A finite-dimensional construction of a max-stable process for spatial extremes

Brian J. Reich[1], and **Benjamin A. Shaby**[2]

North Carolina State University Department of Statistics Technical Reports #2633

February 28, 2011

Abstract

From heat waves to hurricanes, often the environmental processes that are the most critical to understand probabilistically are extreme events. Such extremal processes manifestly exhibit spatial dependence. Max-stable processes are a class of asymptotically-justified models that are capable of representing spatial dependence among extreme values. While these models satisfy modeling requirements, they are limited in their utility because their corresponding joint likelihoods are unknown for more than a trivial number of spatial locations, preventing, in particular, Bayesian analyses. In this paper we propose an approximation to the Gaussian extreme value process (GEVP) that, critically, is amenable to standard MCMC and inclusion in hierarchical models. We show that this model is max-stable and approximates the GEVP arbitrarily well. The proposed model also leads to a non-stationary extension, which we use to analyze the yearly maximum temperature in the southeast US for years 1983–2007.

KEY WORDS: Bayesian hierarchical modeling, generalized extreme value distribution, max-stable process, non-stationarity.

[1]Department of Statistics, North Carolina State University
[2]Department of Statistical Science, Duke University

1 Introduction

Extreme weather events can have a dramatic effect of socio-economic well-being. For example, high temperatures can adversely affect energy consumption, agriculture, and public health. Statistical techniques are crucial to accurately quantifying the likelihood of extreme events and monitoring changes in their frequency and intensity. Extreme events are by definition rare, therefore estimation of local climate characteristics can be improved by borrowing strength across nearby locations. While methods for univariate extreme data are well-developed, modeling spatially-referenced extreme data is an active area of research.

Max-stable processes (de Haan and Ferreira (2006)) are the natural infinite-dimensional generalization of the univariate generalized extreme value (GEV) distribution. Just as the only limiting distribution of the maximum of independent univariate random variables is the GEV, the maximum of independent copies of any stochastic process can only converge to a max-stable process. Max-stable process models for spatial data may be constructed using the spectral representation of de Haan (1984). Max-stable processes were first used for spatial analysis by Smith (1990). Since then, a handful of subsequent spatial max-stable process models have been proposed, notably that of Schlather (2002) and Kabluchko et al. (2009), who proposed a more general construction that includes several other known models as special cases. Applications of spatial max-stable processes include Coles (1993), Buishand et al. (2008), and Blanchet and Davison (2011).

Because closed-form expressions for the likelihoods associated with spatial max-stable

processes are not available, parameter estimation and inference is problematic. Taking advantage of the availability of bivariate densities, Padoan et al. (2010) suggest maximum pairwise likelihood estimation and asymptotic inference based on a sandwich matrix (Godambe and Heyde (1987)). The problem of spatial prediction, conditional on observations, for max-stable random fields (analogous to Kriging for Gaussian processes) has proven difficult. The recent conditional sampling algorithm of Wang and Stoev (2010) is capable of producing both predictions and prediction standard errors for most spatial max-stable models of practical interest, subject to discretization errors that can be made arbitrarily small.

Bayesian estimation and inference for max-stable process models has been elusive. Approximate Bayesian methods based on asymptotic properties of the pairwise likelihood function are possible. Ribatet et al. (2010) use an estimated sandwich matrix to adjust the Metropolis ratio within an MCMC sampler, while Shaby (2010) rotates and scales the MCMC sample post-hoc. Bayesian models that are not based on max stable processes have been used for analysis of extreme values with spatial structure. Cooley et al. (2007) uses a hierarchical model with a conditionally-independent generalized Pareto likelihood, incorporating all spatial dependence through Gaussian process priors on the generalized Pareto likelihood parameters. Spatial dependence has also been achieved through Bayesian Gaussian copula models (Sang and Gelfand (2010)) and through a more flexible copula based on a Dirichlet process construction (Fuentes et al. (2010)).

In this paper, we propose a finite-dimensional construction for spatial extremes that approximates the Gaussian extreme value process (GEVP) of Smith (1990). We show that the resulting model is max-stable and approximates the GEVP arbitrarily well. Unlike proper max-stable process models that require the use of pairwise likelihood approximations, our finite-dimensional construction permits analysis of the joint distribution of all observations. In addition, because our construction specifies a joint model, it permits straightforward predictions at unobserved locations. By modeling extreme data as the sum of the true process and small-scale Gaussian errors, the model can be analyzed using standard MCMC methods. Also, the proposed framework models the spatial process using a lower-dimensional representation, which leads to efficient computing for large spatial data sets. Implementing the model in a fully-Bayesian framework allows for incorporation of prior information, natural uncertainty assessment for model parameters, and spatial prediction for unmeasured locations.

The finite-dimensional construction of a max-stable model is also conducive to modeling non-stationarity spatial data. The non-stationary extension we propose is also max-stable. Similar to the construction of a non-stationary Gaussian process of Higdon et al. (1999), our non-stationary max-stable process allows the range of spatial association to be a continuous spatial process, so that the degree of spatial association varies spatially.

The paper proceeds as follows. Section 2 proposes the finite-dimensional construction for stationary data. This model is extended to non-stationary case in Section 3. Computing

details are given in Section 4. In Section 5, we conduct a simulation study to examine the frequentist properties of the model and the effects of truncation. In Section 6, we use the proposed method to analyze yearly maximum temperature in the southeast US. Section 7 concludes.

2 A finite-dimensional construction for spatial extremes

Let $\gamma(\mathbf{s})$ be the extreme value at spatial location $\mathbf{s} \in \mathcal{D}$. The stochastic process $\gamma(\mathbf{s})$ is max-stable if there exist constants $a_n(\mathbf{s})$ and $b_n(\mathbf{s})$ so that the process

$$\gamma^*(\mathbf{s}) = \frac{\max\{\gamma^{(1)}(\mathbf{s}), ..., \gamma^{(n)}(\mathbf{s})\} - b_n(\mathbf{s})}{a_n(\mathbf{s})} \tag{1}$$

is identical in law to γ, where $\gamma^{(1)}(\mathbf{s}), ..., \gamma^{(n)}(\mathbf{s})$ are independent copies of the process. If $\mathcal{D}$ is a single point, then this defines the generalized extreme value distribution for $\gamma(\mathbf{s})$. If $\mathcal{D} = \{\mathbf{s}_1, ..., \mathbf{s}_N\}$, then this defines the multivariate extreme value distribution for the vector $[\gamma(\mathbf{s}_1), ..., \gamma(\mathbf{s}_N)]$.

There is no loss of generality in transforming $\gamma(\mathbf{s})$ to $\theta(\mathbf{s})$ with unit Fréchet marginals. Smith (1991) proposes the GEVP model

$$\theta(\mathbf{s}) = \max\{h_1 K(\mathbf{s}|\mathbf{u}_1, \Sigma), h_2 K(\mathbf{s}|\mathbf{u}_2, \Sigma), ...\}, \tag{2}$$

where $\{(h_1, \mathbf{u}_1), (h_2, \mathbf{u}_2), ...\}$ follows a Poisson process with intensity $\lambda(h, \mathbf{u}) = h^{-2} I(h > 0)$, and K is a kernel function standardized so that $\int K(\mathbf{s}|\mathbf{u}, \Sigma) d\mathbf{u} = 1$ for all $\mathbf{s}$. A useful analogy is to think of $\theta(\mathbf{s})$ as the maximum rainfall at site $\mathbf{s}$, generated as the maximum over a countably-infinite number storms. The k^{th} storm has center $\mathbf{u}_k \in \mathcal{R}^2$, intensity $h_k > 0$, and spatial range given by $K(\mathbf{s}|\mathbf{u}_k, \Sigma)$. A common choice for K is the anisotropic Gaussian density $K(\mathbf{s}|\mathbf{u}, \Sigma) = \frac{1}{2\pi}|\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{s}-\mathbf{u})^T \Sigma^{-1} (\mathbf{s}-\mathbf{u})\right]$. Under this model, the joint distribution at locations $\mathbf{s}_1, ..., \mathbf{s}_n$ is

$$\mathrm{P}\left[\theta(\mathbf{s}_1) < c_1, ..., \theta(\mathbf{s}_n) < c_n\right] = \exp\left[-\int \max_i \left\{\frac{K(\mathbf{s}_i|\mathbf{u}, \Sigma)}{c_i}\right\} d\mathbf{u}\right]. \tag{3}$$

In particular, $\theta(\mathbf{s})$ has the unit Fréchet marginal for all $\mathbf{s}$ so that $\mathrm{P}[\theta(\mathbf{s}) < c] = \exp(-1/c)$.

Using the transformation

$$\gamma(\mathbf{s}) = \mu + \frac{\sigma}{\xi}\left[\theta(\mathbf{s})^{\xi} - 1\right], \tag{4}$$

the marginal distribution of $\gamma(\mathbf{s})$ is the generalized extreme value (GEV) distribution with location μ, scale $\sigma > 0$, and shape ξ, denoted $\gamma(\mathbf{s}) \sim \mathrm{GEV}(\mu, \sigma, \xi)$. The GEV distribution has support

$$\gamma(\mathbf{s}) \in \begin{cases} (-\infty, \mu - \sigma/\xi) & \xi < 0 \\ (-\infty, \infty) & \xi = 0 \\ (\mu - \sigma/\xi, \infty) & \xi > 0. \end{cases} \tag{5}$$

Three types of extreme value distributions are special cases of the GEV: the Gumbel

($\xi \to 0$), Fréchet ($\xi > 0$), and Weibull ($\xi < 0$).

2.1 A finite approximation

We approximate this process by assuming the knots $\mathbf{u}_k$ are restricted to a set of $M < \infty$ points $\mathcal{G} = \{\mathbf{v}_1, ..., \mathbf{v}_M\}$ covering the spatial domain of interest. That is, we assume that $\{(h_1, \mathbf{u}_1), (h_2, \mathbf{u}_2), ...\}$ follows a Poisson process with intensity $\lambda(h, \mathbf{u}) = h^{-2} I(h > 0) I(\mathbf{u} \in \mathcal{G})$. Since $\max_k \{h_k K(\mathbf{s}|\mathbf{v}_j, \Sigma)\} = \max_k \{h_k\} K(\mathbf{s}|\mathbf{v}_j, \Sigma)$ for all $\mathbf{s}$, of the points with $\mathbf{u}_k = \mathbf{v}_j$, only the point with the largest h enters the maximum, which gives the approximation

$$\tilde{\theta}(\mathbf{s}) = \max\{z_1 K(\mathbf{s}|\mathbf{v}_1, \Sigma), ..., z_M K(\mathbf{s}|\mathbf{v}_M, \Sigma)\}, \tag{6}$$

of (2), where $z_j = \max\{h_k | \mathbf{u}_k = \mathbf{v}_j\}$. In contrast with $\theta(\mathbf{s})$ in (2), $\tilde{\theta}(\mathbf{s})$ is the maximum of a finite number of terms. Section 2.2 discusses selecting the knot locations $\mathbf{v}_j$.

