled) have differences higher than 2500. County 1: Ware (EBF: 9690, GSK: 8934), County 2: Clinch (EBF: 6693, GSK: 6520), and County 3: Charlton (EBF: 6084, GSK: 6794)

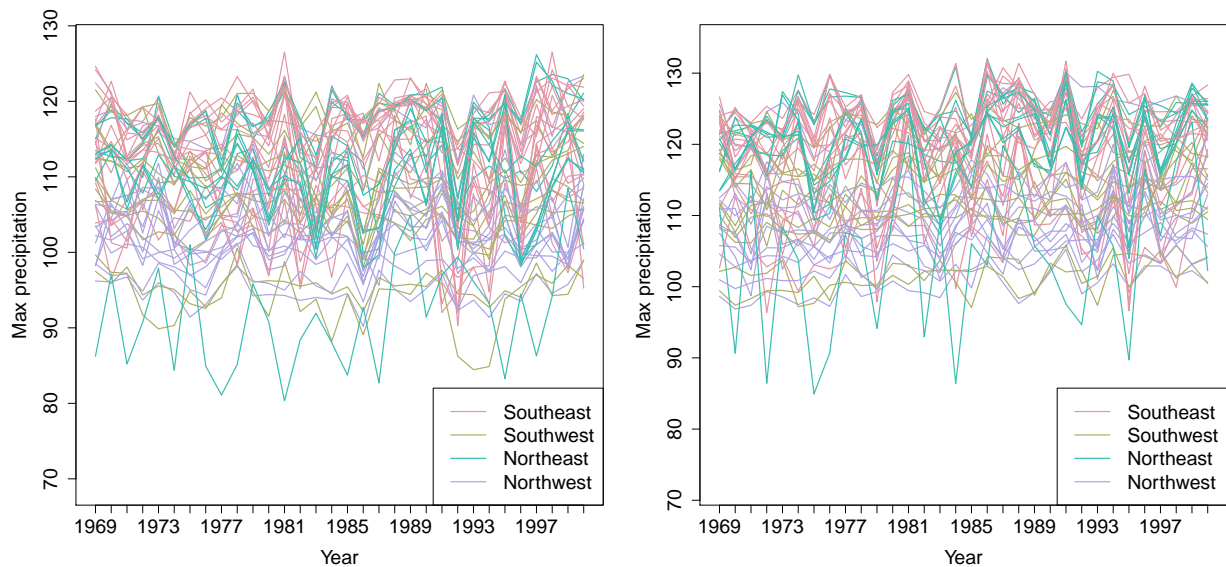


Figure 10: Time series of yearly max precipitation for current (1969 – 2000) (left). Time series of yearly max precipitation for future (2039 – 2070) (right).

for the combined data are given in Figure 11. We also plot the cumulative sum of the contributions for each

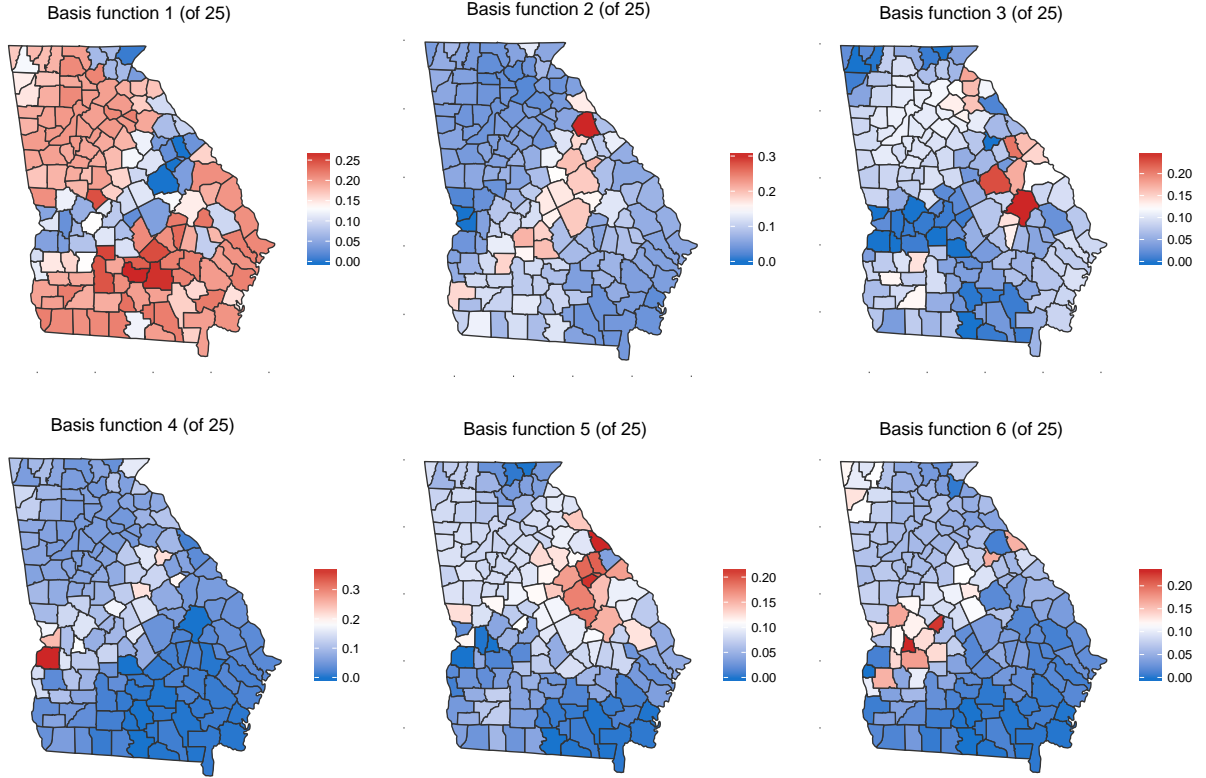


Figure 11: First six EBFs for the combined precipitation data.

basis function in 12. As a comparison, we provide the first six principal components of the fire data along with the cumulative sum of the first 25 eigenvalues in Figure 19.

