

# Continuous Spatial Process Models for Spatial Extreme Values

HUIYAN SANG AND ALAN E. GELFAND

We propose a hierarchical modeling approach for explaining a collection of point-referenced extreme values. In particular, annual maxima over space and time are assumed to follow Generalized Extreme Value (GEV) distributions, with parameters  $\mu$ ,  $\sigma$ , and  $\xi$  specified in the latent stage to reflect underlying spatio-temporal structure. The novelty here is that we relax the conditional independence assumption in the first stage of the hierarchical model, an assumption which has been adopted in previous work. This assumption implies that realizations of the surface of spatial maxima will be everywhere discontinuous. For many phenomena including, e.g., temperature and precipitation, this behavior is inappropriate. Instead, we offer a spatial process model for extreme values which provides mean square continuous realizations, where the behavior of the surface is driven by the spatial dependence which is unexplained under the latent spatio-temporal specification for the GEV parameters. In this sense, the first stage smoothing is viewed as fine scale or short range smoothing while the larger scale smoothing will be captured in the second stage of the modeling. In addition, as would be desired, we are able to implement spatial interpolation for extreme values based on this model. A simulation study and a study on actual annual maximum rainfalls for a region in South Africa are used to illustrate the performance of the model.

**Key words:** Copula Gaussian process models; generalized extreme value distribution; maximum precipitation surfaces; spatial interpolation.

## 1 INTRODUCTION

Extreme value analysis finds wide application in areas such as environmental science (e.g., Lou Thompson et al. (2001)), financial strategy of risk management (e.g., Dahan and Mendelson (2001)) and biomedical data processing (Roberts (2000)). Analysis of extremes in spatial data has become of increasing interest as

---

Huiyan Sang is Assistant Professor, Department of Statistics, Texas A&M University, Texas, USA (e-mail: huiyan.sang@duke.edu). Alan E. Gelfand is Professor, Department of Statistical Science, Duke University, NC, USA (e-mail: alan@stat.duke.edu)

we seek to study surfaces of extremes arising from, say daily weather data or pollution data. We acknowledge that, in studying such surfaces, the extremes need not occur at the same times. So, spatial dependence in extremes is anticipated to be weaker than in the daily data itself but is still of interest.

In this article, we analyze spatial extremes for precipitation events. The motivating data are annual maxima of daily precipitations derived from daily weather station records over South Africa. It is not uncommon to find extreme climate events driven by more than one spatial scale, say, large regional forcings and small scale local forcings. Therefore, an attractive modeling approach in such cases would have the potential to characterize the multi-scale dependence between locations for extreme values of the spatial process. Additionally, spatial interpolation is a problem of interest. Specifically, given observed extreme values, the goal is to learn about the predictive distribution of the unobserved extreme value at an unmonitored location. Prediction of extreme values is useful in monitoring quantitative risk of extreme climate events. Moreover, prediction methods in extreme climate studies have the potential to be applied in statistical downscaling techniques for extremes. Also, interpolation schemes can be used to develop explanatory variables in spatial regression settings. For instance, extreme climate may be more influential with regard to performance of plants at given locations than say average climate.

By now modeling of extremes has received considerable attention in the literature. Many existing approaches have been developed, following the path of utilizing extreme value distribution theory (EVT) (see, e.g., Coles (2001); Beirlant (2004)). One of the challenging issues in extending this work to spatial extreme value modeling lies in the need for multivariate extreme value techniques in high dimensions. Most of the multivariate extreme value theory to date is feasible for low dimensional extreme values. For example, the logistic type of multivariate extreme value distribution (see Coles and Tawn (1991)) has a parametric form too restrictive to capture general dependence structures for large dimension extremes. Some nonparametric models such as the tilted Dirichlet model (Coles (1993)) require integration over the spectral function, which is typically infeasible, especially for high dimensions.

Recently, there has been some work focusing on spatial (or spatial temporal) characterization of extreme values (see, e.g., Kharin and Zwiers (2005), Cooley et al. (2007) and Sang and Gelfand (2008)), including several papers discussing spatial interpolation for extreme values (see, e.g., Cooley et al. (2008) and Buisson et al. (2008)). Cooley et al. (2007) and Sang and Gelfand (2008) use a Bayesian hierarchical model

which assumes the extreme value at each location follows a univariate EVT distribution and the parameters of this distribution follow a spatial model in order to capture spatial dependence. de Haan and Pereira (2006) proposed a stationary max-stable process as an alternative way to introduce spatial dependence. They have applied this max-stable process to predict the distributions of extreme rainfalls across the Netherlands. However, the integration in the distribution function of realizations of their proposed spatial process is intractable, providing a major computational obstacle for handling the dimensionality needed in practice (e.g.,  $n = 200$  in our application).

The contribution of this paper is to extend the hierarchical modeling approach developed in Sang and Gelfand (2008). Their model assumes first stage conditional independence which, at the data level, yields an everywhere discontinuous spatial surface for the spatial maximums. We introduce a fine scale spatial model for the maxima to provide mean square continuous realizations. We retain the second stage specifications of Sang and Gelfand (2008) which allow for spatial dependence in the parameters of the GEV. Thus, we achieve, for example, larger scale spatial dependence in the GEV location parameters with very local dependence in the extreme surface itself. This first stage process model is created through a copula approach where the copula idea is applied to a Gaussian spatial process using suitable transformation. In addition, we demonstrate how to apply this smoothed spatial process model to implement spatial interpolation for extreme values. Though our focus here is on use of this strategy to create a spatial GEV process, we note that it is more generally applicable to create process models yielding smooth realizations for other spatial contexts where we seek marginal distributions in a specific family.

We acknowledge that our Gaussian copula model falls into the category of asymptotic independence rather than the customary extreme value theory models which yield asymptotically dependent random vectors (see e.g., Ledford and Tawn (1996) and Ledford and Tawn (1997)). In application, it may be difficult for the data to distinguish these two asymptotic cases<sup>1</sup> and, more importantly, the Gaussian copula model does enable us to work with high dimensional settings.

The format of the paper is as follows. In Section 2 we briefly review extreme value distribution theory and the hierarchical modeling approach for spatial extreme values proposed in Sang and Gelfand (2008). In Section 3, we develop the spatial Gaussian copula model and discuss its properties. In Section 4 we

---

<sup>1</sup>The approach of Heffernan and Tawn (2004) can handle both cases but is limited to the bivariate setting.

employ this spatial process model for extreme values as a first stage specification in a hierarchical Bayesian modeling framework. Model implementation is illustrated in Section 5 using MCMC methods. In Section 6, we illustrate the proposed approach with a simulated dataset as well as a real data example of precipitation extremes in South Africa. Finally, Section 7 concludes with a brief discussion including future work.