Appendix A.1 shows that z_j, $j = 1, ..., M$, is unit Fréchet. Since this distribution does not depend on M, $\tilde{\theta}(\mathbf{s})$ is stochastically increasing with M because M terms enter the model. To ensure the appropriate marginals, we standardize the process as

$$\theta(\mathbf{s}) = w(\mathbf{s})^{-1} \max\left\{z_1 K(\mathbf{s}|\mathbf{v}_1, \Sigma), ..., z_M K(\mathbf{s}|\mathbf{v}_M, \Sigma)\right\}, \tag{7}$$

where $w(\mathbf{s}) = \sum_{j=1}^{M} K(\mathbf{s}|\mathbf{u}_j, \Sigma)$.

Using these weights, we ensure the desired marginal distributions, as stated in the

following theorem.

Theorem 1. *The marginal distribution of* $\theta(\mathbf{s})$ *is unit Fréchet for all* $\mathbf{s}$.

The proof is given in Appendix A.2. As a corollary, the transformation in (4) leads to GEV marginals for $\gamma(\mathbf{s})$.

In addition to matching the desired marginal distribution, the joint distribution has attractive properties.

Theorem 2. *The spatial process* $\theta(\mathbf{s})$ *is max-stable.*

Proof. Let $\theta^{(i)}(\mathbf{s}) = w(\mathbf{s})^{-1} \max\left\{z_1^{(i)} K(\mathbf{s}|\mathbf{v}_1, \Sigma), ..., z_M^{(i)} K(\mathbf{s}|\mathbf{v}_M, \Sigma)\right\}$, $i = 1, ..., n$, be independent replications of the process. Then

$$\begin{aligned} \theta^*(\mathbf{s}) &= \max\{\theta^{(1)}(\mathbf{s}), ..., \theta^{(n)}(\mathbf{s})\}/n \\ &= w(\mathbf{s})^{-1} \max\left\{z_1^* K(\mathbf{s}|\mathbf{v}_1, \Sigma), ..., z_M^* K(\mathbf{s}|\mathbf{v}_M, \Sigma)\right\}, \end{aligned} \tag{8}$$

where $z_j^* = \max\{z_j^{(1)}, ..., z_j^{(n)}\}/n$. Since the z_j^* are unit Fréchet, θ and θ^* have the same distribution. □

This gives a finite-dimensional construction of a max-stable spatial process with GEV marginals that is conducive to MCMC sampling, as discussed in Section 4. The joint distributions of the model approximate the GEVP model's joint distributions arbitrarily

well for large M. Appendix A.3 shows that the joint distribution function for the finite model is

$$\mathrm{P}\left[\theta(\mathbf{s}_1) < c_1, ..., \theta(\mathbf{s}_n) < c_n\right] = \exp\left[-\sum_{j=1}^{M} \max_i \left\{\frac{K(\mathbf{s}_i|\mathbf{v}_j, \Sigma)}{w(\mathbf{s}_i)c_i}\right\}\right]. \tag{9}$$

Clearly, this approaches (3) as $M \to \infty$.

Spatial dependence for extremes is often summarized using the extremal coefficient $\vartheta(\mathbf{s}, \mathbf{s}+\boldsymbol{\delta})$ (Smith (1990)), defined via the relationship

$$\mathrm{P}\left[\theta(\mathbf{s}) < c, \theta(\mathbf{s}+\boldsymbol{\delta}) < c\right] = \mathrm{P}\left[\theta(\mathbf{s}) < c\right]^{\vartheta(\mathbf{s}, \mathbf{s}+\boldsymbol{\delta})}. \tag{10}$$

If $\theta(\mathbf{s})$ and $\theta(\mathbf{s}+\boldsymbol{\delta})$ are completely dependent, $\vartheta(\mathbf{s}, \mathbf{s}+\boldsymbol{\delta}) = 1$, and if $\theta(\mathbf{s})$ and $\theta(\mathbf{s}+\boldsymbol{\delta})$ are independent, $\vartheta(\mathbf{s}, \mathbf{s}+\boldsymbol{\delta}) = 2$. Appendix A.3 shows that the extremal coefficient for the gridded process (7) is

$$\vartheta(\mathbf{s}, \mathbf{s}+\boldsymbol{\delta}) = \sum_{j=1}^{M} \max\left\{\frac{K(\mathbf{s}|\mathbf{v}_j, \Sigma)}{w(\mathbf{s})}, \frac{K(\mathbf{s}+\boldsymbol{\delta}|\mathbf{v}_j, \Sigma)}{w(\mathbf{s}+\boldsymbol{\delta})}\right\}. \tag{11}$$

Due to the finite grid approximation, the extremal coefficient is non-stationary. For example, $w(\mathbf{s})$ may not equal $w(\mathbf{s}+\boldsymbol{\delta})$ if $\mathbf{s}$ is close to a knot and $\mathbf{s}+\boldsymbol{\delta}$ is not. However, as M increases the extremal coefficient approaches the stationary extremal coefficient of the GEVP process.

Another quantity of interest in the analysis of spatial extremes is the maximum value

achieved over the entire spatial region of interest, that is, $\gamma_{(1)} = \max_{\mathbf{S}}\{\gamma(\mathbf{s})\}$. This has a convenient expression for the finite approximation. Assuming the M knots are on a regular grid with spacing Δ between each column and row, for large M, $w(\mathbf{s}) = \sum_{j=1}^{M} K(\mathbf{s}|\mathbf{u}_j, \Sigma) \approx \Delta^{-2}$, and thus $\max_{\mathbf{S}}\{K(\mathbf{s}|\mathbf{v}_j, \Sigma)/w(\mathbf{s})\} \approx \Delta^2 \max_{\mathbf{S}}\{K(\mathbf{s}|\mathbf{v}_j, \Sigma)\}$ for all j. The maximum of the Gaussian kernel is $\frac{1}{2\pi}|\Sigma|^{-1/2}$, achieved at $\mathbf{s} = \mathbf{v}_j$. Therefore, for the Gaussian kernel

$$\begin{aligned} \max_{\mathbf{s}}\{\theta(\mathbf{s})\} &= \max\left\{z_1 \frac{K(\mathbf{s}|\mathbf{v}_1, \Sigma)}{w(\mathbf{s})}, ..., z_M \frac{K(\mathbf{s}|\mathbf{v}_M, \Sigma)}{w(\mathbf{s})}\right\} \\ &\approx \frac{M\Delta^2}{2\pi}|\Sigma|^{-1/2}\frac{\max\{z_1, ..., z_M\}}{M} = \phi\frac{z_{(1)}}{M}, \end{aligned} \tag{12}$$

where $\phi = \frac{M\Delta^2}{2\pi}|\Sigma|^{-1/2}$ and $z_{(1)} = \max\{z_1, ..., z_M\}$. Therefore $\gamma_{(1)} \approx \tilde{\gamma}_{(1)} = \mu + \frac{\sigma}{\xi}\left[(\phi\frac{z_{(1)}}{M})^{\xi} - 1\right]$. For the finite construction, this approximation of $\gamma_{(1)}$ is the maximum of a finite number of model parameters, thus computing its posterior distribution is straight-forward using MCMC output. Also, since $z_{(1)}/M$ follows a unit Fréchet, $\tilde{\gamma}_{(1)} \sim \text{GEV}(\mu + \frac{\sigma}{\xi}[\phi^{\xi} - 1], \sigma\phi^{\xi}, \xi)$.

2.2 Knot selection

Figure 1a plots the extremal coefficient for a several sets of knots. In this example, the M knots are the $\sqrt{M} \times \sqrt{M}$ regular square grid covering $[-1, 1]^2$, the kernel is Gaussian with $\Sigma = \tau^2 I_2$, and the kernel bandwidth set to $\tau = 0.1$. The figure plots the extremal coefficient between the center $\mathbf{s} = (0, 0)^T$ and $\mathbf{s} + \boldsymbol{\delta} = (\delta, 0)^T$ for various distances δ and spacings between gridpoints $\Delta = 2/(\sqrt{M} - 1)$, along with the extremal coefficient

of the GEVP model, $2\Phi(\delta/0.2)$, where Φ is the standard normal distribution function. The extremal coefficient approaches the extremal coefficient of the GEVP model as the grid spacing decreases; the extremal coefficient with grid spacing $\Delta = 0.105$ provides a reasonable approximation, and the extremal coefficient with $\Delta = 0.079$ is nearly identical to the extremal coefficient of the exact model.

Figure 1b suggests a rule of thumb for selecting the number of knots required to approximate the GEVP process. This plot gives the smallest grid spacing Δ for each bandwidth τ so that the average absolute distance between the true and approximate extremal coefficients over the interval $\delta \in (0, 5\tau)$ is no larger than either 0.005, 0.01 or 0.02. This suggests that the grid spacing should be slightly less than the bandwidth to provide a good approximation. The number of knots required to satisfy this criteria is very large for small bandwidths. For example, a regular grid with spacing $\Delta = 0.01$ requires $M = 40,401$ to cover $[-1, 1]^2$. Therefore, an accurate approximation may be computationally challenging if the spatial domain is large relative to the spatial range of the process. The number of knots required drops quickly as the bandwidth increases.

Rather than selecting the knots to be fixed on a regular grid, an alternative is to assume the knots are unknown with a prior over the spatial domain of interest. We compare these approaches in Sections 5 and 6. We find that often random knot locations provide a better approximation with far fewer knots than the regular grid.

2.3 Measurement error model

In this section we describe a method for Bayesian inference using the approximate model proposed in Section 2.1. Given the GEV parameters μ, σ, and ξ and the kernel function, the process at n locations $\{\gamma(\mathbf{s}_1), ..., \gamma(\mathbf{s}_n)\}$ is completely determined by the intensities $\{z_1, ..., z_M\}$. Therefore, the likelihood for $\{\gamma(\mathbf{s}_1), ..., \gamma(\mathbf{s}_n)\}$ requires a map from $\{\gamma(\mathbf{s}_1), ..., \gamma(\mathbf{s}_n)\}$ to $\{z_1, ..., z_M\}$. This map may not exist, for example if $M < n$, and generally does not have a closed form. This is common in dimension reduction methods for Gaussian process models (for example, Higdon et al. (1999), Banerjee et al. (2008), and Cressie and Johannesson (2008)).

