# Nonparametric Spatial Models for Extremes: Application to Extreme Temperature Data. *

Montserrat Fuentes, John Henry, and Brian Reich

SUMMARY

Estimating the probability of extreme temperature events is difficult because of limited records across time and the need to extrapolate the distributions of these events, as opposed to just the mean, to locations where observations are not available. Another related issue is the need to characterize the uncertainty in the estimated probability of extreme events at different locations. Although the tools for statistical modeling of univariate extremes are well-developed, extending these tools to model spatial extreme data is an active area of research. In this paper, in order to make inference about spatial extreme events, we introduce a new nonparametric model for extremes. We present a Dirichlet-based copula model that is a flexible alternative to parametric copula models such as the normal and $t$-copula. This presents the most flexible multivariate copula approach in the literature, and allows for nonstationarity in the spatial dependence of the extremes. The proposed modelling approach is fitted using a Bayesian framework that allow us to take into account different sources of uncertainty in the data and models. We apply our methods to annual maximum temperature values in the east-south-central United States.

## 1 Introduction

Extremely hot summers can drastically reduce agricultural production, increase energy consumption, and lead to hazardous health conditions. Thus, understanding and predicting

*M. Fuentes is a Professor of Statistics at North Carolina State University (NCSU). Tel.: (919) 515-1921, Fax: (919) 515-1169, E-mail: fuentes@stat.ncsu.edu. J. Henry is a postdoctoral fellow at the Department of Statistics at NCSU. B. Reich is an Assistant Professor of Statistics at NCSU. The authors thank the National Science Foundation (Henry, DMS-0354189; Fuentes DMS-0706731, DMS-0353029), the Environmental Protection Agency (Fuentes, R833863), and National Institutes of Health (Fuentes, 5R01ES014843-02) for partial support of this work. *Key words:* Dirichlet processes, extreme temperatures, nonstationarity, return levels, spatial models.

the spatial and temporal variability and trends of extreme weather events is crucial for the protection of socio-economic well-being. Quantifying extremely high surface air temperature changes, and trends in extremes is also crucial for understanding global warming and mitigating its regional impact. Often of interest to scientists is the $n$-year return level for annual maximum of daily temperatures, which is defined as the quantile $z_n$ (from the distribution of the extreme temperatures) which has probability $1/n$ of being exceeded in a particular year. Thus, to calculate return levels, we need to have a good characterization of the distribution of the extreme temperatures. Return values represent rare events, for instance, the twenty year return value is likely to occur only a few times over the course of an individual's lifetime. The probability of an extreme event under non-stationary conditions across time depends on the rate of change of the distribution as well as on the rate of change of the frequency of their occurrence.

Tools for statistical modeling of univariate extremes are well-developed. However, extending these tools to model spatial extreme data is an active area of research. One of the challenging issues in spatial extreme value modeling is the need for spatial extreme value techniques in high dimensions, since most of the multivariate extreme value theories only work well for low dimensional extreme values. In this paper innovative and general statistical methods for modelling of extreme events are proposed, to produce maps of temperature return levels, to estimate trends and variability of extreme temperature events, and to provide uncertainty measures. We introduce a new framework to characterize extremes, a nonparametric Dirichlet process (DP) copula approach. This DP copula defines the most flexible type of copula framework that we currently have in the literature.

Recently, there has been some work focusing on spatial characterization of extreme values (e.g. Kharin and Zwiers, 2005, Cooley et al., 2007, Sang and Gelfand, 2009, Zhang et al., 2008), including papers discussing spatial interpolation for extreme values (e.g. Cooley et al., 2008, and Buishand and Zhou, 2008). Sang and Gelfand (2009) used a Bayesian hierarchical model, which assumes that the annual maxima at each location follows a one-dimensional GEV distribution and that the parameters of this distribution are varying according to latent spatial models capturing the spatial dependence. Nonstationarity refers to spatial dependence that is a function of location, rather than just relative position of observations. To account for nonstationarity in univariate extreme events in an approach popularised by Davison and Smith (1999), the model parameters are modelled as functions of covariates. Eastoe and Tawn (2009) and Eastoe (2009) suggest an alternative approach for spatial nonstationary extremes, the nonstationarity in the whole dataset is first modelled and removed, using a preprocessing technique. Then, the extremes of the pre-processed (transformed) data are modelled using the approach of Davison and Smith (1990), giving a model with both pre-processing and tail parameters. We introduce here new continuous spatial models for extreme values to account for spatial dependence which is unexplained by the latent spatial specifications for the distribution parameters, characterizing also the potential lack of stationarity across space and time. This is the first time that the pre-processing and tail-parameters are analyzed simultaneously using a fully Bayesian approach to account for all sources of uncertainty.

Although there are methods in the literature for extremes (i.e. Eastoe (2009)) to account for spatial correlation between nearby stations, the high-dimensional joint distributions induced by that type of models are restrictive. The most popular approach is a copula which models spatial correlation in a latent space and projects to the data scale in a way that preserves the desired marginal at each site. For instance, the Gaussian copula is asymptotically (as the threshold increases) equivalent to the independent copula. In this work we present a very flexible copula framework, a nonparametric copula for extreme temperatures. We also introduce measures to characterize complex spatial dependence in these extreme temperatures, allowing the extremal coefficient function, commonly used to study the extreme value dependence structure, to be space dependent. This extremal coefficient function is threshold invariant for max stable distributions. However, our temperature data seem to have a threshold specific tail dependence structure. The nonparametric model for extremes presented here has tail dependence that is allowed to be threshold-specific. Furthermore, this new nonparametric spatial framework introduced here to model extremes for annual maximum temperature, is not max-stable, has marginals that are Generalized Extreme Value (GEV) distributions, and it is flexible enough to characterize extreme events with complex spatio-temporal dependence structures.

This paper is organized as follows. In Section 2, we introduce flexible measures to characterize dependence in extremes, in particular to explain extreme dependence for spatial nonstationary and threshold dependent extreme processes. In Section 3, we present copula-based spatial extreme models. In Section 4, we introduce a new nonparametric copula framework, a DP copula. In Section 5, we present some simulation studies to evaluate the performance of the new nonparametric model proposed here. In Section 6, we apply our methods to maximum annual temperature data. We finish in Section 7 with some conclusions and final remarks.

# 2 Measures of spatial dependency for extremes

We assume $X_t(s)$, the recorded maximum temperature amount at location $s$ on year $t$, follows a marginal GEV distribution. The GEV distribution function at each site $s$ in a given domain $D$, is given by

$$F_s(x;\mu_t,\sigma_t,\xi_t) = \exp\left[-\left\{1+\frac{\xi_t(s)(x-\mu_t(s))}{\sigma_t(s)}\right\}_+^{-1/\xi_t(s)}\right], \tag{1}$$

where $\mu_t(s)$ is the location parameter, $\sigma_t(s)$ is the scale, and $\xi_t(s)$ is the shape. The GEV distribution includes three distributions as special cases (Fisher and Tippett, 1928): the Gumbel distribution if $\xi_t(s) = 0$, the Fréchet distribution with $\xi_t(s) > 0$, and the Weibull with $\xi_t(s) < 0$. The distribution's domain also depends on $\xi_t(s)$; the domain is $(-\infty, \infty)$ if $\xi_t(s) = 0$, $(\mu_t(s) - \sigma_t(s)/\xi_t(s), \infty)$ if $\xi_t(s) > 0$, and $(-\infty, \mu_t(s) - \sigma_t(s)/\xi_t(s))$ if $\xi_t(s) < 0$. In modelling extreme (air and sea) temperature values using GEV marginals, we typically obtain negative shape parameters (see, eg. Gilleland and Katz, 2006)

We model the GEV parameters $\mu_t(s)$, $\sigma_t(s)$, and $\xi_t(s)$ in (1) as a spatial processes that characterize spatial nonstationarity in the marginal distribution of the temperature extremes. An example of one of the models used for the GEV parameters is

$$\mu_t(s) = \alpha_\mu(s) + \beta_\mu(s)t,$$

where $\alpha_\mu(s)$ and $\beta_\mu(s)$ are spatial processes with a covariance that is a function of the distance between stations and other parameters. In Section 5 we discuss in more detail the structure and models used for the spatial GEV parameters.

We use the copula framework to characterize the spatial dependence in the extreme temperatures that is left after accounting for the spatial structure of the GEV parameters. We refer to this as *residual spatial dependence*, though note that these are not residuals in the sense of explaining the structure left after removing a trend. Furthermore, we will estimate simultaneously the copula parameters and the GEV parameters. Standard models for spatial extremes tend to assume that conditioning on the GEV parameters we have spatial independence, in this work we would like to characterize this potential spatial dependence that remains in the model after accounting for spatial GEV parameters. We define the GEV residuals, $Y_t(s)$, using the following probability integral transformation

$$Y_t(s) = \left\{1 + \frac{\xi_t(s)(X_t(s) - \mu_t(s))}{\sigma_t(s)}\right\}^{1/\xi_t(s)}, \tag{2}$$

After this transformation, $Y_t(s)$ has a standard Fréchet distribution function ($F_s(y) = e^{-1/y}$). To simplify notation, throughout this paper we describe our GEV residuals using standard Fréchet distributions, but in the application section, as part of our hierarchical Bayesian framework we use the relationship between $Y_t$ and $X_t$ given in (2) at any given location $s$, to obtain the GEV distributions with space-dependent parameters. For the GEV residuals, since we have replications across time, to simplify the notation throughout the next sections, we write $Y(s)$ dropping the subindex $t$.

After the integral transformation, the resulting process $Y_t(s)$ has stationary marginals, but its spatial dependence is not necessarily stationary. The scope of the paper is to introduce a very innovative model to be able to characterize nonstationary spatial dependence for a process that has marginals GEV with spatial-varying parameters. In the next section we present measures to characterize complex extreme dependence, allowing for nonstationarity and threshold-specific dependence structure.

## 2.1 Extremal coefficient

The association between extreme events is often summarized, rather than using a correlation function, using the extremal coefficient. If the vector $(Y(s_1), \ldots, Y(s_m))'$ follows an $m-$variate extreme value distribution where the univariate margins are identically distributed, the extremal coefficient, $\vartheta$, between sites $s_1, \ldots, s_m$ is given by

$$P(\max(Y(s_1), \ldots, Y(s_m)) < u) = (P(Y(s_1) < u))^{\vartheta}$$

for all $u \in \mathcal{R}$. The extremal coefficient was introduced by Smith (1990). The extremal coefficient $\vartheta$ can be interpreted as the number of independent variables involved in an $m-$variate distribution. $\vartheta$ takes values in $[1, m]$ where $\vartheta = 1$ refers to complete dependence, and $\vartheta = m$ to independence. Commonly, extremal coefficients are used for max-stable processes, in which the extremal coefficient does not depend on $u$. However, even for non-max-stable processes, the extremal coefficient is an intuitive way to summarize the relationship between extreme values via joint exceedence probabilities.

Consider the joint distribution for $Y(s_1)$ and $Y(s_2)$ with a extremal coefficient that satisfies

$$P(\max(Y(s_1), Y(s_2)) < u) = (P(Y(s_1) < u))^{\vartheta(u)}, \tag{3}$$

and $\vartheta(u) = \vartheta(1)$, for all $u$. Then, we name $\vartheta(u)$ a *threshold-independent extremal coefficient*. In contrast, if $\vartheta(u)$ depends on $u$ it is called in this paper a *threshold-specific extremal coefficient*. Max stable processes cannot have threshold-specific dependence structure (Beirlant et al., 2004, p255).

To illustrate the potential need for models that would allow for a threshold-specific dependency, in Figure 1 we plot the estimated extremal coefficient for annual maximum temperatures in Florida (FL) for what we define *warm years* versus *cold years* within the period from 1978 to 2007. For this exploratory analysis, the extremal coefficient was estimated using the spatial Gaussian coupla in (5) fitted using a Bayesian approach. The 95% bands presented in this graph, are from the posterior distribution of the extremal coefficient. Here, we took the 30 annual maximum temperatures at each site and averaged across space to obtain a value for each year, thus warm years are the 15 years with the largest spatial-average maximum temperature values, and the cold years are the other 15 years. The significant difference in the extremal coefficient for the different type of years illustrates the potential need of models like the one introduced in this paper in Section 4 that allow for a threshold-specific dependence structure in the extremes for the temperature data. Our model is flexible enough to accommodate different type of dependence for different types of years simultaneously, without using a rule to split the years into different groups.

The spatial dependency structure of extremes may change with location. We introduce a nonstationary extremal coefficient function to characterize nonstationary spatial dependency structures in extremes. For notational simplicity, we temporarily assume the extremal coefficient is threshold-independent. We define a *stationary* extremal function, $\theta(s_1, s_2)$, as the extremal coefficient between locations $s_1$ and $s_2$, that depends on $s_1$ and $s_2$ only through their vector distance $s_1 - s_2$, for any $s_1, s_2 \in D$. Thus,

$$P(\max(Y(s_1), Y(s_2)) < 1) = (P(Y(s_1) < 1))^{\theta(s_1, s_2)},$$

and there is a function $\theta_0$, such that,

$$\theta(s_1, s_2) = \theta_0(s_1 - s_2).$$

This stationary extremal function was introduced by Schlather and Tawn, 2003. Here, we extend this function to a nonstationary setting. A extremal function $\theta(s_1, s_2)$ that can not be reduced to a function of $s_1 - s_2$ is called in this paper a *nonstationary extremal function*.

