Exploration and inference in spatial extremes using empirical basis functions

Samuel A Morris[1], Brian J Reich[1], and Emeric Thibaud[2]

August 13, 2016

Abstract

words...

Key words: non-stationary data analysis; max-stable process; spectral representation; dimension reduction.

[1]North Carolina State University
[2]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). Max-stable processes were also used to model spatial exceedances over a high threshold (Thibaud et al., 2013; Huser and Davison, 2014). 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). 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., under review). 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 propose an empirical basis function (EBF) approach that builds on a finite truncation of the spectral representation, and develops a method-of-moments estimator for the underlying spatial processes. Unlike PCA/EOFs, but similar to dictionary learning (Mairal et al., 2014) and non-negative matrix factorizations (Lee and Seung, 1999), the EBFs are not orthogonal. Nonetheless these spatial functions can be plotted for exploratory analysis to reveal important spatial trends. In addition to exploratory analysis, we show that the EBFs can be used for Bayesian inference on the marginal parameters at each location, modeling spatial dependence, and to test for covariate effects. 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 datsets 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. Section 3 describes the algorithm used to estimate the spatial basis functions, and Section 4 describes the model fit using Markov chain Monte Carlo (MCMC) methods. In Section 5 we demonstrate the use of the EBFs for an analysis of wildfire data in Georgia and precipitation data in the eastern U.S. Lastly in Section 6 we give some summary conclusions and a brief discussion of the findings.

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 notation 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 $T(\mathbf{s}) = -\infty$.

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

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

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

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

$$Z(\mathbf{s}) = \bigvee_{l=1}^{\infty} B(\mathbf{s}, \mathbf{k}_l) A_l \tag{2}$$

where the functions $B(\mathbf{s}, \mathbf{k}_l)$ satisfy $B(\mathbf{s}, \mathbf{k}_l) > 0$ for all $\mathbf{s}$ and $\int B(\mathbf{s}, \mathbf{k}_l)\, \mathrm{d}\mathbf{k}_l = 1$ for all $\mathbf{s}$, and $(\mathbf{k}_l, A_l)$ for $l = 1, \ldots, \infty$ are a Poisson process with intensity measure $\mathrm{d}A\ \mathrm{d}\mathbf{k}/A^2$. In many representations of max-stable process, such as Smith (1990) and Reich and Shaby (2012), the $\mathbf{k}_l$ are spatial locations that represent the center of process l; however, in our proposed method the basis functions are not associated with one particular location and so to simplify notation we let $B_l(\mathbf{s}) = B(\mathbf{s}; \mathbf{k}_l)$

To arrive at a low-rank model, we assume there are a finite and known number of spatial basis functions $B_1(\mathbf{s}), \ldots, B_L(\mathbf{s})$ that explain the important spatial variation in the process. As in de Haan's expansion,

the 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 that realizations of Z are exactly functions of L basis functions, we include independent error variables $\epsilon(\mathbf{s})$ to capture variation not explained by the $B_l(\mathbf{s})$. We follow Reich and Shaby (2012) and decompose $Z(\mathbf{s})$ as $Z(\mathbf{s}) = \theta(\mathbf{s})\varepsilon(\mathbf{s})$ where $\theta(\mathbf{s})$ is a spatial process and $\varepsilon(\mathbf{s}) \overset{\text{iid}}{\sim} \text{GEV}(1, \alpha, \alpha)$ is independent error. The spatial component is

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

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; Appendix A.2) distribution $A_l \overset{\text{iid}}{\sim} \text{PS}(\alpha)$, then $Z(\mathbf{s})$ is max-stable and has unit Fréchet marginal distributions.

Extremal spatial dependence for max-stable processes can be summarized by the extremal coefficient (Schlather and Tawn, 2003; EC) $\vartheta(\mathbf{s}_1, \mathbf{s}_2) \in [1, 2]$, where

$$\text{Prob}[Z(\mathbf{s}_1) < c, Z(\mathbf{s}_2) < c] = \text{Prob}[Z(\mathbf{s}_1) < c]^{\vartheta(\mathbf{s}_1, \mathbf{s}_2)}. \tag{4}$$

For the PS random effects model the EC has the form

$$\vartheta(\mathbf{s}_1, \mathbf{s}_2) = \sum_{l=1}^{L} \left[B_l(\mathbf{s}_1)^{1/\alpha} + B_l(\mathbf{s}_2)^{1/\alpha} \right]^{\alpha}. \tag{5}$$

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

3 Estimating the basis functions

To estimate the extremal coefficient function, we consider the process at n_s spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_{n_s}$ and n_t times $t = 1, \ldots, n_t$. The basis functions are fixed over time, but the random effects and errors are independent over time. That is

$$Z_t(\mathbf{s}) = \theta_t(\mathbf{s})\epsilon_t(\mathbf{s}) \quad \text{where} \quad \theta_t(\mathbf{s}) = \left(\sum_{l=1}^{L} B_l(\mathbf{s})^{1/\alpha} A_{lt} \right)^{\alpha}, \tag{6}$$

$A_{lt} \overset{\text{iid}}{\sim} \text{PS}(\alpha)$, and $\epsilon_t(\mathbf{s}) \overset{\text{iid}}{\sim} \text{GEV}(1, \alpha, \alpha)$. Denote $Y_t(\mathbf{s}_i) = Y_{it}$, $B_l(\mathbf{s}_i) = B_{il}$, $T(\mathbf{s}_i) = T_i$, and $\vartheta(\mathbf{s}_i, \mathbf{s}_j) = \vartheta_{ij}$.

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

(1) Obtain an initial estimate of the extremal coefficient for each pair of locations, $\hat{\vartheta}_{ij}$.

(2) Spatially smooth these initial estimates $\hat{\vartheta}_{ij}$ using kernel smoothing to obtain $\tilde{\vartheta}_{ij}$.

(3) Estimate the spatial dependence parameters by minimizing the difference between model-based coefficients, ϑ_{ij}, and smoothed coefficients, $\tilde{\vartheta}_{ij}$.

The first-stage estimates are obtained using an empirical estimate as follows. To estimate the spatial 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 the U_{it} are approximately uniform at each location. Then for some extreme probability $q \in (0, 1)$, solving (4) suggests the estimate

$$\hat{\vartheta}_{ij}(q) = \frac{\log[Q_{ij}(q)]}{\log(q)}, \tag{7}$$

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

sites are less than q. Since all large q give valid estimates, we average over a grid of q with $q_1 < \cdots < q_{n_q}$

$$\hat{\vartheta}_{ij} = \frac{1}{n_q} \sum_{j=1}^{n_q} \hat{\vartheta}_{ij}(q_j). \tag{8}$$

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

$$\tilde{\vartheta}_{ij} = \frac{\sum_{u=1}^{n_s} \sum_{v=1}^{n_s} w_{iu} w_{jv} \hat{\vartheta}_{uv}}{\sum_{u=1}^{n_s} \sum_{v=1}^{n_s} w_{iu} w_{jv}}, \tag{9}$$

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

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

$$\sum_{i<j} \left(\tilde{\vartheta}_{ij} - \vartheta_{ij} \right)^2 = \sum_{i<j} \left(\tilde{\vartheta}_{ji} - \sum_{l=1}^{L} \left[B_{il}^{1/\hat{\alpha}} + B_{jl}^{1/\hat{\alpha}} \right]^{\hat{\alpha}} \right)^2 \tag{10}$$

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 (10) does not have a closed form, we use block coordinate descent to obtain $\hat{\mathbf{B}}$. We cycle through spatial locations and update the vectors $\left(\hat{B}_{i1}, \ldots, \hat{B}_{iL} \right)$ conditioned on the values for the other location and repeat until convergence. At each step, we use the restricted optimization routine in the `R` function `optim`. This algorithm gives estimates of the B_{il} at the n_s data locations, but is easily extended to all $\mathbf{s}$ for spatial

prediction. The kernel smoothing step ensures that the estimates for $\hat{B}_{il}$ are spatially smooth, and thus interpolation of the $\hat{B}_{il}$ gives spatial functions $\hat{B}_l(\mathbf{s})$.