As with the Gaussian process dimension reduction methods, we model the observations as the sum of the true spatial process and small-scale error. Let $y(\mathbf{s}_i)$ be the observation at location $\mathbf{s}_i$, $i = 1, ..., n$. Then $y(\mathbf{s}_i) = \gamma(\mathbf{s}_i) + \varepsilon_i$, where $\varepsilon_i \overset{iid}{\sim} F$ for some error distribution F. Assuming Gaussian errors $\varepsilon_i \overset{iid}{\sim} \mathrm{N}(0, \sigma_e^2)$ permits the hierarchical model

$$\begin{aligned} y(\mathbf{s}) &\sim \mathrm{N}\left(\mu + \frac{\sigma}{\xi}\left[\theta(\mathbf{s})^{\xi} - 1\right], \sigma_e^2\right) \\ \theta(\mathbf{s}) &= w(\mathbf{s})^{-1} \max\left\{z_1 K(\mathbf{s}|\mathbf{v}_1, \Sigma), ..., z_M K(\mathbf{s}|\mathbf{v}_M, \Sigma)\right\} \\ z_j &\overset{iid}{\sim} \mathrm{GEV}(0, 1, 1). \end{aligned} \tag{13}$$

As described in Section 4, MCMC sampling is possible for this model.

In addition to computational benefits, this measurement error model may be more

attractive in many settings than directly modeling the responses using the GEVP model. This specification has the usual interpretation that $\gamma(\mathbf{s})$ is the true process with $y(\mathbf{s})$ a noisy measurement of $\gamma(\mathbf{s})$. Extreme data are often measured with error, and interest lies primarily in the true process $\gamma(\mathbf{s})$. For example, the temperature data analyzed in Section 6 is smooth over space but rounded to the nearest degree (F). We note that this separation of signal and noise is not possible in non-spatial extreme models without strong prior information about σ_e. However, in the presence of strong spatial association, it is possible to identify both $\gamma(\mathbf{s}_i)$ and ε_i by borrowing strength across nearby locations.

This model assumes the errors ε_i are additive. Extreme data are often collected as the maximum of several noisy observations, e.g., the yearly maximum is estimated as the sample maximum of 365 daily observations. This sampling design does not necessarily produce additive errors for extreme values, even if the daily errors are additive. However, if the variation in the true process is much larger than the measurement error variance, it is reasonable to assume that the errors are additive. In this case, the day that produces the largest true and observed values are likely to be the same, and thus additive daily error leads to additive error in the maximum.

To explore the conditions that lead to near independence between true process and the errors, we conduct a small simulation study. We generate the true daily values at days $t = 1, ..., n_t$ at a single location as either $\mu_t \overset{iid}{\sim}$ N(0,1) or $\mu_t \overset{iid}{\sim}$ Expo(1) so that $\text{Var}(\mu_t) = 1$, and the daily errors as $\epsilon_t \overset{iid}{\sim} \text{N}(0, \sigma^2)$. Figure 2 plots the distribution of the difference

between observed and true yearly maximum $\varepsilon = \max\{\mu_1+\epsilon_1, ..., \mu_{n_t}+\epsilon_{n_t}\} - \max\{\mu_1, ..., \mu_{n_t}\}$ by the quantiles of true yearly maximum $\theta = \max\{\mu_1, ..., \mu_{n_t}\}$ for 100,000 simulated data sets with $n_t = 100$ and various values of σ. The assumption of near independence between θ and ϵ is validated for $\sigma = 0.01$. The error distribution is very slightly affected by θ for $\sigma = 0.1$, and begins to erode with $\sigma = 0.25$, especially for small θ. Processes with small relative measurement error are common in environmental applications. Therefore, we pursue this hierarchical model for its modeling flexibility and computational convenience.

3 A non-stationary model for spatial extremes

The finite-grid approximation in Section 2 is computationally convenient. In addition, it provides a framework for capturing complicated spatial dependence structure, such as nonstationarity. In this section, we propose a nonstationary spatial model that maintains the interpretable GEV marginal distributions.

3.1 Infinite-dimensional model

As suggested by Figure 1a, the extremal coefficient in (11) is nonstationary. However, the nonstationarity is only an artifact of the gridding process and dissipates as the number of knots increases. Here we propose models with systematic nonstationarity by specifying a different bandwidth for each knot, and smoothing the bandwidth across space with a

second spatial prior. That is, we enhance the infinite-dimensional model (2) by assuming the kernel given the knot locations is $K[\mathbf{s}|\mathbf{u}, \Sigma(\mathbf{u})]$. For simplicity, we assume an isotropic model $\Sigma(\mathbf{u}) = \tau(\mathbf{u})^2 I_2$. The log-bandwidths $\eta(\mathbf{u}) = \log[\tau(\mathbf{u})]$ could be modelled as a low-dimensional function over $\mathbf{u}$, or as a Gaussian process.

In order to ensure unit Fréchet marginals, we must standardize the process as

$$\theta(\mathbf{s}) = w(\mathbf{s})^{-1} \max\{h_1 K[\mathbf{s}|\mathbf{u}_1, \Sigma(\mathbf{u}_1)], h_2 K[\mathbf{s}|\mathbf{u}_2, \Sigma(\mathbf{u}_2)], ...\}, \tag{14}$$

where $w(\mathbf{s}) = \int K[\mathbf{s}|\mathbf{u}, \Sigma(\mathbf{u})] d\mathbf{u}$. Using this construction, Appendix A.4 proves the following theorem.

Theorem 3. *For the non-stationary model defined by (14) , the marginal distribution of* $\theta(\mathbf{s})$ *is unit Fréchet for all* $\mathbf{s}$*.*

Appendix A.4 also shows that the extremal coefficient is

$$\vartheta(\mathbf{s}, \mathbf{s} + \boldsymbol{\delta}) = \int \max \left\{ \frac{K[\mathbf{s}|\mathbf{u}, \Sigma(\mathbf{u})]}{w(\mathbf{s})}, \frac{K[\mathbf{s} + \boldsymbol{\delta}|\mathbf{u}, \Sigma(\mathbf{u})]}{w(\mathbf{s} + \boldsymbol{\delta})} \right\} d\mathbf{u}. \tag{15}$$

Clearly, the extremal coefficient can vary spatially if the log-bandwidth process η varies throughout the spatial domain. Appendix A.4 also shows that the maximum of $\theta(\mathbf{s})$ over

the spatial domain $\mathcal{D}$ is Fréchet with scale

$$\phi(\mathcal{D}) = \int \max_{\mathbf{s}\in\mathcal{D}} \left\{ \frac{K(\mathbf{s}|\mathbf{u}, \Sigma(\mathbf{u}))}{w(\mathbf{s})} \right\} d\mathbf{u}. \tag{16}$$

Therefore, the maximum value of $\gamma(\mathbf{s})$ over $\mathcal{D}$ is GEV with location $\mu + \frac{\sigma}{\xi}[\phi(\mathcal{D})^{\xi} - 1]$, scale $\sigma\phi(\mathcal{D})^{\xi}$, and shape ξ.

3.2 Finite-dimensional construction

The approximation to the nonstationary model (14) is

$$\theta(\mathbf{s}) = w(\mathbf{s})^{-1} \max\left\{ z_1 K[\mathbf{s}|\mathbf{v}_1, \Sigma(\mathbf{v}_1)], ..., z_M K[\mathbf{s}|\mathbf{v}_M, \Sigma(\mathbf{v}_M)] \right\}, \tag{17}$$

where $w(\mathbf{s}) = \sum_{j=1}^{M} K[\mathbf{s}|\mathbf{v}_j, \Sigma(\mathbf{v}_j)]$ and $z_j \overset{iid}{\sim}$ GEV(0,1,1). Appendix A.2 shows that the marginal distribution of $\theta(\mathbf{s})$ remains unit Fréchet, and thus the marginal of $\gamma(\mathbf{s})$ in (4) remains GEV. Also, following (8), non-stationary process resulting form this finite-dimensional construction clearly remains max-stable. The extremal coefficient becomes

$$\vartheta(\mathbf{s}, \mathbf{s}+\boldsymbol{\delta}) = \sum_{j=1}^{M} \max\left\{ \frac{K[\mathbf{s}|\mathbf{v}_j, \Sigma(\mathbf{v}_j)]}{w(\mathbf{s})}, \frac{K[\mathbf{s}+\boldsymbol{\delta}|\mathbf{v}_j, \Sigma(\mathbf{v}_j)]}{w(\mathbf{s}+\boldsymbol{\delta})} \right\}. \tag{18}$$

Unlike the limiting extremal coefficient in Section 2, as the number of knots increases, this approaches (15) and thus remains non-stationary.

4 MCMC details

We perform MCMC sampling for the model in (13) using R (`http://www.r-project.org/`). Assuming conjugate priors, the full conditional distribution of μ, σ, and σ_e^2 are normal, truncated normal, and inverse gamma, respectively. These parameters are updated using Gibbs sampling. The shape parameter ξ and, assuming $\Sigma = \tau^2 I_2$, the bandwidth τ (or $\tau(\mathbf{v}_j)$ for the non-stationary model) do not have conjugate full conditionals and are updated using Metropolis sampling. For sampling purposes we transform to $\log(\tau)$.

The primary computational challenge is updating the latent z_j. This is challenging because z_j has a very heavy-tailed prior and complicated full conditional. The full conditional is non-standard because only the maximum of the $z_j K(\mathbf{s}|\mathbf{v}_j, \Sigma)$ enters the likelihood of $y(\mathbf{s})$. Therefore, the likelihood of $y(\mathbf{s})$ is flat in z_j for z_j less than the maximum of $z_k K(\mathbf{s}|\mathbf{v}_k, \Sigma)/K(\mathbf{s}|\mathbf{v}_j, \Sigma)$, $k \neq j$, and is non-differentiable at this point. The full conditional posterior includes the product of several such terms, resulting its complicated form.

Since the prior is heavy-tailed, we transform to $Z_j = \log(z_j)$. We use Metropolis-Hastings sampling for the Z_j. Usual random-walk Metropolis sampling was unsuccessful because the full conditional standard deviation varied considerably depending on the frequency that each term has a non-flat likelihood. Therefore, we use a normal candidate distribution with mean $Z_j^{(t-1)}$ and standard deviation $\varepsilon(Z_j^{(t-1)})$, where $Z_j^{(t-1)}$ is the value

of Z_j from the previous iteration. We select

$$\varepsilon(Z_j^{(t-1)}) = \begin{cases} e_1, & Z_j^{(t-1)} \leq 0 \\ e_2, & 0 < Z_j^{(t-1)} \leq 2 \\ e_3, & 2 < Z_j^{(t-1)} \leq 4 \\ e_4, & Z_j^{(t-1)} > 4, \end{cases} \tag{19}$$

where 0, 2, and 4 are roughly the 40^{th}, 90^{th}, and 98^{th} percentiles of Z_j's prior, respectively. The standard deviations have initial values $e_1 = 2$, $e_2 = 1$, $e_3 = 0.5$, and $e_4 = 0.1$ and are adaptively turned during the burn-in period to give acceptance rates near 0.4 for each level of $Z_j^{(t-1)}$. For models with random knot locations we take a similar approach for updating the knots, using Metropolis-Hastings sampling with a Gaussian candidate distribution centered on the previous knot location and with standard deviation depending on the current value of Z_j as in (19).