## 2 HIERARCHICAL MODELING FOR SPATIAL EXTREMES

Extreme value theory begins with a sequence  $Y_1, Y_2, \dots$  of independent and identically distributed random variables and, given  $n$ , asks about the distribution for  $M_n = \max\{Y_1, \dots, Y_n\}$ . If the distribution of the  $Y_i$  is specified, the exact distribution of  $M_n$  is known. In the absence of such specification, extreme value theory considers the existence of  $\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq y\right) \equiv F(y)$  for two sequences of real numbers  $a_n > 0, b_n$ . If  $F(y)$  is a non degenerate distribution function, it belongs to either the Gumbel, the Fréchet or the Weibull class of distributions, which can all be usefully expressed under the umbrella of the GEV distribution, i.e., the family of distributions with c.d.f

$$G(y; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi \left(\frac{y - \mu}{\sigma}\right)\right]^{-1/\xi}\right\} \quad (2.1)$$

for  $\{y : 1 + \xi(y - \mu)/\sigma > 0\}$ . In (2.1),  $\mu \in \mathbb{R}$  is the location parameter,  $\sigma > 0$  is the scale parameter and  $\xi \in \mathbb{R}$  is the shape parameter. It is easy to see that if  $Y \sim GEV(\mu, \sigma, \xi)$ , then  $Z = (1 + \xi \frac{Y - \mu}{\sigma})^{1/\xi}$  follows a unit Fréchet distribution, with distribution function  $\exp(-z^{-1})$ , i.e., a  $GEV(1, 1, 1)$ . Therefore, in studying multivariate extreme value distributions, it is common to restrict the marginals to be standard Fréchet distributions.

Working with annual maxima over some region  $D$ , let  $Y(\mathbf{s})$  denote, say the annual maximum of daily highest temperature at location  $\mathbf{s}$ . It is plausible to assume that  $Y(\mathbf{s})$  approximately follows a GEV distribution with parameters  $\mu(\mathbf{s})$ ,  $\sigma(\mathbf{s})$  and  $\xi(\mathbf{s})$ , respectively. Then, second stage models are specified for the  $\mu$ 's,  $\sigma$ 's and  $\xi$ 's. As with spatial generalized linear models (see, e.g., Diggle et al. (1998)) and with customary Gaussian process models where a nugget is introduced, the implication is that the  $Y(\mathbf{s})$  are conditionally independent given the second stage parameters (spatial random effects). Similarly, Sang and Gelfand (2008) imposed a conditional independence assumption in the first stage for the response data given the  $\mu(\mathbf{s})$ ,  $\sigma(\mathbf{s})$  and  $\xi(\mathbf{s})$ .

However, one might question the desirability of this conditional independence assumption. Despite the fact that large scale spatial dependence may be accounted for in the GEV parameter specifications, there may still remain unexplained small scale spatial dependence in the extreme data itself. In different words, many extreme climate events such as temperatures and rainfalls are anticipated to be smooth across space.

In this regard, consider interpolation under conditional independence. Suppose  $Y(\mathbf{s}) | (\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})) \sim \text{GEV}(\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s}))$ . Given a new site  $\mathbf{s}_0$ , we are interested in the predictive distribution of  $Y(\mathbf{s}_0)$  conditional on all the observed annual maxima. This distribution is given by:

$$\begin{aligned} (Y(\mathbf{s}_0) | \mathbf{Y}) &\sim \int P(Y(\mathbf{s}_0) | \mu(\mathbf{s}_0), \sigma(\mathbf{s}_0), \xi(\mathbf{s})) \\ &\quad \times P((\mu(\mathbf{s}_0), \sigma(\mathbf{s}_0), \xi(\mathbf{s})) | \mu, \sigma, \xi, \Omega) \\ &\quad \times P(\mu, \sigma, \xi, \Omega | \mathbf{Y}) d\mu d\sigma d\xi d\Omega \end{aligned} \quad (2.2)$$

where  $(\mu, \sigma, \xi) = (\{\mu(\mathbf{s}_i)\}, \{\sigma(\mathbf{s}_i)\}, \{\xi(\mathbf{s}_i)\}, i = 1, \dots, n)$  and  $\Omega$  denotes any remaining parameters in the model. Under sampling based model fitting, expression (2.2) suggests that  $Y(\mathbf{s}_0)$  should be sampled from its predictive distribution by composition. We first obtain the posterior samples  $(\mu(\mathbf{s}_0), \sigma(\mathbf{s}_0), \xi(\mathbf{s}_0))$  conditional on the posterior draws of  $\mu, \sigma, \xi, \Omega$ . In the next step, the  $Y(\mathbf{s}_0)$  are drawn independently, one-for-one given the samples of  $(\mu(\mathbf{s}_0), \sigma(\mathbf{s}_0), \xi(\mathbf{s}_0))$ .

Even assuming continuous process realizations for the GEV parameters, the conditional independence assumption imposed in the first stage will result in everywhere discontinuous prediction surfaces. More precisely, the first stage of the hierarchical model can be written as:

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \frac{\sigma(\mathbf{s})}{\xi(\mathbf{s})} (Z(\mathbf{s})^{\xi(\mathbf{s})} - 1) \quad (2.3)$$

where  $Z(\mathbf{s})$  follows a standard Fréchet distribution. We may view  $Z(\mathbf{s})$  as the “standardized residual” in the first stage GEV model. The conditional independence assumption is equivalent to the assumption that the  $Z(\mathbf{s})$  are *i.i.d.* So, again, even if the surface for each model parameter is smooth, we will obtain a nonsmooth realized surface under the conditional independence assumption. Our goal is to remove this assumption in order to obtain mean square continuous surface realizations.

We conclude this section with a related point. It can be argued that we can avoid the foregoing concerns by building a space-time model for the daily data and then inferring about the the annual maximums under

this model. Let  $\mathbf{W} = \{W_l(\mathbf{s}_i, t)\}$  denote all of the daily data and let  $Y_l(\mathbf{s}_i) = \max_t W_l(\mathbf{s}_i, t)$ , i.e., the annual maximum at location  $\mathbf{s}_i$  for year  $l$ . Then, under a space-time model, spatial interpolation for  $Y_l(\mathbf{s}_0)$  would be provided by the predictive distribution,  $f(Y_l(\mathbf{s}_0)|\mathbf{W})$ . Instead, our approach models  $\mathbf{Y} = \{Y_l(\mathbf{s}_i), i = 1, 2, \dots, n; l = 1, 2, \dots, T\}$  and, thus provides interpolation using  $f(Y_l(\mathbf{s}_0)|\mathbf{Y})$ . While it seems that the former should be preferred, it requires more model specification than the latter and also, we will find that the needed computation will become intractable.

Moreover, a serendipitous benefit may accrue to the latter. For the former, we would interpolate the  $\mathbf{W}$  to achieve interpolation to  $\{W_l(\mathbf{s}_0, t)\}$  and, thus, to obtain predictive samples of  $Y_l(\mathbf{s}_0)$ ; we would be “averaging” and then taking the maximum. Working with the  $Y_l(\mathbf{s}_i)$ , we are taking the maximums and then averaging to achieve the interpolation. The latter is expected to produce stochastically larger distributions since the maximum of averages is at most the average of maximums. But, with professed interest in interpolating extremes and recognizing that interpolation is expected to smooth, the former may mask or underestimate “realized” extremes; we prefer the latter.

### 3 A SMOOTH SPATIAL PROCESS FOR EXTREME VALUES

We wish to introduce spatial dependence to the  $Z(\mathbf{s})$  while retaining Fréchet marginal distributions. A frequent strategy for introducing dependence subject to specified marginal distributions is through copula models. With a stochastic process of  $Z(\mathbf{s})$ 's, we need to apply the copula approach to a stochastic process. The Gaussian process, which is determined through its finite dimensional distributions, offers the most convenient mechanism for doing this. With a suitable choice of correlation function, mean square continuous surface realizations result for the Gaussian process, hence for transformed surfaces under monotone transformation. In other words, through transformation of a Gaussian process, we can obtain a continuous spatial process of extreme values with standard Fréchet marginal distributions.