Figure 2 presents the pairwise estimated extremal coefficient function (and 95% posterior bands) for annual maximum temperatures using data from Georgia (GA) and Tennessee (TN), using the spatial Gaussian copula in (5). The extremal functions are significantly different for both states, which indicates lack of stationarity in the spatial dependence of extreme temperatures. The extremal coefficient is plotted as a function of distance and also threshold. The results suggest stronger spatial correlation in TN than in GA for the extreme temperatures at larger distances. In GA the extremal function takes the value 2 (independence) after few kilometers, which is a reflection of having a more spatial heterogeneous geographic domain than in TN, in particular due to the presence of a large city, Atlanta.

In Section 4 we introduce a nonparametric extension of the copula approach presented in Section 3 that can be used to generate non-stationary dependence structure in extremes and threshold-specific extremal functions.

# 3 Copula-based multivariate extreme models

## 3.1 Spatial Gaussian copula

The Gaussian copula function (e.g. Nelsen, 1999) is defined as $C_\rho(u,v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$, where $u, v \in [0,1]$, $\Phi$ denotes the standard normal cumulative distribution function (CDF), and $\Phi_\rho$ denotes the CDF of the standard bivariate Gaussian distribution with correlation parameter $\rho$ (off-diagonal element in the covariance function). If we use a Gaussian copula to characterize the bivariate dependence structure between extremes at two locations $s_1$ and $s_2$, then, we have $(Y(s_1), Y(s_2)) \stackrel{\mathrm{d}}{=} (G^{-1}\Phi(Z(s_1)), G^{-1}\Phi(Z(s_2)))$ where $Z(s_1)$ and $Z(s_2)$ are standard normal r.v.s with correlation $\rho$, and $G^{-1}$ is the inverse marginal distribution function for $Y(s_1)$ and $Y(s_2)$. The distribution function of $(Y(s_1), Y(s_2))$ is given by $H(Y(s_1), Y(s_2), \rho) = \Phi_\rho\left(\Phi^{-1}G(Y(s_1)), \Phi^{-1}G(Y(s_2)\right)$. The marginal distribution of $Y(s_1)$ and $Y(s_2)$ remains $G$.

We generalize the bivariate case (with two sites, $s_1$ and $s_2$), to a set of sites $\{s_1, \ldots, s_m\}$, using a spatial copula; good references for multivariate copulas are Joe (1997), and Nelsen (1999). The spatial copula introduces a latent Gaussian process $Z(s)$ with mean zero, unit variance, and spatial covariance $\mathrm{cov}(Z(s_1), Z(s_2)) = \rho_Z(s_1, s_2)$. $F_Z$ denotes the multivariate distribution function $MVN(0, \Sigma)$, where $\Sigma = [\rho_Z(s_i, s_j)]_{i,j=1}^m$, of the spatial process $Z$. Then $T(s) \stackrel{(\mathrm{def})}{=} \Phi(Z(s)) \sim$ Unif(0,1). To relate the latent and data processes, let $G$ be the CDF of the standard Fréchet distribution. Then,

$$Y(s) = G^{-1}(T(s)) \sim G. \tag{4}$$

$T(s)$ determines the $Y(s)$'s percentile, and since the $T(s)$ have spatial correlation matrix (via $Z(s)$), the outcomes also have spatial correlation. Given the covariance function $\rho_Z$ of the latent process $Z$ we can derive the Gaussian copula $C_Z$ for the distribution function of $Z$

$$C_Z(u_1, \ldots, u_m) = F_Z(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_m)),$$

where $(u_1, \ldots, u_m) \in [0,1]^m$. Let $F_Y$ denote the multivariate distribution of $Y$, then

$$F_Y(y_1, \ldots, y_m) = C_Z(G(y_1), \ldots, G(y_m)) = F_Z(\Phi^{-1}G(y_1), \ldots, \Phi^{-1}G(y_m)), \tag{5}$$

where $(y_1, \ldots, y_m) \in \mathcal{R}^m$.

If the spatial covariance $\rho_Z$ is stationary, i.e. $\text{cov}(Z(s_1), Z(s_2)) = \rho_Z(s_1 - s_2)$ then the resulting extremal function between $s_1$ and $s_2$ will be also stationary. Since,

$$\theta(s_1, s_2) = \theta_0(s_1 - s_2) = -\log\left(F_Z(\Phi^{-1}G(1), \Phi^{-1}G(1)\right),$$

only depends on $s_1$ and $s_2$ through its vector distance, because $F_Z$ has a stationary covariance.

If the spatial covariance $\rho_Z$ is nonstationary, then the resulting extremal function is non-stationary. The extremal function could be also made threshold-specific, by calculating the $\vartheta(s_1, s_2; u)$ function that satisfies (3), then, we have,

$$\vartheta(s_1, s_2; u) = -u\log\left(F_Z(\Phi^{-1}G_{s_1}(u), \Phi^{-1}G_{s_2}(u)\right)).$$

## 3.2 Limiting copula for the Gaussian copula

Consider iid random vectors $\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(N)}$, where $\mathbf{Y}^{(i)} = (Y^{(i)}(s_1), \ldots, Y^{(i)}(s_m))$, with joint marginal distribution function $F$, and define $\mathbf{M}_N$ the vector of the componentwise maxima (the $j^{\text{th}}$ component of $\mathbf{M}_N$ is the maximum of the $j^{\text{th}}$ component over all $N$ observations). We say that $F$ is in the maximum domain of attraction (eg. Nasri-Roudsari, 1996) of the distribution function $H$, if there exist sequences of vectors $\mathbf{A}_N > 0$ and $\mathbf{B}_N \in \mathcal{R}^d$, such that

$$\lim_{N\to\infty} P\left(\frac{M_{N,1} - B_{N_1}}{A_{N,1}} \leq y_1, \ldots, \frac{M_{N,m} - B_{N_m}}{A_{N,m}} \leq y_m\right) = \lim_{n\to\infty} F^N(\mathbf{A}_N\mathbf{y} + \mathbf{B}_N) = H(\mathbf{y}).$$

A non-degenerate limiting distribution $H$ is known as a multivariate extreme value distribution. The unique copula $C_0$ of the limit $H$ is an extreme value (EV) copula.

The Gaussian copula, for $\rho < 1$, is attracted to an independent EV copula (eg. Demarta and McNeil, 2004). It is straightforward to calculate the extremal coefficient for the independent EV copula, we have, $\vartheta(s_1, s_2; u) = 2$, for all values of $u$ and all pair of locations $s_1$ and $s_2$. Then, based on this asymptotic result, when a spatial Gaussian copula is used to characterize the distribution of extreme values, this distribution may not offer much flexibility to characterize complex dependence in the tails.

In Figure 3, we plot the extremal coefficient function for a Gaussian copula, $\vartheta_\rho(s_1, s_2; u)$, evaluated at different values of $u$ and $\rho$. On the horizontal axis we plot the quantiles levels $e^{-1/u}$. When $\rho = 1$, the distribution is degenerate, and $\vartheta_1(s_1, s_2; u) = 1$ for all values of $u$, in contrast, $\rho = 0$, corresponds to the independent case and the extremal coefficient is always 2. Similar to the independent case, for $\rho = 0.5$, $\vartheta_{0.5}(s_1, s_2; u)$, converges to 2 for large values of $u$. The asymptotic theory refered in this Section shows this will be the case for all $\rho < 1$.

If the pairwise extremal coefficients equal 2, then, the extremal coefficient for $m$ variables equals necessarily $m$ (Tiago de Oliveira, 1975). Thus, the multivariate (spatial) Gaussian copula may not be able to characterize complex tail spatial dependence structures, since asymptotically it does not allow for tail dependence.

# 4 A Dirichlet process copula model

In this section we introduce an extension of the spatial Gaussian copula model that is more flexible and should better capture the phenomenon of dependent extreme values. Instead of assuming a Gaussian distribution for the copula or any other distribution (e.g. a $t$-distribution), we introduce a Bayesian nonparametric representation of the copula. Bayesian nonparametric methods avoid dependence on parametric assumptions by working with probability models on function spaces, in other words, by using infinitely-many parameters. The Bayesian framework allow us to properly characterize the uncertainty in the specification of the distribution of the latent process, and be able to make valid inference.

In particular, we use the spatial Dirichlet process (DP) priors, described in the next section. The Gaussian copula is a particular case, but we allow for other distributions beyond normal. This DP model provides a random joint distribution for a stochastic process of random variables. The fitting of this type of stick breaking (SB) model is fairly straightforward using Markov chain Monte Carlo (MCMC) methods. In the Appendix we review the Dirichlet process and SB models.

## 4.1 Spatial Dirichlet process

A random probability distribution, $F$, has a stick-breaking prior (Sethuraman, 1994) if

$$F \stackrel{d}{=} \sum_{i=1}^{K} p_i \delta_{\theta_i}, \tag{6}$$

where $\delta_z$ denotes a Dirac measure at $z$, $p_1 = V_1$, $p_i = V_i \prod_{j<i}(1 - V_j)$ where $V_1, \ldots, V_{K-1}$ are independent with $V_i \sim \text{Beta}(a_i, b_i)$ and $\theta_1, \ldots, \theta_K$ are independent draws from a centering (or base) distribution $H$. A particular case of the SB prior is the Dirichlet process prior,

characterized by $\nu H$, where $H$ is the base distribution and $\nu$ is a positive scalar (often called the mass or spread parameter), the DP prior arises when $V_i$ follows a Beta$(1, \nu)$ for all $i$.

In order to make this wide class of nonparametric priors useful for our spatial context, we need to index it by space. More generally, we can attempt to introduce dependencies on time or other covariates (leading to nonparametric regression models). Most of the (rather recent) literature in this area follows the ideas in MacEachern (1999), who considered allowing the masses, $V = (V_1, V_2, \dots)$, or the so-called "location parameters", $\theta = (\theta_1, \theta_2, \dots)$, of the atoms to follow a stochastic process defined over the domain. This leads to so-called Dependent Dirichlet (DDP) processes and a lot of this work concentrates on the "single-$p$" DDP model where only the locations, $\theta$, follow stochastic processes. An application to spatial modelling is developed in Gelfand et al. (2005) by allowing the location parameters $\theta$ to be drawn from a random field (a Gaussian process). Other spatial extensions are introduced by Griffin and Steel (2006), Reich and Fuentes (2007), Dunson and Park (2008), and An et al. (2009).

The idea in Gelfand et al. (2005) is to introduce a spatial dependence through the locations, by indexing $\theta$ with the site $s$ and making $\theta(s)$ a realization of a random field, with $H$ being the distribution of a stationary Gaussian process. The joint density of the transformed residuals $\mathbf{Z} = (Z(s_1), \dots, Z(s_m))$ is a location mixture of normals with density function of the form

$$\sum_{i=1}^{K} p_i N_m(\mathbf{Z}|\boldsymbol{\theta}_i, \tau^2 \mathbf{I}_m), \tag{7}$$

using (6), where $\boldsymbol{\theta}_i = (\theta_i(s_1), \dots, \theta_i(s_m))$, and $N_m(\mathbf{Z}|\boldsymbol{\theta}_i, \tau^2\mathbf{I}_m)$ denotes a m-dimensional normal density function evaluated at $\mathbf{Z}$, with mean $\boldsymbol{\theta}_i$ and covariance matrix $\tau^2\mathbf{I}_m$. By varying $\boldsymbol{\theta}_i$ the density function representation in (7) allows for a large amount of flexibility.

## 4.2 DP copula model

The spatial Dirichlet process copula introduces a latent process $Z$, such that in year $t$, for $t = 1, \dots, T$, the joint density of $\mathbf{Z} = (Z(s_1), \dots, Z(s_m))$ at $m$ locations $(s_1, \dots, s_m)$, *given*, $H^m$, the m-random probability measure of the spatial part and $\tau^2$, the nugget component, $f(\mathbf{Z}|H^m, \tau^2)$, is almost surely of the form (a countable mixture of normals)

$$f_Z = \sum_{i=1}^{\infty} p_i N_m(\mathbf{Z}|\boldsymbol{\theta}_i, \tau^2 \mathbf{I}_m), \tag{8}$$

where $N_m(.|\lambda, \Sigma)$ denotes the m-variate normal density with mean vector $\lambda$, and covariance $\Sigma$, the vector $\boldsymbol{\theta}_i = (\theta_i(s_1), \dots, \theta_i(s_m))$, $p_1 = V_1$, $p_i = V_i \prod_{j<i}(1 - V_j)$, $V_i \sim Beta(1, \nu)$,

$$\boldsymbol{\theta}_i | H^m \sim_{ind} H^m,$$

and $H^m = DP(\nu H_0^m)$, $H_0^m = N_m(.|\mathbf{0}_m, \Sigma)$. We denote $F_Z$ the distribution of $Z$ associated to the density in (8).

