# Spatio-Temporal Models for Large-scale Indicators of Extreme Weather

Matthew J. Heaton Matthias Katzfuss Shahla Ramachandar
Kathryn Pedings Yi Li Eric Gilleland
Elizabeth Mannshardt-Shamseldin Richard L. Smith

December 1, 2009

The changing global climate has sparked an interest in how these changes are affecting the intensity and frequency of extreme weather events such as thunderstorms and tornadoes because these extreme events pose a significant threat to life, property, and economic stability. This article uses and evaluates several spatio-temporal statistical extreme value models to model extreme weather from reanalysis data observed across the continental United States and Mexico. The models find that the intensity of extreme weather is particularly high for the central United States. Additionally, the intensity of extreme weather is increasing over time but the amount of increase may not be practically significant.

KEY WORDS: Severe storm; Reanalysis data; Extreme value; Coregionalization; Point Process.

## 1. INTRODUCTION

Changes in frequency and intensity of extreme weather events (e.g. thunderstorms and tornadoes) as a result of global climate change is of great concern as these events pose a significant threat to life, property, and economic stability. Developing models which are able to assess temporal and spatial trends in extreme events is, therefore, important in understanding the effects of these events. Modeling extreme events, however, is difficult for several reasons. For example,

Matthew J. Heaton, Department of Statistical Science, Duke University, matt@stat.duke.edu; Matthias Katzfuss, Department of Statistics, The Ohio State University, katzfuss@stat.osu.edu; Shahla Ramachandar, Department of Mathematical Sciences, University of Texas-Dallas, shahla@utdallas.edu; Kathryn Pedings, Department of Mathematics, College of Charleston, kepeding@edisto.cofc.edu; Yi Li, Department of Mathematics, Duke University, yi.li@duke.edu; Eric Gilleland, National Center for Atmospheric Research, ericg@ucar.edu; Elizabeth Mannshardt-Shamseldin, Department of Statistical Science, Duke University, elizabeth@stat.duke.edu; Richard L. Smith, Department of Statistics and Operations Research, University of North Carolina-Chapel Hill,rls@email.unc.edu

the coarse resolution of current climate models renders them incapable of capturing fine scale extreme events. Thus, modeling large-scale indicators of extreme weather is often considered as a viable alternative to modeling the extreme events directly (see Gilleland et al. 2008). For example, Brooks et al. (2003) found that a combination of convective available potential energy ($CAPE$) in Joules/Kg and wind shear ($WS$) in meters/second are associated with atmospheric conditions that exist during the occurrence of thunderstorms and tornadoes.

Even when using large-scale indicators of extreme events, modeling extreme weather can be difficult due to its highly skewed distribution. For example, Figure 1 displays a histogram of extreme values of $WmSh \equiv WS \times W_{max} \in \mathbb{R}^+$ from the reanalysis data set (described in Section 2) where the transformation $W_{max} = \sqrt{2 \times CAPE}$ allows $W_{max}$ to be on the same scale as $WS$. Needless to say, because the events of interest lie in the tail of the distribution, appropriate statistical models for extreme events need to account for heavy tails such as those displayed in Figure 1.

In an effort to understand the spatial and temporal trend of extreme weather, this article develops statistical models for the large scale indicator of extreme weather $WmSh$. By modeling $WmSh$, rather than $WS$ and $CAPE$ jointly, the outcome is univariate which greatly reduces the complexity of the statistical models. Specifically, this article poses several spatio-temporal statistical models which account for varying degrees of spatial and temporal trends in $WmSh$. A special distinction should be made in that models for $WmSh$ are not models for extreme weather, but, rather, an indicator of conditions for extreme weather to occur. Thus, the models considered in this article are for large-scale *indicators* of extreme weather which can be useful for monitoring and predicting conditions where extreme weather events can occur. The data under consideration are reanalysis data from the National Center for Atmospheric Research (NCAR) and the National Center for Environmental Prediction (NCEP) reanalysis project. Section 2 provides a description of the reanalysis project and the data set used in this article.

The remainder of this article is outlined as follows. Section 3 provides a review of univariate extreme value theory including the generalized extreme value (GEV) distribution for block maxima, the generalized Pareto distribution (GPD) for threshold exceedences, and a point process characterization of extremes (PPE). An exploratory data analysis is performed in Section 4 using the GEV and the GPD approaches. Based on the results of the exploratory analysis, Section 5 suggests three Bayesian hierarchical models for extreme weather in increasing complexity. The results of the fitted models are discussed in Section 6. A discussion of the work undertaken in this article, along with possible future research directions, are provided in Section 7.

## 2. THE NCAR/NCEP REANALYSIS DATASET

The reanalysis project (see Kalnay et al. 1996) is a collaborative effort between the National Center of Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR). The main purpose of this project was to produce relatively high-resolution global analyses of atmospheric fields over a long time period. Covering the years 1957 through 2002, the reanalysis data contains observations from each time index synthesized using a static data assimilator. The component datasets included global rawinsonde data, comprehensive ocean-atmosphere data, aircraft data, surface land synoptic data, satellite sounder data, special sensing microwave/imager data (surface wind speeds), and satellite cloud drift winds. The different stages of the project can be outlined as data preparation, preprocessing, assimilation module, reanalysis output and the climate data assimilation system (CDAS) for future use. For a complete description of the project see Kalnay et al. (1996) and see Kistler et al. (2001) for the related documentation.

This article uses reanalysis data of $CAPE$ and $WS$ measured at 884 grid points on a $1.4 \times 1.4$ degree grid across the United States and adjacent regions over the 42-year period 1958 - 1999. A few days over this time span are not available leading to 15,300 total measurements of $CAPE$ and $WS$ at each location. This article uses the transformation $WmSh = WS \times W_{max}$ to reduce the dimensionality of the data. For a brief description of the reanalysis data on a global scale, see Brooks et al. (2003).

## 3. REVIEW OF EXTREME VALUE THEORY

Let $X$ be a random variable with distribution function $F_X(\cdot \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of parameters and $x_1, \ldots, x_n$ are $n$ $iid$ observations from $F_X(\cdot \mid \boldsymbol{\theta})$. When considering extreme events, the primary concern is with the tail behavior of $F_X(\cdot \mid \boldsymbol{\theta})$, hence only those observations that are considered "extreme" are used for the analysis. As such, the use of distributions in the exponential family as a model for extreme events is not appropriate. Additionally, substituting a plug-in estimator, say $\widehat{\boldsymbol{\theta}}$, into $F_X(\cdot \mid \boldsymbol{\theta})$ frequently results in significant discrepancies between $\mathbb{P}\text{r}(X \geqslant u \mid \widehat{\boldsymbol{\theta}})$ and $\mathbb{P}\text{r}(X \geqslant u \mid \boldsymbol{\theta})$ for large $u$. Rather than working with $F_X(\cdot \mid \boldsymbol{\theta})$ directly, extreme value theory works with families of extreme value distributions to directly model the tail behavior.

### 3.1 The Generalized Extreme Value Distribution for Block Maxima

Consider the extreme event defined by $M_n = \max(x_1, \ldots, x_n)$. If $F_X(\cdot \mid \boldsymbol{\theta})$ is known, then the distribution function of $M_n$ can be calculated directly as $\mathbb{P}\text{r}(M_n \leqslant m) = F_X(m \mid \boldsymbol{\theta})^n$. This

distribution function, however, is not practically useful because $\mathbb{P}\text{r}(M_n \leqslant m) \to 0$ as $n \to \infty$ for any $m$ such that $F_X(m \mid \boldsymbol{\theta}) < 1$. If normalizing sequences $a_n$ and $b_n > 0$ exist such that $\mathbb{P}\text{r}((M_n - a_n)/b_n \leqslant m) \to G(m; \mu, \sigma, \xi)$, where $G(m; \mu, \sigma, \xi)$ is a non-degenerate distribution, then $G(m; \mu, \sigma, \xi)$ is given by,

$$G(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{1}$$

where $\mu$ is a location parameter, $\sigma > 0$ is a scale parameter, $\xi$ is a shape parameter, and $(x)_+ = \max(0, x)$. The distribution function (1) is referred to as the generalized extreme value (GEV) distribution. The shape parameter $\xi$ controls the tail behavior of the distribution. For $\xi > 0$ the tail of $G$ decays polynomially, when $\xi \to 0$ the tail decays exponentially, and when $\xi < 0$ the distribution has an upper bound (Coles 2001).