## 5.5 Results for precipitation analysis

Figure here with BS and QS for the precipitation data

Based on the cross-validation results, we run a full analysis using all of the data with  $L = 25$ . Figure 13 and Figure 14 give the county by county posterior means of  $\beta_{1,\text{time}}$  and  $\beta_{2,\text{time}}$  for EBF and GSK respectively. Figure 15 and Figure 16 give the county by county posterior  $P(\beta_{1,\text{time}} > 0)$  and  $P(\beta_{2,\text{time}} > 0)$  for EBF and GSK respectively.

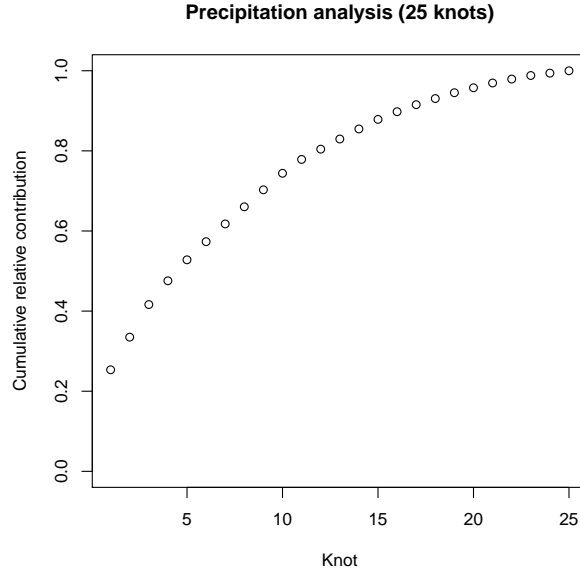


Figure 12: Cumulative sum of contributions  $v_1, \dots, v_{25}$ .

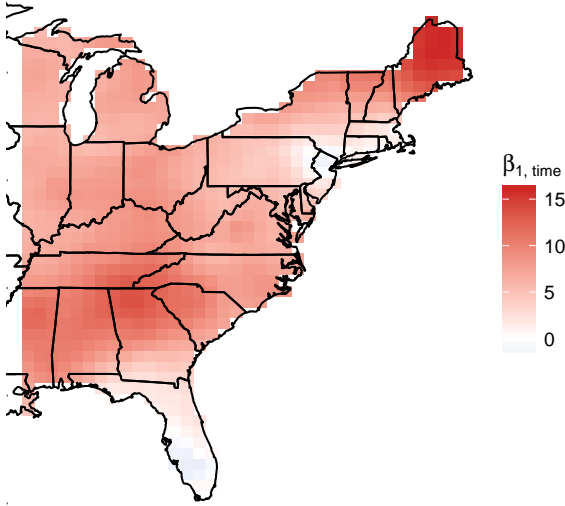
Table 2: Average Brier scores ( $\times 100$ ) for selected thresholds, quantile scores for selected quantiles, and timing for current precipitation analysis.

	Process	Brier Scores ( $\times 100$ )		Quantile Scores		Time (in hours)
		$q(0.95)$	$q(0.99)$	$q(0.95)$	$q(0.99)$	
L = 5	EBF GSK					
L = 10	EBF GSK					
L = 15	EBF GSK					
L = 20	EBF GSK					
L = 25	EBF GSK					
L = 30	EBF GSK					
L = 35	EBF GSK					
L = 40	EBF GSK					

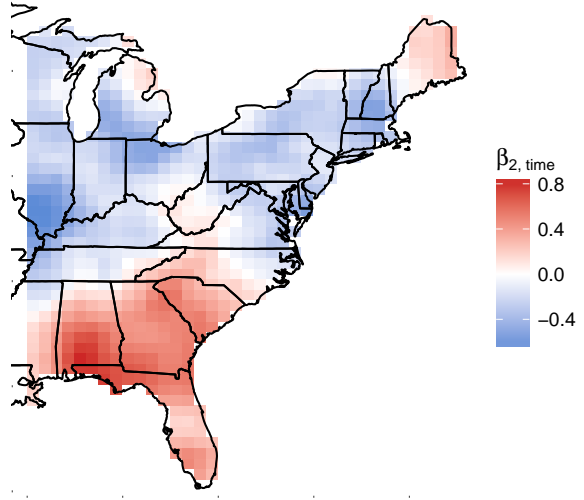
Table 3: Average Brier scores ( $\times 100$ ) for selected thresholds, quantile scores for selected quantiles, and timing for future precipitation analysis.