Then, $T(s) = H_s(Z(s)) \sim \text{Unif}(0,1)$, where $H_s$ is the CDF corresponding to the density in (8) for $Z(s)$. The copula $C_Z$ for the distribution function of $Z(s_1), \ldots, Z(s_m)$ is (conditioning on the $\boldsymbol{\theta}_i$ components),

$$C_Z(u_1, \ldots, u_m) = F_Z(H_{s_1}^{-1}(u_1), \ldots, H_{s_m}^{-1}(u_m)),$$

where $u_1, \ldots u_m \in [0,1]^m$. Then, $Y(s) \sim G^{-1}(T(s)) \sim G$. $G$ is the CDF of the standard Fréchet distribution. Using the relationship in (2), we allow the marginal distributions to be GEVs with space-dependent parameters, by incorporating a change of variable ($Y$ to $X$ in (2)) in the likelihood function. The multivariate distribution of $Y$ is

$$F_Y(y_1, \ldots, y_m) = C_Z(G(y_1), \ldots, G(y_m)).$$

*Spatial dependence for spatial extremes using the DP copula.*

The spatial dependence induced by $F_Z$ (conditioning on the $\boldsymbol{\theta}_i$ components) is in general nonstationary, we have

$$\text{cov}(Z(s_i), Z(s_j)|\boldsymbol{\theta}_i, \tau^2\mathbf{I}_m) = \sum_l p_l\theta_l(s_i)\theta_l(s_j) - \left\{\sum_l p_l\theta_l(s_i)\right\}\left\{\sum_l p_l\theta_l(s_j)\right\}.$$

Then, the resulting extremal function is flexible enough to allow for nonstationary

$$\vartheta(s_1, s_2) = -\log\left(F_Z(H_{s_1}^{-1}G(1), H_{s_2}^{-1}G(1))\right),$$

for all $u$, since the covariance in $F_Z$ can be nonstationary. The extremal function is also threshold-specific,

$$\vartheta(s_1, s_2; u) = -u\log\left(F_Z(H_{s_1}^{-1}G(u), H_{s_2}^{-1}G(u))\right).$$

Since $F_Z$ is a multivariate distribution, the results above can be extended and calculated simultaneously for any number of sites $\{s_1, \ldots, s_m\}$.

In Figure 3 we present the extremal coefficient $\vartheta$ for different quantiles levels, $P[Y < u] = e^{-1/u}$, for a copula that is a mixture of normal distributions. This is a simplified version of the mixture copula proposed in this section, we present it here as an illustration of the flexibility that this mixture copula framework offers to explain tail dependence structures. There are three mixture copula densities in Figure 3 with $\sum_{j=1}^{K} p_j \text{N}_2(x|\mu_j\mathbf{1}_2, \Sigma_j)$, where $x$ is a 2-vector, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$ is a K-dimensional vector, the weights are $p_j = 1/K$, and $\Sigma_j$ is the $2 \times 2$ correlation matrix with off-diagonal element $\rho_j$. For two of these three mixtures, $K = 10$, and $\boldsymbol{\mu}$ is a 10-dimensional vector with equally spaced values between $-3$ and 3, and they have in one case $\boldsymbol{\rho} = (0, \ldots, 1)$, where $\boldsymbol{\rho}$ is a 10-dimensional vector with equally spaced values between 0 and 1, such as the $\text{j}^{th}$ component of $\boldsymbol{\rho}$ is $\rho_j$, and in the other $\boldsymbol{\rho} = (1, \ldots, 0)$. In one case the extremal coefficient increases for large values of $u$, while in the other we have the reverse situation. In Figure 3 we present another mixture with $K = 2$, mean zero,

and $\boldsymbol{\rho} = (0, 1)$. In this case the extremal coefficient for large values of $u$ coverges to 1.5 (as justified next), a situation a Gaussian copula could never characterize.

Let us calculate the pairwise dependence between $Y(s_1)$ and $Y(s_2)$ using the mixture in Figure 3 with $K = 2$. Denote $P[Y(s_i) < u] = p = e^{-1/u}$ for $i = 1, 2$. We have $u = \Phi^{-1}G(u)$, and $P[Y(s_1) < u, Y(s_2) < u] = .5p(1 + p)$. Therefore,

$$\vartheta(u) = \frac{\log(F(\Phi^{-1}G(u), \Phi^{-1}G(u)))}{\log(F(\Phi^{-1}G(u)))} = \frac{\log(.5p(1+p))}{\log(p)}$$

And, by l'Hopital rule

$$\lim_{u\to\infty} \vartheta(u) = \lim_{p\to\mathbf{1}} \vartheta(u) = \lim_{p\to\mathbf{1}} \frac{(1+2p)/(p(1+p))}{1/p} = 1.5$$

Thus, for this particular mixture we obtain that the limiting value of the extremal coefficient for large quantile levels is 1.5.

As we increase the number of mixture componentes ($K$) we allow for more flexibility in the tail dependence, ultimately, in the mixture presented in this Section with $K = \infty$, we can obtain all different type of shapes for the extremal coefficient as a function of $u$. Though in practice, it might be useful to consider finite approximations to the infinite stick-breaking process. Dunson and Park (2008) study the asymptotic properties of truncation approximations to the infinite mixture, while Papaspiliopoulos and Roberts (2008) introduce an elegant computational approach to work with an infinite mixture for Dirichlet processes mixing. From a practical point of view, working with a finite number of mixture components should be enough to characterize complex spatial patterns and dependence structures, see e.g. Reich and Fuentes (2007). The mixture aspect of the model (having more than one component) is what facilitates the dependence structure to be so flexible, as clearly presented in Figure 4. In particular, having in the mixture a term that is a Gaussian process with a covariance that has an off-diagonal value of 1 ($\rho = 1$), makes $\vartheta \in (1, 2)$.

# 5 Simulation study

In this section we conduct a simulation study to illustrate the effect of modeling the joint spatial distribution on estimation the marginal GEV parameters. For each simulated data set, we generate $n = 20$ spatial locations randomly on $[0, 1]^2$, and then generate $T = 50$ independent (over time, not space) replications of the spatial process. The marginal GEV parameters with linear trend are denoted

$$\begin{aligned} \mu_t(s) &= \alpha_\mu(s) + U_t\beta_\mu(s) \\ \log(\sigma_t(s)) &= \alpha_\sigma(s) + U_t\beta_\sigma(s) \\ \xi_t(s) &= \alpha_\xi(s) + U_t\beta_\xi(s) \end{aligned} \tag{9}$$

where $U_t$ is the temporal variable, $t$, standardized by subtracting the sample mean of the values {1,...,T} and dividing by their sample standard deviation. We generate $M = 100$ data sets from each of six simulation designs:

1. $\alpha_\mu(s) = 0$; $\beta_\mu(s) = 1/2$; $Z_t \sim N(0, \Sigma(\rho_1))$
2. $\alpha_\mu(s) = 0$; $\beta_\mu(s) = 1/2$; $Z_t \sim N(0, \Sigma(\rho_2))$
3. $\alpha_\mu(s) = 0$; $\beta_\mu(s) = 1/2$; $Z_t|g_t \sim N(\mu_{g_t}, \Sigma(\rho_{g_t}))$, $P(g_t = g) = 1/3$ for $g \in \{1, 2, 3\}$
4. $\alpha_\mu(s) = \left(s_1 + \sqrt{s_1 s_2} - 1\right)$; $\beta_\mu(s) = \frac{1}{2}\left(s_2 + \sqrt{s_1 s_2} - 1\right)$; $Z_t \sim N(0, \Sigma(\rho_1))$
5. $\alpha_\mu(s)\alpha_\mu(s) = \left(s_1 + \sqrt{s_1 s_2} - 1\right)$; $\beta_\mu(s) = \frac{1}{2}\left(s_2 + \sqrt{s_1 s_2} - 1\right)$; $Z_t \sim N(0, \Sigma(\rho_2))$
6. $\alpha_\mu(s) = \left(s_1 + \sqrt{s_1 s_2} - 1\right)$; $\beta_\mu(s) = \frac{1}{2}\left(s_2 + \sqrt{s_1 s_2} - 1\right)$; $Z_t|g_t \sim N(\mu_{g_t}, \Sigma(\rho_{g_t}))$, $P(g_t = g) = 1/3$ for $g \in \{1, 2, 3\}$

where $s = (s_1, s_2)$, $\Sigma(\rho)$ is the $n \times n$ covariance matrix of the latent process $Z_t$ with $(i, j)$ element $\exp(-||s_i - s_j||/\rho)$, $\mu = (-2, 0, 2)$, $\rho_j = (0.01, 0.3, 1)$, and for all designs $\alpha_\sigma(s) = 1$, $\alpha_\xi(s) = 0.1$, and $\beta_\sigma(s) = \beta_\xi(s) = 0$. The first three designs have the same marginal GEV distribution at each location, but different copulas used to generate the data. The first design is the Gaussian copula with a weak spatial correlation, the second design is the Gaussian copula with moderate spatial correlation, and the third design is a mixture of three normals, with a different spatial correlation for each mixture component. Designs 4-6 have spatially-varying location parameters.

For each data set we fit three copula models: the independent copula ("Indep"), the usual single-component Gaussian copula ("Gauss"), and the DP mixture copula of Section 4.4 ("DP"). The so-called here independent copula, would correspond to the null model, a model without residual dependence. The Gaussian copula could be considered a particular case of the DP mixture copula with just one component. In terms of the parameters added by the DP mixture, every time we add a mixture component we add an extra spatial process to be estimated. For all copula models, we model the GEV parameters as constant across space for Designs 1-3, and allow the location parameters, but not the scale or shape parameters, to vary spatially for Designs 4-6. The GEV parameters held constant across space have $N(0, 10^2)$ priors. The GEV parameters allowed to vary spatially have Gaussian process priors with $E(\alpha_\mu(s)) = \bar{\alpha}_\mu$ and $Cov(\alpha_\mu(s), \alpha_\mu(s')) = \tau^2_{\alpha_\mu} \exp(-||s - s'||/\rho_{\alpha_\mu})$, with $\bar{\alpha}_\mu \sim N(0, 10^2)$ and $\tau^2_{\alpha_\mu} \sim$ InvG(0.1,0.1). We use the same prior model introduced here for $\alpha_\mu(s)$ for the other GEV space-dependent parameters in (9). All the spatial range parameters $\rho$ have U(0,10) priors. For the DP model we take the spread parameter $\nu \sim$ Gamma(1,1), and we use 10 terms in the DP mixture model. The priors are independent.

For each model we compute the posterior mean of the GEV parameters and we obtain the mean square error. We present here the MSE for $\alpha_\mu(s)$. We denote $\hat{\alpha}_\mu(s)^{(sim)}$ the posterior

mean of $\alpha_\mu(s)$ for data set number $sim$, and compute the mean square error

$$MSE_{\alpha_\mu} = \frac{1}{Mn} \sum_{sim=1}^{M} \sum_{i=1}^{n} (\alpha_\mu(s_i) - \hat{\alpha}_\mu(s_i)^{(sim)})^2 \tag{10}$$

where $\alpha_\mu(s)$ is the true value used to generate the data. In addition, we report the coverage probabilities of the 95% intervals, averaged across space and simulated data set. The results are given in Tables 1 and 2.

The first design has weak spatial association, and all three model have similar MSE and coverage. This illustrates that the complex spatial models can reduce to the independence copula if appropriate. The non-spatial model has large MSE and small coverage for the spatial data generated by Designs 2 and 3, and the usual Gaussian copula performs poorly for Design 3 with non-Gaussian latent spatial data. For example, the MSE for the location intercept $\alpha_\mu$ is 2.80 (0.39) for the DP copula, compared to 5.86 (0.79) for the independent copula, and 6.46 (0.81) for the Gaussian copula. also, the coverage probabilities for $\alpha_\mu$ are 0.91 for the DP copula, compared to 0.56 for the independent copula and 0.74 for the Gaussian copula. Therefore, failing to adequately model the underlying spatial process can adversely affect estimation of the GEV parameters, and thus estimates of return levels.

Designs 4-6 have spatially-varying location parameters. The results for all three models are similar for Design 4 with weakly-correlated residuals. For Designs 5 and 6 with spatially correlated residuals, the copula models do not reduce MSE compared to the independence model. This may be due to a lack of identifiability between the spatially-varying location parameters and the spatially-correlated residuals. However, the copula models have considerably smaller MSE for the shape and scale parameters and generally have higher coverage probability than the independence copula model. For the final design with a non-Gaussian latent spatial process, the nonparametric Bayesian model outperforms the usual Gaussian copula model.

Convergence is monitored using trace plots of the deviance and several parameters. We generated a data set from design 6 and fit the full Bayesian DP copula model. The trace plots of the deviance (measure of overall model convergence) and the spatial range $\rho_k$ of the covariance for the spatially-varying coefficients (the most challenging parameter in terms of convergence) are presented in Figure 4. The deviance converges after 1000 iterations, which the burn-in used in these simulations.