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

$$v_l = \frac{1}{n_s} \sum_{i=1}^{n_s} \hat{B}_{il}. \tag{11}$$

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. The order of the terms is arbitrary, and so we reorder the terms so that $v_1 \geq \cdots \geq v_L$.

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

$$\mu_t(\mathbf{s}) = \beta_{1,\text{int}}(\mathbf{s}) + \beta_{1,\text{time}}(\mathbf{s})t \tag{12}$$

$$\log[\sigma_t(\mathbf{s})] = \beta_{2,\text{int}}(\mathbf{s}) + \beta_{2,\text{time}}(\mathbf{s})t \tag{13}$$

where

$$\begin{aligned} \beta_{1,\text{int}}(\mathbf{s}) &\sim \text{N}(\mu_{1,\text{int}}\mathbf{1}, \sigma^2_{1,\text{int}}\boldsymbol{\Sigma}) & \beta_{1,\text{time}}(\mathbf{s}) &\sim \text{N}(\mu_{1,\text{time}}\mathbf{1}, \sigma^2_{1,\text{time}}\boldsymbol{\Sigma}) \\ \beta_{2,\text{int}}(\mathbf{s}) &\sim \text{N}(\mu_{2,\text{int}}\mathbf{1}, \sigma^2_{2,\text{int}}\boldsymbol{\Sigma}) & \beta_{2,\text{time}}(\mathbf{s}) &\sim \text{N}(\mu_{2,\text{time}}\mathbf{1}, \sigma^2_{2,\text{time}}\boldsymbol{\Sigma}) \end{aligned} \tag{14}$$

are Gaussian process priors and $\mathbf{\Sigma}$ 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 ξ 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

$$Y_{it}|\theta_{it}, \Theta_{it} \overset{\text{ind}}{\sim} \text{GEV}(\mu^*_{it}, \sigma^*_{it}, \xi^*), \tag{15}$$

where $\mu^*_{it} = \mu_{it} + \frac{\sigma_{it}}{\xi}(\theta^{\xi}_{it} - 1)$, $\sigma^*_{it} = \alpha\sigma_{it}\theta^{\xi}_{it}$, and $\xi^* = \alpha\xi$. Therefore, the conditional likelihood conveniently factors across observations; marginalizing over the random effect θ_{it} induces extremal spatial dependence. To focus on the extreme values above the local threshold T_i, we use the censored likelihood

$$d(y; \theta_{it}, \Theta_{it}, T_i) = \begin{cases} F(y; \mu^*_{it}, \sigma^*_{it}, \xi^*) & y \leq T_i \\ f(y; \mu^*_{it}, \sigma^*_{it}, \xi^*) & y > T_i, \end{cases} \tag{16}$$

where F and f are the GEV distribution and density functions, respectively, defined in Appendix A.1.

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

$$\begin{aligned} Y_{it}|\theta_{ij} &\overset{\text{indep}}{\sim} d(y; \theta_{it}, \Theta_{it}, T_i) \\ \theta_{it} &= \left(\sum_{l=1}^{L} \hat{B}^{1/\hat{\alpha}}_{il} A_{lt}\right)^{\hat{\alpha}} \quad \text{where} \quad A_{lt} \overset{\text{iid}}{\sim} PS(\hat{\alpha}) \\ \mu_{it} &= \beta_{1,\text{int}}(\mathbf{s}_i) + \beta_{1,\text{time}}(\mathbf{s}_i)t \\ \log(\sigma_{it}) &= \beta_{2,\text{int}}(\mathbf{s}) + \beta_{2,\text{time}}(\mathbf{s})t. \end{aligned} \tag{17}$$

We estimate parameters $\Theta = \{A_{lt}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \xi\}$ 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 Appendix A.2. The hyperparameters $\mu_{1,\text{int}}, \mu_{1,\text{time}}, \mu_{2,\text{int}}, \mu_{2,\text{time}}$ and $\sigma^2_{1,\text{int}}, \sigma^2_{1,\text{time}}, \sigma^2_{2,\text{int}}, \sigma^2_{2,\text{time}}$ 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 $q_1, \ldots, q_{n_q}$, the kernel bandwidth ϕ, 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 $T_i, \ldots, T_{n_s}$. For this we use spatially smoothed sample quantiles. That is, we set T_i to the 0.95 quantile of the Y_{it} and its five nearest neighbors.

5 Data analysis

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

5.1 Gaussian kernel basis functions

To provide a comparison of our model with another approach, we also fit a model that uses standardized Gaussian kernels for the spatial basis functions (Reich and Shaby, 2012). In this method, Reich and Shaby introduce a set of $\mathbf{k}_1, \ldots, \mathbf{k}_L$ spatial knots and use standardized Gaussian kernel functions (GSK; see Ap-

pendix A.3) instead of using EBFs for the $\hat{B}_l(\mathbf{s})$. For the comparison between EBF and GSK methods, we use the same number of basis functions. We obtain estimates of the kernel bandwidth $\hat{\rho}$ and spatial dependence $\hat{\alpha}$, using the same least squares minimization as with the EBF method, and treat these as fixed in the MCMC.

5.2 Analysis of extreme Georgia fires

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

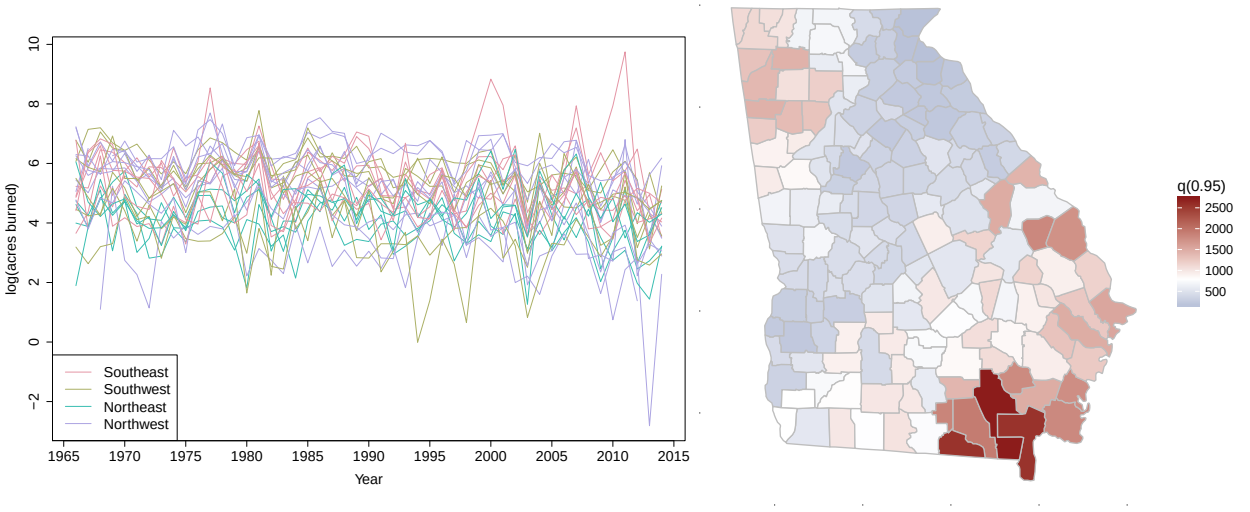

Figure 1: Time series of log acres burned for 25 randomly selected counties with colors coding the county's quadrant (left), and spatially smoothed threshold values, T_i for each county (right).