In practice, the GEV distribution is used to obtain probabilities for block maxima where the blocks are generally determined by some natural division such as years. For example, if $x_{ki}$ is the $i^{th}$ observation in block $k$, the parameters in (1) would be fit using the block maxima $M_k = \max(x_{k1}, \ldots, x_{kn})$. The main disadvantage of the block maxima approach for extremes is that much of the observed data is discarded, which may be a problem if data are already scarce.

### 3.2 The Generalized Pareto Distribution for Threshold Exceedances

To avoid the loss of data issue encountered in the block maxima approach, consider the extreme event $X > x + u \mid X > u$ for a large threshold $u$ such that all $X > u$ is retained for analysis. Defining $Y \equiv X - u \mid X > u$, extreme value theory finds, under certain conditions, $\mathbb{P}\text{r}(Y \leqslant y) \to H(y; \psi, \xi)$ as $u \to \sup\{x : F_X(x \mid \boldsymbol{\theta}) < 1\}$, where,

$$H(y; \psi, \xi) = 1 - \left[1 + \xi\frac{y}{\psi}\right]_+^{-1/\xi}, \tag{2}$$

$\psi > 0$ is a scale parameter, and $\xi$ is the shape parameter. The distribution given by (2) is referred to as the generalized Pareto distribution (GPD). The GPD and the GEV distribution are related to each other in that $\psi = \sigma + \xi(u - \mu)$, and $\xi$ in the GEV distribution is theoretically the same as that in the GPD.

The GPD is used to characterize extreme events defined by exceedances over a high threshold $u$. The advantage of using the GPD instead of the GEV distributions to model extremes is that the GPD retains all the data which fall over the threshold $u$. Thus the amount of data used to fit a GPD may be markedly higher than if only the block maxima are considered. The disadvantage of using the GPD is that $\mathbb{P}\text{r}(Y \leqslant y) \to H(y; \psi, \xi)$ does not hold for a particular choice of $u$, and the

resulting estimated parameters are threshold dependent as shown through the reparameterization from GPD scale $\psi = \sigma + \xi(u - \mu)$ to GEV scale. However, asymptotic theory suggests selecting a large enough $u$ such that $\Pr(Y \leqslant y) \approx H(y; \psi, \xi)$. In practice, $u$ is typically estimated from the data using mean residual life plots (see Coles 2001, Section 4.3.1) or, as is the approach used in this paper, setting $u$ to be some large quantile of the observed values. Alternatively, Cooley et al. (2007) use the fact that estimates of $\xi$ should be stable beyond a sufficiently large $u$ and fit (2) using several different values $u_1, \ldots, u_K$ to obtain $\widehat{\xi}_1, \ldots, \widehat{\xi}_K$ and then choose $u$ to be the minimum $u_i$ such that $\widehat{\xi}_i \approx \widehat{\xi}_{i+1} \approx \cdots \approx \widehat{\xi}_K$. Regardless of the approach, once chosen, $u$ is treated as fixed and (2) is assumed to be a good approximation to $\Pr(Y \leqslant y)$ for the remainder of the analysis. An alternative to selecting a single $u$ was proposed by Behrens et al. (2004) which estimated $u$ probabilistically by assigning a prior distribution to $u$ and estimating $u$ via Markov chain Monte Carlo (MCMC) methods, but their results were inconclusive, and thus not employed here.

### 3.3 A Point Process Characterization of Extremes

Smith (1989) merged the GEV approach and the GPD approach for extremes by developing a point process characterization for extremes (PPE). Instead of viewing the value of extreme events as a univariate outcome, consider the bivariate process $(t, Z)$, where $t$, the time at which $Z \equiv X \mid X > u$ is observed, is also a random variable. Under certain normalization criteria, Smith (1989) showed that the bivariate process $(t, Z)$ behaves as a non-homogeneous Poisson process with intensity,

$$\lambda(t, z) = \frac{1}{\sigma}\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]_+^{-1/\xi - 1}, \tag{3}$$

where $\mu$, $\sigma$, and $\xi$ are the corresponding parameters of the GEV distribution. Taking time to be discrete and considering regions of the form $\mathcal{D} = \{1, \ldots, T\} \times (u, \infty)$, then following the arguments outlined in Chapter 7 of Coles (2001), the likelihood of the observed data $(t_i, z_i)$ is,

$$\begin{aligned} L(\mu, \sigma, \xi \mid \{t_i, z_i\}) &= \exp\left\{-\int_{\mathcal{D}} \lambda(t, z) dz dt\right\} \prod_i \lambda(t_i, z_i) \\ &= \exp\left\{-T\left[1 + \xi\left(\frac{u - \mu}{\sigma}\right)\right]_+^{-1/\xi}\right\} \prod_i \lambda(t_i, z_i), \end{aligned} \tag{4}$$

where $T$ is the number of years of observed data. The main advantage of using the point process is that the GEV parameters (which easily extend to include spatial and temporal effects) are retained while still using all observations that exceed the threshold $u$. Additionally, because the PPE is defined in terms of the GEV parameters, parameter estimates are not tied to the threshold $u$ as in the GPD approach.

The parameters in (1), (2), and (4) can be difficult to interpret. Thus, extreme value models often interpret the estimated parameters in terms of $r$-year return levels. Intuitively, the $r$-year return level is defined as the value so extreme that it is only exceeded once every $r$ years. Mathematically, the $r$-year return level, $RL_r$, solves the equation $\mathbb{P}\text{r}(X > RL_r) = (rn_y)^{-1}$ for $RL_r$, where $n_y$ is the number of observations taken per year. For example, using the GPD, $RL_r$ is given by,

$$RL_r = u + \frac{\psi}{\xi}((n_y \times r \times \zeta_u)^{\xi} - 1), \tag{5}$$

where $\zeta_u = \mathbb{P}\text{r}(X > u)$. When using the point process approach, calculation of $RL_r$ relies on the identity $\psi = \sigma + \xi(u - \mu)$.

## 4. EXPLORATORY DATA ANALYSIS

The primary goal of this paper is to develop an extreme value model for large-scale indicators of ideal conditions for extreme weather events. Specifically, the goal is to build an extreme value model for $WmSh = WS \times W_{max}$, where $W_{max} = \sqrt{2 \times CAPE}$, wind shear ($WS$) measures the vertical shear of the tropospheric horizontal winds in meters/second (m/s), and $CAPE$ is the convective available potential energy of the atmosphere measured in Joules/Kg. By modeling $WmSh$, rather than $WS$ and $CAPE$ jointly, the outcome is now univariate which greatly reduces the complexity of the models.

An exploratory analysis of the data is first conducted to expose any spatial and temporal patterns, so that these patterns can be appropriately accounted for in the models proposed in Section 5. The need to account for spatial correlation is extant, as weather is inherently spatial and does not affect regions in space independently. What is unknown is the degree to which the spatial structure needs to be accounted for in a well-posed statistical model. Not so obvious, however, is whether or not a temporal component is needed. Indeed, the behavior of $WmSh$ over time is one of the primary research questions of this article. Thus, an exploratory analysis is useful to investigate what degree of spatial and temporal effects need to be accounted for in the modeling.