Copulas have received much attention and application in the past two decades (see, Nelsen (2006) for a review). In our context, the idea of using a Gaussian copula to construct a bivariate extreme value distribution is discussed in, e.g., Coles et al. (1999), Schlather and Tawn (2003), and Poon et al. (2004)). Consider a random vector distributed according to a standard bivariate Gaussian distribution with correlation  $\rho$ . The

Gaussian copula function is defined as follows:  $C_\rho(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$  where the  $u, v \in [0, 1]$ ,  $\Phi$  denotes the standard normal cumulative distribution function and  $\Phi_\rho$  denotes the cumulative distribution function of the standard bivariate Gaussian distribution with correlation  $\rho$ . The bivariate random vector  $(X, Y)$  having GEV distributions as marginals, denoted as  $G_x$  and  $G_y$  respectively, can be given a bivariate extreme value distribution using the Gaussian copula as follows. Let  $(X, Y) = (G_x^{-1}\Phi(X^*), G_y^{-1}\Phi(Y^*))$ , where  $G_x^{-1}$  and  $G_y^{-1}$  are the inverse marginal distribution functions for  $X$  and  $Y$  and  $(X^*, Y^*) \sim \Phi_\rho$ . Then the distribution function of  $(X, Y)$  is given by  $H(X, Y; \rho) = C_\rho(\Phi(X^*), \Phi(Y^*))$  and the marginal distributions of  $X$  and  $Y$  are still  $G_x$  and  $G_y$ .

A key point to make is that we introduce the Gaussian copula solely to create a spatial process with Fréchet marginals, hence yielding bivariate distributions as above. We do not proceed from the class of limiting bivariate distributions for a pair of maxima, each having a Fréchet marginal. The latter is a well studied problem with a long history (see, e.g., Beirlant (2004), for details and development), allowing both asymptotic dependence and independence. It is characterized through a representation which is parameterized by a c.d.f. on  $[0, 1]$  with mean  $1/2$ . This representation does not include the former except in a limiting sense; the Gaussian copula implies asymptotic independence and the representation for asymptotic independence is achieved through a particular choice of c.d.f. <sup>2</sup> More precisely, Coles et al. (1999) discuss several measures of extreme dependence using either the bivariate tail function or the conditional tail function. For the Gaussian dependence model, using approximations to integral expressions, they are able to compute the limiting behavior of these measures and recognize it as asymptotic independence.

In this regard, again, our goal is to specify a tractable spatial process model for extremes. We are not interested in using the limiting distribution theory for maxima associated with a Gaussian copula since, from above, this would return us to the conditional independence assumption we seek to avoid. Furthermore, we are not trying to diagnose whether asymptotic dependence or independence is an appropriate assumption. This is a challenging issue both formally (see Figure 6 of Coles et al. (1999), and associated discussion) and empirically (see Coles et al. (1999), Section 3.4) but addressing it in the context of a spatial process specification is beyond the scope of the current work.

---

<sup>2</sup>This representation has nothing to do with the Gaussian assumption. It only says that, were we to use the limiting theory to create an approximate distribution for the pairwise maxima, this distribution would be a product of Fréchet distributions.

Formalizing the details, suppose we start with a standard spatial Gaussian process  $Z^*(\mathbf{s})$  with mean  $\mathbf{0}$ , variance 1, and correlation function  $\rho(\mathbf{s}, \mathbf{s}'; \theta)$ . The standard Fréchet spatial process is the transformed Gaussian process defined as:

$$Z(\mathbf{s}) = G^{-1}\Phi(Z^*(\mathbf{s})) \quad (3.1)$$

where  $\Phi$  is the distribution function of a standard normal distribution, and  $G$  is the distribution function of a standard Fréchet distribution. It is clear that  $Z(\mathbf{s})$  defined in (3.1) is a valid stochastic process since it is induced by a strictly monotone transformation of a Gaussian process. Indeed, this standard Fréchet process is completely determined by  $\rho(\mathbf{s}, \mathbf{s}'; \theta)$ .

More precisely, suppose we observe extreme values at a set of sites  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ . The realizations  $\mathbf{Z}^* = (Z^*(\mathbf{s}_1), \dots, Z^*(\mathbf{s}_n))$  follow a multivariate normal distribution which is determined by  $\rho$ . Let  $\mathbf{Z} = (G^{-1}\Phi(Z^*(\mathbf{s}_1)), \dots, G^{-1}\Phi(Z^*(\mathbf{s}_n)))$  and  $\mathbf{Z}^* = (Z^*(\mathbf{s}_1), \dots, Z^*(\mathbf{s}_n))$ . Given  $\rho(\mathbf{s}, \mathbf{s}'; \theta)$ , we obtain the Gaussian copula  $C_{\mathbf{Z}^*}$  for the distribution function of  $\mathbf{Z}^*$  as:

$$C_{\mathbf{Z}^*}(u_1, \dots, u_n) = F_{\mathbf{Z}^*}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)) \quad (3.2)$$

where  $(u_1, \dots, u_n) \in [0, 1]^n$  and  $F$  is the multivariate distribution function of  $\text{MVN}(\mathbf{0}, H)$  with  $H = [\rho(\mathbf{s}_i, \mathbf{s}_j; \theta)]_{i,j=1}^n$ .

Let  $F(\mathbf{Z})$  denote the multivariate distribution of  $\mathbf{Z}$ . Then  $F(\mathbf{Z}) = C_{\mathbf{Z}^*}(\Phi^{-1}G(z_1), \dots, \Phi^{-1}G(z_n))$ , where  $\mathbf{z} = (z_1, \dots, z_n) \in \mathbb{R}^n$ . It is clear that, under  $F(\mathbf{z})$ , the marginal distribution for each of the  $Z(\mathbf{s}_i)$  is standard Fréchet, i.e.,  $GEV(1, 1, 1)$ .

Useful properties of the transformed Gaussian process include:

- Joint, marginal and conditional distributions are all immediately obtained from standard multivariate normal distribution theory once the covariance structure of the process has been specified.
- There are numerous choices for the correlation function. In fact, the Matérn class (see Stein (1999)) with smoothness parameter greater than 0 assures that process realizations are mean square continuous.
- Efficient computational algorithms that have been established for Gaussian processes can be utilized after inverse transformation.

- The transformation in (3.1) retains the stationarity property. If the Gaussian process  $Z^*(\mathbf{s})$  is a stationary spatial process with a valid correlation function  $\rho(\mathbf{s} - \mathbf{s}', \theta)$ , then  $Z^*(\mathbf{s})$  is also strongly stationary, which implies that  $Z(\mathbf{s})$  is a strongly stationary process.
- Evidently,  $\rho$  characterizes the dependence in the Fréchet process through  $\rho(\mathbf{h}) = E(G^{-1}\Phi(Z^*(\mathbf{s}))(G^{-1}\Phi(Z^*(\mathbf{s} + \mathbf{h})))$ . However, since  $G$  has no moments, non-moment based dependence metrics may be considered through the bivariate c.d.f. or p.d.f.
- Evidently, the transformed Gaussian approach is not limited to extreme value analysis. For any spatial response data set where, for each location  $\mathbf{s}$ , we seek to have a common marginal distribution  $G$ , or, more generally, a specified marginal distribution function,  $G_{\mathbf{s}}$  at  $\mathbf{s}$ ,  $G_{\mathbf{s}}^{-1}\Phi(Z^*(\mathbf{s}))$  described in (3.1) provides a valid spatial process to achieve this.