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 possible to increase q_1, but we set $q_1 = 0.90$ to increase the stability when estimating $\hat{\vartheta}_{ij}$. 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, separately by county, by subtracting the site's median and dividing by the site's interquartile range. Denote the standardized data by $\widetilde{\mathbf{Y}}_i$. Then we combine all sites together and plot a mean residual plot for $\widetilde{Y}_{it}, i = 1, \ldots, n_s$ and $t = 1, \ldots, n_t$. 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).

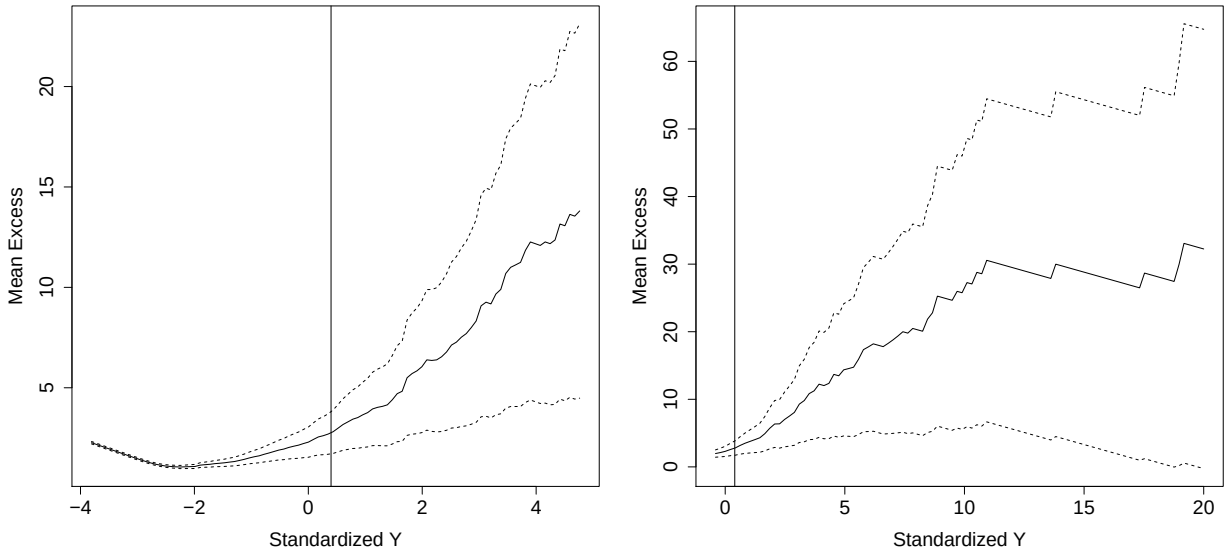

Figure 2: Mean residual plot for the data pooled across counties after standardizing using the county's median and interquartile range. The two panels show different ranges on the x-axis and include a vertical line at the sample 95th percentile.

The empirical basis functions for the analysis can be used to help explore spatial dependence in the extremes. The first six EBFs for the wildfire data along with the cumulative sum of the contributions for $v_1, \ldots, v_{25}$ are given in Figure 3. The first EBF roughly separates the Southeastern Plains (blue) from the coastal region in the southeast and mountains regions in the northeast (red). The remaining EBFs further partition the plains. As a comparison, we provide the first six principal components of the fire data along

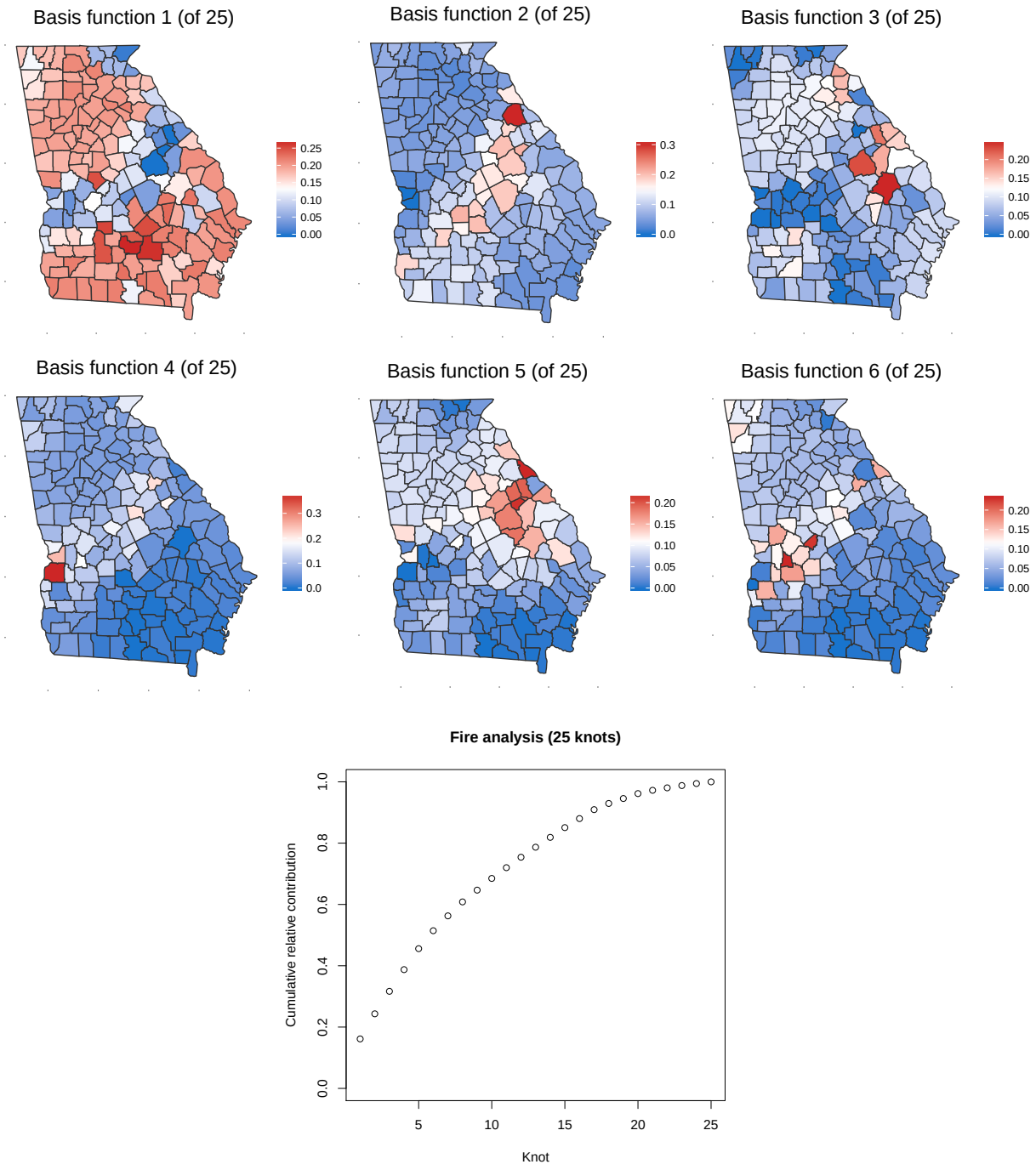

Figure 3: First six EBFs for the Georgia fire data and the cumulative sum of contributions $v_1, \ldots, v_{25}$.

with the cumulative sum of the first 25 eigenvalues in Appendix B.