	Process	Brier Scores ( $\times 100$ )		Quantile Scores		Time (in hours)
		$q(0.95)$	$q(0.99)$	$q(0.95)$	$q(0.99)$	
L = 5	EBF GSK					
L = 10	EBF GSK					
L = 15	EBF GSK					
L = 20	EBF GSK					
L = 25	EBF GSK					
L = 30	EBF GSK					
L = 35	EBF GSK					
L = 40	EBF GSK					

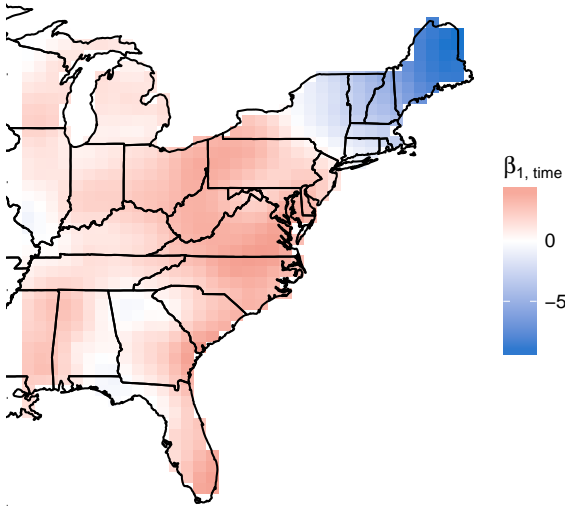
EBF: Current Posterior Mean of  $\beta_{1, \text{time}}$



EBF: Current Posterior Mean of  $\beta_{2, \text{time}}$



EBF: Future Posterior Mean of  $\beta_{1, \text{time}}$



EBF: Future Posterior Mean of  $\beta_{2, \text{time}}$

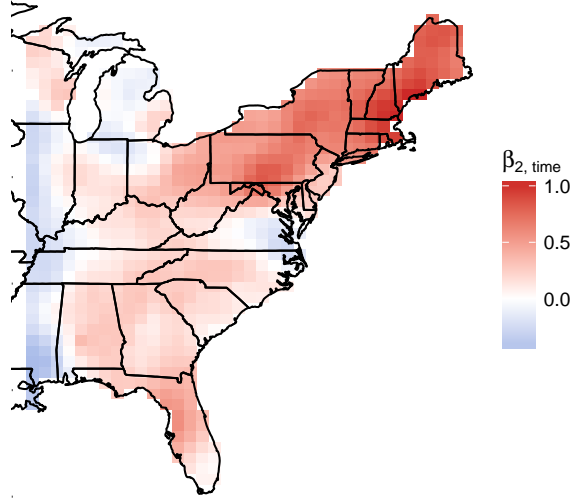
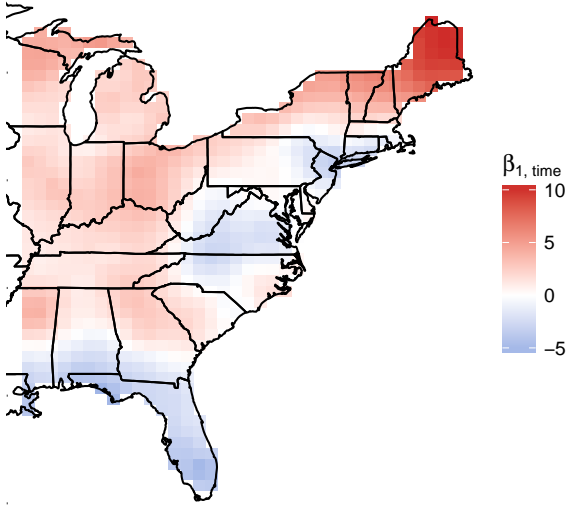
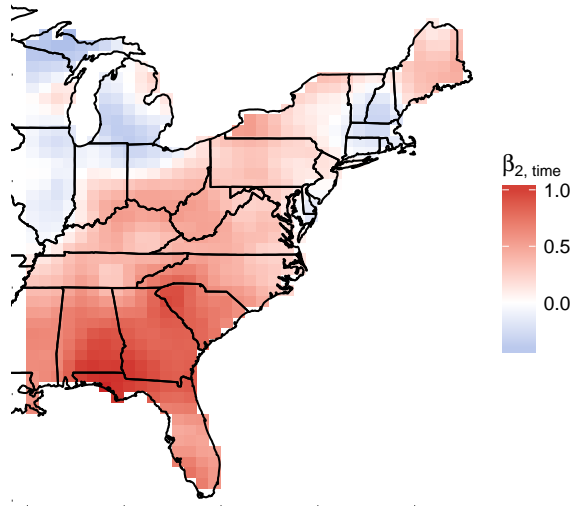


Figure 13: Posterior mean of  $\beta_{1, \text{time}}$  (left) and  $\beta_{2, \text{time}}$  (right) for current (top) and future (bottom) precipitation data using EBF.

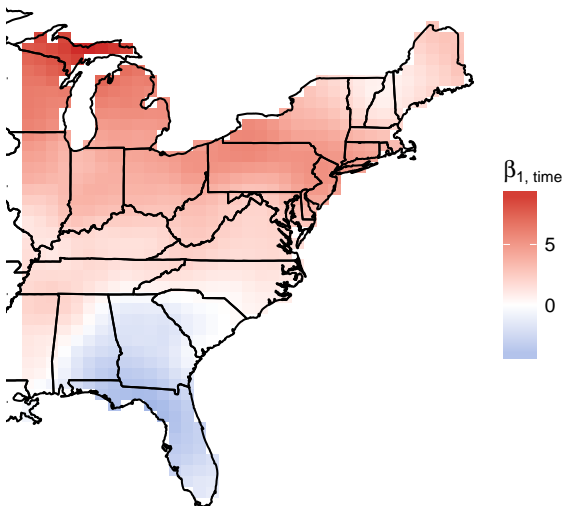
GSK: Current Posterior Mean of  $\beta_{1, \text{time}}$



