

Spatial modeling for risk assessment of extreme values from environmental time series: A Bayesian nonparametric approach

Athanasios Kottas ^{*}, Ziwei Wang and Abel Rodríguez

*Department of Applied Mathematics and Statistics, University of California,
Santa Cruz, U.S.A.*

Abstract: We propose an approach to modeling and risk assessment for extremes of environmental processes evolving over time and recorded at a number of spatial locations. We follow an extension of the point process approach to analysis of extremes under which the times of exceedances over a given threshold are assumed to arise from a non-homogeneous Poisson process. To achieve flexible shapes and temporal heterogeneity for the intensity of extremes at any particular spatial location, we utilize a logit-normal mixture model for the corresponding Poisson process density. A spatial Dirichlet process prior for the mixing distributions completes the nonparametric spatio-temporal model formulation. We discuss methods for posterior simulation, using Markov chain Monte Carlo techniques, and develop inference for spatial interpolation of risk assessment quantities for high-level exceedances of the environmental process. The methodology is tested with a synthetic data example and is further illustrated with analysis of rainfall exceedances recorded over a period of 50 years from a region in South Africa.

Keywords: Dirichlet process mixture model; Gaussian process; non-homogeneous Poisson process; rainfall precipitation; spatial Dirichlet process

**Correspondence to: Athanasios Kottas, Department of Applied Mathematics and Statistics, Baskin School of Engineering, MS: SOE2, University of California, 1156 High Street, Santa Cruz, CA 95064, U.S.A. E-mail: thanos@ams.ucsc.edu*

1 INTRODUCTION

Extreme value analysis plays a key role in the environmental sciences. Extreme natural phenomena, such as severe droughts, unusually low temperatures or torrential rains, are rare but catastrophic events, which can result in large economic losses and high cost in human life. Therefore, risk analysis to quantify the uncertainty associated with such extreme events is both scientifically relevant and practically important for effective environmental policy making.

In extreme value analysis for environmental problems interest lies in very large or very small values of variables associated with a physical process, which is typically recorded over both time and space. Statistical inference and prediction for rare events is complicated by the fact that observations corresponding to the center of the distribution, which are the most abundant, carry little information about the tails. This implies particular challenges for the study of dynamical variations of the process under study. For instance, rainfall records could show a steady average behavior over time, while the amount of rainfall of the largest storms may be increasing. Capturing also spatial dependence for processes observed at a number of monitoring stations adds to the challenge for modeling and inference.

The literature on modeling extremes associated with independent and identically distributed observations is fairly well developed; see, e.g., Kotz and Nadarajah (2000) and Coles (2001). With the supporting theoretical results dating to Fisher and Tippett (1928), the traditional approach is to model blockwise maxima using the generalized extreme value distribution. Another commonly used approach involves modeling the exceedances over a given threshold using a generalized Pareto distribution (Pickands, 1975; Davison and Smith, 1990). The related Bayesian literature comprises mainly parametric models (e.g., Coles and Powell, 1996; Stephenson and Tawn, 2004), an exception being Tressou (2008) where nonparametric mixtures of Pareto distributions are used to model threshold exceedances.

There is a relatively smaller collection of modeling methods for extremes from stochastic

processes evolving over time and space, although this is an active research area in the more recent literature. The Bayesian paradigm offers clear advantages in this setting, since it allows exploration of flexible hierarchical model formulations and proper incorporation of full predictive uncertainty. The main theme of Bayesian modeling approaches has been to extend in a hierarchical fashion the parametric distributions used in extreme value analysis. In particular, the observed block maxima or threshold exceedances are typically assumed to arise conditionally independent from the generalized extreme value or generalized Pareto distribution, respectively, with temporally and spatially dependent parameters. Common approaches to introduce the spatio-temporal dependence to the parameters include dynamic linear models (Huerta and Sansó, 2007) and Gaussian processes (Cooley et al., 2007; Sang and Gelfand, 2009). The approach in Sang and Gelfand (2010) fits within the same framework, but relaxes the conditional independence assumption in the first stage of the hierarchical model. Other more recent contributions include copula-based semiparametric methods (Fuentes et al., 2012) and hierarchical modeling based on max-stable processes (Reich and Shaby, 2012).

In this paper, we utilize the point process approach to analysis of extremes, an approach that encompasses the more commonly used methods based on the generalized extreme value or generalized Pareto distributions. This approach is based on a non-homogeneous Poisson process (NHPP) model for the exceedances over a high threshold and the time of their occurrence. The theoretical framework has been introduced by Pickands (1971), whereas applications can be found in, e.g., Smith (1989), Coles and Tawn (1996), and Cooley and Sain (2010).

We build on the nonparametric modeling framework from Wang et al. (2011), where a mixture model for the NHPP intensity was developed to overcome the restrictive aspects of the standard parametric form, most notably, the homogeneity for the intensity of exceedance times. Here, we focus on the time dimension under the bivariate NHPP approach and study the practically important extension of spatial modeling for the exceedance time intensities. Our objective is to retain inferential flexibility for the temporal intensity while incorporating nonparametric

spatial dependence into the modeling. To this end, we represent the NHPP density at each site through a mixture of logit-normal kernels, and use a spatial Dirichlet process for the mixing distributions to drive the nonparametric (non-Gaussian and non-stationary) spatial dependence. A prior probability model for the spatial surface of total exceedance intensities completes the model specification. We develop methods for Markov chain Monte Carlo (MCMC) posterior simulation, and for spatial interpolation of risk assessment quantities for high-level exceedances.

Our illustrative data analysis involves rainfall exceedances, using data from the Cape Floristic Region located in the southwestern coastline of South Africa. The times of exceedances are based on daily precipitation records from monitoring stations across South Africa between year 1950 to 1999. However, the proposed methodology is generally applicable to modeling and spatial prediction of threshold exceedances from different types of environmental variables.

The outline of the manuscript is as follows. In Section 2, we develop the spatial nonparametric modeling approach, including an overview of the relevant background, details of the model formulation, and methods for posterior inference. Section 3 illustrates the methodology using a simulated data example and the rainfall data. Finally, Section 4 concludes with discussion.

2 METHODS

We begin in Section 2.1 with a brief overview of the point process approach to analysis of extremes, and of a general framework for Bayesian nonparametric modeling under this approach. Section 2.2 develops the nonparametric spatial model for extremes from environmental time series. Implementation details regarding posterior simulation are discussed in Section 2.3.

2.1 Bayesian nonparametric point process modeling for extremes

Consider a sequence $\{X_j : j = 1, \dots, r\}$ of i.i.d. random variables with distribution function F_0 , where j denotes the period over which the observation on X_j is collected. If we restrict atten-

tion to the observations that fall above a given threshold u , the original sample is thinned to a bivariate point pattern, where each pair comprises the time and corresponding value of the threshold exceedance. This point pattern yields a realization from a two-dimensional point process $\{N(A) : A \subset \mathcal{A} = \{1, \dots, r\} \times [u, \infty)\}$. Pickands (1971) showed that the limiting form of this point process as $u \rightarrow \infty$ is a bivariate NHPP with intensity function $\sigma^{-1} \{1 + \xi \sigma^{-1}(y - \mu)\}_+^{-1/\xi-1}$, where $z_+ = \max\{z, 0\}$. Here, μ , σ and ξ are location, scale, and shape parameters, respectively, and y is the argument for the threshold exceedance. The shape parameter ξ is determined by the tail behavior of F_0 . In particular, if F_0 has polynomial tails then $\xi > 0$, in which case we say that F_0 is in the Fréchet domain of attraction.

Note that the point process approach provides a comprehensive framework for extreme value analysis, since it implies both the generalized extreme value distribution for blockwise maxima and the generalized Pareto for the conditional distribution of threshold exceedances. Moreover, it has the practically important advantage of using information about all data points above the given threshold, rather than just one value per time period. However, a key restriction of the parametric modeling framework arises from the form of the asymptotic NHPP intensity, which is homogeneous in time, whereas recording threshold exceedances often induces clustering.

In Wang et al. (2011), we have extended the point process approach for modeling exceedances from a general stochastic process $\{X_t : t \in [0, T]\}$ evolving over time and observed in a (bounded) time interval. Hereinafter, we will use $[0, 1]$ to represent the time period of interest; inference on the original time scale can be obtained through straightforward transformation. The pairs $\{(t_i, Y_i) : i = 1, \dots, n\}$, where t_i is the time at which the i -th exceedance over threshold u occurred and $Y_i \equiv X_{t_i}$ is the value of the i -th exceedance, are again taken to be a realization from a NHPP with support on $\mathcal{A} = [0, 1] \times [u, \infty)$. However, to provide more flexible inference than the asymptotic parametric form, a nonparametric mixture model is formulated for the NHPP intensity function, $\lambda(t, y)$. The approach is based on methodology originally developed in Kottas (2006) and Kottas and Sansó (2007) for temporal and spatial NHPP intensities, respectively,

and more recently extended by Taddy and Kottas (2012) to marked NHPPs.

The model formulation utilizes the representation of a NHPP intensity function through a density function and a parameter that defines the total intensity over the observation window. Specifically, $\lambda(\cdot) = \gamma f(\cdot)$, where $\gamma \equiv \int_{\mathcal{A}} \lambda(t, y) dt dy$ is the total intensity of exceedances, and $f(\cdot)$ is a density function on \mathcal{A} , which fully controls the shape of the intensity function. Hence, a rich prior for the NHPP intensity can be constructed through a nonparametric mixture model for the NHPP density, $f(t, y) \equiv f(t, y; G) = \int k(t, y | \theta) dG(\theta)$. Here, $k(t, y | \theta)$ is a kernel density on \mathcal{A} indexed by parameter vector θ , and G is a random mixing distribution.