Given the basis function estimates, we run the MCMC for 35,000 iterations using a burnin period of 25,000 iterations. We consider models fit with both EBF and GSK, and fit the model using $L = 5, 10, \ldots, 40$. Timings for each setting of L for 1,000 iterations are given in Table 1. These timings come from a single core of an Intel Core i7-5820K Haswell-E processor, using the OpenBLAS optimized BLAS library (`http://www.openblas.net`).

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 (QS) for quantile level q^* is given by $2\{I[Y(\mathbf{s}) > \hat{q}(\mathbf{s})] - q^*]\}\{\hat{q}(\mathbf{s}) - Y(\mathbf{s})\}$ where $\hat{q}(\mathbf{s})$ is the estimated q^*th quantile at site $\mathbf{s}$, and $I[\cdot]$ is an indicator function. The Brier score (BS) for predicting an exceedance of a level c at site $\mathbf{s}$ is given by $\{I[Y(\mathbf{s}) > c] - \hat{P}[Y(\mathbf{s}) > c]\}^2$. 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 1. For the data, the BS and QS are fairly similar for all number of basis functions and the EBF versus GSK. This is perhaps due to the spatial variation in the marginal distribution explaining most of the spatial variation. In fact, using the EBFs, the estimate of residual dependence for the fire data is $\hat{\alpha} = 0.861$ ($\alpha = 1$ is residual independence).

Based on the cross-validation results, we run a full analysis using all of the data with $L = 25$. Despite the GSK model having slightly smaller BS and QS values, we use the EBF method. Figure 4 gives posterior summaries for three quantities of interest. We provide maps of the linear time trend in the GEV location

Table 1: Average Brier scores ($\times 100$), average quantile scores for $q(0.95)$ and $q(0.99)$, and time (in minutes) for 1,000 iterations for fire analysis.

		Brier Scores ($\times 100$)		Quantile Scores		
	Process	$q(0.95)$	$q(0.99)$	$q(0.95)$	$q(0.99)$	Time
L = 5	EBF	4.269	1.653	107.904	67.345	1.16
	GSK	4.244	1.644	103.822	64.046	1.18
L = 10	EBF	4.328	1.675	107.075	66.300	1.56
	GSK	4.170	1.665	104.214	64.589	1.47
L = 15	EBF	4.330	1.690	108.381	67.671	1.86
	GSK	4.214	1.654	104.490	65.201	1.83
L = 20	EBF	4.346	1.697	107.389	66.957	2.18
	GSK	4.174	1.646	104.671	65.430	2.20
L = 25	EBF	4.263	1.650	106.656	64.913	2.54
	GSK	4.216	1.661	104.208	64.468	2.55
L = 30	EBF	4.328	1.678	106.265	64.674	2.90
	GSK	4.228	1.660	104.143	64.443	2.92
L = 35	EBF	4.329	1.671	106.817	65.002	3.29
	GSK	4.256	1.663	105.016	64.920	3.32
L = 40	EBF	4.284	1.653	106.621	64.753	3.70
	GSK	4.233	1.666	105.301	64.932	3.56

($\beta_{1,\text{time},i}$), and GEV log scale ($\beta_{2,\text{time},i}$) and a map of $\Delta Q90_i = Q90_{i,2014} - Q90_{i,1965}$ the change in the mean of the posterior distribution of $Q90_{i,t}$ between $t_1 = 1965$ and $t_2 = 2014$. We also provide maps of the posterior probability that each of the three terms is positive.

We construct the posterior distribution of the the estimated 90th quantile $Q90_{i,t}$ using the GEV parameters as follows. Let $Q90_{i,t}^{(j)}$ be the estimated 90th quantile at site i for time t at iteration j. We first compute

$$\begin{aligned} \mu_{i,t}^{(j)} &= \beta_{1,\text{int},i}^{(j)} + \beta_{1,\text{time},i}^{(j)} t \\ \log\left(\sigma_{i,t}^{(j)}\right) &= \beta_{2,\text{int},i}^{(j)} + \beta_{2,\text{time},i}^{(j)} t. \end{aligned} \tag{18}$$

Let $Q90_{i,t}^{(j)} = \mu_{i,t}^{(j)} + \sigma_{i,t}^{(j)} F^{-1}\left(0.90, \xi^{(j)}\right)$ where $F^{-1}(q, \xi)$ is the inverse distribution function of the GEV$(1, 1, \xi)$ distribution evaluated at the qth quantile. Finally, let $Q90_{i,t}$ be posterior mean of $Q90_{i,t}^{(j)}$. To obtain the posterior probability of seeing an increase over time, we take the posterior distributions of each parameter of interest at the two time points. Consider two time points $t_1 < t_2$. Let $\varphi_t^{(j)}$ be the parameter of interest at iteration j and time t. We then take the posterior mean of $I\left[\varphi_{t_2}^{(j)} > \varphi_{t_1}^{(j)}\right]$, for the posterior probability of seeing an increase in φ from time t_1 to t_2.

These results suggest that there has been a slight increase in the amount of acres burned over time. When looking at the variability of the acres burned over time, we see some patterns that generally correspond to landcover in Georgia (see `http://narsal.uga.edu/` for landcover maps). In the high intensity urban areas (e.g. the Piedmont region) and croplands, we see a decrease in variability as well as very low P$[\Delta Q90 > 0]$. However, in the more forested areas (e.g. Blue Ridge mountains, Okefenokee Swamp), we see increased variability over time as well as a high P$[\Delta Q90 > 0]$. This suggests that the 10-year return level for wildfires in regions with larger tree cover is increasing over time which corresponds with increased

drought and higher temperatures.

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). Because the data are block maxima, we set $T = -\infty$.

For this dataset, to estimate the EBFs, we use the combined current and future data. The first six EBFs for the combined data along with the cumulative sum of the contributions for $v_1, \ldots, v_{25}$ are given in Figure 6. 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 Appendix B. For the precipitation data, we run the MCMC for 25,000 iterations using a burnin period of 15,000 iterations. We consider models fit with both EBF and GSK, and fit the model using $L = 5, 10, \ldots, 40$. Timing for each setting of L is given in Table 2 for 1,000. The timings are obtained using the same machine as for the fire analysis.

5.5 Precipitation results

We use 5-fold cross-validation to assess the predictive performance of a model. For each method, we randomly select 80% of the observations across counties and years to be used as a training set to fit the model. The remaining 20% of sites and years are withheld for testing model predictions. As with the fire analysis, we use Brier scores and quantile scores to compare model performance. The Brier and quantile scores for the current and future precipitation data analysis are given in Table 2. For these data, we observe more variation in the scores across the number of basis functions and generally an advantage in using EBF