GSK: Current Posterior Mean of  $\beta_{2, \text{time}}$



GSK: Future Posterior Mean of  $\beta_{1, \text{time}}$



GSK: Future Posterior Mean of  $\beta_{2, \text{time}}$

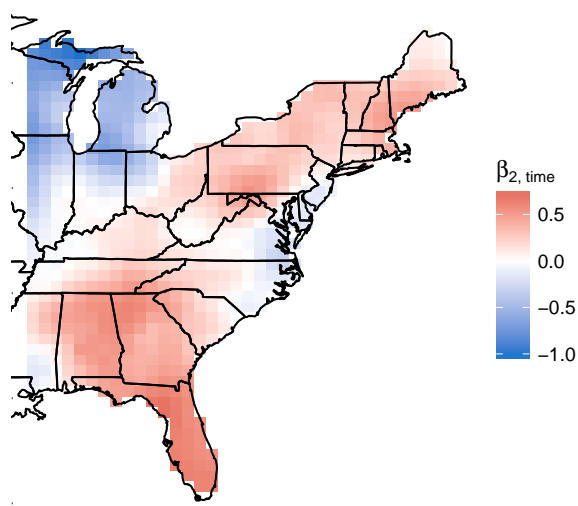
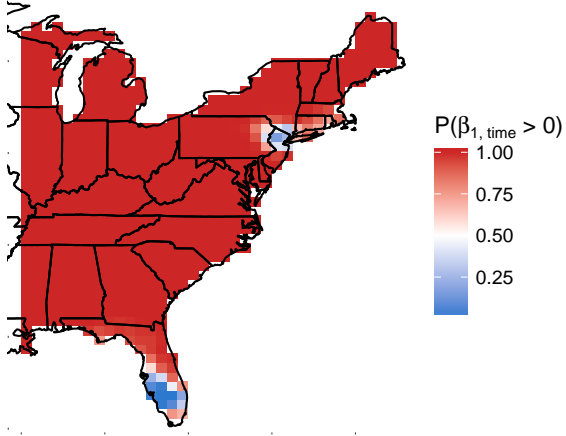
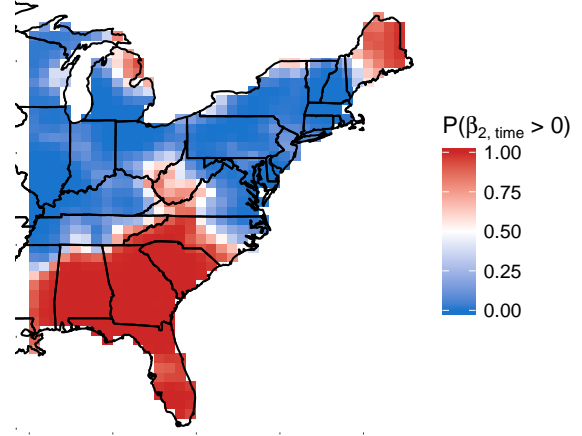


Figure 14: Posterior mean of  $\beta_{1, \text{time}}$  (left) and  $\beta_{2, \text{time}}$  (right) for current (top) and future (bottom) precipitation data using GSK.

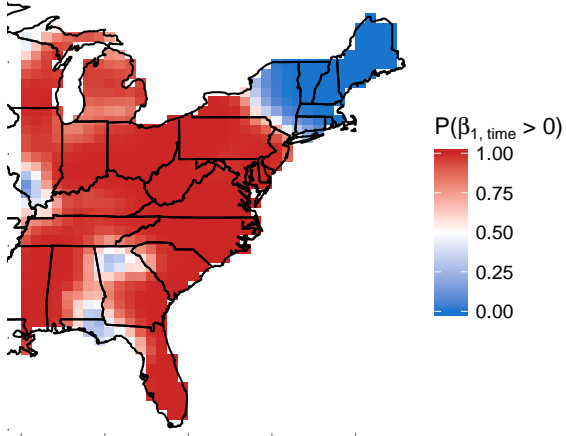
EBF: Current Posterior  $P(\beta_{1, \text{time}} > 0)$



EBF: Current Posterior  $P(\beta_{2, \text{time}} > 0)$



EBF: Future Posterior  $P(\beta_{1, \text{time}} > 0)$



EBF: Future Posterior  $P(\beta_{2, \text{time}} > 0)$

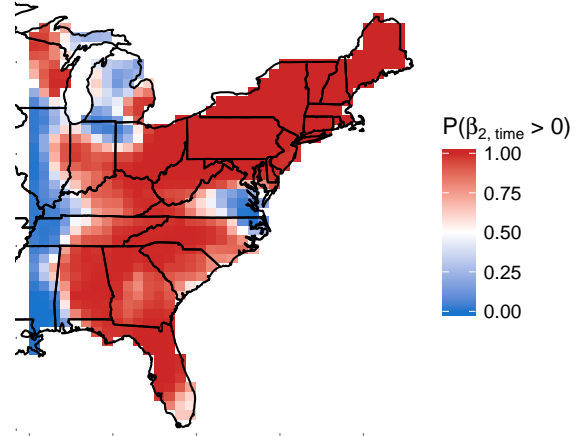
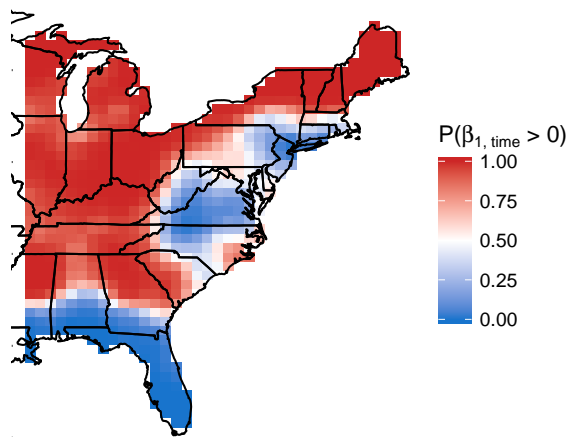
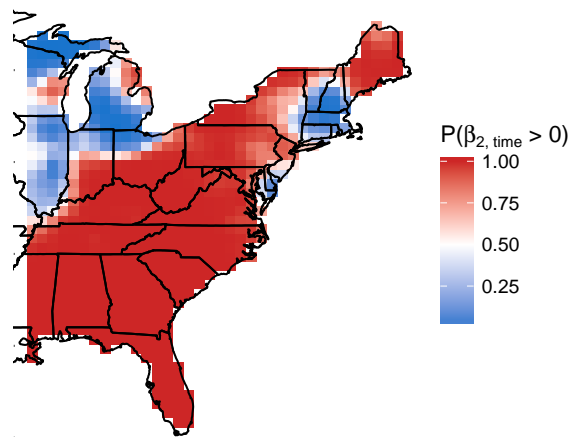