The choice of the Dirichlet process (DP) prior (Ferguson, 1973) for the mixing distribution G results in a DP mixture model for $f(t, y; G)$ yielding access to well established prior model properties and methods for posterior simulation. We write $G \sim \text{DP}(\alpha, G_0)$ to denote that a DP prior is assigned to random distribution G , where G_0 is the DP centering distribution and α controls how close G is to G_0 ; large values of α result in small variability in DP realizations. The DP constructive definition (Sethuraman, 1994) is revealing of its structure and will also be key for the later development of the nonparametric spatial model. According to this definition, if $G \sim \text{DP}(\alpha, G_0)$, G admits an (almost sure) representation of the form $\sum_{l=1}^{\infty} w_l \delta_{\vartheta_l}$, where the point masses, $\{\vartheta_1, \vartheta_2, \dots\}$, form an i.i.d. sample from G_0 , and the corresponding weights arise through a stick-breaking construction. Specifically, $w_1 = v_1$ and for $l \geq 2$, $w_l = v_l \prod_{r < l} (1 - v_r)$, where $\{v_1, v_2, \dots\}$ is another i.i.d. sample from a $\text{Beta}(1, \alpha)$ distribution (also, independent of the $\{\vartheta_1, \vartheta_2, \dots\}$ sequence). The discreteness of DP realizations is an asset in the context of non-parametric mixing for applications where clustering of the observations is practically relevant as in, e.g., density estimation, classification, and regression. In particular, the precision parameter α controls the number of effective distinct mixture components (e.g., Escobar and West, 1995). For instance, for density estimation problems with moderately large sample sizes, n , a useful approximation to the prior expectation for the number of clusters is given by $\alpha \log\{(\alpha + n)/\alpha\}$.

Regarding the choice of the DP mixture kernel, care is needed to balance flexible inference

and the implied tail behavior of the underlying stochastic process marginal distributions. Wang et al. (2011) argue for a product kernel form $k(t, y) = k_1(t)k_2(y)$, with a beta density for $k_1(t)$, and $k_2(y) = \sigma^{-1} \{1 + \xi\sigma^{-1}(y - u)\}^{-1/\xi - 1}$, for $y \geq u$, i.e., a generalized Pareto density with location parameter set to the threshold value u . The DP mixing is with respect to both parameters of the beta kernel component, and the scale, $\sigma > 0$, and shape, $\xi > 0$, parameter of the generalized Pareto component. Under this specification, it can be shown that the corresponding marginal distribution for X_t belongs to the Fréchet domain of attraction, that is, the nonparametric mixture prior models an underlying stochastic process with heavy tailed behavior.

2.2 The modeling approach

The nonparametric mixture modeling framework outlined in Section 2.1 combines the appealing features of the point process approach to extreme value analysis with the inferential power of Bayesian nonparametric prior models. To our knowledge, the approach proposed in Wang et al. (2011) provides the first attempt to fully nonparametric modeling for extremes from a single time series, with flexible resulting inference for the joint intensity of extremes, the marginal intensity over time, and for different types of return level curves.

Here, we study more general spatio-temporal data structures involving threshold exceedances from environmental processes observed at multiple spatial locations over a certain time interval (which, again, without loss of generality is transformed to $[0, 1]$). More specifically, let $\mathcal{S} \subset \mathbb{R}^2$ be the geographic region under study, and $\mathbf{s}_{\text{obs}} = (\mathbf{s}_1, \dots, \mathbf{s}_m)$ the m distinct locations in \mathcal{S} where the process is observed. Hence, the full data set comprises $\{(t_i(\mathbf{s}_j), Y_i(\mathbf{s}_j)) : i = 1, \dots, n_j; j = 1, \dots, m\}$, where $n_j \equiv n_{\mathbf{s}_j}$ is the number of threshold exceedances at location \mathbf{s}_j , $t_i(\mathbf{s}_j)$ is the time at which the i -th exceedance occurred at location \mathbf{s}_j , and $Y_i(\mathbf{s}_j)$ is the value of that exceedance. For such problems, it is of interest to explore spatial modeling extensions for the NHPP intensity of extremes while retaining the flexibility of a fully nonparametric inference framework. This

is a non-trivial extension and, in this work, we take the first step in this direction by focusing on the time dimension under the point process approach.

We therefore consider only the times of threshold exceedances $\{t_i(\mathbf{s}_j) : i = 1, \dots, n_j\}$ from each observed spatial location $\mathbf{s}_j \in \mathcal{S}$, for $j = 1, \dots, m$. Following the definition of the bivariate NHPP assumed under the general approach, for any generic location $\mathbf{s} \in \mathcal{S}$, the point pattern $\{t_i(\mathbf{s}) : i = 1, \dots, n_{\mathbf{s}}\}$ is a realization from a temporal NHPP on $[0, 1]$. The corresponding temporal intensity function at location \mathbf{s} is denoted by $\lambda_{\mathbf{s}}(t)$, where this is the appropriate marginal of the bivariate NHPP intensity $\lambda_{\mathbf{s}}(t, y)$. Here, we seek to develop a nonparametric prior model for $\{\lambda_{\mathbf{s}}(t) : t \in [0, 1]; \mathbf{s} \in \mathcal{S}\}$, that is, for a collection of temporal NHPP intensities evolving over (continuous) space. The key inferential objectives are twofold: to allow general time-inhomogeneous shapes for the intensity of threshold exceedances at each specific spatial location; and to enable flexible inference for these spatially varying temporal intensities and for implied risk assessment functionals. The implicit assumption is that of a smooth evolution of the intensities across space, although the proposed model for the spatial dependence is nonparametric relaxing both of the customary assumptions of Gaussianity and stationarity.

Regarding the choice of threshold u , we view its specification as a component of scientific or policy making considerations for the particular problem at hand. Hence, for any substantive application of the methodology, the threshold would be chosen in consultation with the domain experts. The threshold value can be site-specific, and this is how we envision the model to be applied in general settings. Since the examples of Section 3 involve a small geographic region, we work with a constant threshold value across space for our illustrative data analyses, and without loss of generality, retain the non-spatially varying notation for the threshold.

2.2.1 Mixture modeling for the temporal intensity of threshold exceedances

To build the prior model for $\{\lambda_{\mathbf{s}}(t) : t \in [0, 1]; \mathbf{s} \in \mathcal{S}\}$, we follow the strategy discussed briefly in Section 2.1. In particular, for any spatial location $\mathbf{s} \in \mathcal{S}$, we utilize the decomposition of

the intensity function into the total intensity $\gamma_{\mathbf{s}}$ and the NHPP density function $f_{\mathbf{s}}(t)$ on $[0, 1]$, such that $\lambda_{\mathbf{s}}(t) = \gamma_{\mathbf{s}} f_{\mathbf{s}}(t)$. Here, $\gamma_{\mathbf{s}} = \int_0^1 \lambda_{\mathbf{s}}(t) dt$, where $\gamma_{\mathbf{s}} < \infty$ based on the NHPP definition that imposes local integrability for the intensity function. Now, for any observed point pattern $\{t_i(\mathbf{s}) : i = 1, \dots, n_{\mathbf{s}}\}$ of exceedance times at location \mathbf{s} , the NHPP likelihood can be written as

$$L(\lambda_{\mathbf{s}}(\cdot)) \equiv L(\gamma_{\mathbf{s}}, f_{\mathbf{s}}(\cdot)) \propto \exp(-\gamma_{\mathbf{s}}) \gamma_{\mathbf{s}}^{n_{\mathbf{s}}} \prod_{i=1}^{n_{\mathbf{s}}} f_{\mathbf{s}}(t_i(\mathbf{s})). \quad (1)$$

The full likelihood requires an extension of (1) to include the data from all locations, but this expression highlights the practical utility of the $(\gamma_{\mathbf{s}}, f_{\mathbf{s}}(\cdot))$ representation for the NHPP intensity. Namely, it allows us to build the model for the spatially varying intensities through a nonparametric prior model for spatially dependent densities. A prior model for the spatial surface $\{\gamma_{\mathbf{s}} : \mathbf{s} \in \mathcal{S}\}$ will also be needed, but owing to the factorization in (1), the estimation of its parameters proceeds independently of the model for the NHPP densities.

We propose a mixture model formulation for the spatially varying NHPP densities, $f_{\mathbf{s}}(t) \equiv f(t; G_{\mathbf{s}}) = \int k(t | \theta) dG_{\mathbf{s}}(\theta)$, for $t \in [0, 1]$ and $\mathbf{s} \in \mathcal{S}$. Here, $k(t | \theta)$ is the parametric kernel density supported by the unit interval, and $G_{\mathbf{s}}$ is the random mixing distribution indexed by spatial location \mathbf{s} . To meet our inferential goals, we need an appropriate kernel density that enables general, possibly multimodal shapes for the mixture density at any location, as well as a nonparametric prior model for the (uncountable) collection of mixing distributions $G_{\mathcal{S}} = \{G_{\mathbf{s}} : \mathbf{s} \in \mathcal{S}\}$ that allows flexible inference for spatial interpolation of the intensity of extremes.

Regarding the mixture kernel, the Beta distribution (used in Kottas, 2006; Wang et al., 2011) is a natural choice given the range of shapes the Beta density achieves, and the fact that it is directly bounded to $[0, 1]$. However, the lack of a conditionally conjugate distribution for the parameters of the Beta density makes implementation of posterior simulation challenging even when modeling a single density with a nonparametric mixture of Beta densities. This challenge is exacerbated in terms of both modeling and implementation of inference in our context which

involves a collection of spatially related densities. Hence, we work with a more convenient modeling platform based on a logit-normal kernel,

$$k(t \mid \theta, \tau^2) = (2\pi\tau^2)^{-1/2}t^{-1}(1-t)^{-1} \exp\{-[\log(t/(1-t)) - \theta]^2/2\tau^2\}, \quad t \in [0, 1]. \quad (2)$$

Note that this density arises through the logistic transformation, $t = \exp(z)/(1 + \exp(z))$, of a $N(\theta, \tau^2)$ density for z . As discussed below, this provides a significant advantage in the formulation of the nonparametric prior model for G_S and in MCMC posterior simulation, since we can work with a (spatially dependent) mixture of normals for the logit-transformed exceedance times. The potential drawback of the logit-normal kernel is that it is susceptible to boundary effects due to the logit transformation, $\text{logit}(t) = \log(t/(1-t))$, and the normal distribution tails. However, in practice, the nonparametric mixture structure allows robust inference under both kernel choices provided the data do not maintain high intensity at the edges of the observation window; see, e.g., the empirical comparison in Taddy and Kottas (2012).