For the simulation study in Section 5 and data analysis in Section 6, we generate 30,000 samples and discard the first 10,000 samples as burn-in. Convergence is monitored using trace plots and autocorrelation plots for several representative parameters.

5 Simulation study

We conduct a simulation study to compare methods for selecting the knots $\mathbf{v}_j$ and to study the performance of the approximate method. Each simulated data set has 50 spatial locations drawn uniformly on $[0,1]^2$ and 25 independent (over replication, not space) replications. We simulate only stationary data due to the computational burden of the non-stationary model. The data are generated as

$$y_t(\mathbf{s}) \sim \mathrm{N}\left(\gamma_t(\mathbf{s}), \sigma_e^2\right) \quad \text{and} \quad \gamma_t(\mathbf{s}) = \mu + \frac{\sigma}{\xi}\left[\theta_t(\mathbf{s})^\xi - 1\right], \tag{20}$$

where $y_t(\mathbf{s})$ is the observation for replication t at location $\mathbf{s}$, and θ_t are stationary GEVP fields with Gaussian kernels and diagonal bandwidth matrix $\Sigma = \tau^2 I_2$, generated without approximation using the `R` package `SpatialExtremes`. We consider four simulation designs by varying the GEV parameters μ, σ, and ξ, the error variance σ_e^2, and the bandwidth τ:

1. $\mu = 2$, $\sigma = 3$, $\xi = 0.05$, $\sigma_e^2 = 1.0$, $\tau = 0.1$
2. $\mu = 2$, $\sigma = 3$, $\xi = -0.1$, $\sigma_e^2 = 1.0$, $\tau = 0.1$
3. $\mu = 2$, $\sigma = 3$, $\xi = -0.1$, $\sigma_e^2 = 1.0$, $\tau = 0.2$
4. $\mu = 2$, $\sigma = 3$, $\xi = -0.1$, $\sigma_e^2 = 5.0$, $\tau = 0.1$

We simulated 200 datasets from each of these four simulation designs. For each data set we compare five models. The first ("iid") ignores spatial correlation and assumes $y_t(\mathbf{s}) \overset{iid}{\sim}$

$\mathrm{N}(\gamma_t(\mathbf{s}), \sigma_e^2)$ and $\gamma_t(\mathbf{s}) \overset{iid}{\sim} \mathrm{GEV}(\mu, \sigma, \xi)$. The second ("Gaussian") accounts for spatial correlation using a Gaussian copula, $\gamma_t(\mathbf{s}) = G^{-1}\{\Phi[z_t(\mathbf{s})]\}$, where G is the distribution function of the $\mathrm{GEV}(\mu, \sigma, \xi)$ density, Φ is the standard normal distribution function, and $z_t(\mathbf{s})$ is a latent Gaussian process with mean zero, variance one, and Matérn spatial correlation function. The remaining three models are variations of the stationary approximate model for Section 2 with different choices for the knot locations. We consider the knots fixed on 7×7 ("Fixed49") and 10×10 ("Fixed100") equally-spaced grids covering $[-0.1, 1.1]^2$, as well as 50 free knots ("Free50") with a different set of knot locations for each replication and $\mathrm{Unif}([-0.1, 1.1]^2)$ prior for the knots locations.

For priors we choose $\mu \sim \mathrm{N}(0, 10^2)$, $\sigma \sim \mathrm{N}(0, 10^2)$ truncated on $\mathcal{R}^+$, $\xi \sim \mathrm{N}(0, 0.1^2)$, $\sigma_e^2 \sim$ InvGamma(0.1,0.1), and $\log(\tau) \sim$ N(-2,1) so that τ has prior median 0.14 and prior 95% interval (0.02,0.96). Also, for the Gaussian copula model, the Matérn correlation range ϕ and smoothness ν have priors $\log(\phi) \sim$ N(-2,1) and $\log(\nu) \sim$ N(1,1) to have prior median 2.7 and prior 95% interval (0.4,19) for ν.

We compare methods in terms of both parameter estimation and prediction. We split the 50 sites into 25 for training and 25 for testing. For parameter estimation, we fit the model using only the 25 training locations and report the mean squared error and coverage probability of 95% intervals for the GEV parameters μ, σ and ξ. For all parameters we use the posterior median as the estimate. For prediction, we report prediction mean squared error and coverage probabilities for $\gamma_t(\mathbf{s})$ at the testing locations. To compare methods

of estimating extreme values, we also report the average (over testing locations) posterior probability of $\gamma_t(\mathbf{s})$ greater than 10 (roughly the 92^{nd} percentile for $\xi = 0.05$ and the 85^{th} percentile for $\xi = -0.1$), separately for observations with true $\gamma_t(\mathbf{s})$ greater than ("P>10 true") and less than ("P>10 false") 10.

The results are given in Tables 1 and 2. In terms of both estimation of the GEV parameters and prediction, free knots are superior to fixed knots. Therefore, it appears that the additional flexibility of free knots justifies its computational complexity. The free knot model also generally outperforms the independent and Gaussian copula models. The difference between models is the largest for the GEV shape parameter, ξ. For the shape parameter the coverage probability is often below the nominal level for Gaussian copula model. For prediction, the free knot model has smaller mean squared error and better classification of values above and below 10 than the Gaussian copula. Prediction MSE, and MSE for GEV parameter estimation (especially for the independence model), are the largest for the final design with large measurement error. However, even with large measurement error, the free knot model has good coverage probabilities.

6 Analysis of southeast US maximum temperature

To illustrate the models in Sections 2 and 3, we analyze yearly maximum temperature in the southeast US for years 1983 to 2007. Our objective is to estimate the return levels at each location, along with the appropriate measure of uncertainty, while accounting for

spatial association between observations. The data are downloaded from the National Climate Data Center at `http://www7.ncdc.noaa.gov/CDO/cdo`. The data are rounded to the nearest degree (F), which is accounted for by the measurement error model in Section 2.3. We consider only stations with fewer than 25% missing observations. Figure 3 shows the locations of the $n = 268$ stations.

Figure 3 plots the MLE of the GEV parameters, computed separately at each site with no missing data (since other sites produced errors). There is considerable spatial variation in the location (median standard error is 0.15) and scale (median standard error is 0.41) estimates, with smaller location estimates in mountains of western North Carolina and in coastal regions, and smaller scale estimates in Florida and near the coast. There is less spatial variation in the shape parameter (median standard error is 0.15), and the spatial variation is small relative to the standard errors (Figure 3d). Therefore, we allow the location and scale parameters to vary spatially and assume that the shape is constant for all locations, i.e.,

$$y_t(\mathbf{s}) \sim \mathrm{N}(\gamma_t(\mathbf{s}), \sigma_e^2) \quad \text{and} \quad \gamma_t(\mathbf{s}) = \mu(\mathbf{s}) + \frac{\sigma(\mathbf{s})}{\xi}\left[\theta_t(\mathbf{s})^\xi - 1\right], \tag{21}$$

where $y_t(\mathbf{s})$ is the maximum temperature for year t and location $\mathbf{s}$, and θ_t are independent (over t) GEVP fields with unit Fréchet marginals.

The location process $\mu(\mathbf{s})$ is modeled as a Gaussian process with mean $\beta_0 + \beta_1 X(\mathbf{s})$,

where $X(\mathbf{s})$ is the elevation at location $\mathbf{s}$, and exponential covariance $\text{Cov}[\mu(\mathbf{s}), \mu(\mathbf{s}')] = \delta_1^2 \exp(-||\mathbf{s} - \mathbf{s}'||/\rho_1)$. The scale process is modelled as a zero-inflated spatial process. That is, we assume that $\sigma(\mathbf{s}) = \max\{0, \sigma^*(\mathbf{s})\}$, where $\sigma^*(\mathbf{s})$ is a Gaussian process with mean β_2 and covariance $\text{Cov}[\sigma^*(\mathbf{s}), \sigma^*(\mathbf{s}')] = \delta_2^2 \exp(-||\mathbf{s} - \mathbf{s}'||/\rho_2)$. Under this model, the full conditionals of $\sigma^*(\mathbf{s})$ at each location are conjugate, similar to the scale parameters in Section 4. For priors we assume that $\beta_j \sim \text{N}(0, 100^2)$, $\delta_j^2 \sim \text{InvG}(0.1,0.1)$, and, after scaling the spatial coordinates to $[0, 1]^2$, $\log(\rho_j) \sim \text{N}(\text{-}2,1)$, $l = 1, 2$. The remaining priors are the same as those in Section 5.

We compare several versions of the model by randomly (across both space and time) withholding 10% (n=633) of the observations as testing data. The models are fit to the remaining training data (n=5,646), and we compare predictive models using the same metrics as in Section 5. We compare eight stationary models with different choices for the knots: fixed regular grids covering $[-0.1, 1.1]^2$ with $M = 8^2, 10^2, ..., 18^2$ knots and free knots with $M = 50$ and $M = 100$ knots. For the free knot model the knots are allowed to be different each year and are given uniform priors on $[-0.1, 1.1]^2$. We also fit the nonstationary model with $M = 50$ free knots. We specify a cubic spatial model for the log bandwidth, i.e., $\log[\tau(\mathbf{v})] = b_0 + \mathbf{V}^T\mathbf{b}$, where $\mathbf{V}$ is a complete third-degree polynomial expansion of $\mathbf{v} = (v_1, v_2)^T$, $b_0 \sim \text{N}(\text{-}2,1)$, and the elements $\mathbf{b}$ have independent $\text{N}(0, \sigma_b^2)$ priors. The hypervariance $\sigma_b^2 \sim \text{InvGamma}(0.1,0.1)$.

The test set results are in Table 3. All models have nearly nominal predictive coverage

probability. Prediction mean squared error is the largest for the model with $M = 64$ knots, and is similar for models with at least 144 fixed knots or 50 free knots. Therefore, it appears 144 knots gives a sufficient approximation for these data, and the results are not sensitive to the number of knots above this threshold.