The trace plots for the range paramters are parameterized in terms of the correlation between points separated by 0.5, i.e., $\exp(-0.5/\rho)$. In Figure 4 we plot these trace plots for $\rho_{\alpha_\mu}$ and $\rho_{\beta_\mu}$ which are the range of the covariance for the intercept and slope parameters of the spatially-varying GEV location parameter. The prior 95% interval (interval calculated with the prior distributions) for this is (0.14, 0.95), so there is significant Bayesian learning.

Table 1: 100*MSE (SE) for the simulation study.

| Design | Model | Location int ($\alpha_\mu$) | Scale int ($\alpha_\sigma$) | Shape int ($\alpha_\xi$) | Location slope ($\beta_\mu$) | Scale slope ($\beta_\sigma$) | Shape slope ($\beta_\xi$) |
|---|---|---|---|---|---|---|---|
| 1 | Indep | 1.29 (0.16) | 0.08 (0.01) | 0.06 (0.01) | 1.28 (0.18) | 0.09 (0.01) | 0.07 (0.01) |
| 1 | Gauss | 1.28 (0.16) | 0.08 (0.01) | 0.06 (0.01) | 1.27 (0.17) | 0.09 (0.01) | 0.07 (0.01) |
| 1 | DP | 1.30 (0.17) | 0.08 (0.01) | 0.06 (0.01) | 1.27 (0.17) | 0.09 (0.01) | 0.07 (0.01) |
| 2 | Indep | 5.07 (0.72) | 0.19 (0.03) | 0.14 (0.02) | 5.54 (0.89) | 0.24 (0.04) | 0.12 (0.02) |
| 2 | Gauss | 4.59 (0.69) | 0.17 (0.02) | 0.09 (0.01) | 4.65 (0.73) | 0.18 (0.03) | 0.06 (0.01) |
| 2 | DP | 4.59 (0.68) | 0.17 (0.02) | 0.09 (0.01) | 4.67 (0.75) | 0.18 (0.03) | 0.06 (0.01) |
| 3 | Indep | 5.86 (0.79) | 0.29 (0.04) | 0.13 (0.02) | 5.54 (0.86) | 0.34 (0.05) | 0.17 (0.03) |
| 3 | Gauss | 6.46 (0.81) | 0.31 (0.04) | 0.17 (0.03) | 7.06 (1.17) | 0.49 (0.08) | 0.21 (0.03) |
| 3 | DP | 2.80 (0.39) | 0.20 (0.02) | 0.08 (0.01) | 2.11 (0.32) | 0.16 (0.02) | 0.11 (0.02) |
| 4 | Indep | 5.03 (0.27) | 0.08 (0.01) | 0.05 (0.01) | 4.66 (0.21) | 0.09 (0.01) | 0.06 (0.01) |
| 4 | Gauss | 4.94 (0.28) | 0.08 (0.01) | 0.05 (0.01) | 4.68 (0.19) | 0.09 (0.01) | 0.06 (0.01) |
| 4 | DP | 4.91 (0.27) | 0.08 (0.01) | 0.05 (0.01) | 4.61 (0.19) | 0.08 (0.01) | 0.06 (0.01) |
| 5 | Indep | 9.90 (0.76) | 0.30 (0.04) | 0.13 (0.02) | 8.88 (0.61) | 0.24 (0.03) | 0.13 (0.02) |
| 5 | Gauss | 12.74 (1.50) | 0.32 (0.05) | 0.08 (0.01) | 9.90 (0.70) | 0.14 (0.02) | 0.06 (0.01) |
| 5 | DP | 11.98 (1.30) | 0.36 (0.06) | 0.09 (0.01) | 9.95 (0.69) | 0.15 (0.02) | 0.06 (0.01) |
| 6 | Indep | 10.25 (0.83) | 0.42 (0.06) | 0.20 (0.03) | 10.28 (0.86) | 0.28 (0.04) | 0.18 (0.02) |
| 6 | Gauss | 15.32 (1.15) | 0.43 (0.06) | 0.21 (0.03) | 12.57 (1.01) | 0.37 (0.05) | 0.14 (0.02) |
| 6 | DP | 8.63 (0.73) | 0.27 (0.04) | 0.10 (0.01) | 9.42 (0.92) | 0.20 (0.03) | 0.07 (0.01) |

Table 2: Coverage probabilities for the simulation study.

| Design | Model | Location int ($\alpha_\mu$) | Scale int ($\alpha_\sigma$) | Shape int ($\alpha_\xi$) | Location slope ($\beta_\mu$) | Scale slope ($\beta_\sigma$) | Shape slope ($\beta_\xi$) |
|---|---|---|---|---|---|---|---|
| 1 | Indep | 0.90 | 0.94 | 0.95 | 0.91 | 0.92 | 0.93 |
| 1 | Gauss | 0.90 | 0.94 | 0.95 | 0.93 | 0.92 | 0.91 |
| 1 | DP | 0.90 | 0.93 | 0.95 | 0.93 | 0.93 | 0.90 |
| 2 | Indep | 0.58 | 0.75 | 0.80 | 0.64 | 0.73 | 0.85 |
| 2 | Gauss | 0.92 | 0.97 | 0.90 | 0.92 | 0.92 | 0.94 |
| 2 | DP | 0.93 | 0.98 | 0.92 | 0.91 | 0.94 | 0.95 |
| 3 | Indep | 0.56 | 0.72 | 0.82 | 0.63 | 0.64 | 0.79 |
| 3 | Gauss | 0.74 | 0.83 | 0.78 | 0.80 | 0.67 | 0.76 |
| 3 | DP | 0.91 | 0.92 | 0.95 | 0.95 | 0.94 | 0.91 |
| 4 | Indep | 0.96 | 0.95 | 0.97 | 0.97 | 0.95 | 0.97 |
| 4 | Gauss | 0.97 | 0.96 | 0.98 | 0.98 | 0.94 | 0.97 |
| 4 | DP | 0.97 | 0.96 | 0.97 | 0.97 | 0.95 | 0.97 |
| 5 | Indep | 0.84 | 0.64 | 0.86 | 0.88 | 0.76 | 0.82 |
| 5 | Gauss | 0.88 | 0.90 | 0.93 | 0.89 | 0.93 | 0.97 |
| 5 | DP | 0.89 | 0.89 | 0.93 | 0.90 | 0.95 | 0.96 |
| 6 | Indep | 0.84 | 0.60 | 0.69 | 0.83 | 0.73 | 0.76 |
| 6 | Gauss | 0.86 | 0.71 | 0.78 | 0.87 | 0.73 | 0.82 |
| 6 | DP | 0.89 | 0.87 | 0.93 | 0.89 | 0.92 | 0.97 |

# 6 Application

Extreme temperature events may cause loss of life, injury, property damage, and threaten the existence of some species. Observed and projected climate change has direct implications for the occurrence of these extreme temperature events, and extreme temperatures may be more responsible for changes in natural and human systems than changes in the average weather (Parmesan et al., 2000). The recent report of the government's Climate Change Science Program (CCSP, 2008) states that the greatest impacts of climate change on society and wildlife will be experienced through changes in extreme weather events as global temperatures increase (Vliet and Leemans, 2006). So, the frequency and intensity of many temperature extremes is now changing. For example, in recent decades most of North America has experienced more unusually hot days (Houghton et al, 2001). Systems tend to adapt to their historical range of extremes, in the meantime the impacts of these extreme events are more likely to have negative as opposed to positive impacts on human and biological systems. Thus, it is of paramount importance for climate change adaption planning to accurately quantify this historical range (distribution) of extreme temperature events and monitor its evolution.

The climate models described in the Intergovernmental Panel on Climate Change (IPCC) First Assessment Report (Mitchell et al., 1990) showed that a warmer mean temperature increases the probability of extreme warm days and decreases the probability of extreme cold days. This result has appeared consistently in a number of more recent different climate model configurations (Dai et al., 2001; Yonetani and Gordon, 2001). Using global climate deterministic models, in North America the greatest increase in the 20-year return values of daily maximum temperature (IPCC third assessment report), is found in central and southeast North America, where there is a decrease in soil moisture content. In this paper we study extremes for maximum daily temperatures in this subdomain of interest, south-east-central U.S., and we obtain maps of 20 and 50 year return values, using Bayesian spatial statistical modelling frameworks, rather than climate models. We also present the uncertainty in the obtained return-value maps. In our analysis we allow for nonstationarity across space and time. The probability of an extreme event under nonstationary conditions is going to depend on the rate of change of the distribution as well as on the rate of change of the frequency of their occurrence. Under these nonstationary conditions, the concept of the return period or return level is altered, since the value is highly dependent on the extrapolated period of consideration.

## 6.1 Data

Our application uses surface air daily maximum temperature data produced by the National Climatic Data Center (NCDC) in Asheville, NC. The online data files are available at www.ncdc.noaa.gov/cgi-bin/res40.pl?page=gsod.html.

In this section, we study temperature extremes in the east-south-central and south Atlantic United States over a 30 year period from 1978 to 2007. More specifically, daily surface air temperature records were obtained over the years 1978−2007 from 60 stations located in Alabama (AL), Florida (FL), Georgia (GA), Kentucky (KY), Mississippi (MS), and Tennessee (TN). These stations are shown in Figure 5, and are located within the region with the greatest increase in 20-year return values of daily maximum temperature according to the IPCC Third Assessment Report "Climate change 2001".

In the following section we apply the Dirichlet process copula proposed in this paper to the extreme temperature data, and we compare its performance to the Gaussian copula and to the spatial varying GEV model (independent copula).

## 6.2 DP copula approach for extreme temperatures

We assume that the annual maximum temperature at location $s$ for year $t$, $X_t(s)$, for $t = 1, \ldots, 30$ and $s \in \{s_1, \ldots, s_{60}\}$, follows a GEV distribution with location parameter $\mu_t(s)$, scale parameter $\sigma_t(s)$ and shape parameter $\xi_t(s)$. Thus, we have a marginal GEV distribution for $X_t(s)$ with parameters that are spatial functions

$$\begin{aligned} \mu_t(s) &= \alpha_\mu(s) + \beta_\mu(s)U_t, && (11)\\ \log \sigma_t(s) &= \alpha_\sigma(s) + \beta_\sigma(s)U_t, \text{ and} && (12)\\ \xi_t(s) &= \alpha_\xi(s) + \beta_\xi(s)U_t. && (13) \end{aligned}$$

$U_t$ is the standarized temporal covariate. The GEV spatially-varying coefficient $\alpha_\mu(s)$ in the location parameter has a spatial Gaussian prior with $\mathrm{E}(\alpha_\mu(s)) = \bar{\alpha}_\mu$ and $\mathrm{Cov}(\alpha_\mu(s), \alpha_\mu(s')) = \tau^2_{\alpha_\mu} \exp(-||s-s'||/\rho_{\alpha_\mu})$, with $\bar{\alpha}_\mu \sim \mathrm{N}(0, 10^2)$ and $\tau^2_{\alpha_\mu} \sim \mathrm{InvG}(0.1,0.1)$. We use the same Gaussian prior model for the other GEV spatially varying coefficients. For the spatial Gaussian copula model we introduce a latent process $Z_t$ with prior $N(0, \Sigma(\rho_{Z_t}))$, $\Sigma(\rho_{Z_t})$ is the $60 \times 60$ covariance matrix with $(i, j)$ element $\tau^2_{Z_t} \exp(-||s_i - s_j||/\rho_{Z_t})$. The latent process $Z_t$ would have a DP prior for the DP copula. All the spatial range parameters $\rho$ have U(0,1000km) priors. For the DP model we take the spread parameter $\nu \sim$ Gamma(1,1).

We conduct a 5-fold cross-validation (CV) to compare the performance of the Dirichlet process copula model to the spatial Gaussian copula model and to an independent copula, and to asses how well these models fit the annual maximum temperature described in Section 6.1. The CV is done by splitting the temperature data randomly over space and time into $g = 5$ groups. We consider six different models, summarized in Table 3 and described below, each fit using 4 of the 5 groups of data. These models are then used to predict temperature values for those locations and time points that have been removed. This is repeated so that each group is removed once. We compare the models using the MSE values (and standard errors (SE) of MSEs) between observed and predicted temperature values. Since we have a 5-fold CV, the SEs of the MSEs are obtained by calculating $S/\sqrt{(5)}$, where $S$ is the sample standard deviation of the 5 MSEs for the 5 splits.

| | $\alpha_\mu$ and $\beta_\mu$ spatial | Copula Type | MSE (SE) | Coverage Probability |
|---|---|---|---|---|
| Model 1 | No | Indep. | 3.348 (0.134) | 93.78% |
| Model 2 | No | Gaussian | 3.052 (0.128) | 94.00% |
| Model 3 | No | DP | 3.002 (0.131) | 92.28% |
| Model 4 | Yes | Indep. | 2.623 (0.127) | 94.17% |
| Model 5 | Yes | Gaussian | 1.945 (0.108) | 93.94% |
| Model 6 | Yes | DP | 1.558 (0.096) | 94.78% |

Table 3: Cross validation results using 60 sites for the annual daily max temperature data.