Now, for any spatial location $\mathbf{s} \in \mathcal{S}$, the proposed mixture model for the density of threshold exceedance times is expressed as

$$f_{\mathbf{s}}(t) \equiv f(t; G_{\mathbf{s}}, \tau^2) = \int k(t \mid \theta, \tau^2) dG_{\mathbf{s}}(\theta), \quad t \in [0, 1] \quad (3)$$

where $k(t \mid \theta, \tau^2)$ is given by (2). Therefore, the NHPP density is modeled with a semiparametric mixture based on nonparametric mixing with respect to only the location parameter of the logit-normal kernel. This mixture model formulation strikes a good balance between model flexibility and computational feasibility. Location mixtures of logit-normals can capture non-standard density shapes, including skewness or multimodality; however, this may come at the expense of a larger number of mixture components than what would be needed under the model that includes mixing also with respect to the scale parameter of the logit-normal kernel. Although it

is possible to extend the spatial nonparametric model to include location-scale mixing, this more general representation requires a more complex prior for $G_{\mathcal{S}}$ and more complicated methods for posterior simulation. The scale parameter τ^2 of the kernel can be viewed as a bandwidth parameter, which is estimated from the data based on an inverse-gamma prior.

2.2.2 The spatial nonparametric prior model

To build the spatial dependence in the prior model for the threshold exceedance time densities, $\{f(t; G_{\mathbf{s}}, \tau^2) : \mathbf{s} \in \mathcal{S}\}$, we use a spatial DP prior (Gelfand et al., 2005) for the collection of corresponding mixing distributions $G_{\mathcal{S}} = \{G_{\mathbf{s}} : \mathbf{s} \in \mathcal{S}\}$. The spatial DP defines a nonparametric prior for the distribution of random fields, and it can thus be used to develop semiparametric models for spatial or spatio-temporal data by replacing customary Gaussian process (GP) specifications for spatial random effects distributions. Central to its development is the DP stick-breaking definition discussed in Section 2.1. Under the standard model setting with DP priors, the locations ϑ_l in the constructive definition are either scalar or vector valued, and thus G_0 is supported by a possibly multivariate, albeit finite dimensional, Euclidean space.

To model nonparametrically the distribution of a random field over region $\mathcal{S} \subset \mathbb{R}^2$, G_0 is extended to a parametric stochastic process $G_{0,\mathcal{S}}$ over the region of interest, a natural choice for which is a GP (possibly after transformation of the spatial random effects parameters). Hence, the (almost sure) representation for spatial DP prior realizations becomes

$$G_{\mathcal{S}} = \sum_{l=1}^{\infty} w_l \delta_{\vartheta_{l,\mathcal{S}}}$$

where the $\vartheta_{l,\mathcal{S}} = \{\vartheta_l(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$ are independent realizations from $G_{0,\mathcal{S}}$. We take a GP for $G_{0,\mathcal{S}}$ with constant mean function ζ , constant variance η^2 , and isotropic exponential correlation function, that is, $\text{Corr}(\vartheta_l(\mathbf{s}), \vartheta_l(\mathbf{s}') \mid \rho) = \exp(-\rho \|\mathbf{s} - \mathbf{s}'\|)$, where $\rho > 0$ is the range parameter. As discussed in Section 2.3, the full Bayesian model is completed with priors for the precision

parameter α and for the GP hyperparameters, $\phi = (\zeta, \eta^2, \rho)$.

Therefore, the spatial DP prior model involves a countable mixture of GP realizations with weights defined through stick-breaking as in the standard DP prior. Consequently, for any finite set of spatial locations $(\mathbf{s}_1, \dots, \mathbf{s}_r)$, the spatial DP prior induces a DP prior for the finite collection of mixing distributions $(G_{\mathbf{s}_1}, \dots, G_{\mathbf{s}_r})$; the centering distribution of this DP prior is the r -dimensional normal induced by the GP used for $G_{0,\mathcal{S}}$. This is a key property of the spatial DP prior with respect to both simulation-based model fitting and predictive inference for spatial interpolation. Spatial DPs provide an illustration of dependent DPs (MacEachern, 2000) in that they yield a stochastic process of random distributions, one at each location in \mathcal{S} . These distributions are dependent but such that, at each index value, the distribution is a univariate DP.

Hence, for any location $\mathbf{s} \in \mathcal{S}$, the spatial DP prior yields a location DP mixture of logit-normals following the formulation in (3). The DP mixture model interpretation is also valid for any finite collection of locations, with the additional structure of spatial dependence induced to the threshold exceedance time densities by the spatial dependence in the mixing distributions. It is important to note that the spatial DP generates non-stationary spatial surface realizations with non-Gaussian finite dimensional distributions, even when the centering GP is isotropic. Moreover, if $G_{\mathbf{s}}$ and $G_{\mathbf{s}'}$ denote the marginal distributions at generic locations \mathbf{s} and \mathbf{s}' , then the continuity of the $\vartheta_{l,\mathcal{S}}$ (implied by the exponential correlation function of $G_{0,\mathcal{S}}$) yields that, as the distance between \mathbf{s} and \mathbf{s}' gets smaller, the difference between $G_{\mathbf{s}}$ and $G_{\mathbf{s}'}$ gets smaller. Formally, for any $\varepsilon > 0$, $\lim_{\|\mathbf{s}-\mathbf{s}'\| \rightarrow 0} \Pr(\mathcal{L}(G_{\mathbf{s}}, G_{\mathbf{s}'}) < \varepsilon) = 1$, where \mathcal{L} is the Lévy distance (MacEachern, 2000; Gelfand et al., 2005); see Guindani and Gelfand (2006) for a detailed study of smoothness properties for spatial DP realizations. Hence, the level of dependence between $G_{\mathbf{s}}$ and $G_{\mathbf{s}'}$, and thus between $f(t; G_{\mathbf{s}}, \tau^2)$ and $f(t; G_{\mathbf{s}'}, \tau^2)$, is driven by the distance between the spatial locations. The practical implication is that in predictive inference for spatial interpolation, we learn more from locations \mathbf{s}' nearby \mathbf{s} than from more distant locations, a desirable property for densities that are expected to evolve relatively smoothly across space.

2.2.3 The spatial model for the total intensity of threshold exceedances

To complete the model specification for the spatially dependent intensity functions, we turn to a spatial probability model for the total intensity surface. Here, we work with a GP-based hierarchical specification, using a GP prior for the log-intensity surface $\{\beta(\mathbf{s}) = \log(\gamma_{\mathbf{s}}) : \mathbf{s} \in \mathcal{S}\}$. The first stage arises from the NHPP assumption, namely, the observed exceedance counts are assumed Poisson distributed, $n_j | \beta_j \stackrel{ind.}{\sim} \text{Poisson}(\exp(\beta_j))$, for $j = 1, \dots, m$, where $\beta_j = \log(\gamma_{\mathbf{s}_j})$. The GP prior for $\{\beta(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$ is assumed to have constant mean λ , constant variance κ^2 , and exponential correlation function $\exp(-\psi\|\mathbf{s} - \mathbf{s}'\|)$. Finally, we place a normal prior $N(m_\lambda, S_\lambda^2)$ on λ , an inverse gamma prior $IG(a_{\kappa^2}, b_{\kappa^2})$ (with mean $b_{\kappa^2}/(a_{\kappa^2} - 1)$ provided $a_{\kappa^2} > 1$) on κ^2 , and a uniform prior $\text{Unif}(0, b_\psi)$ on ψ . MCMC posterior simulation, as well as specification of the hyperpriors for λ , κ^2 and ψ , is discussed in the Appendix.

We note that the exceedance counts will typically take small to moderate values. If the particular application involves also a small number of spatial locations (as for the data sets considered in Section 3), the relatively simple model specification discussed above is arguably a suitable choice. For problems involving data from a large number of locations and/or where physical information is available, more structured GP mean functions or non-stationary covariance functions can be entertained. The fact that the prior model for the total intensities is specified independently of that for the NHPP densities is an asset in this respect.

2.3 Posterior simulation and inference for risk assessment

Inference for the intensity of extremes across space requires posterior simulation for: the random mixing distribution $G_{\mathbf{s}}$ over a number of spatial locations \mathbf{s} , including interpolation at new locations; the spatial DP prior hyperparameters; and the parameters of the GP-based spatial model for the total intensity of exceedances.

In general, nonparametric Bayesian inference for related distributions (indexed by time,

space, or covariate values) requires some form of replication, although imbalance in the replicate responses can be handled. In the absence of replication, posterior simulation can be overly sensitive to the prior specification and predictive inference at new index points will inevitably fall back exclusively to the prior. In our context, the replication is provided by the set of threshold exceedance times, $\{t_i(\mathbf{s}_j) : i = 1, \dots, n_j\}$, at each observed site \mathbf{s}_j , $j = 1, \dots, m$.

However, in contrast to earlier applications of dependent nonparametric prior models to geostatistics problems, our observations are random times arising as the events of the underlying NHPP at each site. This aspect of the data structure creates a challenge in matching the observed times to form response vectors across sites. An option is to discretize the time interval under study into time units specified such that at most one exceedance time is included in each time unit from each location; the default choice would be the unit at which the data is recorded, e.g., a day for daily rainfall records. We can then construct time-ordered response vectors, of dimension between 1 and m , that include an entry for all sites for which there was an exceedance at the specific time unit. For the hierarchical data model, the k -th response vector is assigned a vector of mixing parameters, $\theta_k(\mathbf{s}_{\text{obs}}) = (\theta_k(\mathbf{s}_1), \dots, \theta_k(\mathbf{s}_m))$, where the $\theta_k(\mathbf{s}_{\text{obs}})$ arise conditionally independent from the DP prior induced by the spatial DP at the vector of observed locations \mathbf{s}_{obs} . This hierarchical model formulation is along the lines in Gelfand et al. (2005) and Kottas et al. (2008) for independent and temporally dependent replicates, respectively.

Our data examples involve small exceedance counts over the time period of interest; with daily records over 50 years, the realized n_j range roughly between 15 to 250. Hence, constructing the replicates as discussed above results in response vectors with a very small number of entries relative to m . We are thus working with data structures where replication is present, but there is essentially only a single observation at any location. Here, we apply the spatial DP model under this scenario, where the i -th observation at the j -th location, $t_i(\mathbf{s}_j)$, is assigned a vector of mixing parameters, $\theta_{ij}(\mathbf{s}_{\text{obs}}) = (\theta_{ij}(\mathbf{s}_1), \dots, \theta_{ij}(\mathbf{s}_m))$, from which only $\theta_{ij}(\mathbf{s}_j)$ is used in the hierarchical model representation. A similar approach to implementing a dependent DP

prior model for time series problems can be found in Rodriguez and ter Horst (2008).