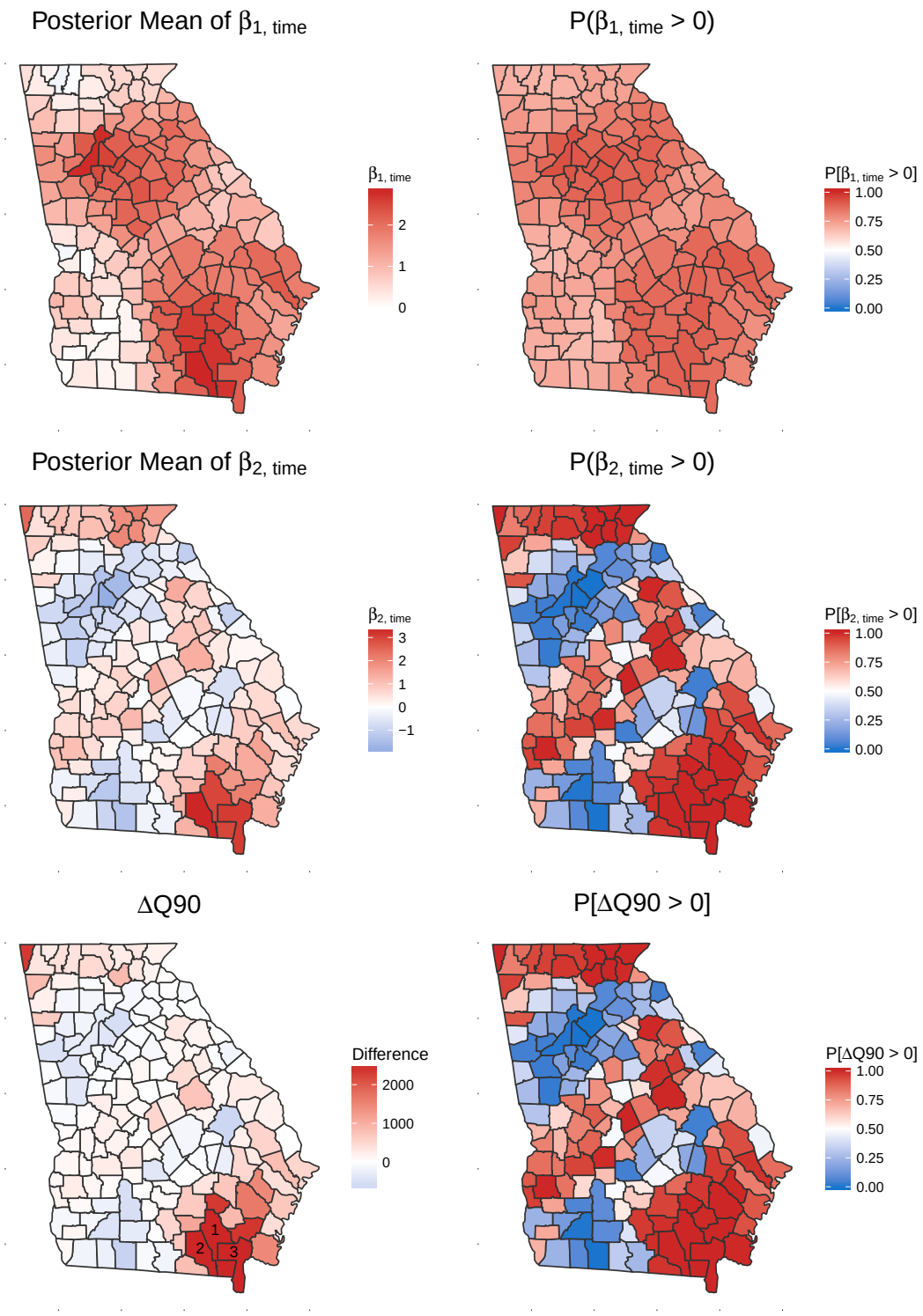

Figure 4: Posterior mean of $\beta_{1,\text{time}}$ (top left), posterior mean of $\beta_{2,\text{time}}$ (middle left), estimate of $\Delta Q90$ (bottom left), P$[\beta_{1,\text{time}} > 0]$ (top right), P$[\beta_{2,\text{time}} > 0]$ (middle right), and P$[\Delta Q90 > 0]$ for fire data using EBF. In three counties (labeled), $\Delta Q90 > 2500$: County 1 - Ware (11,109), County 2 - Clinch (7,128), and County 3 - Charlton (6,545)

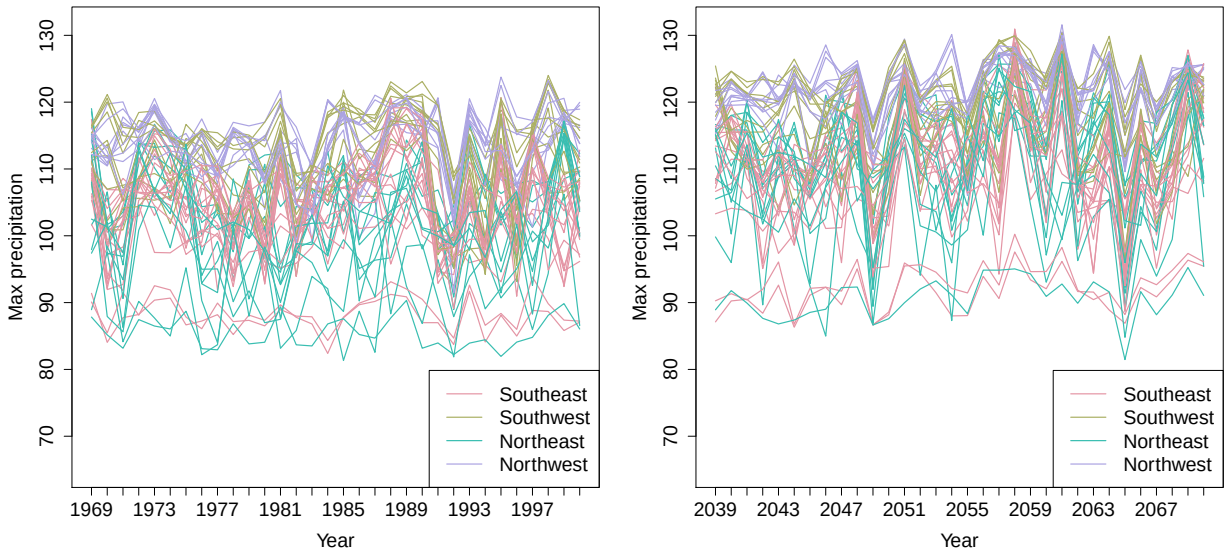

Figure 5: Time series of yearly max precipitation for current (1969 – 2000) (left). Time series of yearly max precipitation for future (2039 – 2070) (right).

over GSK. This is in contrast to the fire analysis, and is likely due to the fact that the spatial dependence is estimated to be much stronger than for the fire analysis. When using the EBFs, the estimate of residual dependence for the precipitation data is $\hat{\alpha} = 0.280$ ($\alpha = 1$ is residual independence).

Based on the cross-validation results, we run a full analysis using all of the data with $L = 25$ and EBF. Figure 7 gives posterior summaries for three quantities of interest. Because we have two separate time periods, current and future, we look at the differences between the estimates for $\hat{\mu}$, $\log(\hat{\sigma})$, and $\hat{q}(0.90)$ between $t_1 = 2000$ and $t_2 = 2070$ where $\hat{\mu}$, $\log(\hat{\sigma})$, and $\hat{q}(0.90)$ are calculated as described in Section 5.3. We plot $\Delta\mu = \hat{\mu}_{2070} - \hat{\mu}_{2000}$, $\Delta \log(\hat{\sigma}) = \log(\hat{\sigma})_{2070} - \log(\hat{\sigma})_{2000}$, and $\Delta Q90 = \hat{q}(0.90)_{2070} - \hat{q}(0.90)_{2000}$, and the estimated probabilites that each are positive.