## 4 MODELING DETAILS

Here, we formalize the hierarchical modeling specifications for extremes in space. We take (2.3) as our first stage model where the  $Z(\mathbf{s})$  follow a standard Fréchet process defined as above, with a stationary correlation function  $\rho(\mathbf{s} - \mathbf{s}'; \theta_{\mathbf{z}})$ . That is, we have a hierarchical model in which the first stage conditional independence assumption is removed. Now, we turn to specification of the second stage processes and, in addition, offer a MCMC approach for the implementation of the proposed hierarchical model at the end of this section.

For the second stage model, specifications for  $\mu(\mathbf{s})$ ,  $\sigma(\mathbf{s})$  and  $\xi(\mathbf{s})$  have to be made with care. In particular, we assume there is spatial dependence for the  $\mu$ 's but that  $\sigma$  and  $\xi$  are constant across the study region. The latter assumptions are limiting but spatial variation in the locations is often of primary interest and a single static spatial dataset is not likely to be able to inform about a process for  $\sigma(\mathbf{s})$  and, even less likely for  $\xi(\mathbf{s})$ . More precisely, suppose we propose the specification  $\mu(\mathbf{s}) = \mathbf{X}(\mathbf{s})'\beta + W(\mathbf{s})$ .  $\mathbf{X}(\mathbf{s})$  is the site-specific vector of potential explanatory variables. The  $W(\mathbf{s})$  are spatial random effects, capturing the effect of unmeasured or unobserved covariates with large operational scale spatial pattern. The most common specification for  $W(\mathbf{s})$  is a zero-centered Gaussian Process determined by a valid covariance function  $C(\mathbf{s}_i, \mathbf{s}_j)$ . Below, we will take  $C(\mathbf{s}_i, \mathbf{s}_j) = \sigma_{\mu}^2 \rho(\mathbf{s}_i - \mathbf{s}_j; \theta_{\mu})$ . Here,  $C$  is apart from  $\rho$  introduced at the

first stage. In fact, from (2.3), plugging in the model for  $\mu(\mathbf{s})$  we obtain

$$Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})'\beta + W(\mathbf{s}) + \frac{\sigma}{\xi}(Z(\mathbf{s})^\xi - 1). \quad (4.1)$$

We see that we have a model for  $Y(\mathbf{s})$  with two sources of spatial error. With distinct correlation functions, from a single sample of spatial maxima, due to the structured spatial dependence, it is possible to learn about the parameters of the two processes producing these error contributions. In particular, we assume that the scale associated with the  $W(\mathbf{s})$  process is coarse scale while that associated with the  $Z^*(\mathbf{s})$  process is more fine scale. For instance, with isotropic correlation functions, we would assume the range for the  $W(\mathbf{s})$  process is greater than that for the  $Z^*(\mathbf{s})$  process (see Section 5 below).

Since interpolation is a key objective, we clarify how to implement it. Given a new location  $\mathbf{s}_0$ , the posterior predictive distribution of  $Y(\mathbf{s}_0)$ , conditional on all the observed annual maxima, is given by

$$\begin{aligned} Y(\mathbf{s}_0)|\mathbf{Y} &\sim \int P(Y(\mathbf{s}_0)|\mu(\mathbf{s}_0), \mu, \sigma, \xi, \theta_\mu, \theta_z, \mathbf{Y}) \\ &\quad \times P((\mu(\mathbf{s}_0)|\mu, \sigma, \xi, \theta_\mu, \theta_z, \mathbf{Y}) \\ &\quad \times P(\mu, \sigma, \xi, \theta_\mu, \theta_z|\mathbf{Y}))d\mu d\sigma d\xi d\theta_\mu d\theta_z. \end{aligned} \quad (4.2)$$

In (4.2), the first two expressions under the integral simplify. The first becomes  $P(Y(\mathbf{s}_0)|\mu(\mathbf{s}_0), \mu, \sigma, \xi, \theta_z, \mathbf{Y})$  and is sampled through  $Z^*(\mathbf{s}_0)|\mathbf{Z}^*$  using (4.1) and (3.2). The second becomes  $P((\mu(\mathbf{s}_0)|\mu, \theta_\mu)$  which, from above, is a conditional normal distribution. And so, posterior samples of  $(\mu, \sigma, \xi, \theta_\mu, \theta_z)$ , with composition, yield predictive samples for  $Y(\mathbf{s}_0)$ . The essential difference between (4.2) and (2.2) is the presence of  $\mathbf{Y}$  in the first term under the integral in the former. This necessitates sample draws from the Gaussian process rather than independent Fréchet draws.

#### 4.1 EXTENSION TO SPATIO-TEMPORAL PROCESSES

Often, we collect space-time data over long periods of time, e.g., many years, and we seek to study say, annual spatial maxima. Now, we are given a set of extremes  $\{Y(\mathbf{s}_i, t), i = 1, \dots, n; t = 1, \dots, T\}$ . The first stage of the hierarchical model now becomes a space time standard Fréchet process for the  $Z(\mathbf{s}, t)$ , again, built from a space-time Gaussian process. Time would naturally be discretized to yield, e.g., annual maxima in which case we need only provide a dynamic Gaussian process model (see, Banerjee et al. (2004)) though

we might view time as continuous in order to explicitly capture space-time interaction, using a valid space-time covariance function, see Stein (2005). In fact, we might assume  $Z_t^*(\mathbf{s})$  and  $Z_{t'}^*(\mathbf{s})$  are two independent Gaussian processes when  $t \neq t'$ . That is, it is often plausible to assume temporal independence since annual block size may be long enough to yield approximately independent annual maximum observations.

Model specifications for  $\mu(\mathbf{s}, t)$ ,  $\sigma(\mathbf{s}, t)$  and  $\xi(\mathbf{s}, t)$  could account for dependence structures both within and across location and time. Exploratory analysis is helpful in terms of learning about temporal trend in these parameters and learning about spatial dependence in these parameters. For example, GEV parameter estimation based on independent GEV models typically can guide us towards formal spatial temporal specifications for these parameters. Once again, interest will likely focus on  $\mu(\mathbf{s}, t)$  with perhaps a spatially varying  $\sigma(\mathbf{s})$  and a constant  $\xi$ . Sang and Gelfand (2008) offers detailed discussion with regard to these latent specifications. We provide an illustration in Section 6.