For these data, the non-stationary model does not improve mean squared error. However, it gives a small reduction in predictive variance and a slight improvement in the classification of temperature above 95 degrees. Figure 4 plots the posterior mean and standard deviation of the spatially-varying bandwidth $\tau(\mathbf{v})$. The bandwidth is much larger in southern part of the spatial domain, especially Florida. Compared to the standard deviation, the difference between the bandwidth in the north and south is statistically significant. The effect of the spatially-varying bandwidth can be seen in the fitted values, that is, the interpolated surface of $\gamma_t(\mathbf{s})$'s posterior mean, in Figure 5. The surfaces are smoother in the south than in the north. In both cases, the estimated mean surface is fairly smooth. The non-spatial error σ_e has posterior mean 1.45 and standard deviation 0.02.

Figure 4 also plots posterior mean of the GEV location and scale parameters. As expected, the location parameter is the lowest near the coast and in the high elevation regions in western North Carolina. The scale parameter is the smallest in the south, agreeing with the exploratory analysis in Figure 3. The GEV shape parameter is held constant over space, and has posterior mean and standard deviation -0.055 and 0.037, respectively. Figure 6 summarizes the posterior of the m-year return levels $RL(m) =$

$\mu(\mathbf{s}) + \sigma(\mathbf{s})\{[-\log(1-1/m)]^{-\xi} - 1\}/\xi$, i.e., the value so that $P[y < RL(m)] = 1/m$. As with the GEV location parameter, the largest 50-year return levels are in Georgia and South Carolina. There is considerable uncertainty in these estimates. As shown in Figure 6, for two locations in Georgia (Figure 6c) and Florida (Figure 6d), the posterior 95% intervals generally have width around 5 degrees (F) for 5-year return levels and slightly more than 5 degrees for 50-year return levels.

7 Discussion

In this paper we propose a new computational approach for spatial max-stable processes. We show via a simulation study that this model provides reliable inference for GEV parameters, as well as prediction for non-monitored locations. We also provide a non-stationary extension, and demonstrate its feasibility using a large temperature dataset.

In addition to nonstationarity, the proposed finite representation of the GEVP field opens the door for several exciting research directions. For example, we have used Gaussian kernels for all analyses. However, this may be overly-restrictive for heterogeneous processes such as precipitation, where the GEVP seems not to fit the data as well as the Schlather (2002) process, at least for the simple underlying marginal structures used in Blanchet and Davison (2011). We believe that the more flexible structures described here, made available through the use of hierarchical modeling, should mitigate the lack of fit seen in Blanchet and Davison (2011). To obtain a more flexible model for spatial association, it should be

possible to replace the Gaussian kernel with any other kernel that integrates to one, that is, any other two-dimensional density function. For large data sets, it may even be possible to estimate the kernel function nonparametrically from the data. Zheng et al. (2010) and Reich and Fuentes (2011) use Bayesian non-parametrics to estimate the spatial covariance function of a Gaussian process. This approach could be extended to the GEVP, using, say, a Dirichlet process mixture prior for the kernel function.

The methods proposed in this paper could also be extended to more complicated dependency structures. For example, we have ignored the temporal dependence in temperature data because the spatial association is far stronger than the temporal association for these data. However, using three-dimensional kernels (two for space, one for time) would give a feasible max-stable model for spatiotemporal data. Also, multivariate spatial data could be modeled by assuming different kernels and intensities for each process, and connecting the processes using a multivariate max-stable process for the intensities.

Acknowledgements

This work is partially supported by National Science Foundation grants DMS-0706731 (Reich) and DMS-0914906 (Shaby) as well as the Statistics and Applied Mathematical Sciences Institute (SAMSI). We also wish to acknowledge several helpful discussions with Richard Smith of the University of North Carolina - Chapel Hill and Alan Gelfand of Duke University.

Appendix A.1 - Proof that z is unit Fréchet

Let the intensities h follow a Poisson process with rate $\lambda(h) = \frac{1}{h^2} I(h > 0)$. Then $N_c = \sum_k I(h_k > c) \sim \text{Poisson}(1/c)$ since the Poisson rate on $[c, \infty]$ is $\int_c^\infty \frac{1}{h^2} dh = 1/c$. Therefore, $\text{P}(z < c) = \text{P}(N_c = 0) = \exp(-1/c)$ and z is unit Fréchet.

Appendix A.2 - Proof of Theorem 1

Proof. Denote $C_j(\mathbf{s}) = w(\mathbf{s})^{-1} K(\mathbf{s}|\mathbf{v}_j, \Sigma)$ and thus $\theta(\mathbf{s}) = \max\{z_1 C_1(\mathbf{s}), ..., z_M C_M(\mathbf{s})\}$. Then

$$\text{P}\left[\theta(\mathbf{s}) < c\right] = P\left[z_1 C_1(\mathbf{s}) < c, ..., z_M C_M(\mathbf{s}) < c\right] = \prod_{j=1}^{M} P\left[z_j < \frac{c}{C_j(\mathbf{s})}\right]. \tag{22}$$

Since $z_j \overset{iid}{\sim} \text{GEV}(0,1,1)$, $P\left[z_j < \frac{c}{C_j(\mathbf{s})}\right] = \exp\left(-\frac{C_j(\mathbf{s})}{c}\right)$ and

$$\text{P}\left[\theta(\mathbf{s}) < c\right] = \exp\left[-\frac{1}{c}\sum_{j=1}^{M} C_j(\mathbf{s})\right]. \tag{23}$$

Since $w(\mathbf{s}) = \sum_{j=1}^{M} K[\mathbf{s}|\mathbf{v}_j, \Sigma(\mathbf{v}_j)]$, $\sum_{j=1}^{M} C_j(\mathbf{s}) = 1$, and so $\text{P}\left[\theta(\mathbf{s}) < c\right] = \exp\left[-\frac{1}{c}\right]$, which is the unit Fréchet distribution function. □

This also applies to the non-stationary model of Section 3.1. For the non-stationary model, $C_j(\mathbf{s}) = w(\mathbf{s})^{-1} K[\mathbf{s}|\mathbf{v}_j, \Sigma(\mathbf{v}_j)]$ and $\sum_{j=1}^{M} C_j(\mathbf{s}) = 1$, as required by (23).

Appendix A.3 - Derivation of the joint distribution

The extremal coefficient is a function of joint distribution function

$$\begin{aligned}
\mathrm{P}\left[\theta(\mathbf{s}_1) < c_1, ..., \theta(\mathbf{s}_n) < c_n\right] &= \prod_{j=1}^{M} P\left[z_j < \min_i \left\{\frac{c_i}{C_j(\mathbf{s}_i)}\right\}\right] \\
&= \exp\left[-\sum_{j=1}^{M} \frac{1}{\min_i \left\{\frac{c_i}{C_j(\mathbf{s}_i)}\right\}}\right] \\
&= \exp\left[-\sum_{j=1}^{M} \max_i \left\{\frac{C_j(\mathbf{s}_i)}{c_i}\right\}\right].
\end{aligned}$$

In particular, for two observations at locations $\mathbf{s}$ and $\mathbf{s}+\boldsymbol{\delta}$,

$$\mathrm{P}\left[\theta(\mathbf{s}) < c, \theta(\mathbf{s}+\boldsymbol{\delta}) < c\right] = \exp\left[-\frac{1}{c}\sum_{j=1}^{M} \max\{C_j(\mathbf{s}), C_j(\mathbf{s}+\boldsymbol{\delta})\}\right].$$

Since θ has Fréchet marginals, $\mathrm{P}[\theta(\mathbf{s}) < c] = \mathrm{P}[\theta(\mathbf{s}+\boldsymbol{\delta}) < c] = \exp(-1/c)$ and the extremal coefficient is

$$\begin{aligned}
\vartheta(\mathbf{s}, \mathbf{s}+\boldsymbol{\delta}) &= \frac{\log\{\mathrm{P}\left[\theta(\mathbf{s}) < c, \theta(\mathbf{s}+\boldsymbol{\delta}) < c\right]\}}{\log\{\mathrm{P}\left[\theta(\mathbf{s}) < c\right]\}} \\
&= \frac{-\frac{1}{c}\sum_{j=1}^{M} \max\{C_j(\mathbf{s}), C_j(\mathbf{s}+\boldsymbol{\delta})\}}{-\frac{1}{c}} \\
&= \sum_{j=1}^{M} \max\{C_j(\mathbf{s}), C_j(\mathbf{s}+\boldsymbol{\delta})\}.
\end{aligned}$$

Appendix A.4 - Proof of Theorem 3

Proof. The integrated intensity on the set $\mathcal{B} = \{(h, \mathbf{u})|hK[\mathbf{s}|\mathbf{u}, \Sigma(\mathbf{u})] > w(\mathbf{s})c\}$ is

$$\int_{-\infty}^{\infty}\int_{w(\mathbf{s})c/K[\mathbf{s}|\mathbf{u},\Sigma(\mathbf{u})]}^{\infty}\frac{1}{h^2}dhd\mathbf{u} = \frac{1}{w(\mathbf{s})c}\int K[\mathbf{s}|\mathbf{u}, \Sigma(\mathbf{u})]d\mathbf{u} = \frac{1}{c}. \tag{24}$$

Therefore, $N_{\mathcal{B}}$, the number of events of the Poisson process $\{(h_1, \mathbf{u}_1), (h_2, \mathbf{u}_2)...\}$ on the set $\mathcal{B}$ is Poisson with mean $1/c$, and thus $P[\theta(\mathbf{s}) < c] = P[w(\mathbf{s})^{-1}\max\{h_1K(\mathbf{s}|\mathbf{u}_1, \Sigma), h_2K(\mathbf{s}|\mathbf{u}_2, \Sigma), ...\} < c] = P(N_{\mathcal{B}} = 0) = \exp(-1/c)$, the unit Fréchet distribution function. □

Similarly, the integrated intensity on the set

$$\mathcal{E} = \left\{(h, \mathbf{u}) \mid h\max\left\{\frac{K[\mathbf{s}|\mathbf{u}, \Sigma(\mathbf{u})]}{w(\mathbf{s})}, \frac{K[\mathbf{s}+\boldsymbol{\delta}|\mathbf{u}, \Sigma(\mathbf{u})]}{w(\mathbf{s}+\boldsymbol{\delta})}\right\} > c\right\}$$

is

$$\frac{1}{c}\int_{-\infty}^{\infty}\max\left\{\frac{K[\mathbf{s}|\mathbf{u}, \Sigma(\mathbf{u})]}{w(\mathbf{s})}, \frac{K[\mathbf{s}+\boldsymbol{\delta}|\mathbf{u}, \Sigma(\mathbf{u})]}{w(\mathbf{s}+\boldsymbol{\delta})}\right\}d\mathbf{u}. \tag{25}$$