More specifically, let $z_{ij} = \text{logit}(t_i(\mathbf{s}_j))$, for $i = 1, \dots, n_j; j = 1, \dots, m$, be the logit-transformed observations. Then, the hierarchical model for the data can be expressed as

$$\begin{aligned} z_{ij} \mid \theta_{ij}(\mathbf{s}_{\text{obs}}), \tau^2 &\stackrel{\text{ind.}}{\sim} \text{N}(\theta_{ij}(\mathbf{s}_j), \tau^2), \quad i = 1, \dots, n_j; j = 1, \dots, m \\ \theta_{ij}(\mathbf{s}_{\text{obs}}) \mid G_{\mathbf{s}_{\text{obs}}} &\stackrel{\text{i.i.d.}}{\sim} G_{\mathbf{s}_{\text{obs}}}, \quad i = 1, \dots, n_j; j = 1, \dots, m \end{aligned}$$

where $G_{\mathbf{s}_{\text{obs}}} \mid \alpha, \phi \sim \text{DP}(\alpha, G_{0, \mathbf{s}_{\text{obs}}})$, that is, the DP prior induced by the spatial DP. Therefore, $G_{0, \mathbf{s}_{\text{obs}}}$ is an m -variate normal distribution with mean vector $\zeta \mathbf{1}_m$ and covariance matrix $\Sigma = \eta^2 R(\rho)$, with $R_{jj'}(\rho) = \exp(-\rho \|\mathbf{s}_j - \mathbf{s}_{j'}\|)$, for $j, j' = 1, \dots, m$. Here, $\mathbf{1}_m$ denotes an m -dimensional vector with all its elements equal to 1.

We use blocked Gibbs sampling (Ishwaran and James, 2001) for MCMC posterior simulation. The approach is based on a truncation approximation to the DP prior for $G_{\mathbf{s}_{\text{obs}}}$ defined through $G_{\mathbf{s}_{\text{obs}}}^N = \sum_{l=1}^N p_l \delta_{\vartheta_l(\mathbf{s}_{\text{obs}})}$, where the $\vartheta_l(\mathbf{s}_{\text{obs}}) = (\vartheta_l(\mathbf{s}_1), \dots, \vartheta_l(\mathbf{s}_m))$ are i.i.d. realizations from $G_{0, \mathbf{s}_{\text{obs}}}$, and the weights $\mathbf{p} = \{p_l : l = 1, \dots, N\}$ are defined using the DP stick-breaking construction subject to the constraint $p_N = 1 - \sum_{l=1}^{N-1} p_l$. The truncation level N can be chosen to any desired level of accuracy, using standard DP properties (e.g., Ishwaran and Zarepour, 2000); $N = 60$ was used for both data examples of Section 3. Then, the model can be fit to the data without the need to impute the mixing parameter vectors $\theta_{ij}(\mathbf{s}_{\text{obs}})$. To this end, we introduce configuration variables $\mathbf{L} = \{L_{ij} : i = 1, \dots, n_j; j = 1, \dots, m\}$, where $L_{ij} = l$, for $l = 1, \dots, N$, if and only if $\theta_{ij}(\mathbf{s}_{\text{obs}}) = \vartheta_l(\mathbf{s}_{\text{obs}})$. Hence, the hierarchical model for the data becomes

$$\begin{aligned} z_{ij} \mid \{\vartheta_l(\mathbf{s}_{\text{obs}})\}, L_{ij}, \tau^2 &\stackrel{\text{ind.}}{\sim} \text{N}(\vartheta_{L_{ij}}(\mathbf{s}_j), \tau^2), \quad i = 1, \dots, n_j; j = 1, \dots, m \\ L_{ij} \mid \mathbf{p} &\stackrel{\text{i.i.d.}}{\sim} \sum_{l=1}^N p_l \delta_l(L_{ij}) \quad i = 1, \dots, n_j; j = 1, \dots, m \end{aligned} \tag{4}$$

where $\vartheta_l(\mathbf{s}_{\text{obs}}) \mid \phi \stackrel{\text{i.i.d.}}{\sim} G_{0, \mathbf{s}_{\text{obs}}}$, for $l = 1, \dots, N$, and the prior for \mathbf{p} , given α , is a generalized Dirichlet distribution (Ishwaran and James, 2001). The hierarchical model is completed with

hyperpriors for the spatial DP parameters: an exponential prior for α , a normal prior $N(m_\zeta, S_\zeta^2)$ for ζ , an inverse gamma prior $IG(a_{\eta^2}, b_{\eta^2})$ for η^2 , and a uniform prior $\text{Unif}(0, b_\rho)$ for ρ . Moreover, an $IG(a_{\tau^2}, b_{\tau^2})$ prior is assigned to τ^2 . The Appendix provides details on MCMC posterior simulation for model (4), as well as on specification of the priors for its hyperparameters.

The resulting posterior samples can be used to extend the inference to spatial interpolation based on a set of M new locations, $\mathbf{s}_{\text{new}} = (\tilde{\mathbf{s}}_1, \dots, \tilde{\mathbf{s}}_M)$. Spatial interpolation for the total intensity surface $\{\gamma_{\mathbf{s}} : \mathbf{s} \in \mathcal{S}\}$ proceeds through standard GP predictive computing based on the implied conditional normal distribution for $(\beta(\tilde{\mathbf{s}}_1), \dots, \beta(\tilde{\mathbf{s}}_M))$ given $(\beta(\mathbf{s}_1), \dots, \beta(\mathbf{s}_m))$. In Section 3, we illustrate with posterior mean estimates for the $\{\gamma_{\mathbf{s}} : \mathbf{s} \in \mathcal{S}\}$ surface.

GP predictive calculations are also central for spatial interpolation of the NHPP densities, in conjunction with the spatial DP structure for the set of mixing distributions that includes the new sites. Specifically, under the DP truncation approximation, $G_{(\mathbf{s}_{\text{obs}}, \mathbf{s}_{\text{new}})}^N = \sum_{l=1}^N p_l \delta_{(\vartheta_l(\mathbf{s}_{\text{obs}}), \vartheta_l(\mathbf{s}_{\text{new}}))}$, where now the $(\vartheta_l(\mathbf{s}_{\text{obs}}), \vartheta_l(\mathbf{s}_{\text{new}}))$ arise independent from an $(m+M)$ -variate normal distribution with mean vector $\zeta \mathbf{1}_{m+M}$ and covariance matrix with structure that extends the one in $G_{0, \mathbf{s}_{\text{obs}}}$. Hence, having obtained posterior samples for the $\vartheta_l(\mathbf{s}_{\text{obs}})$ (and \mathbf{p}), the additional sampling needed to complete the posterior realizations for $G_{(\mathbf{s}_{\text{obs}}, \mathbf{s}_{\text{new}})}^N$ is from M -variate conditional normal distributions to impute $\vartheta_l(\mathbf{s}_{\text{new}})$ given $\vartheta_l(\mathbf{s}_{\text{obs}})$, for $l = 1, \dots, N$.

With posterior samples for the mixing distribution available at any desired set of sites, we can report different types of risk assessment inference. For any (observed or new) site \mathbf{s} , point estimates for the density or intensity of exceedance times can be obtained along with corresponding uncertainty bands. This inference is immediate from the definition of the mixture model for the NHPP density function, $f(t; G_{\mathbf{s}}^N, \tau^2) = \sum_{l=1}^N p_l k(t | \vartheta_l(\mathbf{s}), \tau^2)$, or intensity function, $\gamma_{\mathbf{s}} f(t; G_{\mathbf{s}}^N, \tau^2)$. Using the NHPP definition, we can compute risk surface estimates defined through the probability of a specific number of threshold exceedances within any time interval of interest. For both data examples of Section 3, we illustrate with the probability of at least one exceedance in a given month across a number of years. Letting (t_1, t_2) denote the time inter-

val of interest (e.g., a specific month), the risk surface probability of at least one exceedance is given by $1 - \exp(-\gamma_{\mathbf{s}} \int_{t_1}^{t_2} f(t; G_{\mathbf{s}}^N, \tau^2) dt) = 1 - \exp\{-\gamma_{\mathbf{s}} \sum_{l=1}^N p_l (\int_{t_1}^{t_2} k(t | \vartheta_l(\mathbf{s}), \tau^2) dt)\}$, with each integral term readily computed through a difference of two normal cdf values. Our illustration represents an admittedly narrow example of risk assessment, since inference for the risk surface probability is not accompanied by impact analysis and/or vulnerability evaluation. The results presented here are merely meant to demonstrate the capacity of the nonparametric modeling approach for flexible inference which can potentially be placed in the context of a broader analysis for problems that involve additional information for more general risk assessment.

3 DATA ILLUSTRATIONS

3.1 Synthetic data example

Our first illustration involves simulated data based on the same region and 25 observed sites (Figure 1), time interval (years 1950 – 1999), and time unit (days) as the real data discussed in Section 3.2. The times of exceedances at each site were generated using a two-state, time-inhomogeneous Markov chain, with spatial structure introduced by making the transition probabilities spatially dependent. Specifically, let $v_t(\mathbf{s})$ be an indicator variable such that $v_t(\mathbf{s}) = 1$ if an exceedance occurs at time t and location \mathbf{s} . Then, the data is simulated according to

$$\Pr(v_t(\mathbf{s}) = 1 | v_{t-1}(\mathbf{s}) = k) = \Phi(\mu_{k,t}(\mathbf{s})), \quad k \in \{0, 1\},$$

for $t = 1, \dots, T = 18,262$, and a given $v_0(\mathbf{s})$, where $\Phi(\cdot)$ denotes the standard normal cdf, $\mu_{0,t}(\mathbf{s}) = 0.25 \sin(4\pi t T^{-1}) + \epsilon_0(\mathbf{s})$, and $\mu_{1,t}(\mathbf{s}) = 0.25 \cos(4\pi t T^{-1}) + \epsilon_1(\mathbf{s})$. Here, $\{\epsilon_k(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$, for $k = 0, 1$, are independent realizations from an isotropic GP with mean -2.7 , variance 1, and correlation function $\exp\{-0.2\|\mathbf{s} - \mathbf{s}'\|\}$. The number of realized exceedances across the 25 sites ranges between 14 and 224. The true probability of no exceedances at site \mathbf{s} during

time period $(t_0 + 1, \dots, t_0 + R)$ can be expressed conditional on the state at time t_0 , in particular, when $v_{t_0}(\mathbf{s}) = 0$, it is given by $\prod_{t=t_0+1}^{t_0+R} \{1 - \Phi(\mu_{0,t}(\mathbf{s}))\}$, whereas if $v_{t_0}(\mathbf{s}) = 1$, it is obtained as $\{1 - \Phi(\mu_{1,t}(\mathbf{s}))\} \prod_{t=t_0+2}^{t_0+R} \{1 - \Phi(\mu_{0,t}(\mathbf{s}))\}$.