## 5 BAYESIAN IMPLEMENTATION

Given the complexity of the proposed hierarchical models, we employ MCMC methods to do the model fitting. We begin with prior specifications for the parameters. Customarily, we set  $\beta \sim MVN(\mu_\beta, \Sigma_\beta)$ . In our model, a vague normal prior is assigned to the shape parameter  $\xi$ , and an inverse gamma prior for  $\sigma$ . For the two spatial processes, for convenience, we employ the familiar exponential covariance function, hence introducing range parameters  $\phi_\mu$  and  $\phi_z$ . In addition, we need a variance parameter for the  $W(\mathbf{s})$  process which we denote by  $\sigma_\mu^2$ . In general, it is difficult to identify both the spatial variance and the spatial range parameters for a process (see Zhang (2004)); somewhat informative prior specifications are necessary. With presumed interest in letting the data inform about the spatial variability, we impose informative priors on the ranges based upon the size of the associated domain. Furthermore, since, in our model, spatial random effects exist at two spatial scales, without any constraints on the decay/range parameters, further identifiability issues arise regarding inference about these spatial ranges. Therefore, arguing as we have above, we impose the restriction,  $\phi_\mu > \phi_z$ .

Posterior inference for the model parameters is completed using Gibbs sampling (Gelfand and Smith (1990)) with Metropolis-Hastings updating (Gelman et al. (2004)). Given  $\mu(\mathbf{s})$ , we directly sample  $\beta$  and

$\mathbf{W} = (W(\mathbf{s}_1), \dots, W(\mathbf{s}_n))$ . Parameters without closed form full conditional distributions are updated using Metropolis-Hastings steps.  $\xi$  and  $\log(\sigma)$  are sampled by random walk Metropolis-Hastings with Gaussian proposals. The conditional distribution of the latent spatial component  $\mathbf{Z}^*$  is a high dimensional, non-Gaussian distribution and difficult to sample from using standard Metropolis-Hastings. So, we employ the Metropolis-adjusted Langevin Algorithm (MALA), also known as Langevin-Hastings, to help the mixing of the MCMC in this regard. Gradient information of the posterior distribution is used in making proposal distributions. See Stramer and Tweedie (1999), Christensen et al. (2003), Robert and Casella (2004) for more detailed discussion on this algorithm.

## 6 EXAMPLES

### 6.1 A SIMULATION STUDY

In this section we present a simulation study designed to examine the performance of the transformed Gaussian process model relative to the non-smoothed spatial GEV models. The set up of the study is as follows. Sampling is done at a set of 1200 locations over a  $[0, 10] \times [0, 10]$  rectangle. 300 locations are obtained by sampling using a uniform distribution over the rectangle and 900 locations are specified on a regular lattice. All locations are shown in Figure 1. Let  $\mathbf{Y} = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))$  denote a response vector.  $\mathbf{Y}$  is obtained according to the GEV model discussed above:

$$\begin{aligned} Y(\mathbf{s}) &= \mu(\mathbf{s}) + \frac{\sigma}{\xi}(Z(\mathbf{s})^\xi - 1) \\ \mu(\mathbf{s}) &= \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + W(\mathbf{s}) \\ Z(\mathbf{s}) &= G^{-1}\Phi(Z^*(\mathbf{s})). \end{aligned} \tag{6.1}$$

Realizations of  $W(\mathbf{s})$  are obtained from a Gaussian process with exponential covariance function  $\sigma_\mu^2 \rho(\phi_\mu)$  with range parameter  $\phi_\mu = 4$  and scale parameter  $\sigma_\mu = 1$ . We set  $\xi = 0.5$ ,  $\sigma = 3$ . We include only a constant intercept,  $\mu = 10$ , in the covariate part, yielding the  $\mu(\mathbf{s}_i)$ . The  $Z(\mathbf{s}_i)$  are generated from  $\mathbf{Z}^* = (Z_1^*, \dots, Z_n^*)$ , obtained as part of a realization of a Gaussian process with exponential correlation function  $\rho(\phi_Z)$  and unit spatial scale parameter at observed location set  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ , where  $\phi_Z = 1.4$  is the range parameter. The resulting 1200  $Y(\mathbf{s}_i)$  yield the surface shown at pixel scale in Figure 2.

We hold out the 900 regular lattice locations for validation purpose and fit the model in (6.1) using the simulated extreme values at the 300 uniformly sampled locations. To make comparison, using the same data set, we also fit the corresponding model with the  $Z(\mathbf{s})$  *i.i.d.*:

$$\begin{aligned} Y(\mathbf{s}) &= \mu(\mathbf{s}) + \frac{\sigma}{\xi}(Z(\mathbf{s})^\xi - 1) \\ \mu(\mathbf{s}) &= \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + W(\mathbf{s}) \\ Z(\mathbf{s}) &\sim GEV(1, 1, 1) \quad i.i.d \end{aligned} \quad (6.2)$$

We seek to investigate whether or not the spatial signals in the standardized residuals can be successfully detected by fitting the model with smoothed residuals.

For each model, we ran 200,000 iterations to collect posterior samples after a burn in period of 50,000 iterations, thinning using every fifth iteration. Trace plots of parameters indicate good convergence of the respective marginal distributions. The long burn in period is caused mainly by the slow mixing of  $\mathbf{Z}^*$ , yielding, as expected, longer run times for the model in (6.1) compared with (6.2).

Table 1 displays the posterior means for the parameters and the corresponding 95% credible intervals under the model in (6.1) and the model in (6.2), respectively. All 95% credible intervals cover the true parameter values for the model we simulated from.

We now consider the performance of these models based on prediction for the holdout set of 900 locations. Specifically, we obtain  $\hat{r}$ , the empirical coverage probability of nominal 95% predictive intervals. In addition, using posterior medians, we check the performance of prediction by computing the average absolute predictive errors (*AAPE*) of each model. Given the observed value of  $Y(\mathbf{s})$  in the holdout data set, *AAPE* is given by

$$AAPE = \frac{1}{900} \sum_{i=1}^{900} |\hat{Y}(\mathbf{s}) - Y(\mathbf{s})| \quad (6.3)$$

where  $\hat{Y}(\mathbf{s})$  is the posterior median.

Table 2 compares the predictive performance under the two models. The prediction performance for the model in (6.1) is roughly 25% better than that of the model in (6.2) based on the *AAPE* criterion and empirical coverage is closer to the nominal level. Extreme values with positive shape parameters are heavy tailed. So, we display the predictive median surface for the extreme values as well as the associated surface

for the lengths of the 95% predictive intervals in Figure 3 for the models in (6.1) and (6.2), respectively. Comparing Figure 3 and Figure 2, the predictive surface for the extreme values based on the model in (6.1) recovers the true surface and the local spatial patterns for the simulated extreme values reasonably well. In contrast, the predictive median surface obtained from the model in (6.2) only reflects large scale spatial dependence and fails to capture the local peaks. In addition, on average, the model in (6.2) has longer 95% predictive intervals compared with the model in (6.1).

## 6.2 A REAL DATA EXAMPLE

We consider precipitation data collected at 200 monitoring sites in the Cape Floristic Region (CFR) in South Africa. The Cape Floristic Region is home to the highest density of plant species in the world. It is recognized as one of the world's floristic kingdoms and as a global biodiversity hotspot, including about 9000 plant species, 69% of which are found nowhere else. Most of the study region has a Mediterranean-type climate, typically characterized by cool, wet winters and warm, dry summers. Our "big picture" motivation is the challenging ecological problem of trying to characterize the effect of extreme climate events on the distribution and abundance of species. It is anticipated that extreme climate events, such as drought, heavy rainfall and very high or low temperatures will be consequential in explaining plant performance. This issue is well-discussed for vegetation in South Africa; see, for instance, the website of G. Midgley (<http://www.sanbi.org/gcrg/midgleypubs.pdf>) who has written extensively on this problem.