In Models 1-3, no parameters in (11)-(13) are varying spatially, while models 4-6 have spatially-varying location parameters, $\alpha_\mu$ and $\beta_\mu$, and all other parameters constant across space. Models 1-2 and 4-5 are Gaussian copula models with GEV marginal distributions, while Models 3 and 6 are Dirichlet process (DP) copula models with GEV marginal distributions. We approximated the DP copula density in (8) with a 3-component mixture. The GEV parameters in (11)-(13) that are held constant across space have $N(0, 10^2)$ priors, while those varying spatially have the Gaussian process spatial priors described in this Section.

From Table 3 it is clear that allowing the marginal GEV parameters to vary spatially improves prediction, regardless of the residual correlation model. MSE varies substantially and significantly by the type of copula. Model 4 which ignores residual correlation has 35% larger MSE than Model 5's Gaussian copula. Also, accounting for complex spatial relationships using the flexible DP mixture copula with spatially-varying GEV coefficients gives the smallest MSE of the models considered. This seems to suggest the need for models to characterize the residual dependence, going beyond the *independent copula* model above, after accounting for GEV spatially varying parameters, as well as the need for more flexible copula models than the spatial Gaussian.

In Figure 6 we present a summary of the posterior distributions for some of the GEV parameters using a DP copula, and also using a spatial Gaussian copula. Figure 6a shows the posterior median of the spatially-varying coefficient $\beta_\mu(s)$ in the location parameter $\mu(s)$ of the GEV marginal distribution, using the DP copula framework. There is an significantly increasing trend in the eastern and central part of our domain, in particular in FL. Figure 6b presents the posterior SD for the trend, there is higher variability for this parameter in the northern part of our domain, where there is higher elevation. The temporal trend was not significant for the scale and the shape parameters, so we just focus on the intercept coefficients for those parameters. Figure 6c shows the median of the posterior distribution of the spatially-varying intercept $\alpha_\xi(s)$ for the shape parameter, using the DP copula framework. This parameter is negative for the DP copula, corresponding to the Weibull distribution. However, it is positive for the Gaussian copula (Fréchet distribution), see Figure 6c. A positive shape parameter would correspond to a distribution that, in this case, has very heavy tails, and this causes problems with the spatial prediction, as we will see next when we calculate the return levels. For the scale parameter (graph no shown here), larger

values are obtained in the northern part of our domain, corresponding to areas with larger elevation. The estimated location and scale parameters were very similar using the DP and the Gaussian copulas, the shape parameter however, as we have seen, had opposite signs. A negative shape parameter is more consistent with the current literature and previous analysis of extreme temperatures (e.g. Gilleland and Katz, 2006).

## 6.3 Return levels for extreme temperatures

In the context of modelling extreme temperatures, it is often of interest to obtain differences of the $n$-year return levels across time. We obtain return levels here using the spatial Gaussian copula and the DP copula approaches. Figure 7 presents the posterior distributions for the 20 and 50 year return values for annual extremes of maximum daily temperatures. We use data from 1978 to 2007, and we fix $t$ (time) at the last time point. The return levels using the DP copula have very similar spatial patterns as the sample mean of the extreme temperatures presented in Figure 5. Though, the sample mean is smoother across space. The 20-year return levels using the DP copula are about 1 degree Celsius (°C) higher (Figure 7a) than the sample mean of the extreme temperatures, and the 50-year return levels are about 2 °C higher (Figure 7c). The maximum values for the return levels are obtained in the eastern and central part of our domain, eastern KN and MS, and central parts of AL, and GA, which are the areas that also have higher extreme temperatures. However the patterns were more difficult to explain using the spatial Gaussian copula, since in particular for the 50 year return levels shown in Figure 7d, lower values were concentrated in central AL, which is an area in our subdomain that has the highest extreme temperatures.

Figure 8 presents the median of the posterior distribution for the difference in the 20-year return levels for surface air temperature using the DP copula and the spatial Gaussian copula. The difference in the return values is obtained by calculating the return levels using data from 1978 - 2007, at two different values of the time covariate (t), using t= 30, and t= 20. Figure 8a presents results for the DP copula, the difference in return levels is greater and significant in MS and GA (about 0.5-1 °C), where there is a decrease in soil moisture content (IPCC third report: Houghton et al., 2001). The differences in the return levels using the spatial Gaussian copula, Figure 8b, are very variable and again difficult to interpret, we believe due to the heavy tails induced by the positive shape parameter.

To determine the necessary number of mixture components, we keep increasing $K$ until results do not change by increasing any further the value of $K$. A extensive study of the sensitivity of the results to the selection of prior distributions was conducted, in particular varying the number of mixture components $K$, considering different upper bounds on the spatial range parameters and changing the variance parameters in the GEV coefficients. In terms of the number of mixture components, once we had 3 mixture components the results were very robust, and they were not affected by increasing the number of components. The estimated return levels were no sensitive to the chosen priors for the variance parameters. For the range parameter, the results were robust to the choice of prior as long as the range

upper bond was greater than 100km.

# 7 Discussion

In this work we study the spatial structure of extreme temperature values. We introduce a modelling framework that offers a flexible approach to characterize complex spatial patterns and explain potential nonstationarity in the extremes. We present an extension of copula frameworks using Dirichlet type of mixtures. An advantange of the formulation presented in this paper using nonparametric models is that many of the tools developed for Dirichlet processes can be applied with some modifications. In terms of the computational effort and feasibility of its implementation, the DP copula and the spatial Gaussian copula offer similar challenges, since the main computational inconvenience is working with the spatial covariance matrix.

Multivariate extensions of the nonparametric spatial approaches presented here can be adopted to model simultaneously maximum and minimum extreme temperature values or other extreme weather variables, using, for instance the nonparametric spatial framework proposed by Fuentes and Reich (2008). They could also be applied to spatial daily data with Generalized Pareto marginal distributions.

# References


An, Q., Wang, C., Shterev, I., Wang, E., Carin, L, and Dunson, D. (2009). Hierarchical kernel stick-breaking process for multi-task image analysis. *International Conference on Machine Learning* (ICML), to appear.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004) *Statistics of Extremes. Theory and Applications.* John Wiley and Sons Inc, NJ, USA.

Blackwell and J.B. MacQueen. (1973). Discreteness of Ferguson selections. *Annals of Statistics,* **1**, 356-358.

Buishand, D.H.L., and Zhou, C. (2008). On spatial extremes: with application to a rainfall problem. *Annals of Applied Probability,* **2**, 624-642.

Climate Change Science Program (CCSP) (2008). Weather and Climate Extremes in a Changing Climate Regions of Focus: North America, Hawaii, Caribbean, and U.S. Pacific Islands. U.S. goverments CCSP.

Coles, S.G. (1993). Regional modeling of extreme storms via max-stable processes. *Journal of the Royal Statistical Society, B,* **55**, 797-816.

Coles, S.G., Heffernan, J., and Tawn, J.A. (1999). Dependence measures for extreme value analyses. *Extremes,* **2**, 339-365.

Coles, S.G., and Tawn, J.A. (1996). Modelling extremes of the areal rainfall process. *Journal of the Royal Statistical Society, B,* **58**, 329-347.

Cooley, D. D. Nychka, and P. Naveau. (2007) Bayesian Spatial Modeling of Extreme Precipitation Return Levels. *Journal of the American Statistical Association*, 102, 824-840.

Cooley, D., Naveau, P., and Davis, R. (2008). Dependence and spatial prediction in max-stable random fields. *University of Colorado.*

Dai, A., T.M.L. Wigley, B. A. Boville, J.T. Kiehl, and L.E. Buja. (2001). Climates of the 20th and 21st centuries simulated by the NCAR climate system model. *Journal of Climate,* **14**, 485-519.

Davison, A.C., Smith, R.L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society, B*, **15**, 393-442.

Demarta, S. and A.J. Mcneil, A.J. (2005). The t copula and related copulas. *International Statistical Review*, **73** (1), 111.

Doksum, K. (1974). Tailfree and Neutral Random Probabilities and Their Posterior Distributions. *The Annals of Probability*, **2**, 183-201

Dunson, D.B. and J. H. Park. (2008). Kernel stick-breaking processes. *Biometrika,* **95**, 307-323.

Eastoe, E.F. (2009). A hierarchical model for non-stationary multivariate extremes: a case study of surface-level ozone and $NO_x$ data in the UK. To appear in *Environmetrics.*

Eastoe, E.F., Tawn, J.A. (2009). Modelling non-stationary extremes with application to surface-level ozone. To appear in *Journal of the Royal Statistical Society, C.*

Ferguson TS (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1**, 209-230.

Fisher, R. A., and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.*, 24, 180-190.

Fuentes, M. (2001). A High Frequency Kriging for Nonstationary Environmental Processes. *Environmetrics,* bf 12, 1-15.

Fuentes, M., and Reich, B. (2008). Multivariate Spatial Nonparametric Modelling via Kernel Processes Mixing. Mimeo Series #2622 Statistics Department, NCSU.

http://www.stat.ncsu.edu/library/mimeo.html.

Gelfand A.E., Kottas A., and MacEachern S.N. (2005). Bayesian Nonparametric Spatial Modeling with Dirichlet Process Mixing. *Journal of the American Statistical Association*, **100**, 1021-1035.

Griffin J.E., and Steel, M.F.J. (2006) Order-based dependent Dirichlet processes. *Journal of the American Statistical Association*, **101**, 179-194.

Gilleland, E., and Katz, R.W. (2006). Analyzing seasonal to interannual extreme weather and climate variability with the extremes toolkit.

NCAR tech. report. www.assessment.ucar.edu/pdf/Gilleland2006revised.pdf

Houghton, J.T., Ding, Y., Griggs,D.J., Noguer, M., van der Linden, P.J., Da,X., Maskell, K., and Johnson, C.A. (2001). Climate Change 2001: The Scientific Basis. *Intergovernmental panel on climate change.* (IPCC) Third Assessment Report.

www.grida.no/publications/other/ipcc_tar/

Ishwaran, H. and James, L. (2001). Gibbs-sampling methods for stick-breaking priors, *Journal of the American Statistical Association*, **96**, 161-173.

Joe, H. (1997). *Multivariate models and dependence concepts,* London: Chapman & Hall.

Kharin, V. and Zwiers, F. (2005). Estimating extremes in transient climate change simulations. *Journal of Climate,* **18**, 1156-1173.

MacEachern, S.N. (1998). Computational methods for mixture of Dirichlet process models, in *Practical Nonparametric and Semiparametric Bayesian Statistics*, D. Dey, P. Müller and D. Sinha, eds. New York: Springer-Verlag, pp. 23–44.

MacEachern, S.N. (1999). Dependent nonparametric processes, in *ASA Proceedings of the Section on Bayesian Statistical Science*, Alexandria, VA: American Statistical Association.

Mitchell, J.F.B., S. Manabe,V. Meleshko and T. Tokioka. (1990). Equilibrium climate change and its implications for the future. In Climate Change. The IPCC Scientific Assessment. Contribution of Working Group 1 to the first assessment report of the Intergovernmental Panel on Climate Change, [Houghton, J. L, G. J. Jenkins and J. J. Ephraums (eds)], Cambridge University Press, Cambridge, pp. 137-164. Mitchell, J.F.B., T.C. Johns, J.M. Gregory and S.F.B. Tett,

Nasri-Roudsari, D. (1996). Extreme value theory of generalized order statistics *Journal of Statistical Planning and Inference,* **55**, 281-297.

Nelsen, R. (1999). *An introduction to copulas,* New York: Springer-Verlag.

Papaspiliopoulos O., and Roberts, G. (2008). Retrospective MCMC for Dirichlet process hierarchical models. *Biometrika,* **95**, 169-186.

Parmesan, C., Root, T.L., and Willing, M.R. (2000). *Bulletin of the American Meteorological Society*, **81**, 443-450.

Reich B.J., Fuentes M (2007). A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. *Annals of Applied Statistics,* **1** 249-264.

Sang, H. and Gelfand, A.E. (2009). Hierarchical Modeling for Extreme Values Observed over space and time, Environmental and Ecological Statistics (forthcoming)

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

Sethuraman, J. (1994). A constructive definition of Dirichlet priors, *Statistica Sinica*, **4**, 639–50.

Smith, R.L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript. Tech. report at University of North Carolina, Chapel Hill.

Tiago de Oliveira, J. (1975). Bivariate and multivariate extremal distribution. *Statistical distributions in scientific work,* **1**, 355-361. G. Patil et al., Dr. Reidel Publ. Co.

Vliet, A.J.H. van; Leemans, R. (2006) Rapid species responses to changes in climate require stringent climate protection targets In: Avoiding dangerous climate change / Schellnhuber, H.J., Cramer, W., Nakicinovic, N., Wigley, T., Yohe, G. Cambridge : Cambridge University Press, 135 - 143.

Zhang, J., Craigmile, P.F., and Cressie, N. (2008). Loss Function Approaches to Predict a Spatial Quantile and Its Exceedance Region. *Technometrics,* **50** (2), 216-227.

Yonetani, T. and H.B. Gordon. (2001). Simulated changes in the frequency of extremes and regional features of seasonal/annual temperature and precipitation when atmospheric CO2 is doubled. *Journal of Climate,* **14**, 1765-1779.