Let $x_{td}(\mathbf{s}_l)$ be the observed value of $WmSh$ on day $d$ in year $t$ at location $\mathbf{s}_l$. Furthermore, let $\mu(\mathbf{s}_l)$, $\sigma(\mathbf{s}_l)$, $\psi(\mathbf{s}_l)$ and $\xi(\mathbf{s}_l)$ be the corresponding parameters in (1) and (2) at location $\mathbf{s}_l$ for the $l = 1, \ldots, 884$ spatial locations of the reanalysis data. For the reanalysis data considered here, having a fixed $u$ across all locations would be problematic because each $\mathbf{s} \in \mathcal{S}$ has vastly different values of $WmSh$. Thus, for large $u$, the data at several sites would be thrown out completely. Rather, the threshold is determined to be spatially varying, in that $u(\mathbf{s})$ is fixed at the $95^{th}$ percentile of $\{x_{td}(\mathbf{s}) : t = 1, \ldots, 42; d = 1, \ldots, 365\}$. This value for $u(\mathbf{s})$ was chosen because it is high enough such that exceedances can be considered to be in the tail of the

distribution, while low enough to retain a sufficient portion of the data at each location. A map of this $u(\mathbf{s})$ is shown in Figure 2. Of note in this figure is the long narrow band of consistently higher $WmSh$ values in the central United States extending from South Dakota down through Texas and into Mexico. This band of higher intensity $WmSh$ is accounted for by the models discussed in section 5.

To explore the spatial pattern in the distribution of $WmSh$, each observation is treated as independent across space, and maximum likelihood estimates of the GEV parameters $\mu(\mathbf{s}_l)$, $\sigma(\mathbf{s}_l)$, $\psi(\mathbf{s}_l)$ and $\xi(\mathbf{s}_l)$ are obtained using the extRemes package in R (Gilleland et al. 2004). The estimated 20-year return levels are plotted in Figure 3. The most notable spatial pattern in return levels of Figure 3 occurs in the mid-western United States and along the Gulf of Mexico where the return levels are notably higher. Plots of the maximum likelihood estimates (MLEs) $\widehat{\mu}(\mathbf{s}_l)$, $\widehat{\sigma}(\mathbf{s}_l)$, $\widehat{\xi}(\mathbf{s}_l)$ and $\widehat{\psi}(\mathbf{s}_l)$ (not shown) display similar spatial patterns to those in Figure 3, suggesting that spatial structure could be built into any or all of these parameters.

To investigate possible temporal patterns, the location and scale parameters for the GEV model are extended to include a temporal effect for year, such that $\mu(\mathbf{s}_l)$ and $\sigma(\mathbf{s}_l)$ become $\mu_t(\mathbf{s}_l) = \beta_0(\mathbf{s}_l) + \beta_1(\mathbf{s}_l) \times t$ and $\log(\sigma_t(\mathbf{s}_l)) = \alpha_0(\mathbf{s}_l) + \alpha_1(\mathbf{s}_l) \times t$, respectively, where $t$ denotes the year. Again assuming independence across space, a distinct GEV distribution is fit at each location. It is found that, of the $\widehat{\beta}_1(\mathbf{s}_l)$, 13% are significantly positive and 5% significantly negative, while of the $\widehat{\alpha}_1(\mathbf{s}_l)$, 5% are significantly positive and 1% significantly negative (these tests are performed without multiplicity correction at the $5\%$ significance level and, thus, the overall Type I error rate is inflated). A similar analysis extending $\log(\psi)$ in the GPD distribution to include a temporal trend shows that 25% of the corresponding slope parameters are significant. Together, these results indicate that inclusion of a temporal effect in a statistical model for $WmSh$ may be beneficial, but that this temporal effect should vary across space.

As a final component of the exploratory analysis, consider the dependence between the parameters of the GEV and GPD. Strong positive dependence ($\rho = 0.69$) is exhibited between $\widehat{\mu}(\mathbf{s}_l)$ and $\widehat{\sigma}(\mathbf{s}_l)$, but negative dependence is found between $(\widehat{\mu}(\mathbf{s}_l), \widehat{\sigma}(\mathbf{s}_l))$ and $\widehat{\xi}$ with $\rho = -0.30$ and $\rho = -0.36$, respectively. The correlation between $\widehat{\xi}(\mathbf{s}_l)$ and $\widehat{\psi}(\mathbf{s}_l)$ in the GPD model is observed at $\rho = -0.55$. Thus, incorporating dependence between parameters in a statistical model may be necessary.

## 5. SPATIO-TEMPORAL MODELS

Given the results from the exploratory analysis in Section 4, this section proposes three statistical models for modeling extreme values of $WmSh$, in order of increasing complexity. Model 1 represents a parsimonious model for $WmSh$ with a crude spatial effect and no temporal ef-

fect. Model 2 builds on Model 1 through the use of Gaussian processes to account for the spatial correlation, and Model 3 then builds on Model 2 by introducing a temporal component.

Each model below is presented as a Bayesian hierarchical model and posterior inference is done via Markov chain Monte Carlo (MCMC) sampling. As shown by, among others, Cooley et al. (2007), Sang and Gelfand (2007), Cooley and Sain (2009), and Sang and Gelfand (2009), Bayesian hierarchical modeling for extremes has proven very insightful. Particularly, the use of Bayesian methods for model construction and fitting is preferred here because of the importance of accurately quantifying the uncertainty associated with each parameter. Additionally, after having obtained draws from the joint posterior distribution of model parameters, obtaining the marginal posterior distribution over quantities such as the $r$-year return level is straightforward.

### 5.1 Model 1

Let $u(\mathbf{s}_l)$ be the $95th$ quantile of $\{x_{td}(\mathbf{s}_l)\}_{t,d}$ and $y_{td}(\mathbf{s}_l) \equiv (x_{tk}(\mathbf{s}_l) - u(\mathbf{s}_l))_+$ be the exceedances over the threshold $u(\mathbf{s}_l)$. The first model uses the GPD likelihood given by (2), such that,

$$\begin{aligned} y_{td}(\mathbf{s}_l)|\psi(\mathbf{s}_l), \xi(\mathbf{s}_l) &\overset{iid}{\sim} \text{GPD}\left(\psi(\mathbf{s}_l), \xi(\mathbf{s}_l), u(\mathbf{s}_l)\right), \\ \log(\psi(\mathbf{s}_l)) &= \alpha_\psi + \mathbf{1}_{\mathbf{s}_l \in \mathcal{A}_\psi}\delta_\psi, \\ \xi(\mathbf{s}_l) &= \alpha_\xi + \mathbf{1}_{\mathbf{s}_l \in \mathcal{A}_\xi}\delta_\xi. \end{aligned} \tag{6}$$

Here, the regions $\mathcal{A}_\psi$ and $\mathcal{A}_\xi$ are areas of exceptional values of the GPD-model MLEs of $\psi$ and $\xi$, respectively. They roughly correspond to the mid-western United States and areas along the Gulf of Mexico. The region $\mathcal{A}_\psi$ consists of all locations $\mathbf{s}_l$ for which $\widehat{\psi}(\mathbf{s}_l) > 6.3$, and $\mathcal{A}_\xi$ is defined as all $\mathbf{s}_l$ between $-122^\circ$ and $-88^\circ$ longitude for which $\widehat{\xi}(\mathbf{s}_l) < -0.1$. For all models, the regions $\mathcal{A}_\psi$ and $\mathcal{A}_\xi$ are assumed known. Although the regions here are somewhat arbitrarily chosen, they are based on the band of higher $WmSh$ found in Figure 2. The variables $\delta_\psi$ and $\delta_\xi$ represent the added effects of being in these regions. Certainly, these regions could be estimated through the use of mixtures of distributions but this possibility is not considered here and is left as an open research question.