Note that the synthetic data generating mechanism is completely unrelated to our statistical model. Hence, this simulation example is used to illustrate the flexibility of the nonparametric mixture model to reconstruct risk surfaces from general stochastic processes. In addition, the simulation is intended to demonstrate that, despite its flexibility, the nonparametric model does not overfit the data. Indeed, although the model can potentially capture non-stationary and non-separable behavior, the underlying data generating process is stationary and separable.

We follow the strategy discussed in the Appendix to specify the model hyperpriors. Regarding the spatial DP parameters, we place a normal prior on ζ with mean 0 and variance 10, an $\text{IG}(3, 12)$ prior on η^2 , and a $\text{Unif}(0, 2.34)$ prior on ρ . An exponential prior with mean 3 is assigned to the spatial DP precision parameter α . For the GP-based model for the total intensity surface, we assign a normal prior to λ with mean 3.74 and variance 10, an $\text{IG}(2, 0.52)$ prior to κ^2 , and a $\text{Unif}(0, 2.34)$ prior to ψ . Finally, the scale parameter τ^2 of the logit-normal kernel is assigned an $\text{IG}(3, 3)$ prior. We observed significant prior-to-posterior learning for all the spatial DP hyperparameters, and for τ^2 . As expected, given the nature of the observables for the total intensity surface model and the small number of spatial locations, there was less learning for parameters λ , κ^2 and ψ ; nevertheless, posterior densities for these parameters were noticeably concentrated relative to the corresponding prior densities.

Figure 2 shows an image plot of the true surface for the number of exceedances, computed from simulated data over a grid of sites, along with the model-based point estimate (posterior mean) for the total intensity surface. Note that, even though the data generating process does not imply that the exceedance counts at a given location follow a Poisson distribution, our model provides reasonable estimates capturing the underlying spatial heterogeneity.

Figure 3 presents posterior mean estimates for the risk surface probability of at least one

exceedance in the month of June in three given years (year 1954, 1964, and 1974). These maps illustrate the ability of the model to capture both temporal or spatial heterogeneity. The point estimates generated by the model tend to be smoother than the true surfaces, but capture very well the patterns implied by the underlying stochastic process. To supplement the graphical comparison results with a quantitative measure of model assessment, we report on the coverage of 95% (equal-tail) credible intervals. Based on a grid of 601 spatial locations (including the 25 observed sites), the proportion of 95% credible intervals for the site-specific probability of at least one exceedance that contain the corresponding true value is: 96.3% for June 1954, 93.3% for June 1964, and 94.3% for June 1974. A potential concern for complex Bayesian nonparametric models is that they may overfit the data with undesirable implications in prediction. In this respect, the results above are encouraging, since the data arise from a stochastic mechanism with simpler structure than what the spatial nonparametric mixture model can accommodate.

3.2 Rainfall precipitation data

Here, we present an illustration with rainfall exceedances from data collected in the Cape Floristic Region in South Africa. The Cape Floristic Region is located in the southwestern coastline of South Africa covering roughly 90,000 km². Although it is the smallest of the six recognised floral kingdoms in the world, it has the highest diversity, density and endemism of the flora species. The Cape Floristic Region has a semi-mediterranean climate pattern. In the west of the region, around Cape town and Paarl, the climate is characterized by hot dry summers and cool wet winters. Moving to the east, rainfall tends to be uniformly distributed over the year. Because the entire region lies between the southwestern ocean and the northeastern L-shaped mountain system, known as Cape Fold Mountains, the precipitation varies significantly. Specifically, rainfall ranges from 300 – 500 millimeters in the lowlands and 1,000 – 3,300 millimeters in the mountain areas. A previous analysis of annual rainfall maxima at 1,078 grid cells over the

entire region is presented in Sang and Gelfand (2009). The raw data consists of the daily grid-aggregated precipitation, obtained via certain interpolation techniques (Hewitson and Crane, 2005), based on records at monitoring stations across South Africa between 1950 and 1999.

For an illustrative data example, and considering the topography of the region and the climate pattern discussed above, we work with a subregion of the Cape Floristic Region. In particular, we select 25 sites ($\mathbf{s}_1, \dots, \mathbf{s}_{25}$ in Figure 1) from the southwest coastline area including the city of Cape Town and vicinity; the longitude and latitude of the specific subregion range from (18.5, 19.6) and $(-33.4, -34.4)$, respectively. Moreover, in the interest of cross-validation for spatial prediction, we consider 5 additional sites where data is available, but not used in fitting the model; these sites are denoted by $\tilde{\mathbf{s}}_1, \dots, \tilde{\mathbf{s}}_5$ in Figure 1. To assemble the final data set with the times of exceedances at each site, we set the threshold to $u = 350$ millimeters. The range for the number of exceedances across the 25 observed sites is from 14 to 241.

Given that the region and time interval are the same with the simulated data set, there are similarities in the hyperpriors of the spatial DP model for the NHPP density and of the GP model for the total intensity surface. In particular, we place a normal prior on ζ with mean 0 and variance 10, an $\text{IG}(3, 12)$ prior on η^2 , a $\text{Unif}(0, 2.34)$ prior on ρ , and an exponential prior with mean 3 on α . Moreover, we assign a normal prior to λ with mean 3.95 and variance 10, an $\text{IG}(2, 0.6)$ prior to κ^2 , and a $\text{Unif}(0, 2.34)$ prior to ψ . Finally, τ^2 is assigned an $\text{IG}(3, 3)$ prior. Regarding prior-to-posterior learning for the model hyperparameters, results were consistent with the ones for the synthetic data discussed in Section 3.1.

The posterior mean and 95% uncertainty bands for the exceedance times density at the 25 monitoring sites are plotted in the top 5 rows of Figure 4, while the bottom row shows the predicted density at the five new sites shown in Figure 1. (Note that the bottom row panels include the histograms of the exceedance times, although data at these 5 sites were not used in the model fitting.) In general, the model captures well the heterogeneity of the rainfall exceedance times across space. For the observed sites, the estimates become more accurate with larger number

of realized exceedances. Nonparametric spatial interpolation is illustrated with the estimates at the new sites, where predictive inference is more accurate when interpolating at locations that have a number of monitoring sites nearby; for example, contrast the estimates at sites \tilde{s}_4 and \tilde{s}_5 .

The posterior mean estimate for the total rainfall exceedance intensity is shown in Figure 5 (right panel); as a point of reference, the left panel of Figure 5 includes the image plot of the realized number of exceedances at the 25 monitoring sites. The model estimates a larger intensity of extremes in the central part of the region relative to the northwestern and southeastern parts.

Finally, we report inference for the risk surface probability of at least one exceedance in a particular month at different years. Recall that the climate pattern in the studied region is mediterranean with cool wet winters. Hence, we focus on months when large rainfall is to be expected, and in particular, we choose the month of June. Figure 6 plots the posterior mean estimates at twelve years covering all five decades. The nonparametric mixture model estimates spatially varying risk surfaces with both intensity and shapes changing across years. The overall pattern reveals higher probabilities of at least one exceedance over June in the center of the studied region, with idiosyncratic features in certain years, such as the second mode more clearly seen in June 1954 and again in June 1992.

4 DISCUSSION

We have developed a Bayesian nonparametric model for the analysis of extremes from environmental variables observed over time and across a number of monitoring sites. The methodology builds on the point process approach to extreme value analysis through a nonparametric mixture model for the spatially varying intensities. The modeling approach allows general time-inhomogeneous shapes for the intensity of threshold exceedances at each specific site, as well as nonparametric spatial interpolation for practically important risk assessment functionals. A posterior simulation algorithm to implement such inference has been designed. The model has

been tested with a simulated data set and applied to rainfall exceedances recorded over a time period of 50 years from a subregion of the Cape Floristic Region in South Africa.

Our data examples included a small number of sites given the relatively small size of the geographic region under study. Moreover, the intention was to demonstrate the capacity of the spatial nonparametric mixture model to provide useful inference results under moderate sample sizes. For extreme value analysis applications, the number of observations from each site will typically be small to moderate. However, one can envision practically important scenarios that involve a large number of observed sites (at least, in the thousands). For such cases, standard posterior simulation methods are not practical for implementation of the spatial DP mixture model. Alternative cost-effective MCMC algorithms for large data sets (e.g., Guha, 2010) may provide a platform for expanding the practical utility of the proposed methodology.

Acknowledgements

This research was supported in part by the National Science Foundation under awards SES 1024484 and DMS 0915272. The authors are grateful to Huiyan Sang for providing the data analyzed in Section 3.2. They also wish to thank two reviewers and the Special Issue Editors for useful comments.

References

- Coles, S. G., 2001. An introduction to statistical modeling of extreme values. Springer-Verlag, New York.
- Coles, S. G., Powell, E. A., 1996. Bayesian methods in extreme value modelling: A review and new developments. *International Statistical Review* **64**, 119–136, DOI:10.2307/1403426.
- Coles, S. G., Tawn, J. A., 1996. A Bayesian analysis of extreme rainfall data. *Applied Statistics* **45**, 463–478, DOI:10.2307/2986068.