Figure 15: Posterior  $P(\beta_{1, \text{time}} > 0)$  (left) and  $P(\beta_{2, \text{time}} > 0)$  (right) for current (top) and future (bottom) precipitation data using EBF.

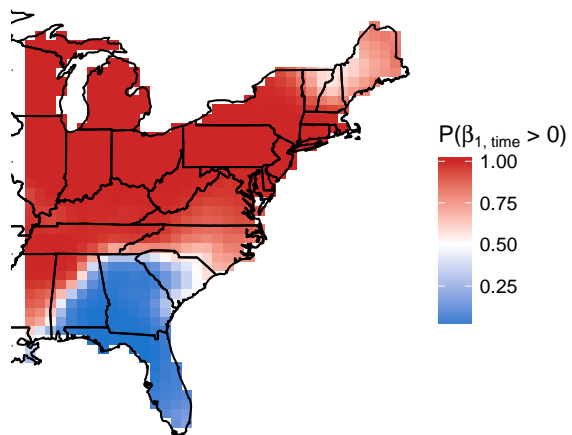
GSK: Current Posterior  $P(\beta_{1, \text{time}} > 0)$



GSK: Current Posterior  $P(\beta_{2, \text{time}} > 0)$



GSK: Future Posterior  $P(\beta_{1, \text{time}} > 0)$



GSK: Future Posterior  $P(\beta_{2, \text{time}} > 0)$

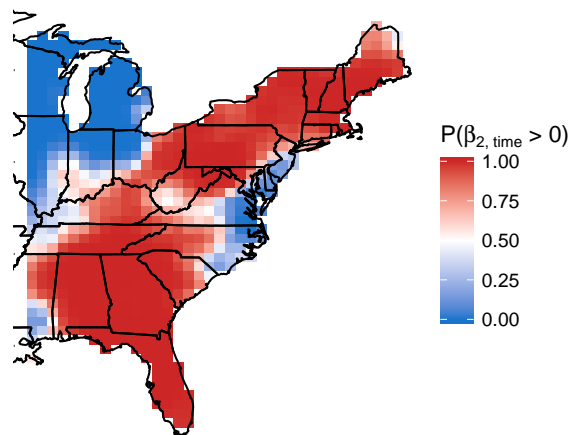


Figure 16: Posterior  $P(\beta_{1, \text{time}} > 0)$  (left) and  $P(\beta_{2, \text{time}} > 0)$  (right) for current (top) and future (bottom) precipitation data using GSK.



207 We also plot the difference in  $\hat{q}(0.90)$ , the estimated 90th quantile, between 2000 and 1969 for current  
 208 and 2070 and 2039 for future. The estimates are obtained in the same way as for the fire analysis and are  
 209 given in Figure 9.

## 210 **6 Conclusions**

## 211 **Acknowledgements**

212 The authors would like to acknowledge Dan Cooley for his helpful suggestions on the manuscript.

## 213 **A Appendices**

### 214 **A.1 Extreme value distributions**

215 The cumulative distribution function for the GEV is  $F(y) = \exp\{-t(y)\}$  where

$$t(y) = \begin{cases} \left[1 + \xi \frac{y - \mu}{\sigma}\right]^{-1/\xi}, & \xi \neq 0 \\ \exp\left\{-\frac{y - \mu}{\sigma}\right\}, & \xi = 0. \end{cases} \quad (17)$$

216 The probability density function for the GEV is given by  $f(y) = \frac{1}{\sigma} t(y)^{\xi+1} \exp\{-t(y)\}$  where  $t(y)$  is  
 217 defined in (17).