We illustrate using the annual maximum of daily rainfalls for the years 1956, 1976, 1996 and 2006. We fit the model in (6.1) for each selected year aiming to compare the distributions of extreme precipitations at different years. In particular, we fit both models in (6.1) and (6.2) for the annual maxima in 2006 for comparison purposes. Models are fitted using two parallel chains. Again, Model (6.1) takes longer to run, completing roughly 200 iterations per minute using Matlab code with dual 2.8 GHz Xeon CPUs and 12GB memory. We ran 20,000 iterations to collect posterior samples after a burn in period of 50,000 iterations, thinning using every fifth iteration. Trace plots of parameters indicate good convergence of the respective marginal distributions.

Table 3 displays the posterior means for the parameters and the corresponding 95% credible intervals under the model in (6.1) for each selected year and the model in (6.2) for the year 2006. In particular, we

have the point estimates of the small scale range parameter  $\phi_Z = 0.042$  and the large scale range parameter  $\phi_\mu = 0.161$  at the year 2006. For the model in (6.2), we only have spatial random effects in the location parameters. The point estimate of  $\phi_\mu$  under this model is 0.079. The posterior estimation for  $\xi$  reveals a small positive value, indicating the GEV distributions of annual maximum rainfalls in CFR have slightly heavy upper tails. The GEV distribution of the annual maxima in 1996 has a significantly larger location parameter and scale parameter compared with other selected years. In fact, the heavy rainfalls recorded in 1996 caused heavy losses to agriculture and infrastructure in the Western Cape province of South Africa (see e.g., van Bladeren (2000)).

Each of the models enables prediction for any new sites in the study region. Annual maximum rainfalls at a new site could be simply obtained by updating samples from (4.2) for the smoothed model and from (2.2) for the non-smoothed model. In fact, we held out the annual maximum rainfalls at 64 new sites from 2006 for validation purposes, in order to compare models in terms of the prediction performance. Again, posterior medians are adopted as the point estimates of the predicted annual maxima because of the skewness of the predictive distribution for  $\mathbf{Y}_{s_0}$ . Again, we study the prediction performance by computing the average absolute predictive errors (*AAPE*) as defined in (6.3) and,  $\hat{r}$ . Table 4 summarizes  $\hat{r}$  and *AAPE* for the two models. The model in (6.1) has the smaller prediction error in terms of *AAPE*. In addition, the  $\hat{r}$  for the model in (6.1) is slightly closer to 0.95 compared with the model in (6.2).

Finally, return levels for the occurrence of extreme events are often of practical interest in climate studies. The return level  $z_p$  is defined as the threshold which is exceeded by the extreme value with probability  $p$  (see Coles (2001)). Equivalently,  $z_p$  can be viewed as a threshold which is such that we expect an exceedance once every  $1/p$  years. Given posterior samples of  $\mu$ ,  $\sigma$  and  $\xi$  and  $\mathbf{Z}^*$  under the smoothed model in (6.1), we infer the return level surface with a return period of 25 years, presenting the estimated  $z_{1/25}$  surfaces for each of the four years in Figure 4. We can see considerable spatial variability in return levels. Of course, return levels lose their importance when the process changes over time. However, we can, for instance, see much higher return levels based upon fitting the 1996 data compared with fitting the 2006 data. Accordingly, plants will be exposed to dramatically different risks if one of these patterns prevails rather than the other in the future.

## 7 DISCUSSION

We have presented a spatial process for extreme values which produces mean square continuous realizations. The joint distribution function of the realizations arises from a transformed Gaussian process and can be expressed explicitly for any given set of locations. We employ this process as a first stage specification in a hierarchical model using a GEV distribution to describe the asymptotic distributions of maxima taken from a time series of daily records. The second stage of our hierarchical models specifies variations in time and large scale dependence in space for the parameters in the GEV distributions. We showed that the overall model is *multi-scale*. In addition to the illustrative analyses we have supplied, we also can use this methodology as a spatial interpolation tool to produce risk maps of extreme values, for example, the risk of having annual maximum of daily highest temperatures greater than 40°.

Extensions for this work will take us to more general spatial temporal contexts as discussed in Section 4.1. Computationally more efficient algorithms will be needed to fit these hierarchical models. Further extensions can involve multivariate data collected at each location, providing multivariate extremes in space and time. Can the extremes for one variable help us to interpolate extremes for another, essentially co-kriging with extremes?

## ACKNOWLEDGEMENTS

The work was supported by NSF DEB 0516198. The first author was also supported by Award Number KUS-CI-016-04 made by King Abdullah University of Science and Technology (KAUST). We are grateful to Bruce Hewitson and Chris Leonard for help in the development of the dataset along with Gabi Hegerl, Andrew Latimer, Anthony Rebelo, John Silander, Jr. and K. Feridun Turkman for valuable discussions.

## REFERENCES

- Banerjee, S., Carlin, B., and Gelfand, A. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall, Boca Raton.
- Beirlant, J. (2004), *Statistics of extremes: theory and applications*, Wiley.
- Buishand, T., de Haan, L., and Zhou, C. (2008), "On spatial extremes: with application to a rainfall problem," *Annals of Applied Statistics*.

- Christensen, O., Roberts, G., and Lund, M. (2003), “Robust MCMC for spatial GLMMs,” *Preprints in Mathematical Sciences*, 23.
- Coles, S. (1993), “Regional Modelling of Extreme Storms via Max-Stable Processes,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 55, 797–816.
- (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer, London.
- Coles, S., Heffernan, J., and Tawn, J. (1999), “Dependence measures for extreme value analyses,” *Extremes*, 2, 339–365.
- Coles, S. and Tawn, J. (1991), “Modelling Extreme Multivariate Events,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 53, 377–392.
- Cooley, D., Naveau, P., and Davis, R. (2008), “Dependence and Spatial Prediction in Max-Stable Random Fields,” *University of Colorado, submitted for 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.
- Dahan, E. and Mendelson, H. (2001), “An Extreme-Value Model of Concept Testing,” *Management Science*, 47, 102–116.
- de Haan, L. and Pereira, T. (2006), “Spatial extremes: Models for the stationary case,” *Ann. Statist.*, 34, 146–168.
- Diggle, P., Tawn, J., and Moyeed, R. (1998), “Model-Based Geostatistics,” *Journal of the Royal Statistical Society (Series C): Applied Statistics*, 47, 299–350.
- Gelfand, A. and Smith, A. (1990), “Sampling-Based Approaches to Calculating Marginal Densities,” *Journal of the American Statistical Association*, 85, 398–409.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004), *Bayesian Data Analysis*, Chapman & Hall, Boca Raton.
- Heffernan, J. and Tawn, J. (2004), “A conditional approach for multivariate extreme values,” *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 66, 497–530.
- Kharin, V. and Zwiers, F. (2005), “Estimating Extremes in Transient Climate Change Simulations,” *Journal of Climate*, 18, 1156–1173.
- Ledford, A. and Tawn, J. (1996), “Statistics for near independence in multivariate extreme values,” *Biometrika*, 83, 169–187.