Figure 1: Extremal coefficient functions for the maximum temperatures in FL using a spatial Gaussian copula, plotted as a function of distance for warm years (red lines) and cold years (blue lines). In this graph we present the median of the posterior distribution for the extremal coefficient (solid lines), as well as 95% posterior bands (dashed lines).

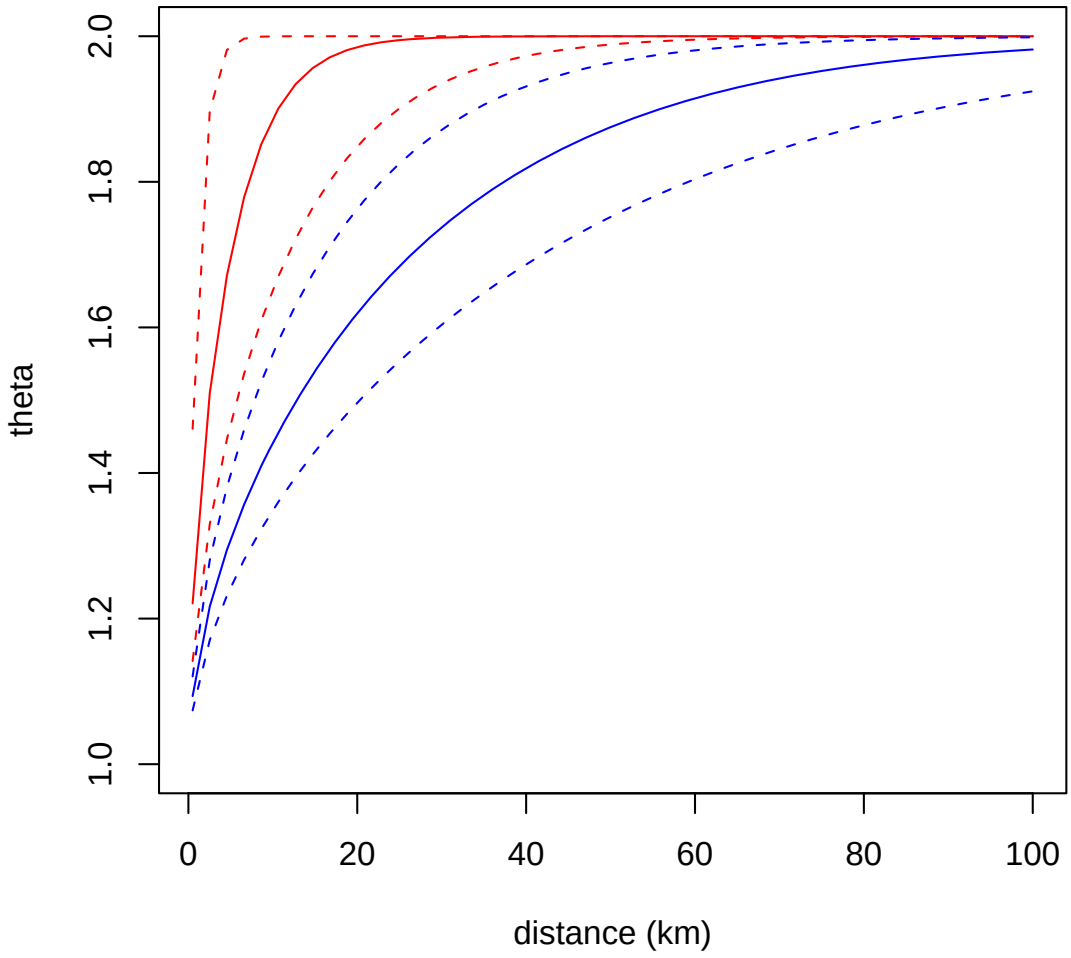


Figure 2: Extremal coefficient functions for the maximum temperatures in GA (black line) and TN (blue line) using a Gaussian copula. In this graph we present the median of the posterior distribution for the extremal coefficient (thick lines), as well as 95% posterior bands (thin lines).

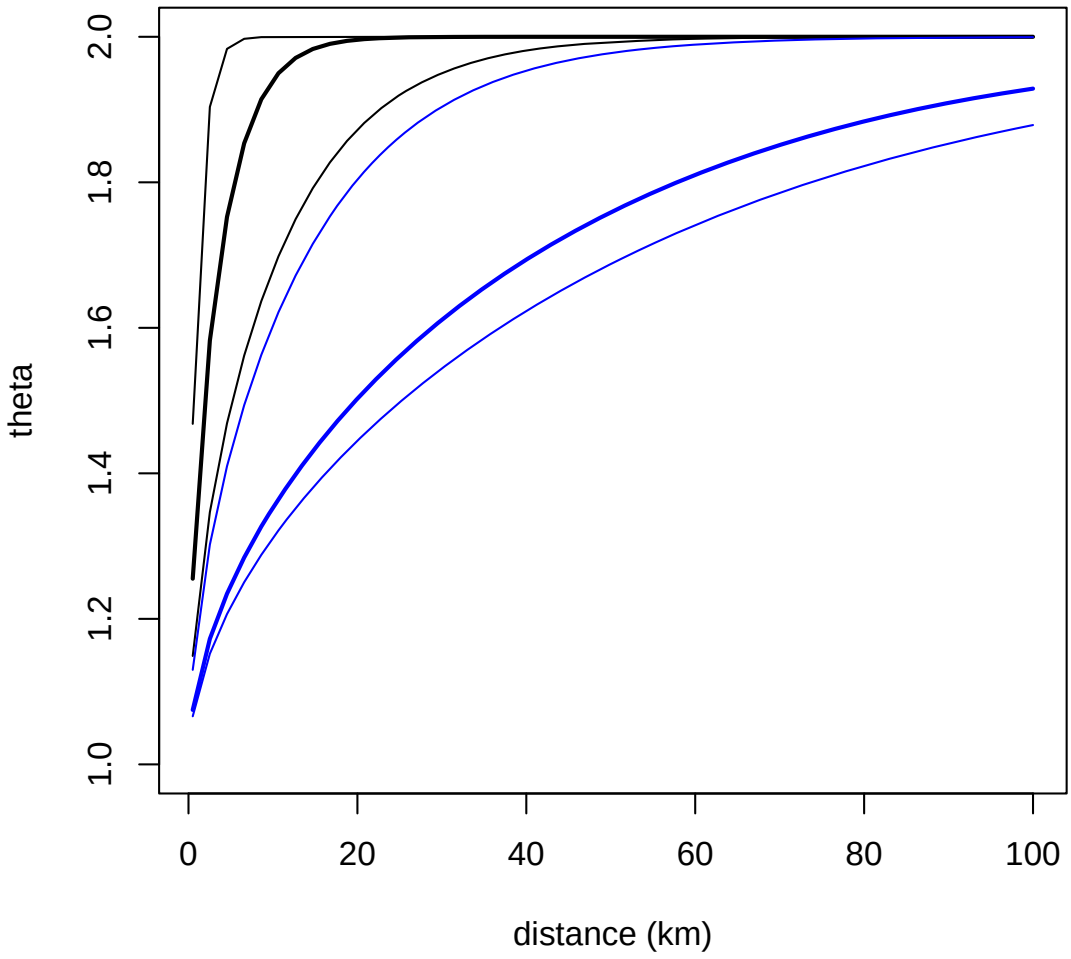


Figure 3: Extremal coefficients for different copula models with standard Fréchet marginals, versus different quantile levels. The thin lines are the extremal coefficients for Gaussian copulas with different spatial correlation parameters $\rho$. The solid lines are the extremal coefficients for mixture of normals copula with different mean and spatial correlation for each term, $\mu_k$ and $\rho(\mu_k)$, respectively.

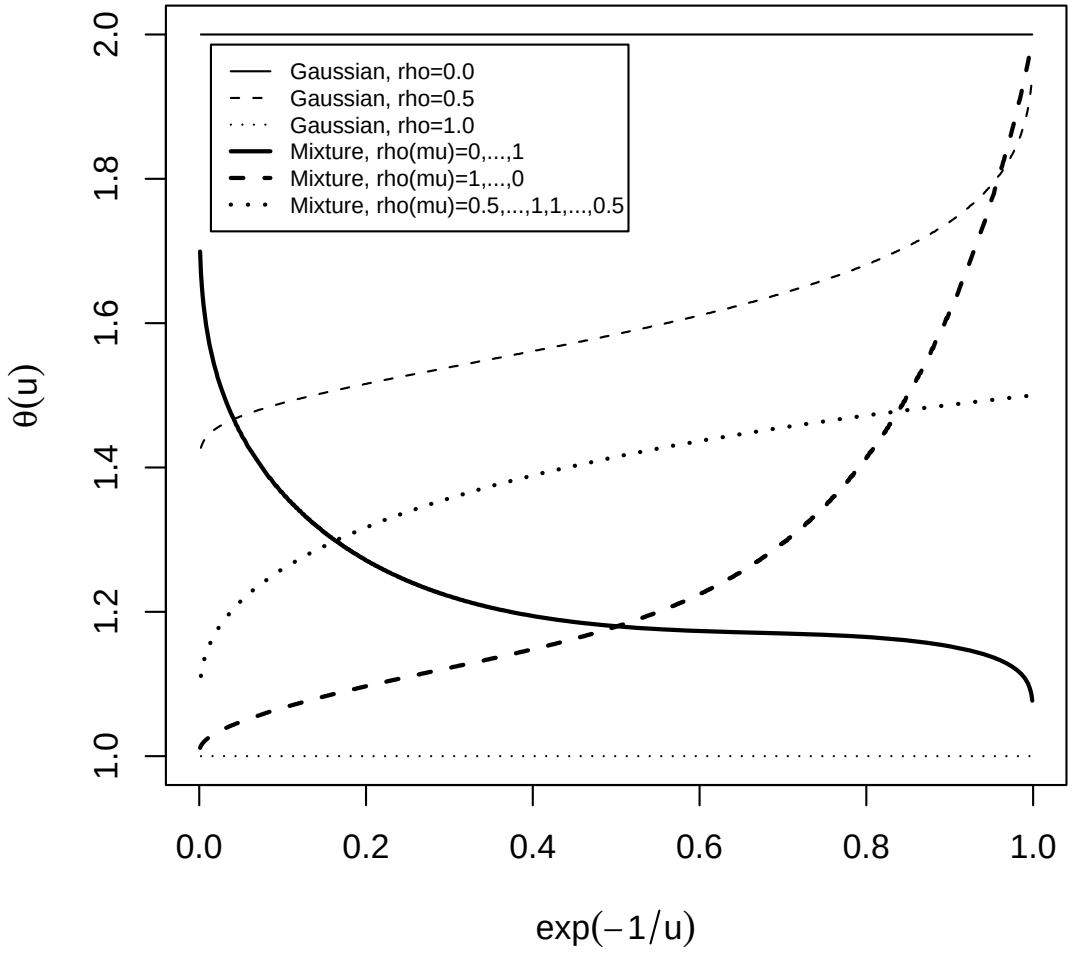
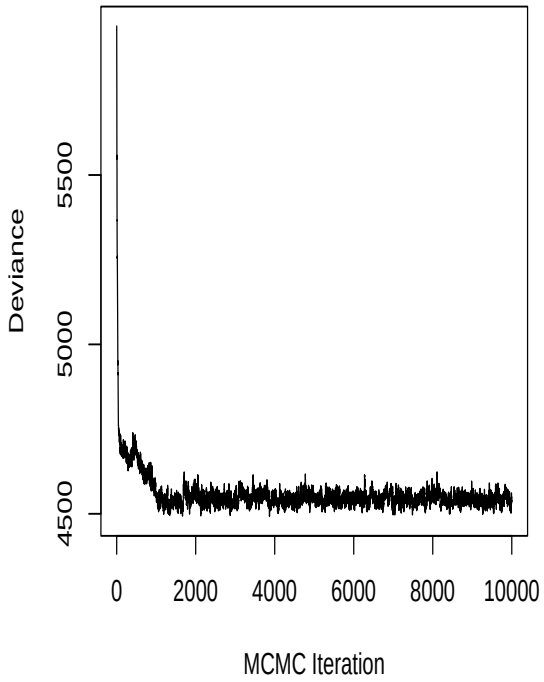
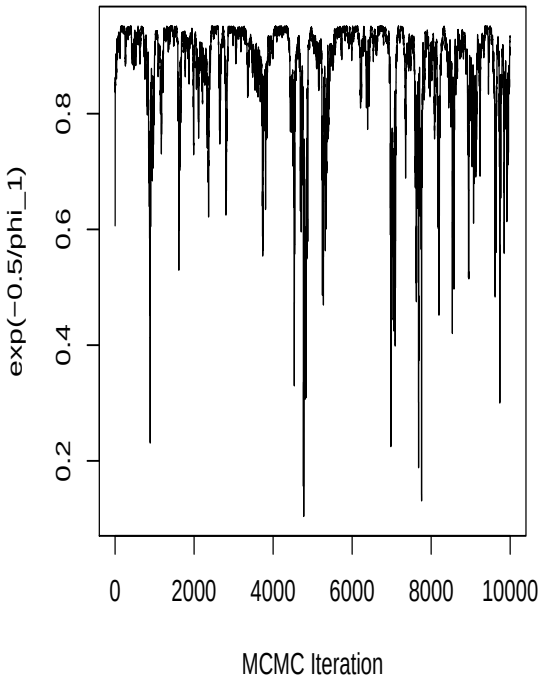


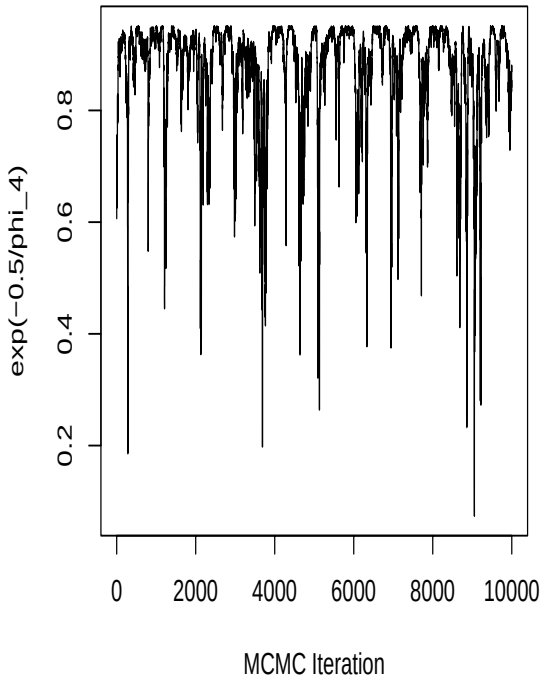
Figure 4: Monitoring convergence, trace plots for the Deviance and range parameters.