- Cooley, D., Nychka, D., Naveau, P., 2007. Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association* **102**, 824–840, DOI:10.1198/016214506000000780.
- Cooley, D., Sain, S. R., 2010. Spatial hierarchical modeling of precipitation extremes from a regional climate model. *Journal of Agricultural, Biological, and Environmental Statistics* **15**, 381–402, DOI: 10.1007/s13253-010-0023-9.
- Damien, P., Wakefield, J., Walker, S., 1999. Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society, Series B* **61**, 331–344, DOI: 10.1111/1467-9868.00179.
- Davison, A., Smith, R., 1990. Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society, Series B* **52**, 393–442.
- Escobar, M., West, M., 1995. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577–588, DOI:10.1080/01621459.1995.10476550.
- Ferguson, T. S., 1973. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**, 209–230, DOI:10.1214/aos/1176342360.
- Fisher, R. A., Tippett, L. H. C., 1928. Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society* **24**, 180–190.
- Fuentes, M., Henry, J., Reich, B. J., 2012. Nonparametric spatial models for extremes: Application to extreme temperature data. *Extremes* **15**, DOI:10.1007/s10687-012-0154-1.
- Gelfand, A. E., Kottas, A. E., MacEachern, S. N., 2005. Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association* **100**, 1021–1035, DOI: 10.1198/016214504000002078.
- Gelman, A., Rubin, D. B., 1992. Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* **7**, 457–511, DOI: doi:10.1214/ss/1177011136.
- Guha, S., 2010. Posterior simulation in countable mixture models for large datasets. *Journal of the American Statistical Association* **105**, 775–786, DOI: 10.1198/jasa.2010.tm09340.
- Guindani, M., Gelfand, A. E., 2006. Smoothness properties and gradient analysis under spatial Dirichlet process models. *Methodology and Computing in Applied Probability* **8**, 159–189, DOI: 10.1007/s11009-006-8547-8.

- Hewitson, B. C., Crane, R. G., 2005. Gridded area-averaged daily precipitation via conditional interpolation. *Journal of Climate* **18**, 41–57, DOI: 10.1175/JCLI3246.1.
- Huerta, G., Sansó, B., 2007. Time-varying models for extreme values. *Environmental and Ecological Statistics* **14**, 285–299, DOI: 10.1007/s10651-007-0014-3.
- Ishwaran, H., James, L. F., 2001. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* **96**, 161–173, DOI: 10.1198/016214501750332758.
- Ishwaran, H., Zarepour, M., 2000. Markov chain Monte Carlo in approximate Dirichlet and Beta two-parameter process hierarchical models. *Biometrika* **87**, 371–390, DOI: 10.1093/biomet/87.2.371.
- Kottas, A., 2006. Dirichlet process mixtures of Beta distributions, with applications to density and intensity estimation. In: *Proceedings of the Workshop on Learning with Nonparametric Bayesian Methods*, 23rd International Conference on Machine Learning. Carnegie Mellon University, Pittsburgh, Pennsylvania.
- Kottas, A., Duan, J. A., Gelfand, A. E., 2008. Modeling disease incidence data with spatial and spatio-temporal Dirichlet process mixtures. *Biometrical Journal* **50**, 29–42, DOI:10.1002/bimj.200610375.
- Kottas, A., Sansó, B., 2007. Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *Journal of Statistical Planning and Inference* **137**, 3151–3163, DOI: 10.1016/j.jspi.2006.05.022.
- Kotz, S., Nadarajah, S., 2000. *Extreme Value Distributions - Theory and Applications*. Imperial College Press.
- MacEachern, S., 2000. Dependent Dirichlet processes. Tech. rep., Department of Statistics, The Ohio State University.
- Pickands, J., 1971. The two-dimensional Poisson process and extremal processes. *Journal of Applied Probability* **8**, 745–756, DOI: 10.2307/3212238.
- Pickands, J., 1975. Statistical inference using extreme order statistics. *The Annals of Statistics* **3**, 119–131, DOI: 10.1214/aos/1176343003.
- Reich, B. J., Shaby, B. A., 2012. A hierarchical max-stable spatial model for extreme precipitation. *Annals of Applied Statistics*, Forthcoming.

- Rodriguez, A., ter Horst, E., 2008. Bayesian dynamic density estimation. *Bayesian Analysis* **3**, 339–366, DOI:10.1214/08-BA313.
- Sang, H., Gelfand, A. E., 2009. Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics* **16**, 407–426, DOI: 10.1007/s10651-007-0078-0.
- Sang, H., Gelfand, A. E., 2010. Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics* **15**, 49–65, DOI: 10.1007/s13253-009-0010-1.
- Sethuraman, J., 1994. A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.
- Smith, R., 1989. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science* **4**, 367–393, DOI:10.1214/ss/1177012400.
- Stephenson, A., Tawn, J. A., 2004. Bayesian inference for extremes: accounting for the three extremal types. *Extremes* **7**, 291–307, DOI: 10.1007/s10687-004-3479-6.
- Taddy, M. A., Kottas, A., 2012. Mixture modeling for marked Poisson processes. *Bayesian Analysis* **7**, 335–362, DOI: 10.1214/12-BA711.
- Tressou, J., 2008. Bayesian nonparametrics for heavy tailed distribution. Application to food risk assessment. *Bayesian Analysis* **3**, 367–392, DOI: 10.1214/08-BA314.
- Wang, Z., Rodríguez, A., Kottas, A., 2011. A nonparametric mixture modeling framework for extreme value analysis. Tech. Rep. UCSC-SOE-11-26, University of California, Santa Cruz, available from <http://www.soe.ucsc.edu/research/technical-reports/ucsc-soe-11-26>.

APPENDIX. IMPLEMENTATION DETAILS

Here, we provide the details for MCMC posterior simulation from the spatial DP model for the NHPP densities as well as the GP model for the total NHPP intensity surface. We also discuss prior specification for the hyperparameters of these models.

Posterior simulation algorithms: Simulation from the posterior distribution of the spatial DP model (4) is based on the blocked Gibbs sampler, including Metropolis-Hastings (M-H) steps. In particular, model parameters are iteratively updated according to the following steps.

- Updating L_{ij} , $i = 1, \dots, n_j; j = 1, \dots, m$. Each L_{ij} is drawn from a discrete distribution on $\{1, \dots, N\}$ with probabilities proportional to $p_l \mathbf{N}(z_{ij} \mid \vartheta_l(\mathbf{s}_j), \tau^2)$, for $l = 1, \dots, N$.
- Updating α and \mathbf{p} . The draws for these parameters are generic for any choice of kernel in the DP mixture model; details are given in Ishwaran and Zarepour (2000).
- Updating $\vartheta_l(\mathbf{s}_{\text{obs}})$, $l = 1, \dots, N$. Let n^* be the number of distinct components in vector \mathbf{L} , and $\mathbf{L}^* = \{L_k^* : k = 1, \dots, n^*\}$ the set of distinct elements. If $l \notin \mathbf{L}^*$, then $\vartheta_l(\mathbf{s}_{\text{obs}})$ is drawn from the normal centering distribution $G_{0, \mathbf{s}_{\text{obs}}}$. If $l \in \mathbf{L}^*$, the posterior full conditional for $\vartheta_l(\mathbf{s}_{\text{obs}})$ is proportional to $\mathbf{N}_m(\vartheta_l(\mathbf{s}_{\text{obs}}) \mid \zeta \mathbf{1}_m, \Sigma) \prod_{\{(i,j): L_{ij}=l\}} \mathbf{N}(z_{ij} \mid \vartheta_l(\mathbf{s}_j), \tau^2)$, a form which results in an m -variate normal distribution.
- Updating the centering GP parameters. The full conditional for ζ can be derived as a normal distribution with mean $\left(n^* \mathbf{1}'_m \Sigma^{-1} \mathbf{1}_m + S_\zeta^{-2}\right)^{-1} \left(\mathbf{1}'_m \Sigma^{-1} \sum_{k=1}^{n^*} \vartheta_{L_k^*}(\mathbf{s}_{\text{obs}}) + m_\zeta S_\zeta^{-2}\right)$ and variance $\left(n^* \mathbf{1}'_m \Sigma^{-1} \mathbf{1}_m + S_\zeta^{-2}\right)^{-1}$. Given their high posterior correlation, we update η^2 and ρ as a block with a joint random walk M-H step based on a bivariate normal proposal distribution (on the log scale for η^2 and the logit scale for ρ/b_ρ). To achieve good mixing, we estimate the proposal covariance matrix from the output of an initial chain based on separate updates for η^2 and ρ , using a M-H step for ρ and sampling η^2 from its full conditional which is available as an inverse gamma distribution.
- Updating τ^2 . The posterior full conditional for τ^2 is an inverse gamma distribution with shape parameter $a_{\tau^2} + 0.5 \sum_{j=1}^m n_j$ and scale parameter $b_{\tau^2} + 0.5 \sum_{j=1}^m \sum_{i=1}^{n_j} (z_{ij} - \vartheta_{L_{ij}}(\mathbf{s}_j))^2$.

Turning to the model of Section 2.2.3 for the total intensity surface, the MCMC posterior sampling steps are as follows.

- Updating β_j , $j = 1, \dots, m$. The posterior full conditional for each β_j is proportional to $\exp(n_j \beta_j - \exp(\beta_j)) p(\beta_j \mid \{\beta_r : r \neq j\})$, where $p(\beta_j \mid \{\beta_r : r \neq j\})$ denotes the normal distribution for β_j , conditional on $\{\beta_r : r \neq j\}$, implied by the GP prior for $\{\beta(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$. Hence, β_j can be updated using slice sampling (as in Example 4 of Damien et al., 1999).
- Updating the GP prior parameters. The GP mean parameter λ is sampled from its normal posterior full conditional distribution, whereas (κ^2, ψ) are updated jointly with a M-H step designed similarly to the one for (η^2, ρ) discussed above.

Convergence of the MCMC algorithms was assessed by visually inspecting the trace plots associated with various parameters of interest, as well as by computing standard diagnostic criteria. For instance, for the spatial DP model hyperparameters, the R statistic values (Gelman and Rubin, 1992) were below 1.1 after 40,000 iterations. All inferences are based on 3,000 posterior samples obtained after discarding the first 50,000 iterations and thinning the remaining 150,000 every 50 observations. Both MCMC algorithms were implemented in the C programming language. The code for the spatial DP mixture model executed at a rate of 1,500 iterations per minute on a 2 GHz Intel Core 2 Duo laptop with 2 GB memory.