To complete the model specification, the prior distributions for the parameters in (6) are taken as $\alpha_\psi \sim N(5.5, 1)$, $\delta_\psi \sim \text{N}(0, 1)$, $\alpha_\xi \sim \text{N}(0, 0.2^2)$, $\delta_\xi \sim \text{N}(0, 0.2^2)$. The prior distributions for $\alpha_\psi$ and $\delta_\psi$ are relatively vague in an attempt to let the data inform the posterior distribution. The prior distributions for $\alpha_\xi$ and $\delta_\xi$ are more informative in order to constrain $\xi$ to reasonable values. Specifically, if $\xi \leqslant -0.5$, the corresponding GPD has a short bounded upper tail which does not seem to match intuition for $WmSh$. Additionally, Smith (1985) demonstrated that for $\xi \leqslant -0.5$, maximum likelihood estimators do not have the standard asymptotic properties. On the other

hand, if $\xi > 0.5$ then the GPD has a long, unbounded upper tail, which is also counter-intuitive for weather. Alternatively, the method of Coles and Tawn (1996) could have been used to develop informative prior distributions for these parameters, but because of the size of the data set any prior will be swamped by the likelihood, hence pursuing such a course was deemed unproductive.

The model given by (6) is, admittedly, a very crude spatial model and certainly does not capture all aspects of the physical process relating to $WmSh$. Nevertheless, (6) has the benefit of being a very parsimonious and easily-fit model which incorporates a clear-cut distinction between inside and outside of the $\mathcal{A}$ regions, as was observed in the exploratory analysis above.

### 5.2 Model 2

For Model 2, assume,

$$\begin{aligned} y_{td}(\mathbf{s}_l)|\psi(\mathbf{s}_l), \xi(\mathbf{s}_l) &\overset{iid}{\sim} \text{GPD}\left(\psi(\mathbf{s}_l), \xi(\mathbf{s}_l), u(\mathbf{s}_l)\right), \\ \log(\psi(\mathbf{s}_l)) &\sim \text{GP}(\mu_\psi, \tau_\psi^2, \phi_\psi), \\ \xi(\mathbf{s}_l) &= \alpha_\xi + \mathbf{1}_{\mathbf{s}_l \in \mathcal{A}_\xi}\delta_\xi. \end{aligned} \tag{7}$$

Here, $\text{GP}(\mu_\psi, \tau_\psi^2, \phi_\psi)$ is a stationary Gaussian process with constant mean $\mu_\psi$ and exponential covariance function

$$Cov(\log(\psi(\mathbf{s}_l)), \log(\psi(\mathbf{s}_k))) = \tau_\psi^2 \exp\{-\phi_\psi \|\mathbf{s}_l - \mathbf{s}_k\|\},$$

where $\|\cdot\|$ represents spherical distance (in miles). The prior distribution for $\mu_\psi$ is assumed to be non-informative such that $\mu_\psi \sim \text{Unif}(-\infty, \infty)$, because experience has shown that the mean of a Gaussian process is typically well-identified. Additionally, a non-informative prior allows the data to inform the posterior distribution. The parameters of the covariance function, $\tau_\psi^2$ and $\phi_\psi$, are more delicate in that these parameters typically require informative prior distributions to be estimated (see Zhang 2004). Thus, the prior distributions are taken as $\tau_\psi^2 \sim \text{IG}(2.1, 3)$ and $\phi_\psi \sim \text{Unif}(0.001, 0.1)$ where $\text{IG}(a, b)$ is the inverse gamma distribution with rate parameter $b$. Following Cooley et al. (2007), the exponential covariance function was assumed, but certainly other classes of covariance functions (e.g. Matern, Normal, etc.) could be used here.

The model given by (7) introduces an added complexity to Model 1 in that the spatial structure of $\psi(\mathbf{s})$ is captured through a Gaussian process. In this way, each location has its own scale parameter, but these scale parameters are correlated in space. Because the parameter $\xi(\mathbf{s})$ is difficult to estimate, it is unlikely that any spatial structure in $\xi(\mathbf{s})$ could be deciphered, and so $\xi(\mathbf{s})$ is left unchanged from Model 1. Indeed, this choice is further justified by Cooley et al. (2007), who fit a GP to $\xi(\mathbf{s})$ only to find that such a model was not preferable.

### 5.3 Model 3

In order to include a temporal effect, the PPE likelihood given by (4) is used for Model 3 such that,

$$\begin{aligned} x_{td}(\mathbf{s}_l)|x_{td}(\mathbf{s}_l) > u(\mathbf{s}_l), \beta_0(\mathbf{s}_l), \beta_1(\mathbf{s}_l), \sigma(\mathbf{s}_l), \xi(\mathbf{s}_l) &\overset{iid}{\sim} \text{PP}(\beta_0(\mathbf{s}_l) + \beta_1(\mathbf{s}_l) \times t, \sigma(\mathbf{s}_l), \xi(\mathbf{s}_l)), \\ (\beta_0(\mathbf{s}_l), \beta_1(\mathbf{s}_l), \log(\sigma(\mathbf{s}_l)))' &\sim \text{GP}_3(\boldsymbol{\mu}_{M_3}, \boldsymbol{\phi}_{M_3}, \boldsymbol{\Gamma}), \\ \xi(\mathbf{s}_l) &= \alpha_\xi + \mathbf{1}_{\mathbf{s}_l \in \mathcal{A}_\xi} \delta_\xi, \end{aligned} \tag{8}$$

where $\text{PP}(\mu, \sigma, \xi)$ denotes the likelihood given by (4), $t = 1, ;42$ is the year, $\text{GP}_3$ is a trivariate Gaussian process induced via coregionalization (Gelfand et al. 2004), $\boldsymbol{\mu}_{M_3} = (\mu_{\beta_0}, \mu_{\beta_1}, \mu_\sigma)'$, $\boldsymbol{\phi}_{M_3} = (\phi_1, \phi_2, \phi_3)'$, and $\boldsymbol{\Gamma}$ is a $3 \times 3$ lower triangular matrix with entries $\gamma_{ij}$. In more detail, Model 3 constructs a multivariate spatial process via,

$$\boldsymbol{\theta}(\mathbf{s}_l) = \begin{pmatrix} \beta_0(\mathbf{s}_l) \\ \beta_1(\mathbf{s}_l) \\ \log(\sigma(\mathbf{s}_l)) \end{pmatrix} = \begin{pmatrix} \mu_{\beta_0} \\ \mu_{\beta_1} \\ \mu_\sigma \end{pmatrix} + \begin{pmatrix} \gamma_{11} & 0 & 0 \\ \gamma_{21} & \gamma_{22} & 0 \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}, \tag{9}$$

where $w_1$, $w_2$, and $w_3$ are independent, zero-mean Gaussian process realizations with spatial decay parameters $\phi_1$, $\phi_2$, and $\phi_3$, respectively. Using this construction, the covariance between location $\mathbf{s}_l$ and $\mathbf{s}_k$ is then $\boldsymbol{\Gamma}\boldsymbol{R}(\mathbf{s}_l, \mathbf{s}_k)\boldsymbol{\Gamma}'$, where $\boldsymbol{R}(\mathbf{s}_l, \mathbf{s}_k)$ is diagonal with $i^{th}$ entry $\exp\{-\phi_i \|\mathbf{s}_l - \mathbf{s}_k\|\}$. The scale for the $w_i$ can be equal to 1, because the $\gamma_{ij}$ provide the scaling on the induced process. The implied joint distribution for $\boldsymbol{\theta} = (\boldsymbol{\theta}'(\mathbf{s}_1), \ldots, \boldsymbol{\theta}'(\mathbf{s}_L))'$ is Gaussian with mean $\boldsymbol{\mu}_{M_3}$ and $Cov(\boldsymbol{\theta}) = \sum_{i=1}^3 \boldsymbol{R}(\phi_i) \otimes \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i'$, where $\boldsymbol{R}(\phi_i)$ is now $L \times L$ with $ij^{th}$ entry $\exp\{-\phi_i \|\mathbf{s}_i - \mathbf{s}_j\|\}$, and $\boldsymbol{\gamma}_i$ is the $i^{th}$ column of $\boldsymbol{\Gamma}$. To complete the model specification, non-informative priors are used for $\mu_{\beta_0}$, $\mu_{\beta_1}$, and $\mu_\sigma$; $Unif(0.001, 0.1)$ priors are assigned to each $\phi_i$; and diffuse normal and inverse-gamma priors are used for the off-diagonal and diagonal elements of $\boldsymbol{\Gamma}$, respectively.