(a) Trace plot for Deviance



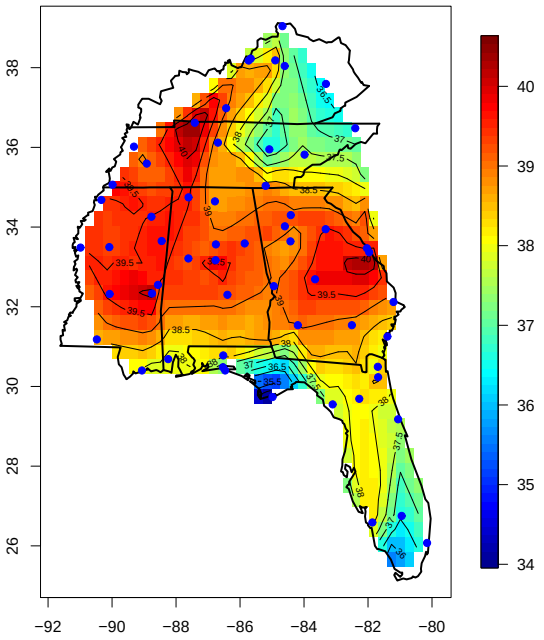

(b) Trace plot for reparameterized range parameter $\rho_{\alpha_\mu}$




(c) Trace plot for reparameterized range parameter $\rho_{\beta_\mu}$

Figure 5: Mean and SD of the yearly maximum surface air temperature values (°C) using data from years 1978-2007. The circles are the observation locations.

(a) Mean of maximum temperatures




(b) SD of maximum temperatures

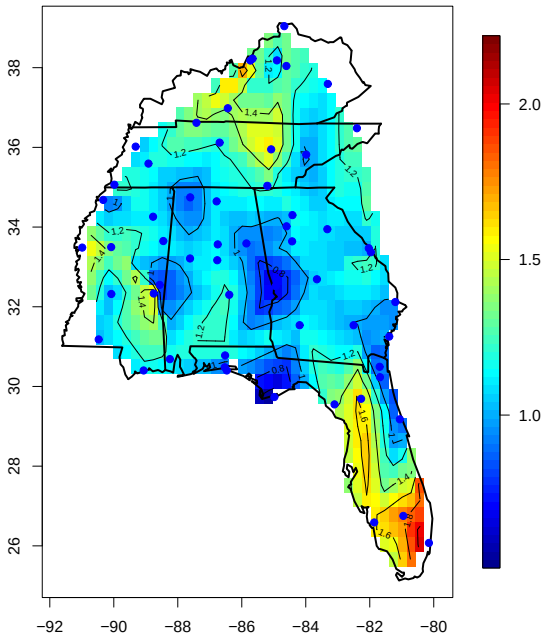
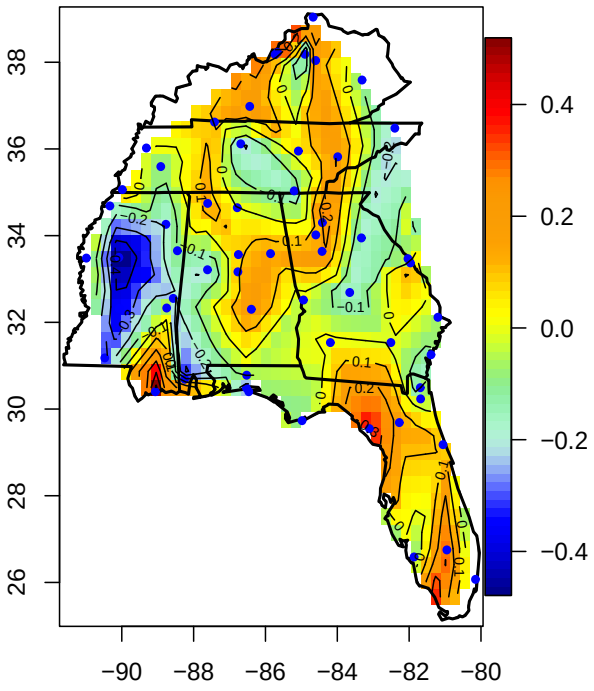
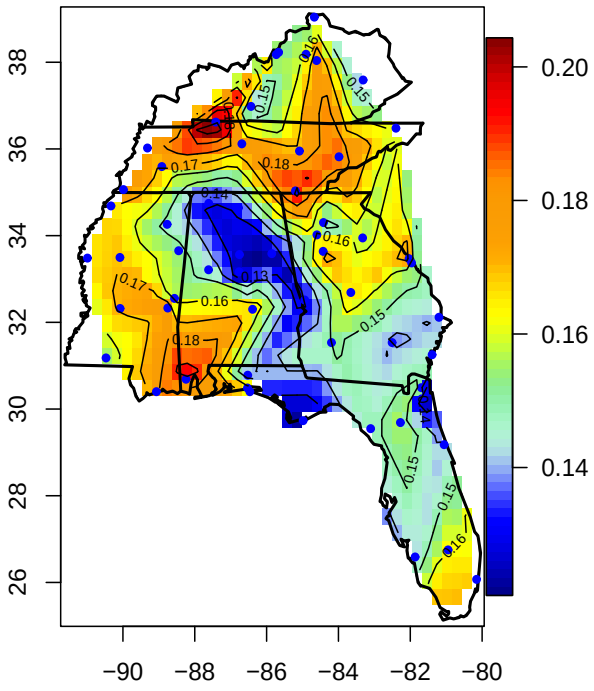


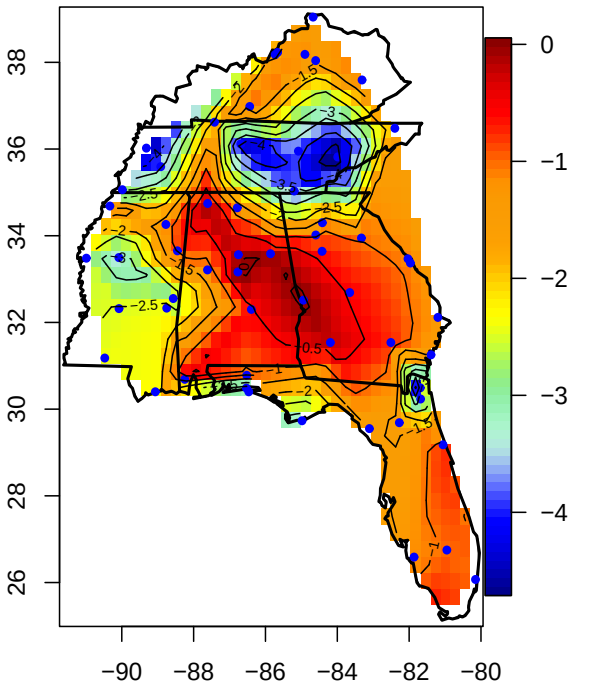
Figure 6: Graphs (a) and (b) show the median and SD of the posterior distribution of the spatially-varying coefficient $\beta_\mu(s)$ that multiplies the temporal trend $U_t$ in the location parameter of the GEV marginal distribution, using the DP copula framework. Graphs (c) and (d) show the median of the posterior distribution of the spatially-varying intercept $\alpha_\xi(s)$ for the shape parameter, using the DP copula framework (left) and Gaussian copula (right).



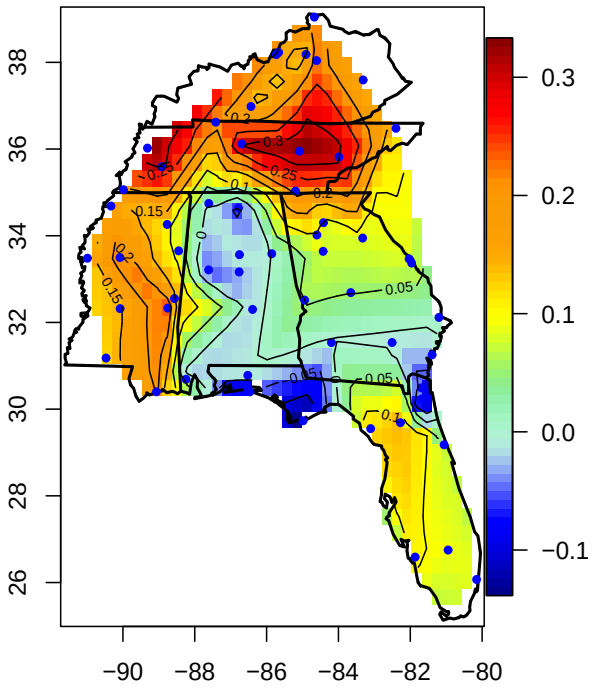

(a) Temporal trend in the location parameter.




(b) Temporal trend in the location parameter (SD).

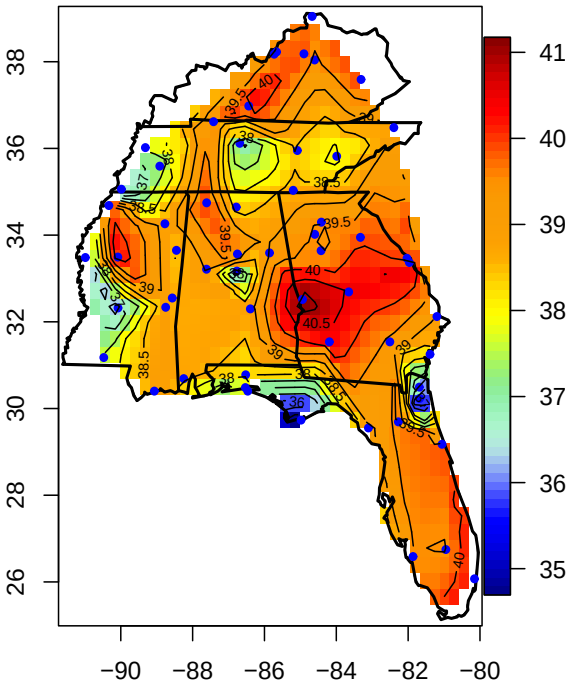



(c) Intercept for the shape parameter (DP copula).

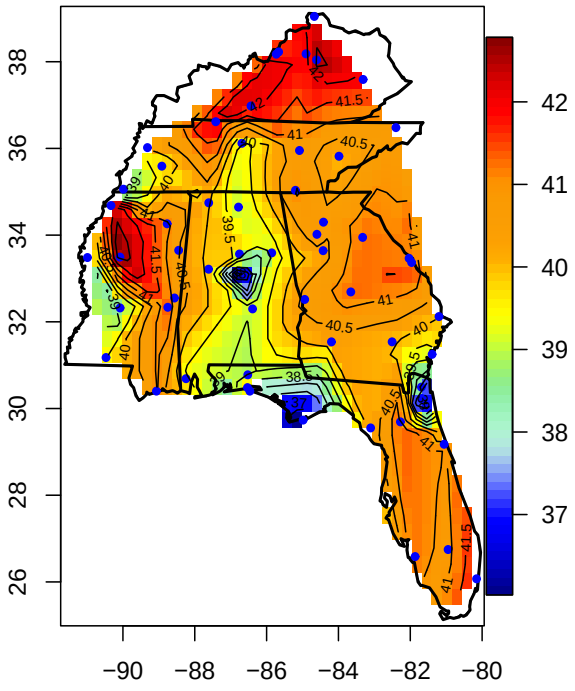



(d) Intercept for the shape (Gaussian copula).