Therefore, $N_{\mathcal{E}}$, the number of events of the Poisson process $\{(h_1, \mathbf{u}_1), (h_2, \mathbf{u}_2)...\}$ on the set $\mathcal{E}$ is Poisson with mean (27), and $P(\theta(\mathbf{s}) < c, \theta(\mathbf{s}+\boldsymbol{\delta}) < c) = P(N_{\mathcal{E}} = 0) =$

$\exp\left[-\frac{1}{c}\int_{-\infty}^{\infty}\max\left\{\frac{K[\mathbf{s}|\mathbf{u},\Sigma(\mathbf{u})]}{w(\mathbf{s})},\frac{K[\mathbf{s}+\boldsymbol{\delta}|\mathbf{u},\Sigma(\mathbf{u})]}{w(\mathbf{s}+\boldsymbol{\delta})}\right\}d\mathbf{u}\right]$. This gives the extremal coefficient

$$\begin{aligned}\vartheta(\mathbf{s},\mathbf{s}+\boldsymbol{\delta}) &= \frac{\log\{\mathrm{P}\left[\theta(\mathbf{s})<c,\theta(\mathbf{s}+\boldsymbol{\delta})<c\right]\}}{\log\{\mathrm{P}\left[\theta(\mathbf{s})<c\right]\}}\\ &= \frac{-\frac{1}{c}\int\max\left\{\frac{K[\mathbf{s}|\mathbf{u},\Sigma(\mathbf{u})]}{w(\mathbf{s})},\frac{K[\mathbf{s}+\boldsymbol{\delta}|\mathbf{u},\Sigma(\mathbf{u})]}{w(\mathbf{s}+\boldsymbol{\delta})}\right\}d\mathbf{u}}{-\frac{1}{c}}\\ &= \int\max\left\{\frac{K[\mathbf{s}|\mathbf{u},\Sigma(\mathbf{u})]}{w(\mathbf{s})},\frac{K[\mathbf{s}+\boldsymbol{\delta}|\mathbf{u},\Sigma(\mathbf{u})]}{w(\mathbf{s}+\boldsymbol{\delta})}\right\}d\mathbf{u}.\end{aligned}$$

Finally, to derive the distribution of the maximum of $\theta(\mathbf{s})$ over the spatial domain $\mathcal{D}$, that is,

$$\begin{aligned}\max_{\mathbf{s}\in\mathcal{D}}\theta(\mathbf{s}) &= \max_{\mathbf{s}\in\mathcal{D}}\max\left\{h_1\frac{K[\mathbf{s}|\mathbf{u}_1,\Sigma(\mathbf{u}_1)]}{w(\mathbf{s})},h_2\frac{K[\mathbf{s}|\mathbf{u}_2,\Sigma(\mathbf{u}_2)]}{w(\mathbf{s})},...\right\}\\ &= \max\left\{h_1\max_{\mathbf{s}\in\mathcal{D}}\frac{K[\mathbf{s}|\mathbf{u}_1,\Sigma(\mathbf{u}_1)]}{w(\mathbf{s})},h_2\max_{\mathbf{s}\in\mathcal{D}}\frac{K[\mathbf{s}|\mathbf{u}_2,\Sigma(\mathbf{u}_2)]}{w(\mathbf{s})},...\right\}\end{aligned} \tag{26}$$

let

$$\mathcal{F}=\left\{(h,\mathbf{u})\mid h\max_{\mathbf{s}\in\mathcal{D}}\left\{\frac{K[\mathbf{s}|\mathbf{u},\Sigma(\mathbf{u})]}{w(\mathbf{s})}\right\}>c\right\}.$$

The integrated intensity on this set is $\phi(\mathcal{D})/c$, where

$$\phi(\mathcal{D})=\int_{-\infty}^{\infty}\max_{\mathbf{s}\in\mathcal{D}}\left\{\frac{K(\mathbf{s}|\mathbf{u},\Sigma(\mathbf{u})}{w(\mathbf{s})}\right\}d\mathbf{u}. \tag{27}$$

Therefore, $N_{\mathcal{F}}$, the number of events of the Poisson process $\{(h_1,\mathbf{u}_1),(h_2,\mathbf{u}_2)...\}$ on the set $\mathcal{F}$ is Poisson with mean $\phi(\mathcal{D})/c$, and $P(\max_{\mathbf{s}\in\mathcal{D}}\theta(\mathbf{s})<c)=P(N_{\mathcal{F}}=0)=\exp\left[-\frac{\phi(\mathcal{D})}{c}\right]$,

the distribution function of a Fréchet with scale $\phi(\mathcal{D})$.

References

Banerjee, S., Gelfand, A., Finley, A. and Sang, H. (2008) Gaussian predictive process models for large spatial data sets. *J. Roy. Statist. Soc. Ser. B*, **70**, 825–848.

Blanchet, J. and Davison, A. C. (2011) Spatial modelling of extreme snow depth. *Ann. Appl. Stat.* Submitted.

Buishand, T. A., de Haan, L. and Zhou, C. (2008) On spatial extremes: with application to a rainfall problem. *Ann. Appl. Stat.*, **2**, 624–642.

Coles, S. G. (1993) Regional modelling of extreme storms via max-stable processes. *J. Roy. Statist. Soc. Ser. B*, **55**, 797–816.

Cooley, D., Nychka, D. and Naveau, P. (2007) Bayesian spatial modeling of extreme precipitation return levels. *J. Amer. Statist. Assoc.*, **102**, 824–840.

Cressie, N. and Johannesson, G. (2008) Fixed rank kriging for large spatial datasets. *J. Roy. Statist. Soc. Ser. B*, **70**, 209–226.

Fuentes, M., Henry, J. and Reich, B. (2010) Nonparametric spatial models for extremes: Applications to exterme temperature data. *Extremes*. Submitted.

Godambe, V. P. and Heyde, C. C. (1987) Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.*, **55**, 231–244.

de Haan, L. (1984) A spectral representation for max-stable processes. *Ann. Probab.*, **12**, 1194–1204.

de Haan, L. and Ferreira, A. (2006) *Extreme value theory*. Springer Series in Operations Research and Financial Engineering. New York: Springer. An introduction.

Higdon, D., Swall, J. and Kern, J. (1999) Non-Stationary Spatial Modeling. In *Bayesian Statistics 6 - Proceedings of the Sixth Valencia Meeting*, 761–768. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, (editors). Clarendon Press - Oxford.

Kabluchko, Z., Schlather, M. and de Haan, L. (2009) Stationary max-stable fields associated to negative definite functions. *Ann. Probab.*, **37**, 2042–2065.

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

Reich, B. J. and Fuentes, M. (2011) Nonparametric Bayesian models for a spatial covariance. *Statistical Methodology*. Accepted.

Ribatet, M., Cooley, D. and Davison, A. (2010) Bayesian inference from composite likelihoods, with an application to spatial extremes. *Extremes*. Submitted.

Sang, H. and Gelfand, A. (2010) Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics*, **15**, 49–65.

Schlather, M. (2002) Models for stationary max-stable random fields. *Extremes*, **5**, 33–44.

Shaby, B. (2010) The open-faced sandwich adjustment for estimating function-based MCMC. Submitted.

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

Wang, Y. and Stoev, S. (2010) Conditional sampling for max-stable random fields. *Arxiv preprint arXiv:1005.0312*.

Zheng, Y., Zhu, J. and Roy, A. (2010) Nonparametric Bayesian inference for the spectral density function of a random field. *Biometrika*, **97**, 238–245.

Table 1: Simulation study results for GEV parameter estimation. Coverage probabilities are the 95% posterior intervals, all entries are multiplied by 100, and standard errors are given in parentheses.

		Mean squared error			Coverage probability		
Design	Model	μ	σ	ξ	μ	σ	ξ
1	IID	5.19 (0.48)	4.20 (0.45)	0.31 (0.04)	79.0	92.0	86.0
	Gaussian	6.22 (0.54)	4.70 (0.41)	0.62 (0.04)	95.0	90.5	60.0
	Fixed49	7.24 (0.62)	11.93 (0.78)	0.51 (0.06)	87.5	68.0	71.5
	Fixed100	5.46 (0.50)	6.02 (0.50)	0.37 (0.04)	93.5	83.0	76.0
	Free50	5.11 (0.44)	3.09 (0.28)	0.14 (0.02)	95.5	95.0	92.5
2	IID	6.49 (0.56)	10.09 (1.50)	0.20 (0.02)	75.0	91.0	91.5
	Gaussian	6.12 (0.55)	3.98 (0.40)	0.30 (0.02)	96.5	89.0	72.5
	Fixed49	6.36 (0.56)	7.41 (0.56)	0.34 (0.03)	92.5	71.5	76.5
	Fixed100	5.88 (0.53)	3.56 (0.32)	0.21 (0.02)	92.5	89.0	83.5
	Free50	5.03 (0.40)	2.71 (0.26)	0.11 (0.01)	95.0	94.5	96.0
3	IID	19.53 (2.60)	30.22 (5.66)	0.43 (0.06)	46.0	75.5	85.0
	Gaussian	18.69 (4.21)	11.24 (1.32)	1.09 (0.05)	92.5	78.5	23.5
	Fixed49	12.02 (1.15)	4.12 (0.43)	0.27 (0.04)	92.5	92.1	80.5
	Fixed100	11.78 (1.15)	4.64 (0.56)	0.32 (0.03)	91.5	91.0	75.5
	Free50	9.76 (0.86)	4.26 (0.44)	0.13 (0.01)	95.0	95.0	95.0
4	IID	32.73 (3.03)	233.06 (18.58)	0.56 (0.04)	91.5	88.5	94.5
	Gaussian	12.78 (1.35)	30.56 (5.03)	0.61 (0.05)	96.0	97.0	93.5
	Fixed49	12.17 (1.38)	22.96 (2.28)	0.47 (0.04)	95.0	95.0	99.0
	Fixed100	12.42 (1.38)	20.57 (1.97)	0.44 (0.04)	95.5	96.5	98.5
	Free50	13.10 (1.38)	21.12 (2.77)	0.42 (0.03)	95.0	96.5	99.0

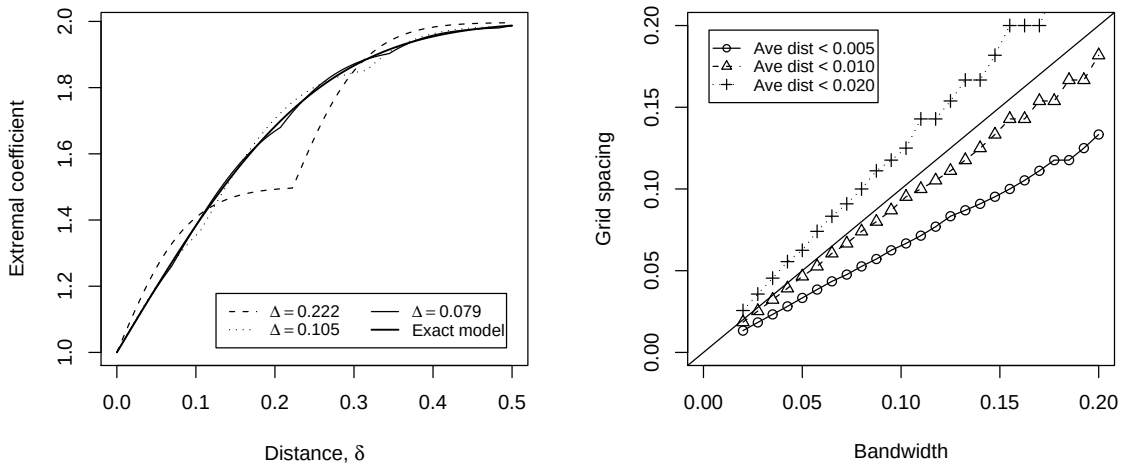
Table 2: Prediction MSE (standard error), coverage probabilities of 90% intervals, and the average posterior probability of greater than 10 for testing observations greater than ("P>10 true") and less than ("P>10 false") 10 for the simulation study. All entries are multiplied by 100.