- (1997), “Modelling dependence within joint tail regions,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 475–499.
- Lou Thompson, M., Reynolds, J., Cox, L., Guttorp, P., and Sampson, P. (2001), “A review of statistical methods for the meteorological adjustment of tropospheric ozone,” *Atmospheric Environment*, 35, 617–630.
- Nelsen, R. (2006), *An Introduction to Copulas*, Springer, New York.
- Poon, S., Rockinger, M., and Tawn, J. (2004), “Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications,” *Review of Financial Studies*, 17, 581–610.
- Robert, C. and Casella, G. (2004), *Monte Carlo Statistical Methods*, Springer, New York.
- Roberts, S. (2000), “Extreme value statistics for novelty detection in biomedical dataprocessing,” *Science, Measurement and Technology, IEE Proceedings-*, 147, 363–367.
- Sang, H. and Gelfand, A. (2008), “Hierarchical Modeling for Extreme Values Observed over Space and Time,” *Environmental and Ecological Statistics (forthcoming)*.
- Schlather, M. and Tawn, J. (2003), “A dependence measure for multivariate and spatial extreme values: Properties and inference,” *Biometrika*, 90, 139–156.
- Stein, M. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, New York.
- (2005), “Space-Time Covariance Functions.” *Journal of the American Statistical Association*, 100, 310–322.
- Stramer, O. and Tweedie, R. (1999), “Langevin-Type Models II: Self-Targeting Candidates for MCMC Algorithms\*,” *Methodology and Computing in Applied Probability*, 1, 307–328.
- van Bladeren, D. (2000), “Flood Event Database for the Western Cape,” *Report for the Department of Water Affairs, Western Cape Region. SRK Consulting, report, 283742*.
- Zhang, H. (2004), “Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics.” *Journal of the American Statistical Association*, 99, 250–262.

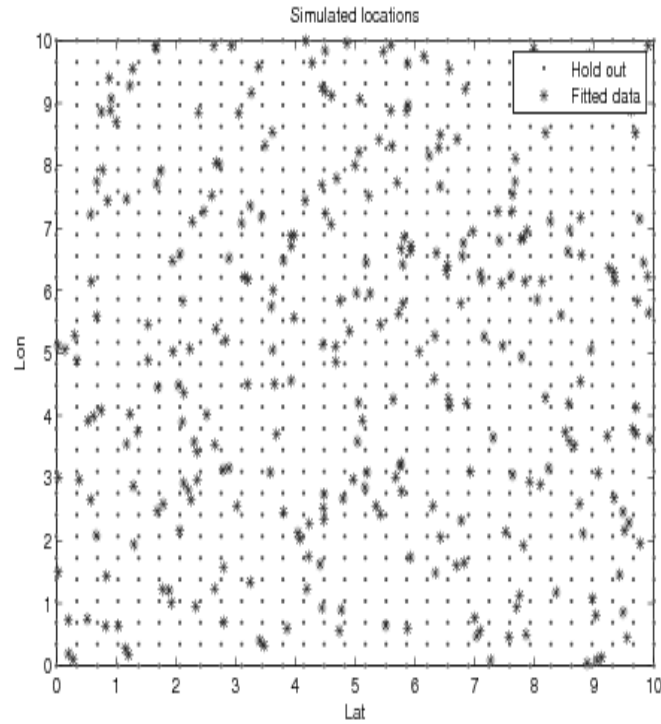


Figure 1. Simulated locations

Table 1. Posterior sample means of parameters and the corresponding 95% credible intervals for Model (6.1) and the model in (6.2).  $\mu$  is the intercept,  $\sigma$  is the scale parameter of GEV, and  $\xi$  is the shape parameter of GEV;  $\phi_\mu$  is the spatial range parameter in the location parameters and  $\phi_z$  is the spatial range parameter in the GEV residuals

	True	Model 6.1 (Smoothed residuals)		Model 6.2 (Non-smoothed residuals)	
		Mean	95% CI	Mean	95% CI
$\mu$	10	9.77	(9.11, 10.46)	9.71	(9.34, 10.52)
$\sigma$	3	2.98	(2.61, 3.40)	3.17	(2.92, 3.53)
$\xi$	0.5	0.47	(0.12, 1.31)	0.42	(0.12, 1.22)
$\phi_\mu$	4	2.66	(0.47, 5.63)	2.81	(0.44, 5.62)
$\phi_z$	1.4	1.51	(0.27, 3.69)		

Table 2. Performance of the model in (6.1) and the model in (6.2) using averaged absolute predictive errors (*AAPE*) and the empirical coverage probability  $\hat{r}$ .

	<i>AAPE</i>	$\hat{r}$
Smoothed GEV	2.65	0.96
Nonsmoothed GEV	3.53	0.97

Table 3. Real data analysis: posterior sample means of parameters and the corresponding 95% credible intervals for the model in (6.1) and (6.2).  $\beta$  is the intercept.  $\sigma^2$  is the scale parameter of GEV and  $\xi^2$  is the shape parameter of GEV

Model	$\beta$	$\sigma^2$	$\xi^2$
Model 6.1(Smoothed GEV)			
$t = 1956$	14.7(13.2,15.9)	6.71 (5.92, 7.46)	0.039 (0.027, 0.087)
$t = 1976$	19.3(18.4,21.0)	7.22 (6.56, 8.00)	0.030 (0.014, 0.081)
$t = 1996$	24.3(21.6,29.2)	14.3 (11.7, 16.4)	0.058 (0.021, 0.124)
$t = 2006$	18.7(16.9,19.9)	5.14 (4.57, 5.38)	0.019 (0.009, 0.057)
Model 6.2 (Non-smoothed GEV)			
$t = 2006$	19.9(18.3, 20.9)	4.93(4.33, 5.21)	0.024 (0.011, 0.059)

Table 4. Real data analysis: performance of the model in (6.1) and (6.2) using averaged absolute predictive errors (*AAPE*) and the empirical coverage probability  $\hat{r}$ .

	<i>AAPE</i>	$\hat{r}$
Smoothed GEV	4.62	0.94
Nonsmoothed GEV	5.60	0.94

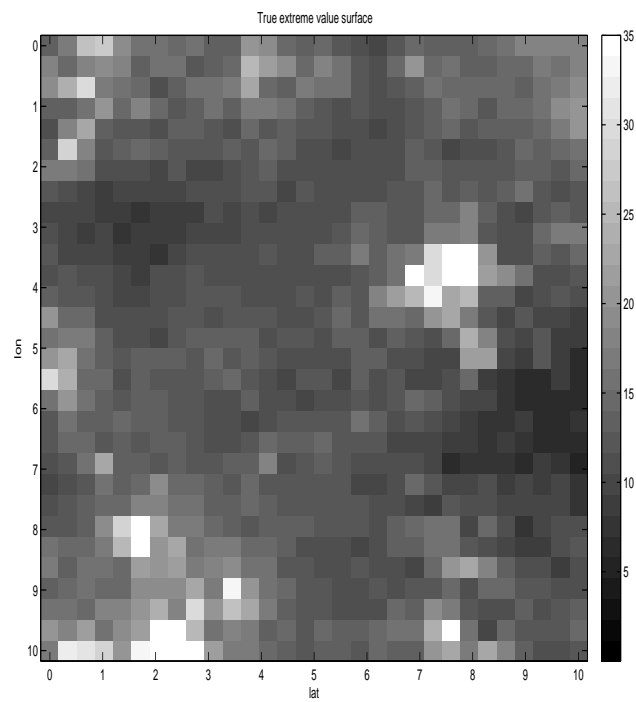


Figure 2. Observed surface (see text for details)