Model 3 is much more complex than either of Model 2 or Model 1 in that the temporal effect coefficients in the location parameter $\mu_t(\mathbf{s}_l) = \beta_0(\mathbf{s}_l) + \beta_1(\mathbf{s}_l) \times t$ are site specific, as was suggested by the exploratory analysis of Section 4. Additionally, the dependence structure of parameters mentioned in Section 4 is explicitly built into the model through the matrix $\boldsymbol{\Gamma}$.

## 6. RESULTS

For each model given above, the adaptive Metropolis algorithm of Haario et al. (2001) is used to obtain 5,000 draws from the posterior distribution of parameters after discarding an initial burn-in of 5,000 draws. Four chains were run for each model and Gelman-Rubin diagnostics (Gelman

and Rubin 1992) indicated that the chains converged with potential scale reduction factors of less than 1.2 for all parameters.

The deviance information criterion (DIC) (see Spiegelhalter et al. 2002) is a measure of model fit that balances goodness-of-fit with model complexity, where smaller values indicate the preferred model. Specifically, DIC $= \bar{D} + p_d$, with $\bar{D} = \mathbb{E}(D(\boldsymbol{\theta}))$ and $p_d = \bar{D} - D(\bar{\boldsymbol{\theta}})$, where $D(\boldsymbol{\theta}) = -2 \times \log -$likelihood is a measure of model fit, $p_d$ is the estimated number of effective parameters in the model, and $\boldsymbol{\theta}$ is a vector of all model parameters. Complex models incur a penalty for the number of parameters, and thus, under DIC, such models might not be preferred. Table 1 displays $\bar{D}$, $p_d$ and $DIC$ for the three models under consideration. According to the DIC, Model 3 is strongly preferred despite having a much higher number of effective parameters.

In addition to DIC, the three spatio-temporal models and a MLE model (i.e. each location was fit independently by maximizing a GPD likelihood at that location) are compared using several other model fit diagnostics based on the predictive distribution for $y_{td}(\mathbf{s}_l)$. Specifically, the models are compared using the predictive root mean square error (PRMSE), coverage probability of predictive intervals, and average width of predictive intervals for both inside and outside of $\mathcal{A}_\xi$. The PRMSE is defined as,

$$\text{PRMSE} = \sqrt{\frac{\sum_{t,d,l}(y_{td}(\mathbf{s}_l) - \tilde{y}_{td}(\mathbf{s}_l))^2}{N}},$$

where $N$ is the total number of exceedances and $\tilde{y}_{td}(\mathbf{s}_l) = \mathbb{E}(Y_{td}(\mathbf{s}_l) \mid \{y_{td}(\mathbf{s}_l)\})$ is the expected value of the predictive distribution for $y_{td}(\mathbf{s}_l)$. For the MLE model, the corresponding model fit diagnostics are calculated assuming the parameters are fixed at the MLEs. The results are also displayed in Table 1. In terms of predictive performance, the results are not as strongly in favor of Model 3 as is the case using DIC. For example, the PRMSE for Model 2 and 3 are essentially the same and only slightly better than the PRMSE obtained under Model 1. However, all the Bayesian models showed improvement over the MLE model. The average width of predictive intervals for Models 2 and 3 is smaller than the average width of predictive intervals for the MLE model and Model 1, especially within region $\mathcal{A}_\xi$.

Both Model 2 and Model 3 more rigorously exploit the spatial dependence in the data, and the results in Table 1 suggest that making use of this correlation is preferred over a less rigorous treatment. Choosing a preferred model between Model 2 and 3 is more difficult. According to DIC, Model 3 is preferred to Model 2, but in terms of predictive performance, the models perform similarly. This is perhaps a result of the large number of "non-practically significant" temporal trend terms in Model 3 (see below). With a small temporal trend term, Model 2 and 3 are very similar, which explains the similarity in predictive performance.

A final interesting comparison is for the estimates of $\xi$ and $\delta_\xi$. The final two rows of Table

1 display 95% credible intervals for $\alpha_\xi$ and $\alpha_\xi + \delta_\xi$ inside and outside of $\mathcal{A}_\xi$. For the MLE model, the interval is the $2.5^{th}$ and $97.5^{th}$ percentile of the MLE's in the corresponding region. Only Model 1 finds that $\alpha_\xi$ is positive indicating an infinitely long tail in $\mathcal{A}_\xi^C$. The other models agree with their estimates of a bounded tail for $WmSh$, which makes sense physically because of constraints preventing $W_{\max}$ and shear from being radically high at the same time.

The spatial pattern and intensity of $r$-year return levels are of practical interest as the return levels provide a measure of how extreme the weather can be over a set period of time. Draws from the posterior distribution $\pi(\text{RL}_r \mid \{y_{td}(\mathbf{s}_l)\}_{t,d,l})$ are obtained one-for-one from the draws of the posterior distribution via (5), where $\zeta_u = 0.05$, because $u(\mathbf{s}_l)$ is defined here as the $95^{th}$ percentile of $\{x_{td}(\mathbf{s}_l)\}$. Figure 4 displays the posterior means of $\text{RL}_{20}$ under Model 2 and 3 for the year 1999. The corresponding plot using the MLE model is shown in Figure 3. The corresponding plot using Model 1 is omitted as Model 2 and 3 are preferred to Model 1. Both models are in general agreement as Model 2 has posterior means of $3977.52$ and $2516.62$ for $\mathbf{s} \in \mathcal{A}_\xi$ and $\mathbf{s} \notin \mathcal{A}_\xi$, respectively, compared with $3994.56$ and $2510.69$ for Model 3.

The estimated temporal trend of extreme weather is displayed in Figure 5, which maps the posterior mean $\widehat{\beta}_1(\mathbf{s}_l)$ over the region of interest. Figure 6 displays the regions where a 95% credible interval for $\beta_1(\mathbf{s}_l)$ does not contain 0. In Figure 6, negative values denote locations for which the 95% credible interval for $\beta_1(\mathbf{s}_l)$ is strictly less than 0, while positive values denote locations where the credible interval is strictly greater than 0. Figures 5 and 6 indicate that the intensity of $WmSh$ is increasing for the majority of the continental United States and parts of the Pacific Ocean. However, these increases are only slight, as $\max_l(\widehat{\beta}_1(\mathbf{s}_l)) \approx 4.5$, which equates to a difference in 20-year return levels of approximately 70 meters per second between 1958 and 1999. Thus, while the yearly trend is positive, the observed increase may not be practically significant in terms of values of $WmSh$ associated with more devastating storms.