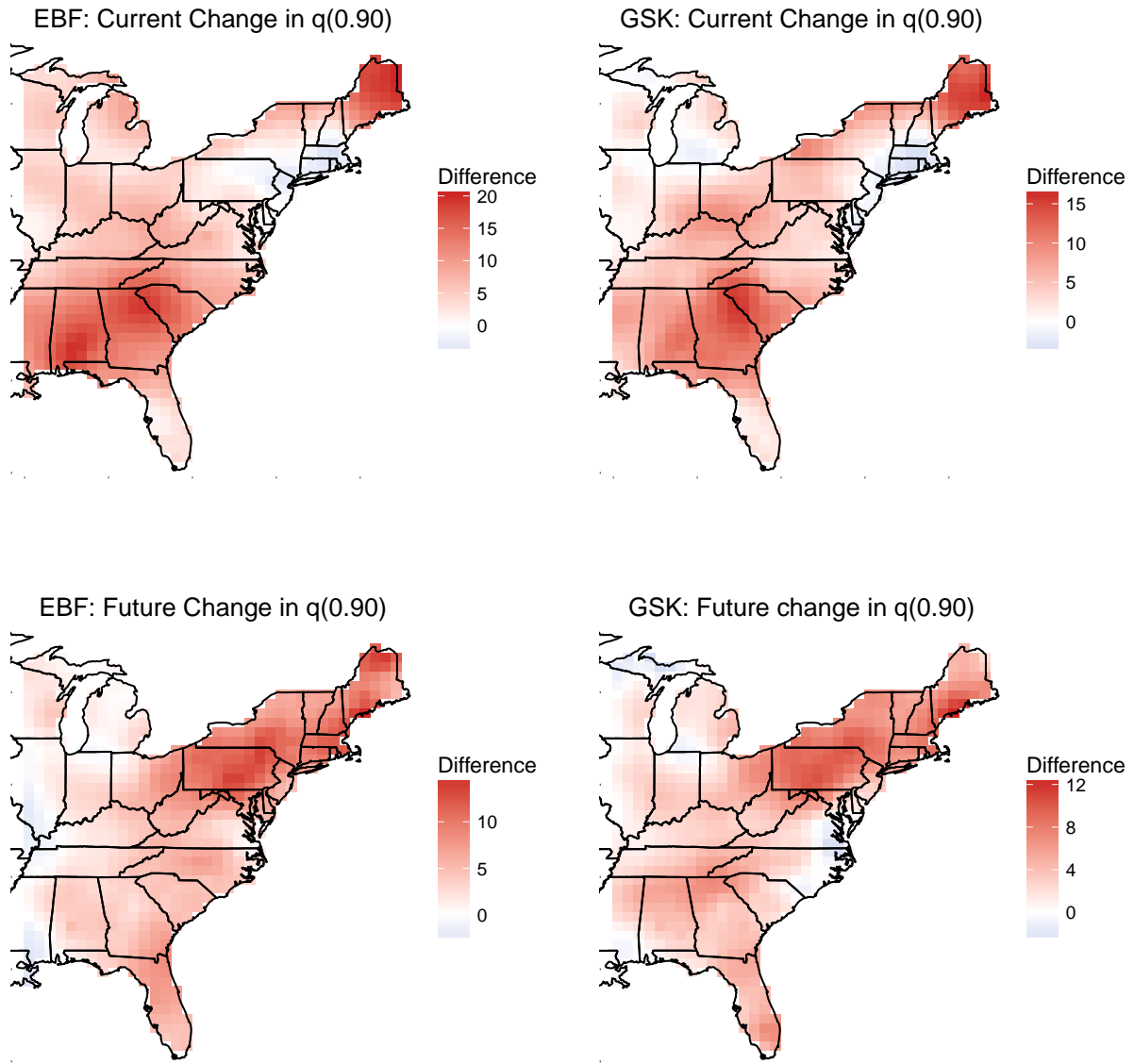


Figure 17: Difference in  $q(0.90)$  for precipitation data between 2000 and 1969 for current (top) and 2070 and 2039 for future (bottom) using EBF (left) and GSK (right).

## 218 **A.2 Grid approximation to PS density**

219 The  $\text{PS}(\alpha)$  density can be challenging to use because it does not have a closed form. From Section 2 of  
 220 (Stephenson, 2009), the density can be expressed as

$$g_1(A) = \int_0^1 g_1(A, B) dB, \quad (18)$$

221 where

$$g_1(A, B) = \frac{\alpha}{1-\alpha} \left(\frac{1}{A}\right)^{1/(1-\alpha)} c(\pi B) \exp \left\{ - \left(\frac{1}{A}\right)^{\alpha/(1-\alpha)} c(\pi B) \right\}, \quad (19)$$

222 with

$$c(\psi) = \left[ \frac{\sin(\alpha\psi)}{\sin(\psi)} \right]^{1/(1-\alpha)} \frac{\sin[(1-\alpha)\psi]}{\sin(\alpha\psi)}. \quad (20)$$

223 Stephenson (2009) presents an auxiliary variable technique to deal with the integral in the density function,  
 224 but we opt to numerically evaluate the integral because it is only one-dimensional. To evaluate the integral,  
 225 we use 50 evenly spaced quantiles of a  $\text{Beta}(0.5, 0.5)$  distribution as the midpoints  $B_1, \dots, B_{50}$ , and then  
 226 use the midpoint rule to evaluate  $\int_0^1 g_1(A, B) dB$ .

## 227 **A.3 Standardized Gaussian kernel functions**

228 Reich and Shaby (2012) use standardized Gaussian kernel functions as their spatial basis functions in the  
 229 low-rank max-stable model. Consider a set of  $\mathbf{k}_1, \dots, \mathbf{k}_L$  spatial knot locations in  $\mathcal{D}^2$ , the region of interest.

230 Then

$$\hat{B}_{il} = \frac{\exp \left\{ -\frac{\|\mathbf{s}_i - \mathbf{k}_l\|^2}{2\rho^2} \right\}}{\sum_{j=1}^L \exp \left\{ -\frac{\|\mathbf{s}_i - \mathbf{k}_j\|^2}{2\rho^2} \right\}} \quad (21)$$

231 where  $\|\cdot\|$  is the Euclidean distance between a site and a knot location.

#### 232 A.4 Principal components for fire and precipitation data

233 As a comparison to the EBFs, Figure 19 gives the first six principal components for the fire data, and Figure 21 gives the first six principal components for the precipitation data.