Design	Model	MSE	Coverage	P>10 true	P>10 false
1	Gaussian	522 (8)	88.3 (0.2)	54.6 (0.6)	3.4 (0.1)
	Fixed49	553 (8)	81.8 (0.2)	57.3 (0.6)	2.9 (0.1)
	Fixed100	498 (8)	85.6 (0.2)	61.2 (0.7)	2.8 (0.1)
	Free50	449 (7)	90.1 (0.1)	66.5 (0.6)	2.9 (0.1)
2	Gaussian	415 (5)	88.8 (0.2)	49.5 (0.8)	2.1 (0.1)
	Fixed49	430 (5)	82.7 (0.2)	53.1 (0.8)	1.6 (0.1)
	Fixed100	395 (5)	86.4 (0.2)	59.0 (0.8)	1.6 (0.1)
	Free50	374 (4)	89.7 (0.1)	63.3 (0.7)	1.7 (0.1)
3	Gaussian	152 (3)	88.7 (0.2)	70.7 (0.9)	1.1 (0.1)
	Fixed49	147 (3)	83.4 (0.2)	75.0 (0.9)	0.8 (0.1)
	Fixed100	137 (3)	85.7 (0.2)	77.0 (0.9)	0.8 (0.1)
	Free50	128 (3)	89.5 (0.2)	80.2 (0.8)	0.7 (0.1)
4	Gaussian	799 (6)	87.8 (0.4)	21.9 (0.7)	3.7 (0.1)
	Fixed49	797 (6)	86.7 (0.4)	25.3 (0.8)	3.2 (0.1)
	Fixed100	793 (6)	87.9 (0.4)	26.1 (0.8)	3.4 (0.1)
	Free50	792 (6)	88.3 (0.5)	26.0 (0.8)	3.5 (0.1)

Table 3: Test-set results for the temperature data. The entries are prediction MSE ("MSE"), average predictive variance ("VAR"), coverage probabilities of 90% intervals ("Coverage"), the average posterior predictive probability of temperature greater than 95 degrees (F), separately for testing observations greater than ("P>95 true") and less than ("P>95 false") 95 degrees (F), and the same average probabilities for 100 degrees (F). All entries are multiplied by 100.

Knots	M	Stationary	MSE	VAR	Coverage	P>95 true	P>95 false	P>100 true	P>100 false
Fixed	64	Yes	246	255	90.7	91.5	20.3	71.2	9.1
Fixed	100	Yes	241	252	91.6	91.5	20.0	71.5	8.9
Fixed	144	Yes	237	250	91.5	91.7	19.7	72.4	9.0
Fixed	196	Yes	235	251	91.9	91.6	19.6	73.0	8.9
Fixed	256	Yes	237	248	91.3	91.7	19.6	72.4	8.8
Fixed	324	Yes	234	250	92.1	91.6	19.6	72.3	8.9
Free	50	Yes	231	245	91.7	91.7	19.4	72.9	8.7
Free	100	Yes	231	246	91.9	91.7	19.5	72.9	8.7
Free	50	No	231	242	91.8	91.9	19.2	72.8	8.7

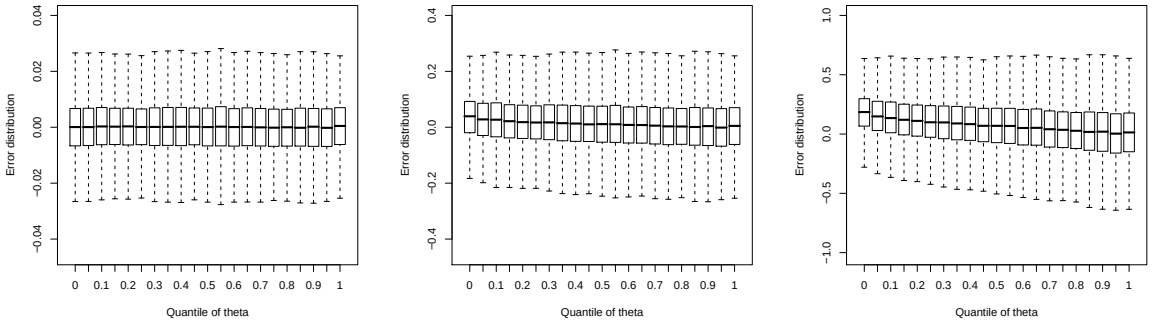
Figure 1: Extremal coefficient for the true and approximate model with a $\sqrt{M} \times \sqrt{M}$ grid of knots with grid spacing Δ covering $[-1, 1]^2$. Panel (a) plots extremal coefficient with bandwidth fixed at 0.1, and Panel (b) plots the largest Δ so that the average distance between the true and approximate extremal coefficient is no larger than thresholds 0.005, 0.01, or 0.02.



(a) Extremal coefficient with bandwidth 0.1 (b) Required grid spacing by bandwidth

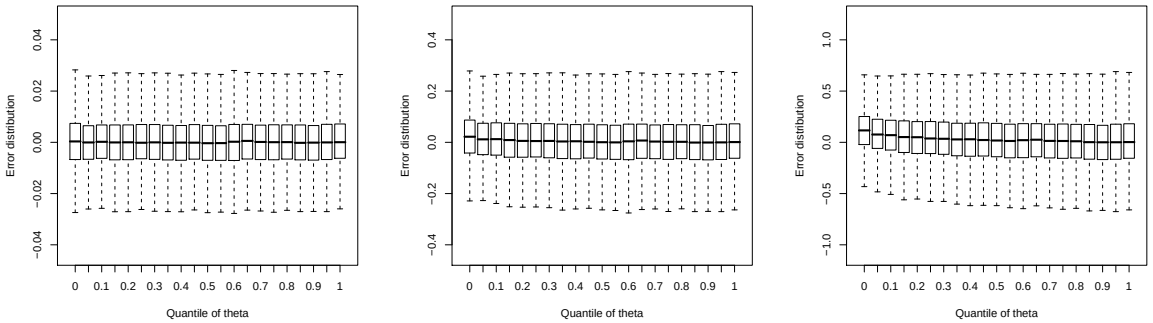
Figure 2: Conditional distribution of the difference between maximum observed and maximum true values (ε) given the maximum true value (θ). Note that the y-axis varies by plot.

(a) $\mu_t \overset{iid}{\sim}$ N(0,1)



$\sigma = 0.01$ $\sigma = 0.10$ $\sigma = 0.25$

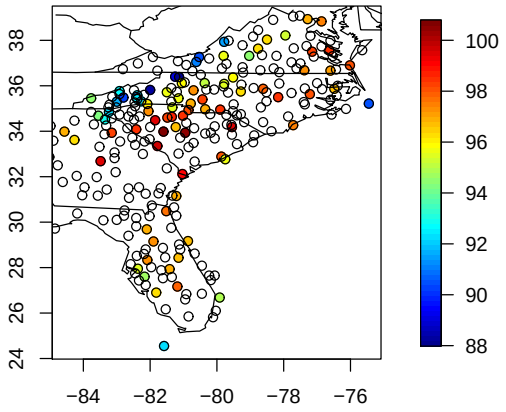
(b) $\mu_t \overset{iid}{\sim}$ Expo(1)



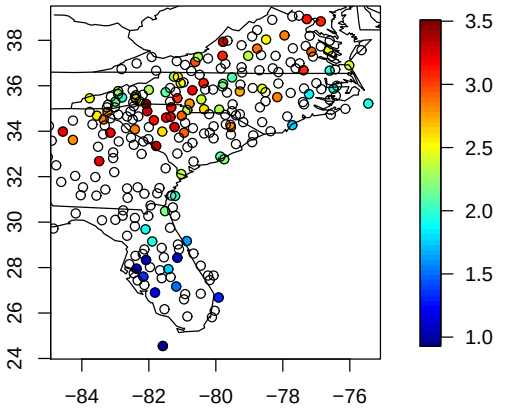
$\sigma = 0.01$ $\sigma = 0.10$ $\sigma = 0.25$

Figure 3: Exploratory analysis of the maximum temperature data. Panels (a) - (c) plot the maximum likelihood estimates of the GEV parameters, computed separately at each station. Empty circles represent stations with missing data, for which the MLEs are not computed. Panel (d) gives the MLE ± twice the standard error for each site for the GEV shape parameter.