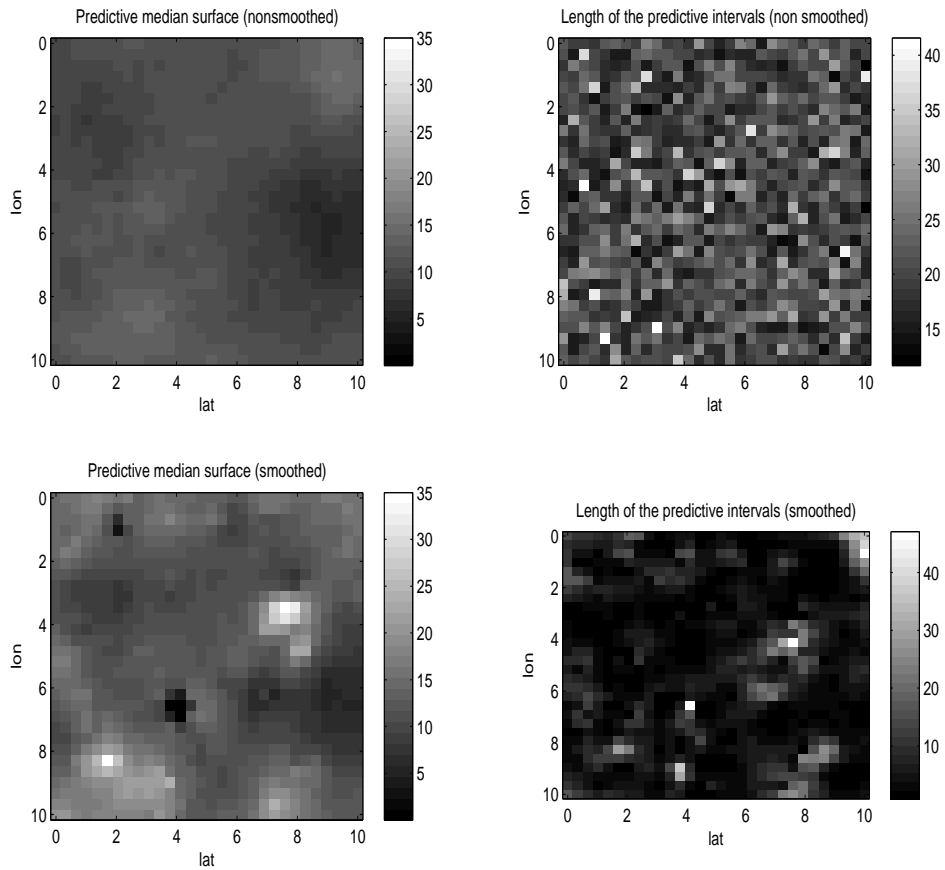


Figure 3. From left to right, top to bottom: (a) The predictive median surface for extreme values using the model in (6.1) (smoothed residuals); (b) The lengths of the 95% predictive intervals for the model in (6.1) (smoothed residuals); (c) The predictive median surface for extreme values using the model in (6.2) (non-smoothed residuals); (d) The lengths of the 95% predictive intervals for the model in (6.2) (non-smoothed residuals)

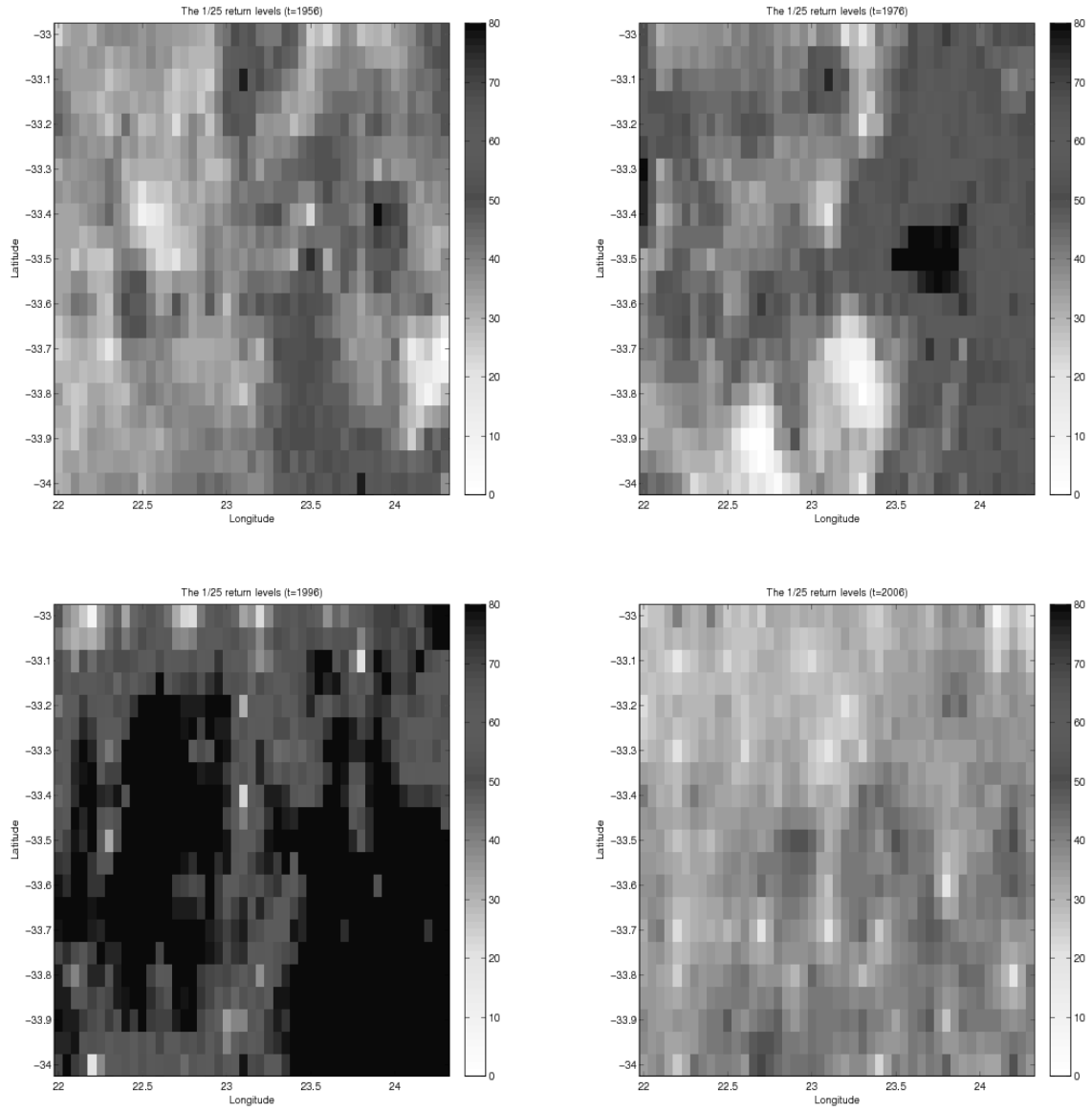


Figure 4. Posterior sample means of the 1/25 return levels for the study region in South Africa