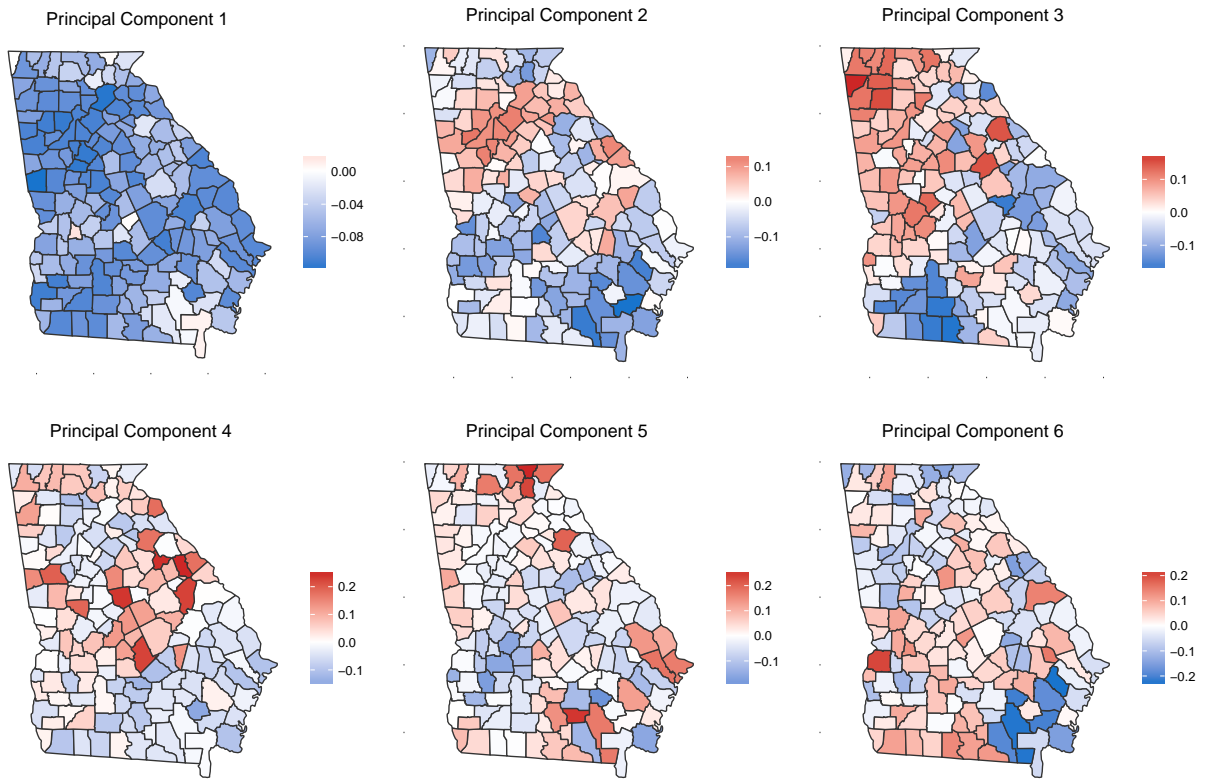


Figure 18: First six principal components for the Georgia fire data.

234

235 One reason why the principal components for the fire data look so different from the EBFs presented in

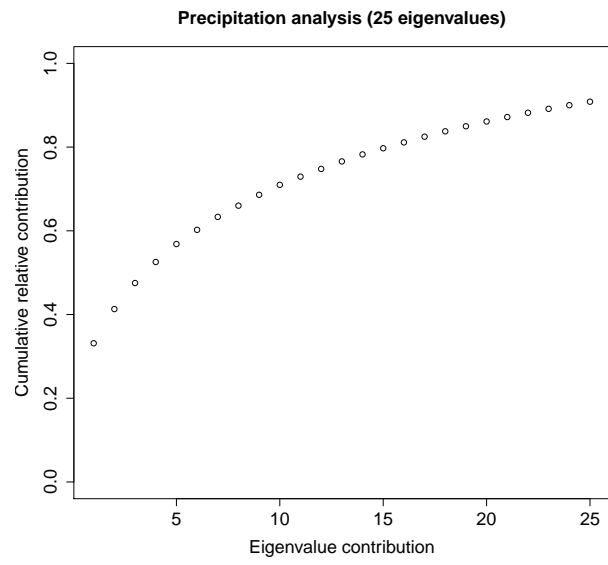


Figure 19: Cumulative sum of the first 25 eigenvalues.