The results seem to suggest that the strength of extreme rain events will increases between 2000 and 2070 as well as greater variability in the northeast region of the U.S. as well as Ohio and parts of the south. There very strong evidence to suggest that most of the eastern U.S. should expect to see an increase in the

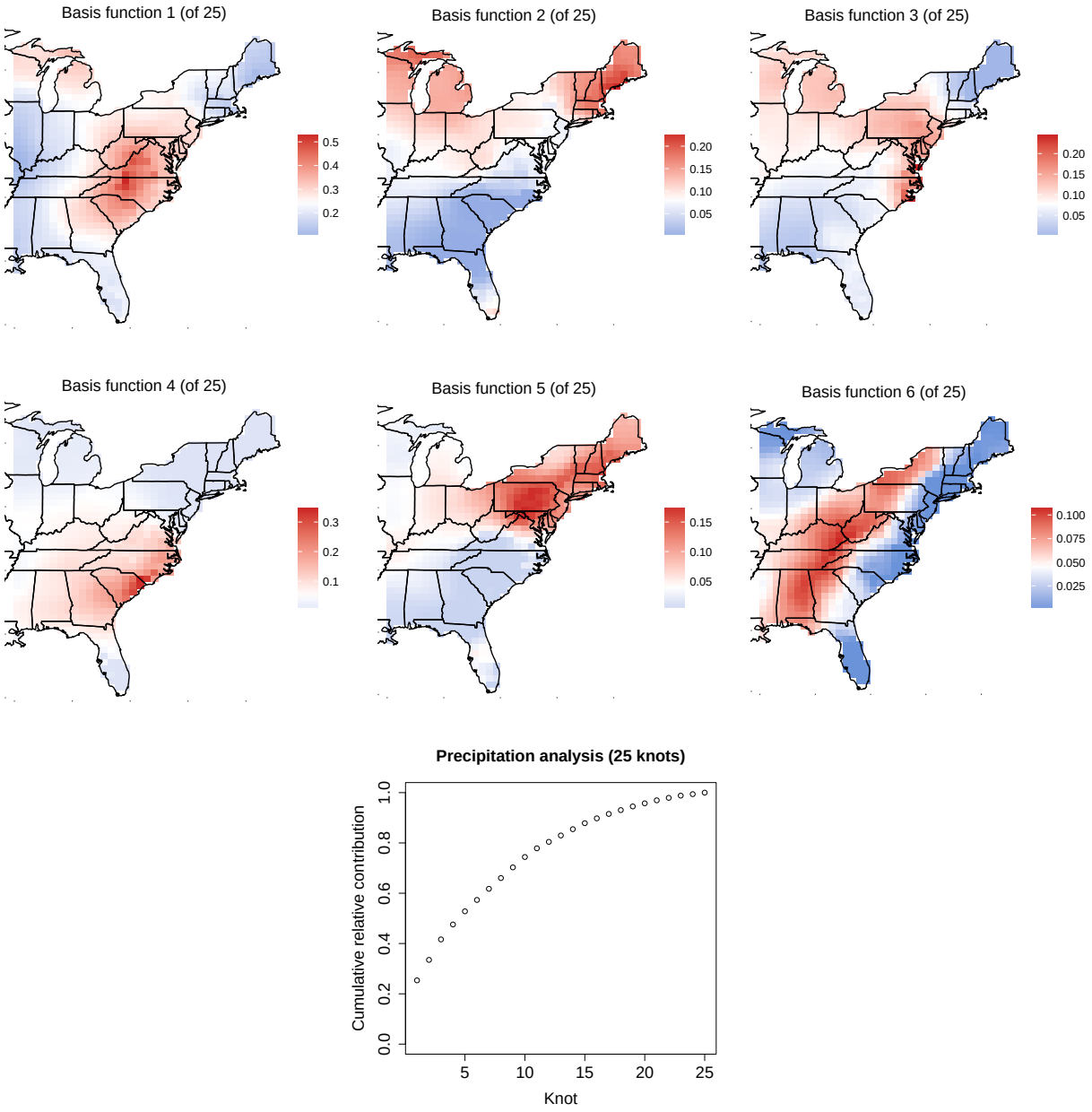

Figure 6: First six EBFs for the combined precipitation data and the cumulative sum of contributions $v_1, \ldots, v_{25}$.

10-year return level between 2000 and 2070. Exceptions to this trend appear in southern parts of Alabama and Mississippi and regions on the border between South Carolina and Georgia which will likely experience a decrease in the 10-year return level.

6 Discussion

In this paper we have proposed new empirical basis functions for a data-driven low-rank approximation to a max-stable process. The basis functions provide researchers with an exploratory data analysis tool to explore maps of extremal dependence over space. The functions can also be used as inputs to an MCMC algorithm for inference and predicitons over space. The results from the data analysis provide evidence to suggest that in the presence of strong spatial dependence as with the precipitation data, the empirical basis functions show an improvement in quantile scores over using knots and standardized Gaussian kernel functions without an increase in the amount of time for computing.

We have used the EBF for exploratory analysis and Bayesian inference. Another possibility is to use the methods to reduce the data under consideration from the actual responses to loadings A_{kt}. That is, given the EBF, one could obtain estimates of the A_{kt} using a separate maximum likelihood estimation for each time point. Time series of the estimated A_{kt} may be used as a fast and simple method to study large-scale spatiotemporal trends.

Table 2: Average Brier scores ($\times 100$), average quantile scores for $q(0.95)$ and $q(0.99)$, and time (in minutes) for 1,000 iterations for precipitation analysis.

		Current					Future				
		BS ($\times 100$)		QS			BS ($\times 100$)		QS		
L	Process	$q(0.95)$	$q(0.99)$	$q(0.95)$	$q(0.99)$	Time	$q(0.95)$	$q(0.99)$	$q(0.95)$	$q(0.99)$	Time
5	EBF	3.813	1.098	0.740	0.203	5.80	3.357	1.112	0.738	0.209	5.82
	GSK	3.854	1.074	0.745	0.204	5.54	3.338	1.101	0.742	0.210	5.49
10	EBF	3.680	1.063	0.698	0.192	6.54	3.148	1.067	0.687	0.198	6.51
	GSK	3.628	1.068	0.705	0.195	6.30	3.088	1.072	0.709	0.201	6.26
15	EBF	3.505	1.065	0.668	0.187	7.27	3.101	1.095	0.661	0.189	7.22
	GSK	3.618	1.051	0.695	0.194	7.05	3.057	1.064	0.697	0.199	7.02
20	EBF	3.400	1.026	0.651	0.182	8.14	3.101	1.087	0.649	0.189	8.11
	GSK	3.552	1.039	0.688	0.192	7.87	3.065	1.062	0.692	0.196	7.85
25	EBF	3.463	1.058	0.650	0.182	9.00	3.003	1.113	0.637	0.185	8.98
	GSK	3.643	1.063	0.679	0.190	8.74	2.039	1.054	0.686	0.196	8.71
30	EBF	3.455	1.038	0.646	0.182	9.89	2.956	1.073	0.630	0.182	9.88
	GSK	3.550	1.046	0.684	0.191	9/60	3.074	1.067	0.685	0.195	9.55
35	EBF	3.471	1.065	0.649	0.183	10.79	3.036	1.114	0.645	0.189	10.79
	GSK	3.608	1.064	0.686	0.194	10.46	3.083	1.075	0.691	0.199	10.43
40	EBF	3.551	1.114	0.652	0.184	11.70	3.050	1.131	0.645	0.189	11.66
	GSK	3.605	1.067	0.686	0.194	11.32	3.104	1.083	0.693	0.199	11.28

Acknowledgements

The authors would like to acknowledge Dan Cooley for his helpful suggestions on the manuscript. The authors' work was partially supported by grants from the Department of the Interior (14-1-04-9), National Institutes of Health (R21ES022795-01A1), the US Environmental Protection Agency (R835228), the National Science Foundation (1107046).

A Appendices

A.1 Extreme value distributions

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

$$t(y) = \begin{cases} \left[1 + \xi \dfrac{y-\mu}{\sigma}\right]^{-1/\xi}, & \xi \neq 0 \\ \\ \exp\left\{-\dfrac{y-\mu}{\sigma}\right\}, & \xi = 0. \end{cases} \tag{19}$$

The probability density function for the GEV is given by $f(y) = \dfrac{1}{\sigma} t(x)^{\xi+1} \exp\{-t(y)\}$ where $t(y)$ is defined in (19).