Prior specification: We follow an approach along the lines in Gelfand et al. (2005) to specify the priors for the hyperparameters of the spatial DP model. In general, we center the normal prior for ζ at 0, and set the shape parameter of inverse gamma priors to small values that yield large (possibly infinite) prior variance. Then, working with a single component of the spatial DP mixture model, the marginal variance for the response on the logit scale can be decomposed into a sum of three terms involving the prior mean of τ^2 , the prior mean of η^2 , and the prior variance of ζ . Hence, with a rough guess at the range of the logit-transformed exceedance times, we can complete the prior specification for ζ , η^2 and τ^2 . To specify the $\text{Unif}(0, b_\rho)$ prior for ρ , we use the *range of dependence* interpretation of this parameter for the centering GP of the spatial DP prior. In particular, under the exponential correlation function, $3/\rho$ is the distance between sites that yields correlation 0.05. The range of dependence is usually assumed to be a fraction of the maximum interpoint distance (say, d_{\max}) over the geographic region under study. Hence, since $3/b_\rho < 3/\rho$, we specify b_ρ such that $3/b_\rho = cd_{\max}$, for $c \leq 1$; $c = 1$ was used as a conservative choice for the data examples of Section 3. Finally, the role α plays in controlling the number of distinct mixture components (as discussed briefly in Section 2.1) can be used to guide the choice of its exponential prior.

A similar prior choice strategy can be used for the GP-based model for the total intensity surface. The approach is the same for the correlation parameter ψ . Here, the marginal mean and variance for the site-specific exceedance counts can be expressed in terms of the λ and κ^2 parameters, using the first two moments of the lognormal distribution (induced for γ_s by the GP prior model). Hence, the prior means for these parameters are specified through proxies for the center and range of the number of exceedances across the region.

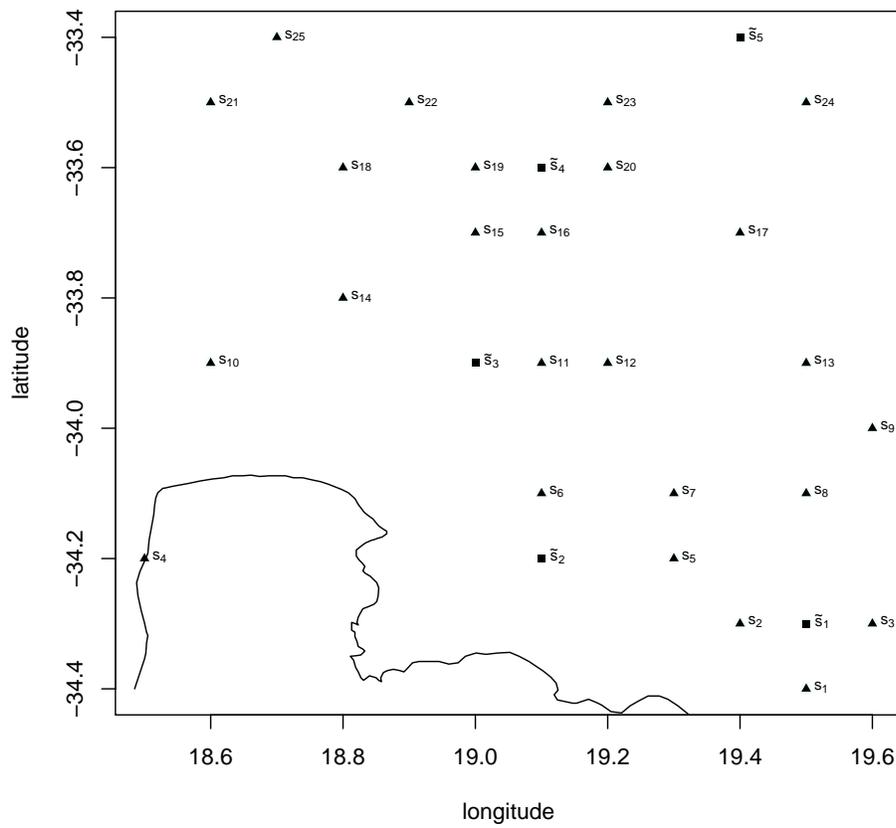


Figure 1: Geographic map of the southwest coastline area subregion of the Cape Floristic Region in South Africa. The map shows the 25 spatial locations (s_1, \dots, s_{25}) which comprise the observed set of sites for the data examples of Section 3, and the 5 new sites ($\tilde{s}_1, \dots, \tilde{s}_5$) used for prediction of the exceedance times density.

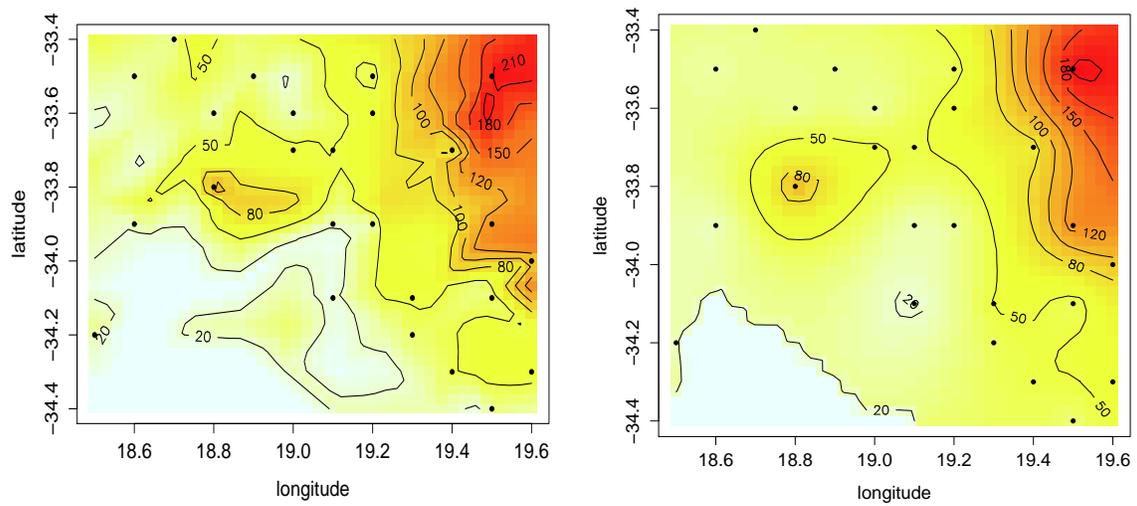
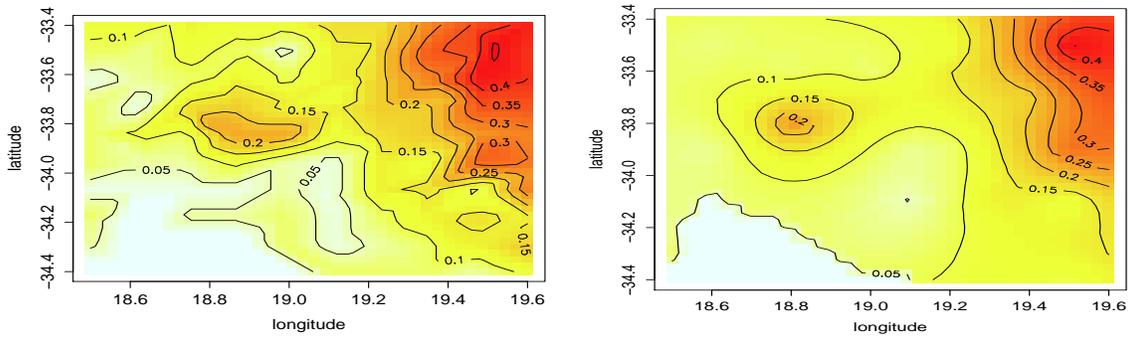
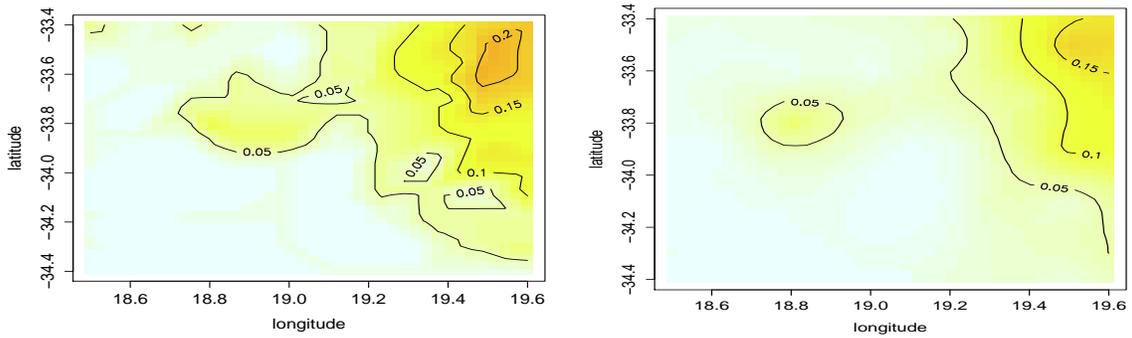


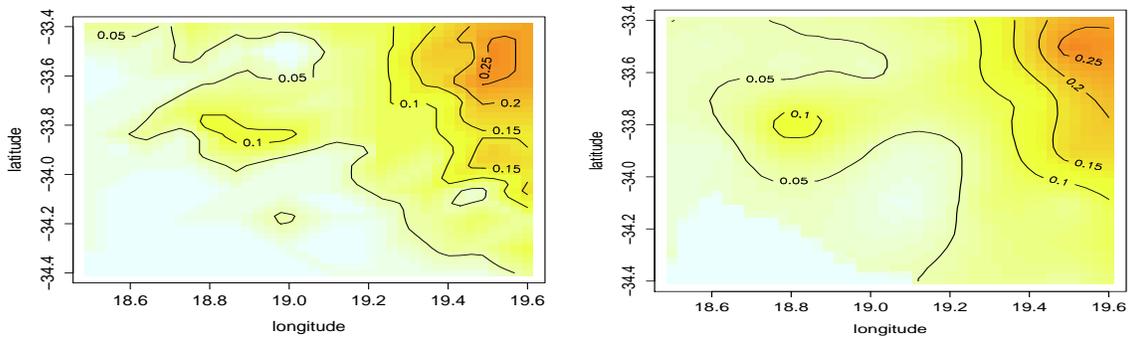
Figure 2: Synthetic data example. Image plot of the true surface for the number of exceedances (left panel) and the posterior mean of $\{\gamma_s : s \in \mathcal{S}\}$ (right panel).



(a) June 1954



(b) June 1964



(c) June 1974

Figure 3: Synthetic data example. True surface (left panels) and posterior mean estimate (right panels) for the probability of at least one exceedance in the month of June for year 1954, 1964, and 1974 (from top to bottom).