## 7. DISCUSSION AND CONCLUSIONS

This article compares a simple MLE extreme value model and three Bayesian hierarchical extreme value models on a spatio-temporal data set of $WmSh$. The results from comparing the models indicate that for the analysis of spatially-distributed extreme values, Bayesian modeling is clearly preferred over frequentist modeling, as it provides a natural way of quantifying uncertainty, and gives more accurate and precise results. For the reanalysis data set under consideration, incorporating spatial dependence on the parameters in the hierarchical models improves model fit and efficiency.

The results from the fitted models regarding an increase in intensity of extreme weather were statistically significant for the southern United States as well as the majority of the western United

States (see Figure 6). However, the temporal trend term on the mean of the extreme value distribution is not statistically different from zero for the northeastern part of the United States, southern Mexico, and the Caribbean islands. Significant negative temporal trends were found for most of the Atlantic ocean. Whether the temporal trend term is positive or negative, the mean degree of increase or decrease is small prompting the question of practical significance (see Figure 5) of the temporal trend.

When considering threshold exceedances, a common practice is to use declustering (see, e.g., Fawcett and Walshaw 2007) techniques to reduce the autocorrelation between observations such that the convergence in distribution to $H(y; \psi, \xi)$ holds. No declustering techniques were use for the reanalysis data used here but the sensitivity of the results to declustering is an active area of research.

While Model 3 is fairly elaborate and exploits spatial correlations for most of its parameters, the spatial dependence assumed for the shape parameter $\xi$ is very crude. As $\xi$ is generally regarded as being difficult to estimate, borrowing strength from nearby locations to estimate it makes intuitive sense. This could be done in a more subtle way via a Gaussian process (see, e.g., Cooley and Sain 2009), but this approach requires more elaborate modeling techniques such as penalized likelihood. A similar comment could be made for the threshold $u(\mathbf{s})$. In this paper, $u(\mathbf{s})$ is simply assumed to be the marginal $95^{th}$ percentile of the data at each location and, therefore, is not smooth across space. It is not clear what the optimal way of exploiting spatial dependencies for this purpose is, and so this question requires further research.

This report only considered modeling $WmSh$. An alternative approach would be to jointly model the variables $W_{\max} = \sqrt{2 \times CAPE}$ and $WS$ as a bivariate extreme. Developing a bivariate model has several challenges in that, while the theory underlying bivariate extremes is well-developed, standard approaches to bivariate modeling are less developed. The increase in dimensionality presents new challenges for model selection, fitting, and validation. Additionally, the strength of dependence between the two variables can, perhaps, diminish at the extreme levels. Additionally, atmospheric system constraints prevent $CAPE$ and $WS$ from becoming arbitrarily large simultaneously. Such system constraints would need to be considered for a bivariate approach. Despite these concerns, a bivariate approach, perhaps following the methods in Sang and Gelfand (2009) or Cooley et al. (2009), may prove useful in predicting and modeling extreme weather.

The data used in this paper are large-scale indicators of extreme weather. Thus, the models presented herein are not modeling extreme weather such as tornadoes and hurricanes directly. While high values of $WmSh$ are typically associated with these extreme weather events, further work (and data) is needed to quantify this relationship more precisely. The question has been partly addressed by Brooks et al. (2003), but the relationship between large-scale indicators and

extreme weather events is still open to further research and inquiry.

## ACKNOWLEDGEMENTS

The majority of this work was done at the 15th Annual Industrial and Mathematical Statistical Modeling (IMSM) Workshop supported by the Center for Research in Scientific Computation (CRSC) and the Statistical and Applied Mathematical Sciences Institute (SAMSI). Some support for this project was provided by the Weather and Climate Impacts Assessment Science Project, which is funded by the U.S. National Science Foundation. The authors wish to thank the organizers of the IMSM workshop. Specifically, Pierre Gremaud, Ilse Ipsen, and Ralph Smith as well as CRSC and SAMSI for financial support of the workshop.

## REFERENCES

Behrens, C. N., Lopes, H. F., and Gamerman, D. (2004), "Bayesian analysis of extreme events with threshold estimation," *Statistical Modelling*, 4, 227–244.

Brooks, H., Lee, J., and Craven, J. (2003), "The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data," *Atmospheric Research*, 67-68, 73–94.

Coles, S. G. (2001), *An Introduction to statistical modeling of extreme values*, Springer-Verlag.

Coles, S. G. and Tawn, J. A. (1996), "A Bayesian analysis of extreme rainfall data," *Applied Statistics*, 45, 463–478.

Cooley, D., Davis, R. A., and Naveau, P. (2009), "The Pairwise Beta Distribution: A Flexible Parametric Multivariate Model for Extremes," Under review.

Cooley, D., Nychka, D., and Naveau, P. (2007), "Bayesian spatial modeling of extreme precipitation return levels," *Journal of the American Statistical Association*, 102, 824–840.

Cooley, D. and Sain, S. R. (2009), "Spatial Hierarchical Modeling of Precipitation Extremes from a Regional Climate Model," Accepted to JABES.

Fawcett, L. and Walshaw, D. (2007), "Improved estimation for temporally clustered extremes," *Environmetrics*, 18, 173–188.

Gelfand, A. E., Schmidt, A. M., Banerjee, S., and Sirmans, C. F. (2004), "Nonstationary multivariate process modeling through spatially varying coregionalization (with discussion)," *Test*, 12, 1–50.

Gelman, A. and Rubin, D. (1992), "Inference from Iterative Simulation using Multiple Sequences," *Statistical Science*, 7, 457–511.

Gilleland, E., Katz, R., and Young, G. (2004), *extRemes: Extreme value toolkit.*, r package version 1.59.

Gilleland, E., Pocernich, M., Brooks, H. E., and Brown, B. G. (2008), "Large-scale indicators for severe weather," Proceedings of the American Statistical Association (ASA) Joint Statistical Meetings (JSM), 3-7 August 2008, Denver, Colorado.

Haario, H., Saksman, E., and Tamminen, J. (2001), "An Adaptive Metropolis Algorithm," *Bernoulli*, 2, 223–242.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., C.Ropelewski, Wang, J., Leetma, A., Reynolds, R., Jenne, R., and Joseph, D. (1996), "The NCEP/NCAR 40-Year Reanalysis Project," *Bulletin of the American Meteorological Society*, 77, 437 –471.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M. (2001), "The NCEP/NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation," *Bulletin of the American Meteorological Society*, 82, 247 –267.

Sang, H. and Gelfand, A. E. (2007), "Hierarchical modeling for extreme values observed over space and time," *Environmental and Ecological Statistics*, 16, 407–426.

— (2009), "Continuous Spatial Process Models for Spatial Extremes," Submitted to JABES.

Smith, R. L. (1985), "Maximum likelihood estimation in a class of non-regular cases," *Biometrika*, 72, 67–90.

— (1989), "Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-level Ozone (with discussion)," *Statistical Science*, 4, 367–393.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), "Bayesian measures of model complexity and fit (with discussion)," *Journal of the Royal Statistical Society, Series B*, 64, 583–639.