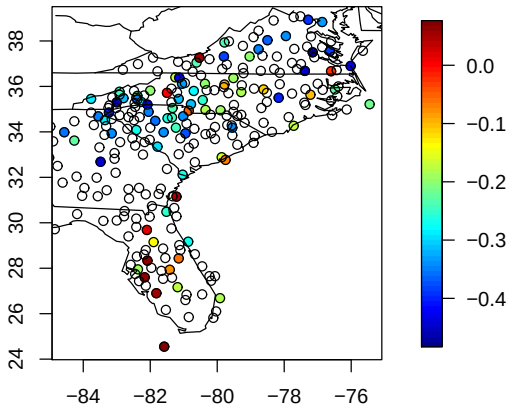
(a) MLE of the GEV location, $\hat{\mu}(\mathbf{s})$



(b) MLE of the GEV scale, $\hat{\sigma}(\mathbf{s})$



(c) MLE of the GEV shape, $\hat{\xi}(\mathbf{s})$



(d) 95% intervals for the GEV shape, $\hat{\xi}(\mathbf{s})$

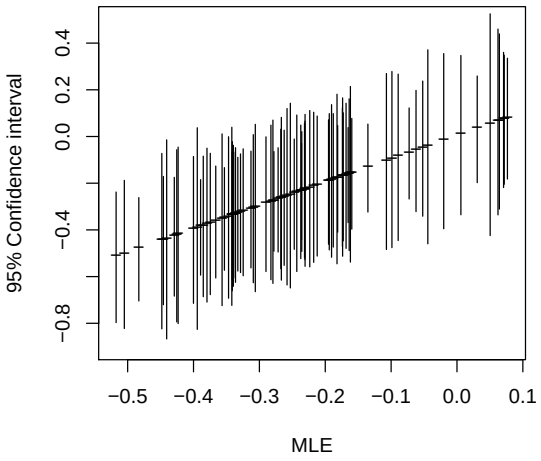
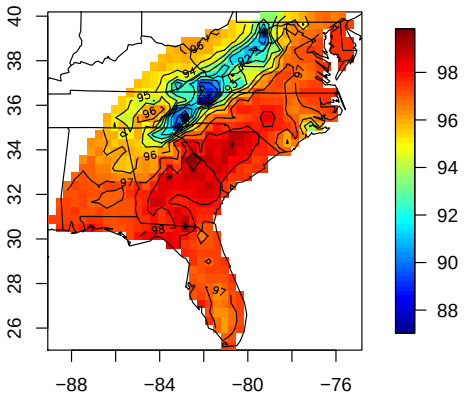
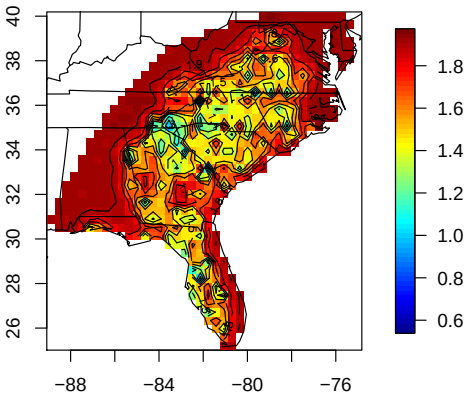


Figure 4: Posterior summaries of the model with 50 free knots and non-stationary residual process. The bandwidth is scaled to the latitude/longitude degree scale.

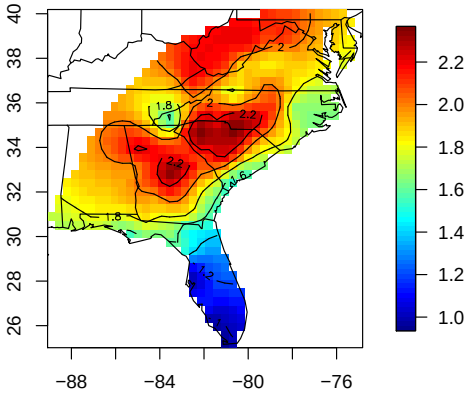
(a) Posterior mean GEV location $\mu(\mathbf{s})$



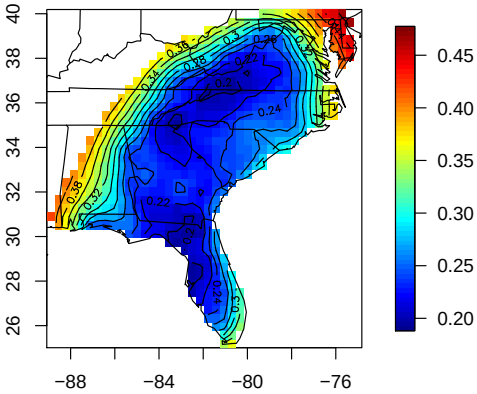
(b) Posterior sd GEV location $\mu(\mathbf{s})$



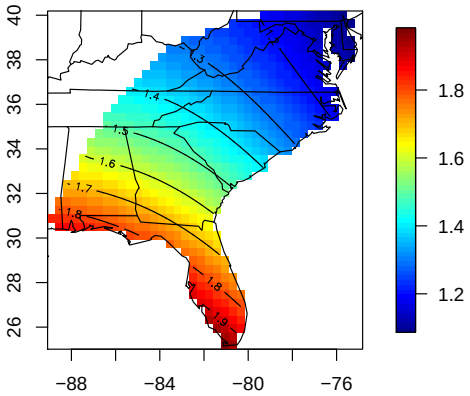
(c) Posterior mean GEV scale $\sigma(\mathbf{s})$



(d) Posterior sd GEV scale $\sigma(\mathbf{s})$



(e) Posterior mean bandwidth $\tau(\mathbf{v})$



(f) Posterior sd bandwidth $\tau(\mathbf{v})$

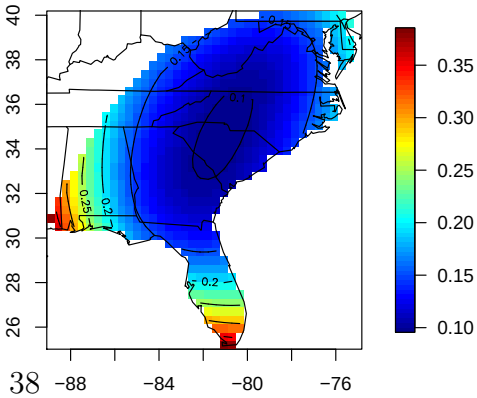
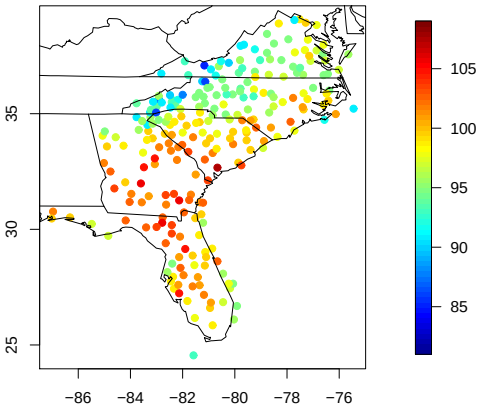
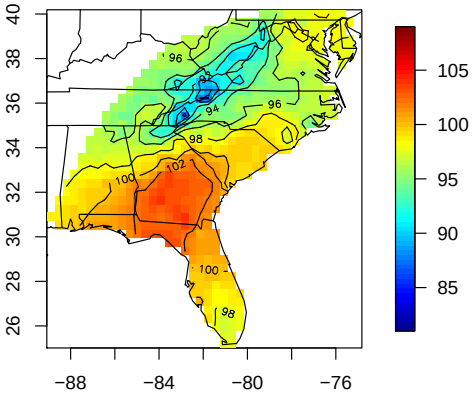


Figure 5: Data and fitted values (posterior mean of $\gamma_t(\mathbf{s})$) for years 1983 and 2007.

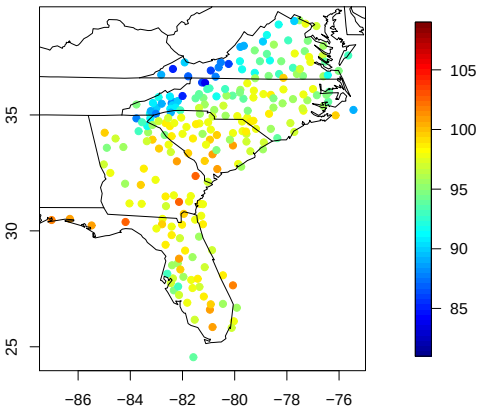
(a) Observed temperature, 1983



(b) Fitted temperature, 1983



(a) Observed temperature, 2007



(b) Fitted temperature, 2007

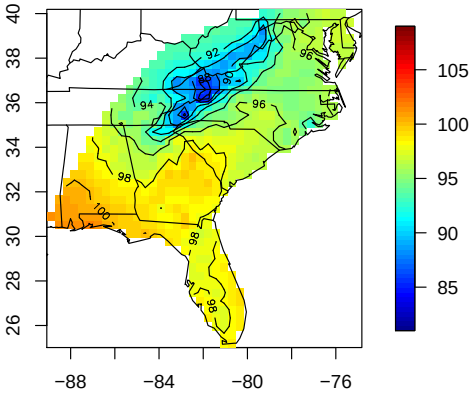
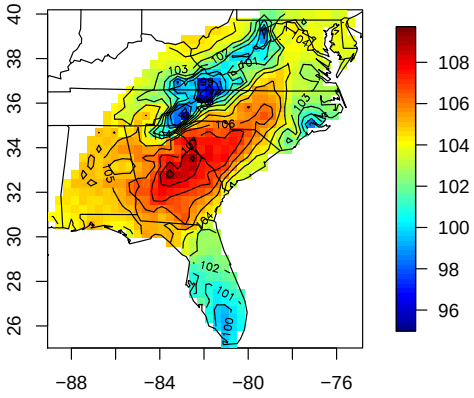
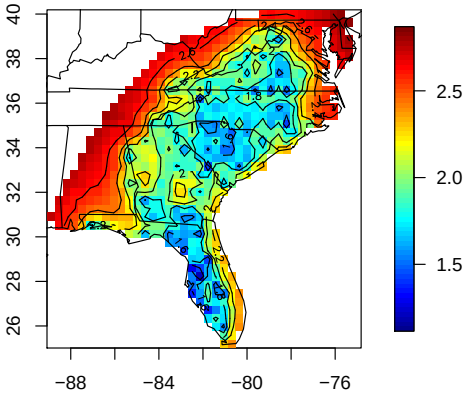


Figure 6: Posterior return levels (degrees F) for the model with 50 free knots and non-stationary residual process. The horizonal lines is Panels (c) and (d) represent the posterior 0.025, 0.25, 0.5, 0.75, and 0.975 quantiles.

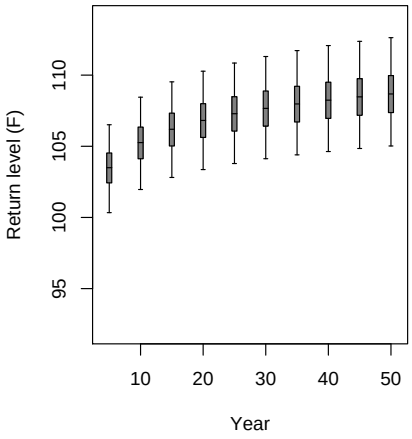
(a) 50-year return level, posterior mean



(b) 50-year return level, posterior sd



(c) Return levels for $\mathbf{s} = (-83.0, 32.9)^T$



(d) Return levels for $\mathbf{s} = (-81.0, 26.9)^T$