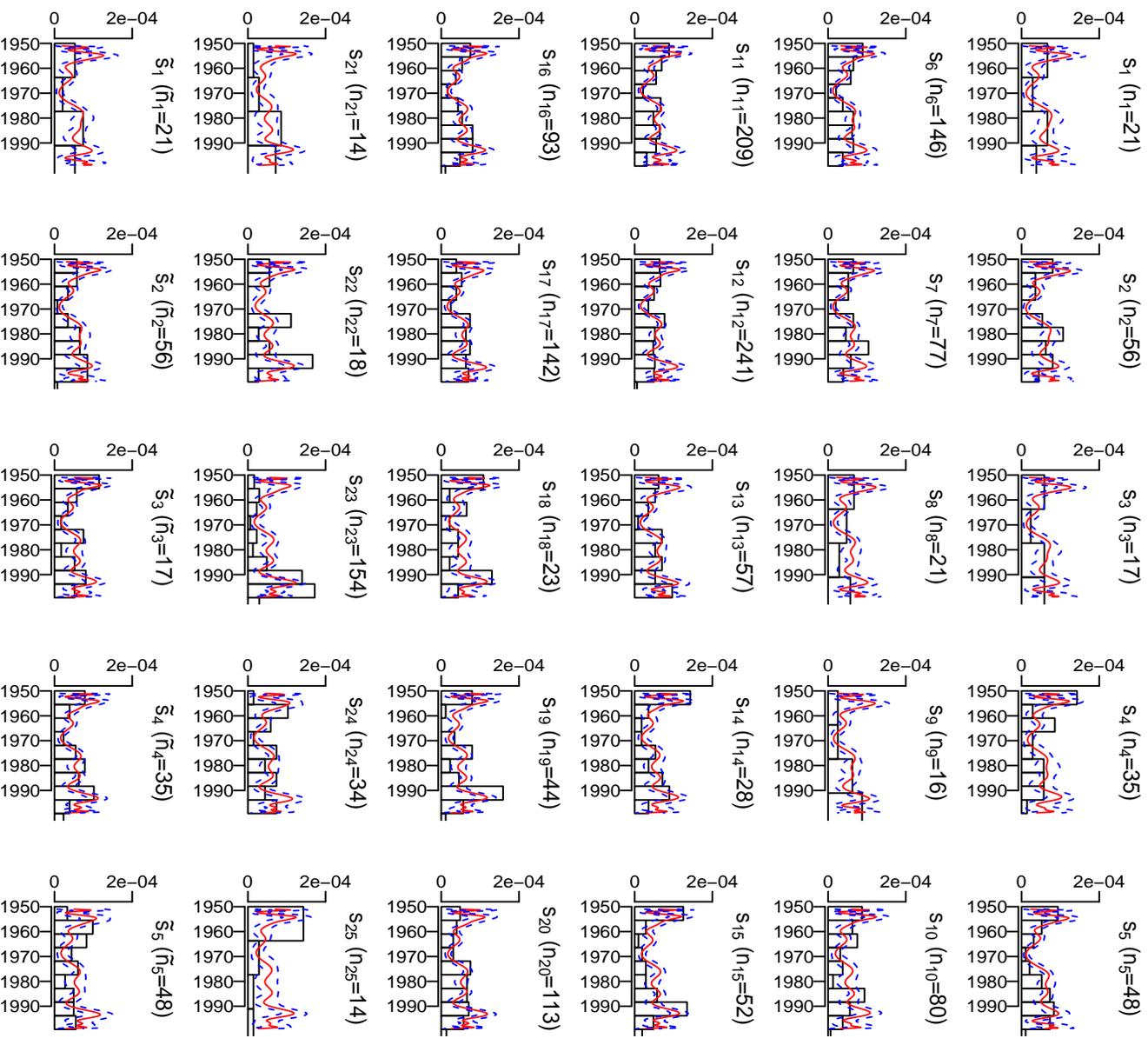


Figure 4: Precipitation data. Posterior mean (red solid line) and 95% interval estimates (blue dashed lines) of the exceedance time density functions at the 25 observed sites (top 5 rows) and at 5 new sites (bottom row). Each panel indicates the corresponding exceedance count and shows a histogram of the observed exceedance times.

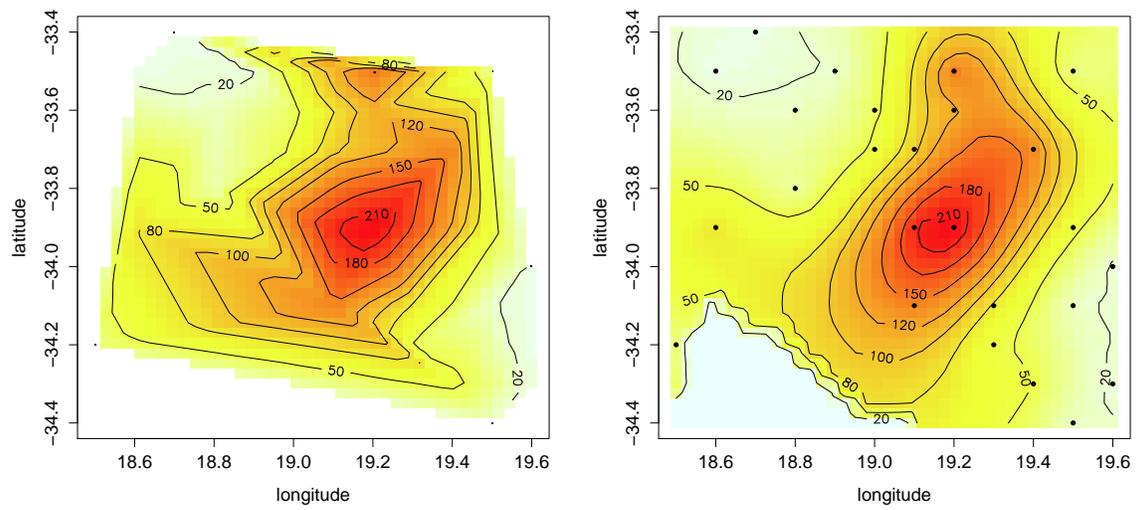


Figure 5: Precipitation data. Image plot of the observed number of exceedances at the 25 monitoring sites (left panel) and the posterior mean estimate of $\{\gamma_s : s \in \mathcal{S}\}$ (right panel).

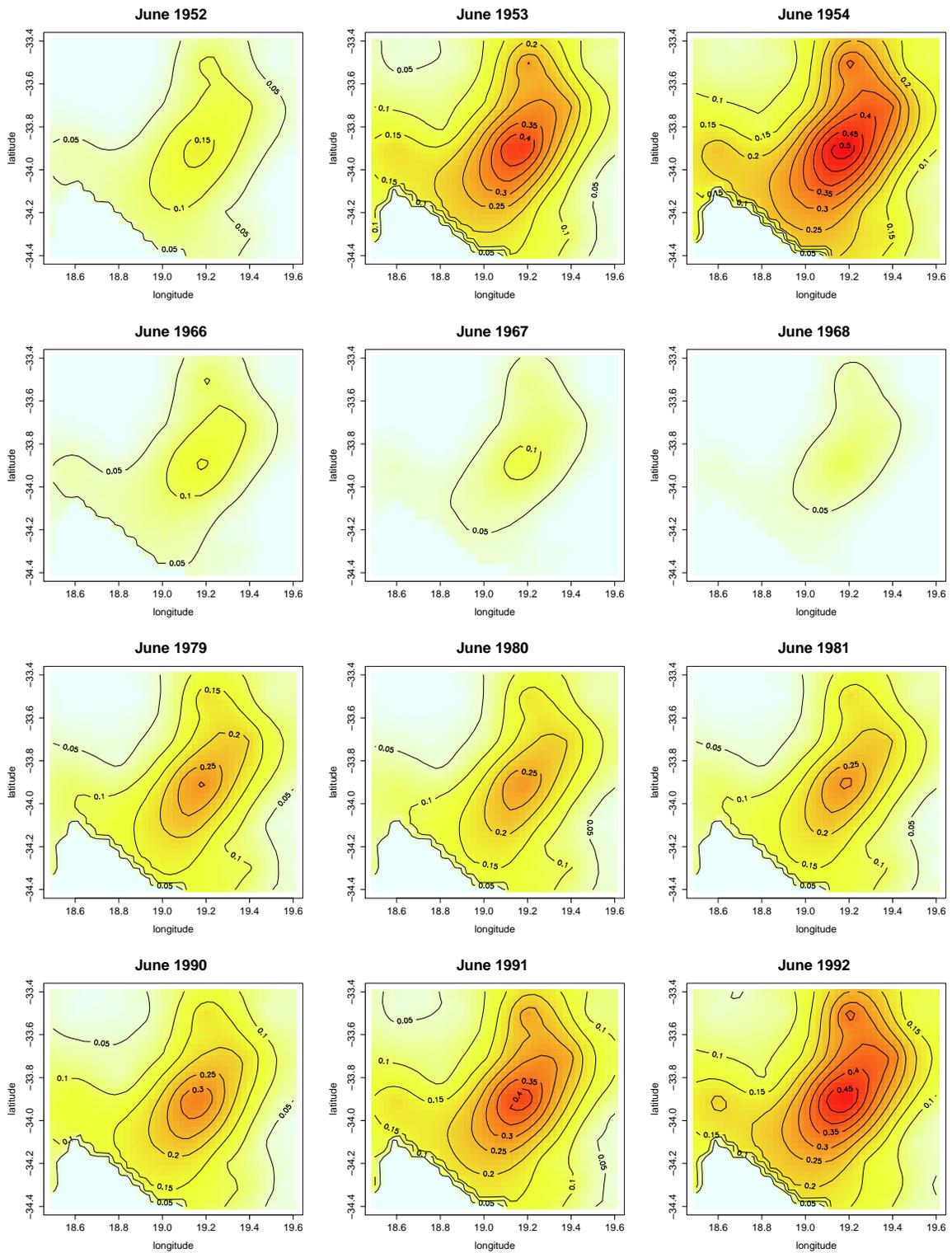


Figure 6: Precipitation data. Posterior mean estimate for the risk surface probability of at least one exceedance in the month of June for twelve years.