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

*Table 1. Model fit diagnostics. The diagnostics $\bar{D}$, $p_d$, and $DIC$ are omitted for the MLE model because these only pertain to Bayesian models.*

| Diagnostic | MLE (GPD) | Model 1 | Model 2 | Model 3 |
|---|---|---|---|---|
| $p_d$ | | 4.43 | 874.98 | 2133.45 |
| $\bar{D}$ | | $3.03 \times 10^6$ | $3.01 \times 10^6$ | $2.85 \times 10^6$ |
| DIC | | $3.03 \times 10^6$ | $3.01 \times 10^6$ | $2.86 \times 10^6$ |
| PRMSE | 394.47 | 323.92 | 313.31 | 313.29 |
| Coverage | 0.95 | 0.94 | 0.95 | 0.95 |
| Width of PI ($\mathcal{A}_\xi$) | 1294.19 | 1046.10 | 584.96 | 578.73 |
| Width of PI ($\mathcal{A}_\xi^C$) | 1368.19 | 1180.42 | 961.30 | 1028.30 |
| 95% Int. for $\alpha_\xi$ | $(-0.05, -0.03)$ | $(0.03, 0.04)$ | $(-0.03, -0.01)$ | $(-0.03, -0.01)$ |
| 95% Int. for $\alpha_\xi + \delta_\xi$ | $(-0.16, -0.14)$ | $(-0.06, -0.05)$ | $(-0.14, -0.13)$ | $(-0.15, -0.14)$ |

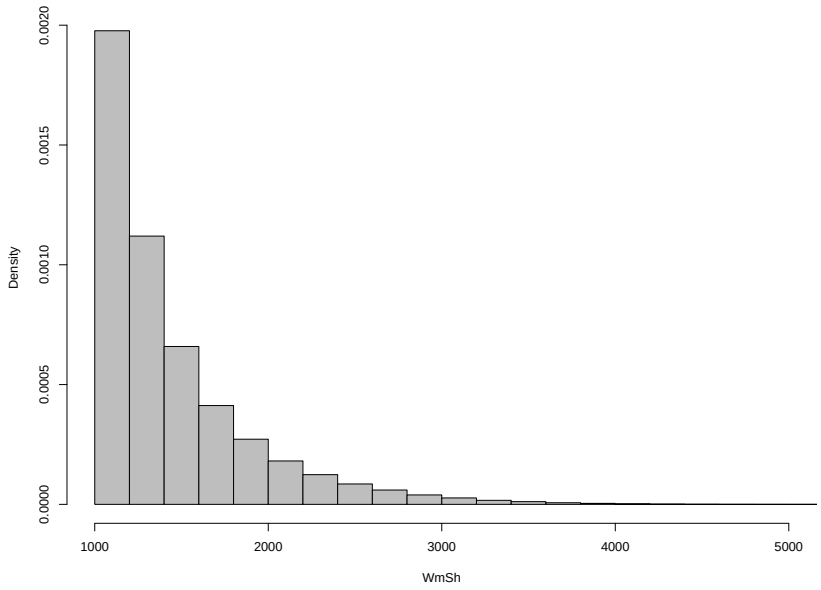


*Figure 1. Histogram of* $1000 < WmSh < 5500$ *for 1958 - 1999 across all locations in the reanalysis data set. Proper statistical models need to account for the heavy upper tail of this distribution. The 5500 upper bound was arbitrarily chosen to accentuate the features of the histogram.*

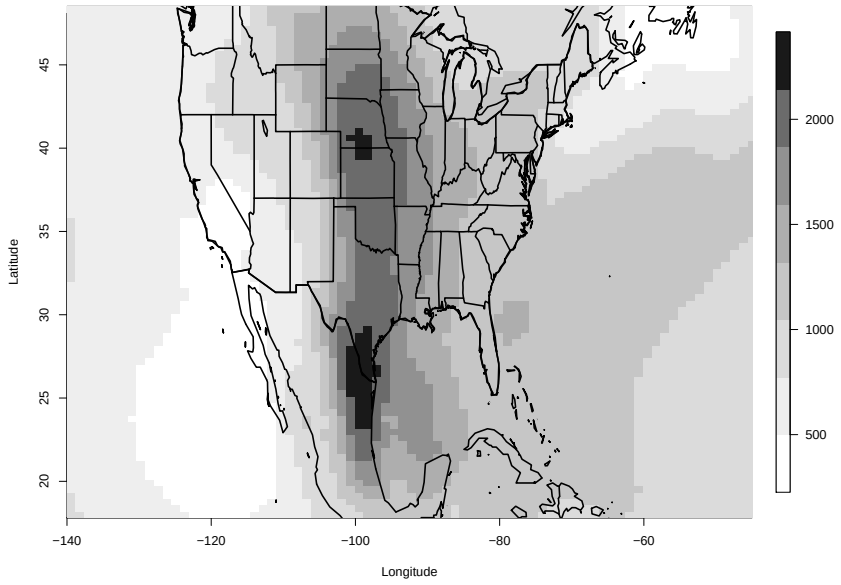


*Figure 2. Thresholds* $u(\mathbf{s})$ *above which* $WmSh$ *excesses are fit to the GPD. These are equivalent to the 95th percentile of* $WmSh$ *at each grid point.*

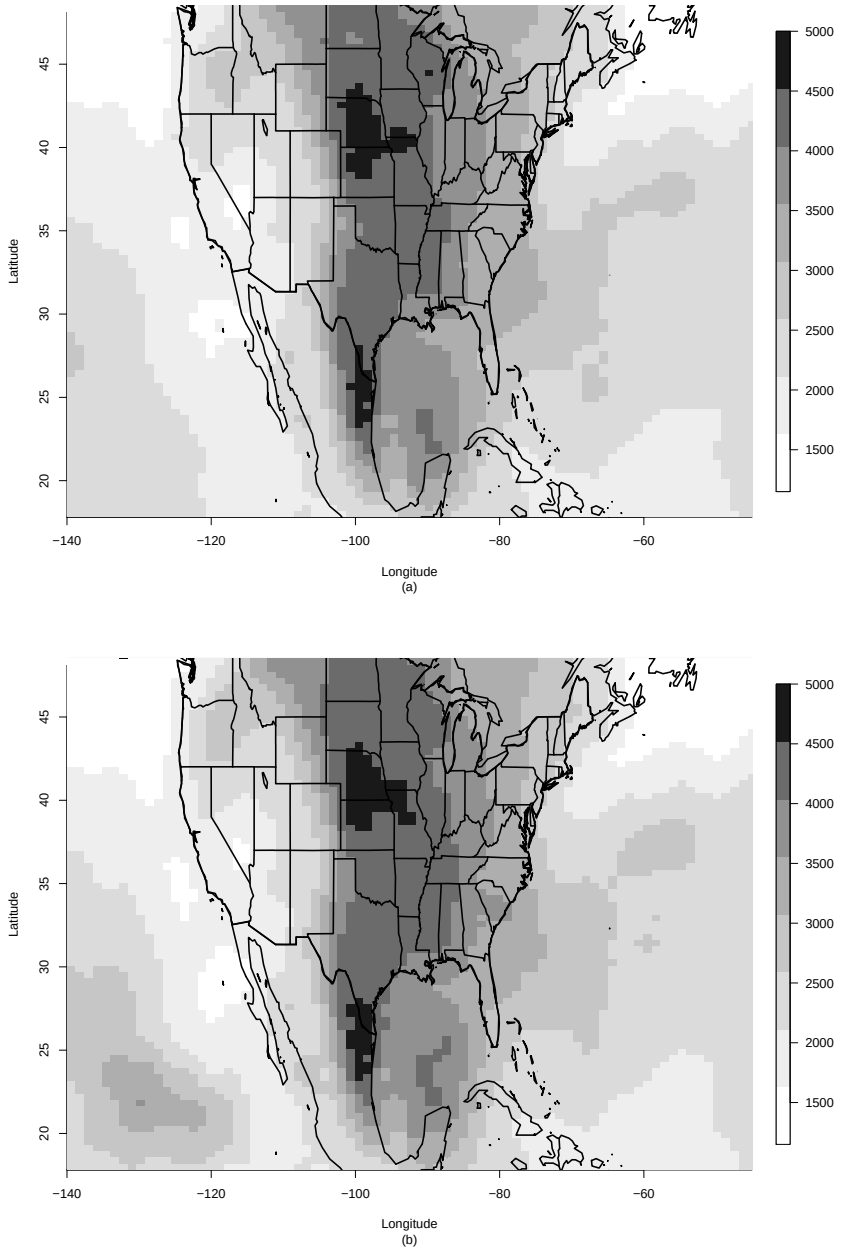


*Figure 3. Maximum likelihood estimates of 20-year return levels using (a) the GEV approach and (b) the GPD approach.*