A.2 Grid approximation to PS density

The PS(α) density can be challenging to use because it does not have a closed form. From Section 2 of (Stephenson, 2009), the density can be expressed as

$$g_1(A) = \int_0^1 g_1(A, B)\, \mathrm{d}B, \tag{20}$$

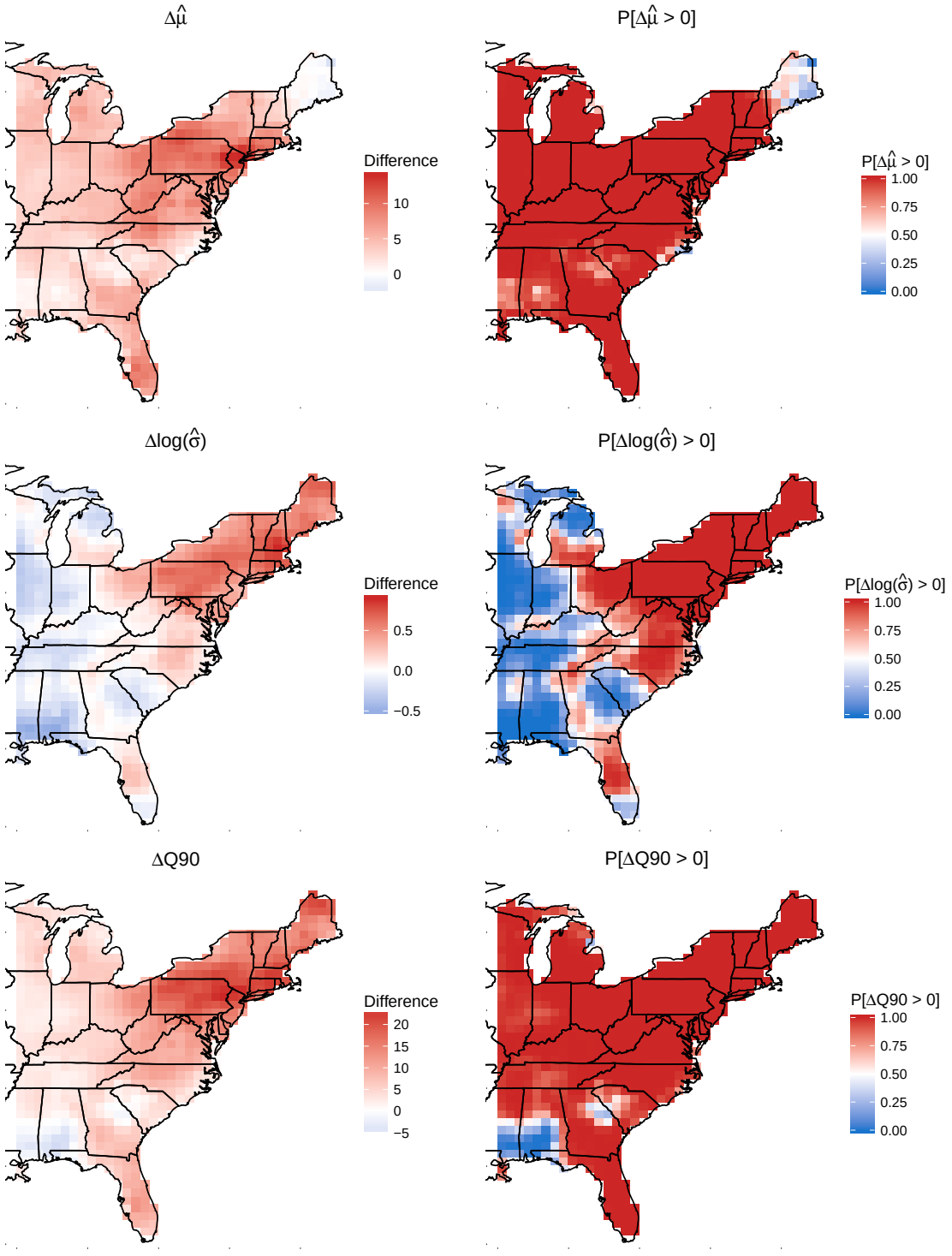

Figure 7: Posterior mean of $\Delta\mu$ (top left), posterior mean of $\Delta \log(\sigma)$ (middle left), estimate of $\Delta Q90$ (bottom left), $\mathrm{P}[\Delta\mu > 0]$ (top right), $\mathrm{P}[\Delta \log(\sigma) > 0]$ (middle right), and $\mathrm{P}[\Delta Q90 > 0]$ between 2000 and 2070 for precipitation data using EBF.

where

$$g_1(A,B) = \frac{\alpha}{1-\alpha}\left(\frac{1}{A}\right)^{1/1-\alpha} c(\pi B)\exp\left\{-\left(\frac{1}{A}\right)^{\alpha/(1-\alpha)} c(\pi B)\right\}, \tag{21}$$

with

$$c(\psi) = \left[\frac{\sin(\alpha\psi)}{\sin(\psi)}\right]^{1/(1-\alpha)} \frac{\sin[(1-\alpha)\psi]}{\sin(\alpha\psi)}. \tag{22}$$

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

A.3 Standardized Gaussian kernel functions

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

$$\hat{B}_{il} = \frac{\exp\left\{-\dfrac{||\mathbf{s}_i - \mathbf{k}_l||^2}{2\rho^2}\right\}}{\displaystyle\sum_{j=1}^{L}\exp\left\{-\frac{||\mathbf{s}_i - \mathbf{k}_j||^2}{2\rho^2}\right\}} \tag{23}$$

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

B Principal components

As a comparison to the EBFs, Figure 8 gives the first six principal components for the fire data, and Figure 9 gives the first six principal components for the precipitation data. These figures show that the EBFs resemble the EOFs for the uncersored precipitation data, but are quite different that the EOF from the censored fire data. This is not surprising considering that the censoring results in only a small subset of data being used to estimate dependence.

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.

Huser, R. and Davison, A. C. (2014) Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76**, 439–461, arXiv:1201.3245.

Lee, D. D. and Seung, S. H. (1999) Learning the parts of objects by non-negative matrix factorizations. *Nature*, **401**, 788 – 791.

Mairal, J., Bach, F. and Ponce, J. (2014) Sparse modeling for image and vision processing. *Foundations and Trends in Computer Graphics and Vision*, **8**, 85 – 283.