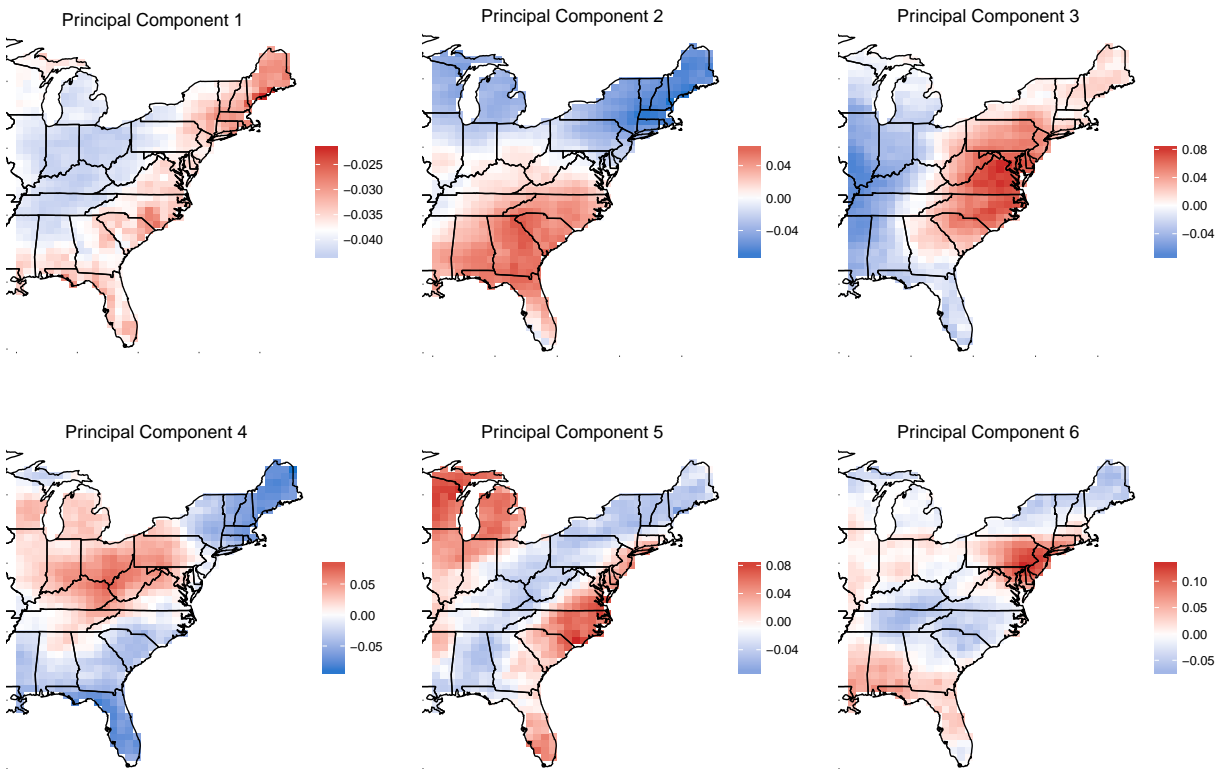


Figure 20: First six principal components for the precipitation data.

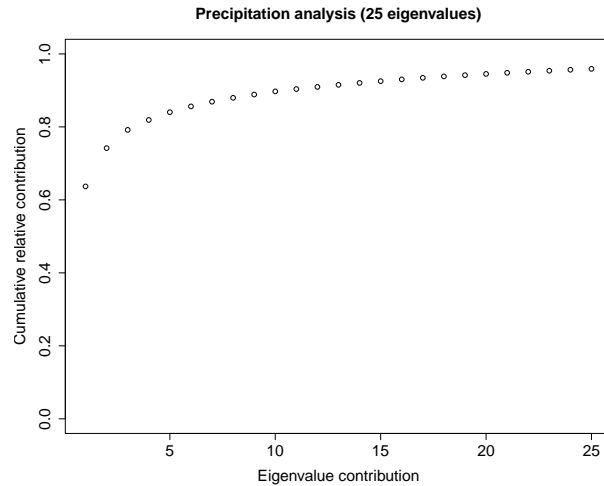


Figure 21: Cumulative sum of the first 25 eigenvalues.

Figure 3 is that the EBFs use censored data whereas the principal components do not.

## References

- Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values*. Lecture Notes in Control and Information Sciences. London: Springer.
- Davison, A. C., Padoan, S. A. and Ribatet, M. (2012) Statistical Modeling of Spatial Extremes. *Statistical Science*, **27**, 161–186.
- Everitt, B. and Hothorn, T. (2008) Principal Components Analysis. In *An introduction to applied multivariate analysis with R*, 21–54. New York, NY: Springer New York.
- Genton, M. G., Ma, Y. and Sang, H. (2011) On the likelihood function of Gaussian max-stable processes. *Biometrika*, **98**, 481–488.
- Gneiting, T. and Raftery, A. E. (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, **102**, 359–378.
- de Haan, L. (1984) A Spectral Representation for Max-stable Processes. *The Annals of Probability*, **12**, 1194–1204.
- de Haan, L. and Ferreira, A. (2006) *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer.
- Hannachi, A., Jolliffe, I. T. and Stephenson, D. B. (2007) Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology*, **27**, 1119–1152.

- 254 Huser, R. and Davison, A. C. (2014) Space-time modelling of extreme events. *Journal of the Royal Statis-*  
255 *tical Society: Series B (Statistical Methodology)*, **76**, 439–461.
- 256 Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-Based Inference for Max-Stable Processes.  
257 *Journal of the American Statistical Association*, **105**, 263–277.
- 258 Reich, B. J. and Shaby, B. A. (2012) A hierarchical max-stable spatial model for extreme precipitation. *The*  
259 *Annals of Applied Statistics*, **6**, 1430–1451.
- 260 Schlather, M. and Tawn, J. A. (2003) A dependence measure for multivariate and spatial extreme values:  
261 Properties and inference. *Biometrika*, **90**, 139–156.
- 262 Smith, R. L. (1990) Max-stable processes and spatial extremes.
- 263 Stephenson, A. G. (2009) High-Dimensional Parametric Modelling of Multivariate Extreme Events. *Aus-*  
264 *tralian & New Zealand Journal of Statistics*, **51**, 77–88.
- 265 Thibaud, E. and Opitz, T. (2015) Efficient inference and simulation for elliptical Pareto processes.  
266 *Biometrika*, **102**, 855–870.
- 267 Wadsworth, J. L. and Tawn, J. A. (2014) Efficient inference for spatial extreme value processes associated  
268 to log-Gaussian random functions. *Biometrika*, **101**, 1–15.
- 269 Wang, Y. and Stoev, S. A. (2011) Conditional sampling for spectrally discrete max-stable random fields.  
270 *Advances in Applied Probability*, **43**, 461–483.