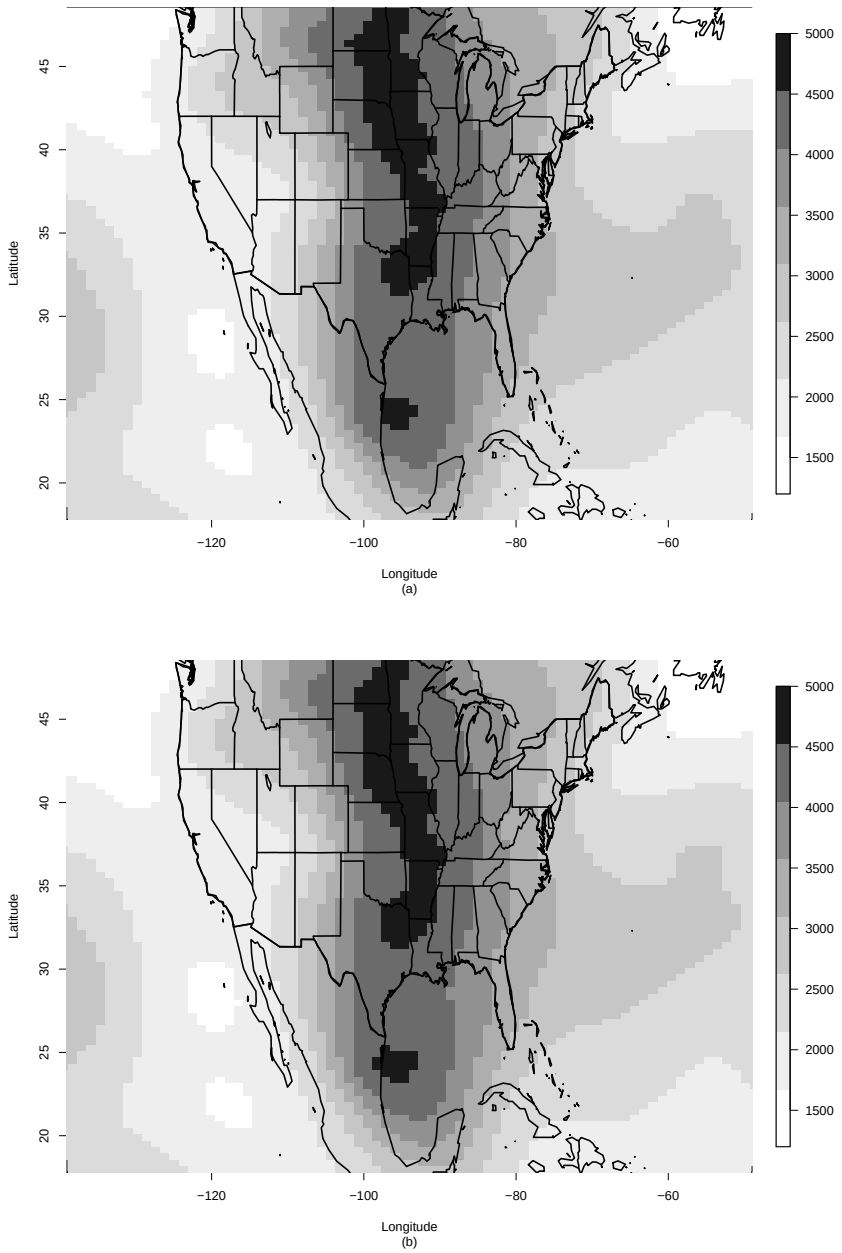


*Figure 4. Posterior mean of $RL_{20}$ for year 1999 as estimated using (a) Model 2 and (b) Model 3.*

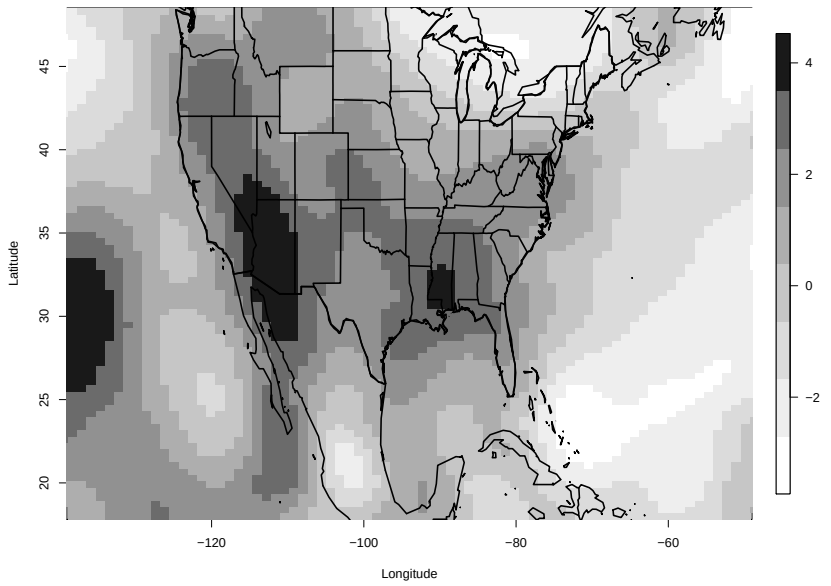


*Figure 5. Map of the temporal coefficient in Model 3 ($\widehat{\beta}_1(\mathbf{s}_l)$). Intensity of $WmSh$ seems to be increasing over time for the south western part of the U.S. as well as the Gulf Coast.*

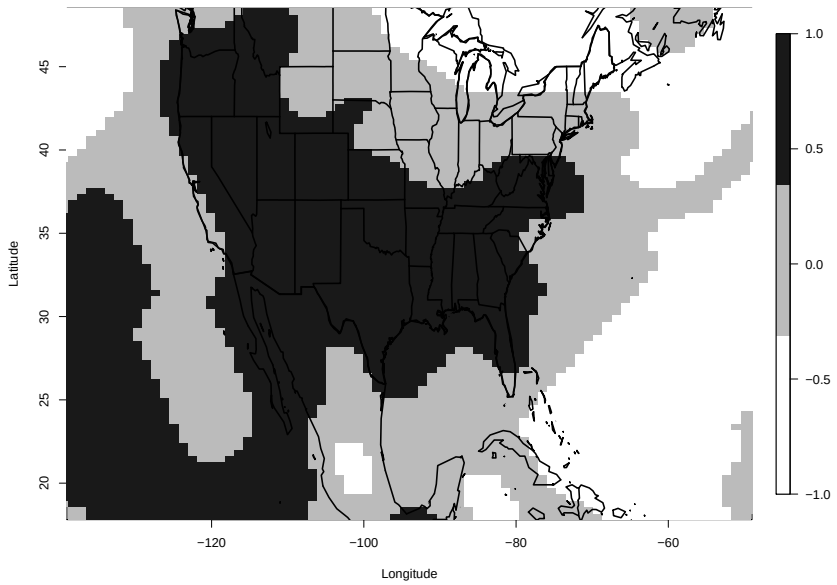


*Figure 6. Significance of the temporal coefficient in Model 3 ($\widehat{\beta}_1(\mathbf{s}_l)$). Values of 1 (-1) indicate a 95% credible interval for $\widehat{\beta}_1(\mathbf{s}_l)$ is strictly greater than (less than) 0 while values of 0 indicate non-significance.*