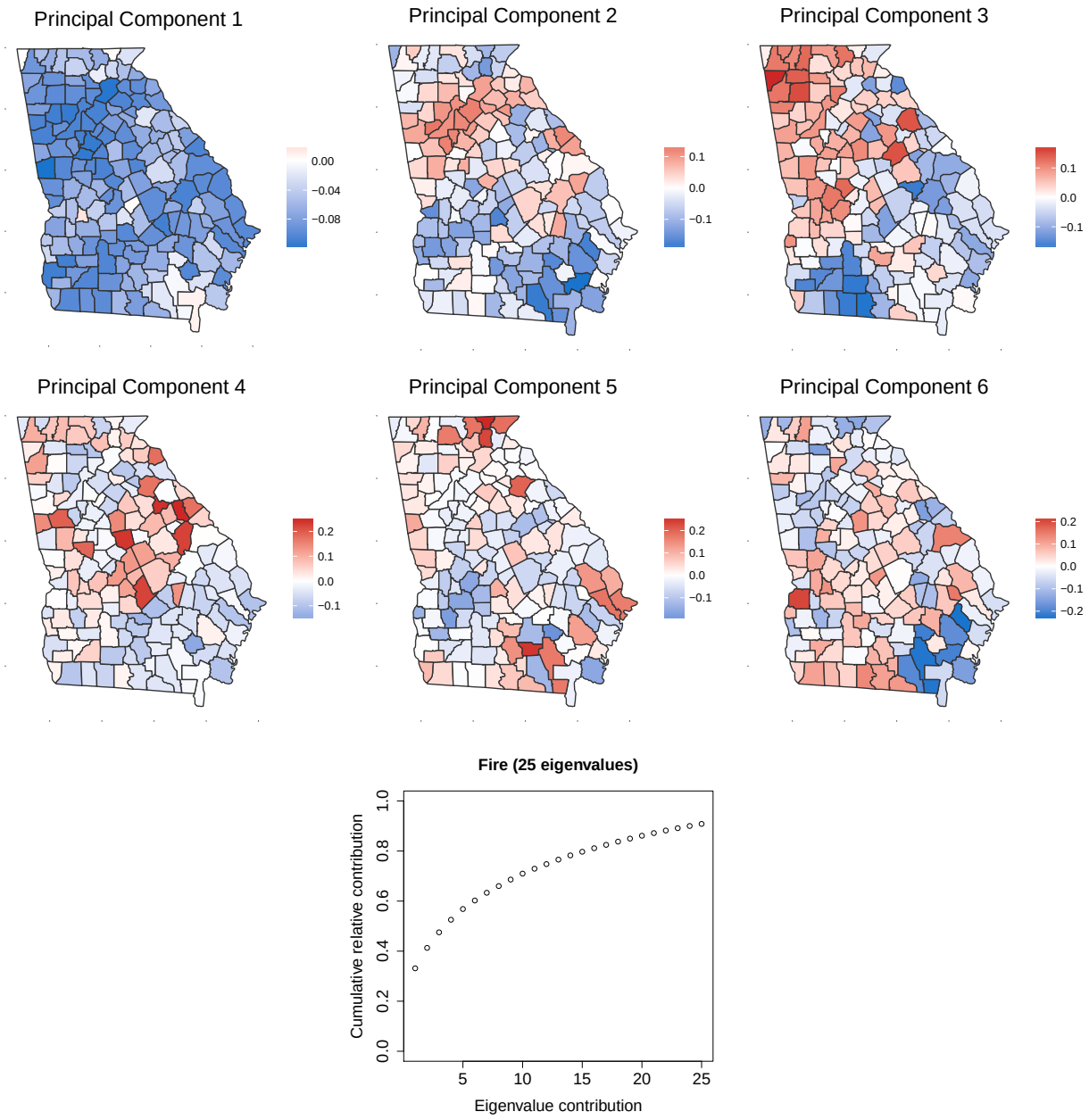

Figure 8: First six principal components and the cumulative sum of the first 25 eigenvalues for the Georgia fire data.

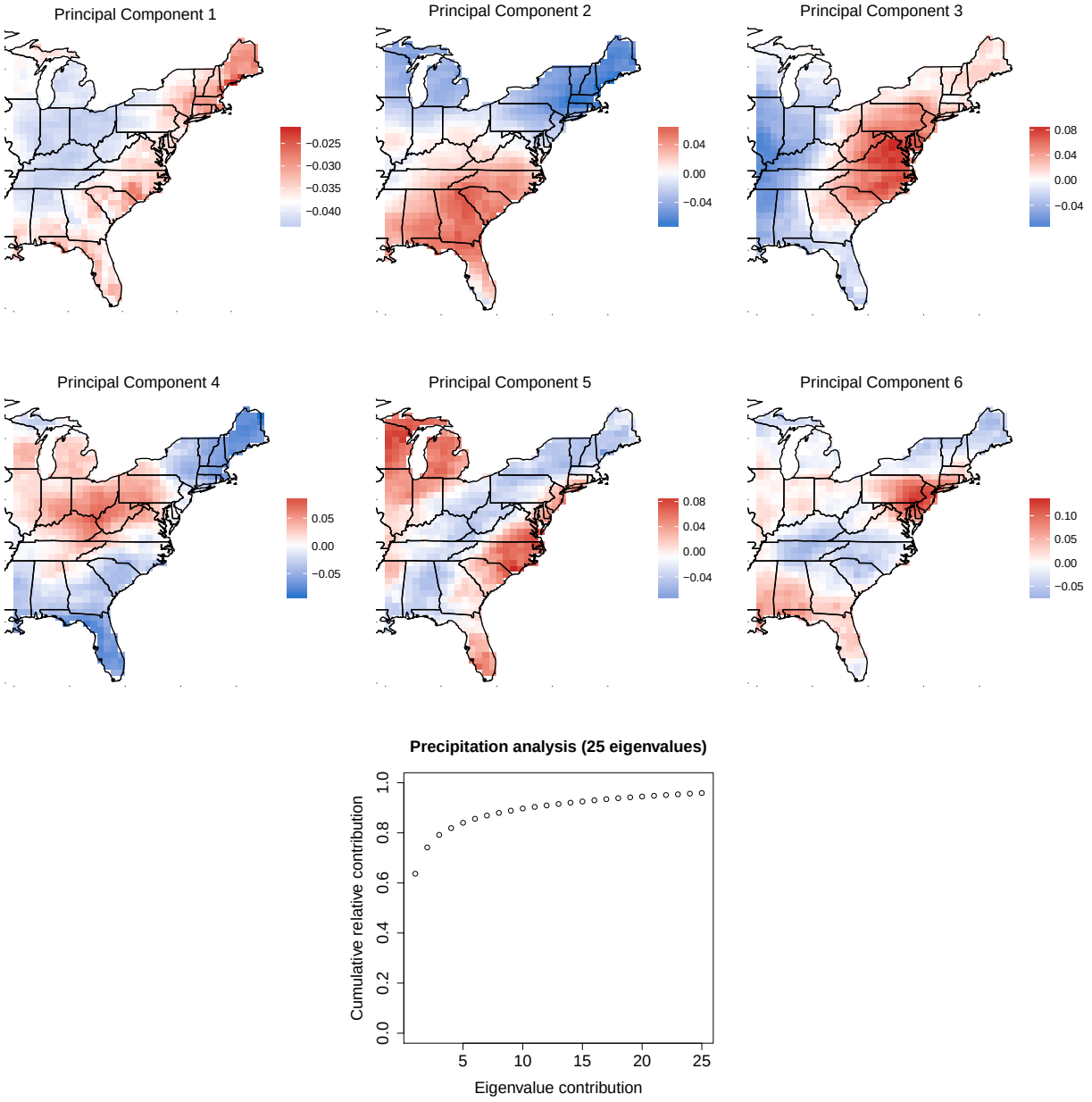

Figure 9: First six principal components and the cumulative sum of the first 25 eigenvalues for the precipitation data.

Morris, S. A., Reich, B. J., Thibaud, E. and Cooley, D. (under review) A space-time skew-*t* model for threshold exceedances. *Biometrics*.

Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, **105**, 263–277.

Reich, B. J. and Shaby, B. A. (2012) A hierarchical max-stable spatial model for extreme precipitation. *The Annals of Applied Statistics*, **6**, 1430–1451.

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

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

Stephenson, A. G. (2009) High-dimensional parametric modelling of multivariate extreme events. *Australian & New Zealand Journal of Statistics*, **51**, 77–88.

Thibaud, E., Mutzner, R. and Davison, A. C. (2013) Threshold modeling of extreme spatial rainfall. *Water Resources Research*, **49**, 4633–4644.

Thibaud, E. and Opitz, T. (2015) Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, **102**, 855–870, arXiv:1401.0168v2.

Wadsworth, J. L. and Tawn, J. A. (2014) Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, **101**, 1–15.

Wang, Y. and Stoev, S. A. (2011) Conditional sampling for spectrally discrete max-stable random fields. *Advances in Applied Probability*, **43**, 461–483, arXiv:1005.0312v2.