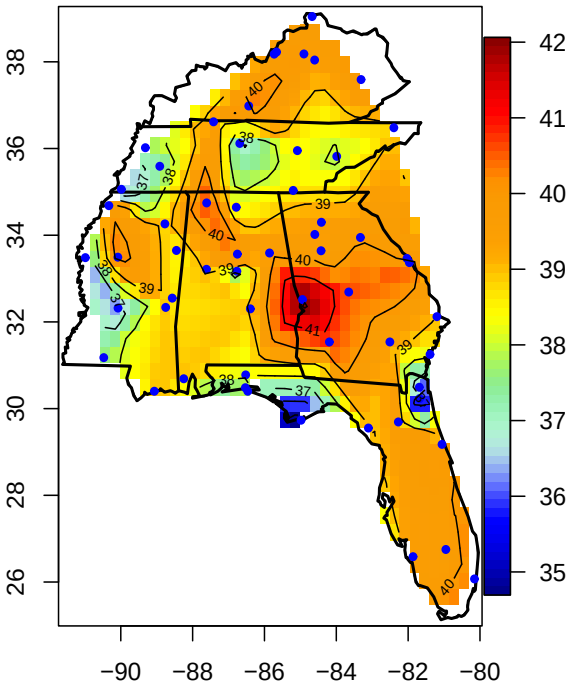
Figure 7: Posterior median for the 20 and 50 year-return (RL) levels fixing $t$ (time) at the last time point, $t = 30$. Graphs (a) and (b) show the 20 year return levels using a DP copula (left), and a spatial Gaussian copula (right). Graphs (a) and (b) show the 50 year return levels using a DP copula (left), and a spatial Gaussian copula (right).



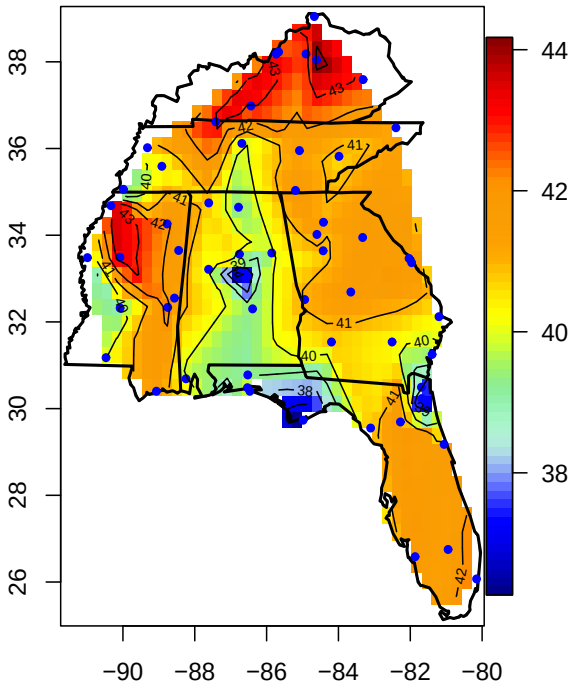

(a) 20-year RL (DP copula).




(b) 20-year RL (Gaussian Copula).

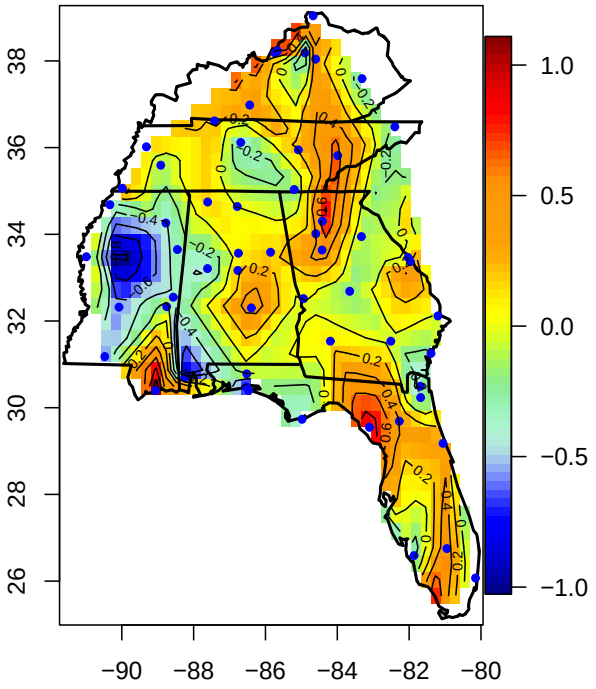



(c) 50-year RL (DP copula).

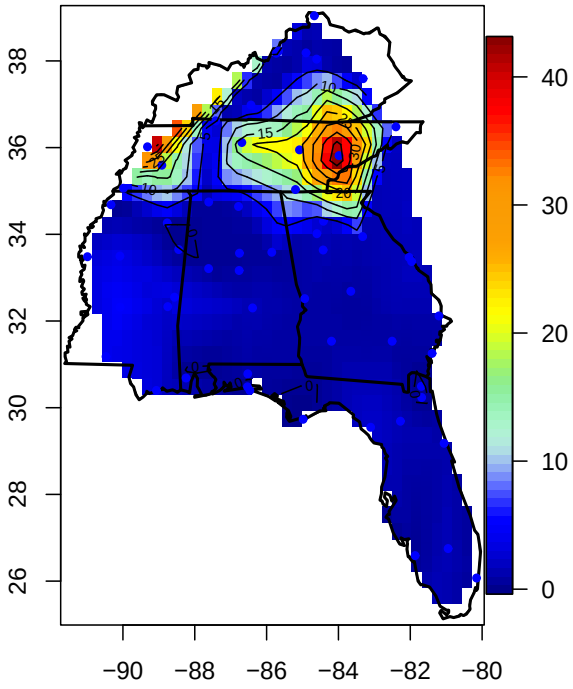



(d) 50-year RL (Gaussian copula).

Figure 8: Graphs (a) and (b) show the median of the posterior distribution for the difference in the 20-year return levels for surface air temperature (°C), using a DP copula (left) and all the available data, and using a Gaussian copula (right). The differences are obtained by calculating the return levels at two different values of the time covariate, using the temporal covariate for year 2007 (when $t = 30$), and for year 1997 (when $t = 20$).




(a) Difference in 20-year RL (DP copula).




(b) Difference in 20-year RL (Gaussian copula).

## APPENDIX A

In this appendix we review the Dirichlet process and the stick breaking prior models.

### A.1 The Dirichlet process

In this section we introduce Dirichlet processes, so we start by explaining the Dirichlet distribution. The Dirichlet distribution is the multivariate generalization of the beta distribution, and conjugate prior of the categorical distribution and multinomial distribution in Bayesian statistics. Each sample from a Dirichlet distribution is itself a distribution on some discrete probability space. Let $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$ be a probability distribution on the discrete space $\chi = \{\chi_1, \chi_2, \ldots, \chi_n\}$, such that, $P(X = \chi_i) = \theta_i$, where $X$ is a random variable in the space $\chi$. The Dirichlet distribution on $\Theta$ is given by

$$P(\Theta|\nu, M) = \frac{\Gamma(\nu)}{\prod_{i=1}^{n} \Gamma(\nu m_i)} \prod_{i=1}^{n} \theta_i^{\nu m_i - 1}$$

where $M = \{m_1, m_2, \ldots, m_n\}$ is the base measure defined on $\chi$ and is the mean value of $\Theta$, and $\nu$ is a precision parameter that explains how concentrated the distribution is around $M$. Both $\Theta$ and $M$ are proper probability distributions, i.e., $\sum \theta_i = \sum m_i = 1$.

If we replace $\theta_i$ by $\Theta(\chi_i)$, and, correspondingly, $m_i$ by $M(\chi_i)$, the Dirichlet distribution on $\chi$ can be written as

$$\Theta(\chi_1), \Theta(\chi_2), \ldots, \Theta(\chi_n) \sim Dir(\Theta, \nu M) \tag{1}$$

where $Dir(.)$ is the Dirichlet density function.

The Dirichlet process is simply an extension of the Dirichlet distribution to continuous spaces. Expression (1) implies the existence of a Dirichlet distribution on every partition of any (possibly continuous) space $\chi$. The Dirichlet Process $\Theta$, represented as $DP(\nu M)$ is the unique distribution over the space of all possible distributions on $\chi$ , such that the relation in (1) holds for every natural number $n$ and every $n$-partition $\{\chi_1, \chi_2, \ldots, \chi_n\}$ of $\chi$. Since $M$ is continuous, one might think that $\Theta$ is a continuous process. However, Blackwell and McQueen (1973) showed that Dirichlet processes are discrete, as they consist of countably infinite point probability masses. This is often not desirable for directly modelling observables that are considered realizations of some continuous process. To avoid this problem, the mixture of Dirichlet processes model (DPM), that we introduce next, is commonly used in practice.

Consider a finite mixture model of the form $Z \sim \sum_{i=1}^{K} p_i f(Z|\theta_i)$. Then, $Z$ is distributed as a mixture of distributions having the same parametric form $f$, but different parameters. The parameters $\theta_i$ are drawn from the same distribution $H_0$. This mixture model can be

expressed hierarchically as follows

$$\begin{aligned} Z|c,\Theta &\sim f(Z|\theta_c) \\ c|p_{1:K} &\sim Discrete(p_1,p_2,\ldots,p_K) \\ \theta_i &\sim H_0(\theta) \\ p_1,p_2,\ldots,p_K &\sim Dir(\nu M) \end{aligned} \tag{2}$$

Here $c$ are the indicators or labels that assign the measurements $Z$ to a parameter value $\theta_c$, and $p_i$ are the mixture coefficients drawn from a Dirichlet distribution. Given the mixture coefficients, the indicator variables are distributed multinomially. The latent indicator variables are used here only to simplify the notation. If the number of components in the mixture is known a priori, the parameters for each component can be drawn from $H_0$ beforehand, and then the Dirichlet distribution would be on $\{\theta_1,\theta_2,\ldots,\theta_K\}$. If we consider the limiting model as $K\to\infty$, then the Dirichlet distribution becomes a Dirichlet process with base measure $M$. For each indicator $c$, drawn conditioning on all the previous indicators from the Multinomial distribution, there is a corresponding $\theta_i$ that is drawn from $H_0$. In the limit as $K\to\infty$, the labels lose their meaning as the space of possible labels becomes continuous. Thus, we discard the use of labels in the model and let the parameters be drawn from a Dirichlet process with base measure $H_0$ instead. Then, the DPM model is represented as

$$\begin{aligned} Z|\theta_i &\sim f(Z|\theta_i) \\ \theta_i|H &\sim H(\theta) \\ H &\sim DP(\nu H_0) \end{aligned} \tag{3}$$

where $DP(\nu H_0)$ is the Dirichlet Process with base measure $H_0$ and spread $\nu$, and $H$ is a random distribution drawn from the DP.

An approach to the construction of a Dirichlet process prior (Fegurson, 1973) is provided by the so-called stick-breaking (SB) prior, discussed next.

### A.2 An alternative representation of the DP: Stick-breaking

An alternative representation of the DP is the SB representation, which we exploit for computation. A random probability distribution, $F$ has a SB prior if

$$F \overset{d}{=} \sum_{i=1}^{K} p_i \delta_{\theta_i}, \tag{4}$$

where $\delta_z$ denotes a Dirac measure at $z$, $p_1 = V_1$, $p_i = V_i \prod_{j<i}(1-V_j)$ where $V_1,\ldots,V_{K-1}$ are independent with $V_i \sim \text{Beta}(a_i,b_i)$ and $\theta_1,\ldots,\theta_K$ are independent draws from a centering distribution $H_0$.

The definition in (4) allows for either finite or infinite $K$ (with the latter corresponding to the conventional definition of nonparametrics). For $K=\infty$ several interesting and well-known processes fall into this class:

The Dirichlet process prior, characterised by $\nu H_0$, where $H_0$ is a distribution and $\nu$ is a positive scalar, arises when $V_i$ follows a Beta$(1, \nu)$ for all $i$.

The Pitman-Yor (or two-parameter Poisson-Dirichlet) process occurs if $V_i$ follows a Beta$(1-a, b+ai)$ with $0 \leq a < 1$ and $b > -a$. As special cases we can identify the Dirichlet process for $a = 0$ and the stable law when $b = 0$.

Stick-breaking priors such as the Dirichlet process almost surely lead to discrete probability distributions. To avoid this problem, the mixture of Dirichlet process model is now the most commonly used specification in practice. Such models assume a continuous model for the observables, given some unknown parameters, and then use a stick-breaking prior as in (4) to model these parameters nonparametrically. An important aspect of these models is that they tend to cluster the observations by assigning several observations to the same parameter values (or atoms of the nonparametric distribution).

Conducting inference with such models relies on MCMC computational methods. One approach corresponds to marginalising out $F$ and using a Polya urn representation to conduct a Gibbs sampling scheme. See MacEachern (1998) for a detailed description of such methods. Another approach (see Ishwaran et al. (2001)) directly uses the stick-breaking representation in (4) and either truncates the sum or avoids truncation through slice sampling or the retrospective sampler proposed by Papaspiliopoulos and Roberts (2008).