ESTIMATING HIGH-RESOLUTION RED SEA SURFACE TEMPERATURE HOTSPOTS, USING A LOW-RANK SEMIPARAMETRIC SPATIAL MODEL

BY ARNAB HAZRA[*] AND RAPHAËL HUSER[†]

Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), [*]*arnab.hazra@kaust.edu.sa;* [†]*raphael.huser@kaust.edu.sa*

In this work, we estimate extreme sea surface temperature (SST) hotspots, that is, high threshold exceedance regions, for the Red Sea, a vital region of high biodiversity. We analyze high-resolution satellite-derived SST data comprising daily measurements at 16,703 grid cells across the Red Sea over the period 1985–2015. We propose a semiparametric Bayesian spatial mixed-effects linear model with a flexible mean structure to capture spatially-varying trend and seasonality, while the residual spatial variability is modeled through a Dirichlet process mixture (DPM) of low-rank spatial Student's t processes (LTPs). By specifying cluster-specific parameters for each LTP mixture component, the bulk of the SST residuals influence tail inference and hotspot estimation only moderately. Our proposed model has a nonstationary mean, covariance, and tail dependence, and posterior inference can be drawn efficiently through Gibbs sampling. In our application, we show that the proposed method outperforms some natural parametric and semiparametric alternatives. Moreover, we show how hotspots can be identified, and we estimate extreme SST hotspots for the whole Red Sea, projected until the year 2100, based on the Representative Concentration Pathways 4.5 and 8.5. The estimated 95% credible region, for joint high threshold exceedances include large areas covering major endangered coral reefs in the southern Red Sea.

1. Introduction. Sea surface temperature (SST) has an immense environmental and ecological impact on marine life and ecosystems worldwide, and in particular for the Red Sea, for example, affecting the survival of endangered animal species, including corals (Reaser, Pomerance and Thomas (2000), Berumen et al. (2013), Lewandowska et al. (2014)), and it also has an important economic impact for neighboring countries which depend on it for their local fisheries and tourism. Hence, the identification of the regions within the Red Sea where SST may exceed high thresholds is a vital concern, and this motivates a proper statistical analysis of (present and future) extreme hotspots from a high-resolution spatiotemporal SST dataset. Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) produces satellite-derived daily SST data at $0.05° \times 0.05°$ resolution (Donlon et al. (2012)). Over the whole Red Sea, daily SST data are available at 16,703 grid cells between 1985–2015, and we consider these data for estimating extreme hotspots.

The most common model in spatial geostatistics is the Gaussian process (GP), due to its appealing theoretical and computational properties (Gelfand and Schliep (2016)). However, fitting an ordinary GP model involves computing the determinant and the inverse of the spatial covariance matrix which is excessively prohibitive in dimensions as high as the Red Sea SST data (here, available at 16,703 grid cells). A variety of methods have been proposed to tackle this problem. These include approaches based on kernel convolutions (Higdon (2002)), low-rank methods using basis functions (Wikle and Cressie (1999)), the predictive process (Banerjee et al. (2008)), approximations of the likelihood in the spectral

domain (Stein (1999), Fuentes (2007)) or by a product of appropriate conditional distributions (Vecchia (1988), Stein, Chi and Welty (2004)), covariance tapering (Furrer, Genton and Nychka (2006), Anderes et al. (2013)) and Markov random fields (Rue and Held (2005), Rue, Martino and Chopin (2009)); see Heaton et al. (2019) for a comparative overview of (some of) these methods. Irrespective of being an ordinary GP or a low-rank GP (LGP) model, the marginal normal density functions are thin-tailed, and, hence, they can heavily underestimate the probabilities of extreme events. Additionally, the tails of multivariate normal distributions lead to independent extremes asymptotically, except in the trivial case of perfect dependence which can result in disastrous underestimation of the simultaneous occurrence probabilities of extreme events (Davison, Huser and Thibaud (2013)). Hence, both GPs and LGPs have been criticized when the main interest lies in the tail behavior. Relaxing the parametric GP assumption, Gelfand, Kottas and MacEachern (2005) propose a flexible nonparametric Bayesian model based on a Dirichlet process mixture (DPM) of spatial GPs in the context of geostatistical analysis; however, Hazra et al. (2018) showed that the joint tail of a finite mixture of GPs also leads to independent extremes. There are more flexible nonparametric spatial models available in the geostatistics literature (see, e.g., Duan, Guindani and Gelfand (2007)), but these are often not computationally suitable for large spatial datasets.

While GPs and related processes are typically used to describe the bulk behavior, models stemming from extreme-value theory are designed to accurately describe the tail behavior. The classical modeling of spatial extremes usually relies on sitewise block maxima or peaks over some high threshold (Smith (1990), Davison, Padoan and Ribatet (2012), Davison and Huser (2015), Davison, Huser and Thibaud (2019), Huser and Wadsworth (2021)). They can be divided into three main categories: (asymptotic) max-stable and Pareto processes (Smith (1990), Padoan, Ribatet and Sisson (2010), Davison, Padoan and Ribatet (2012), Reich and Shaby (2012), Opitz (2013), Thibaud and Opitz (2015), de Fondeville and Davison (2018)), latent variable models (Sang and Gelfand (2009, 2010), Cooley and Sain (2010), Opitz et al. (2018), Castro-Camilo, Huser and Rue (2019)) and subasymptotic models (Huser, Opitz and Thibaud (2017), Morris et al. (2017), Hazra, Reich and Staicu (2020), Huser and Wadsworth (2019)). Max-stable and Pareto processes are asymptotically justified models for spatial extremes, but likelihood computations are usually challenging, even for low or moderate spatial dimensions (see, e.g., Castruccio, Huser and Genton (2016) and Huser et al. (2019)). An exception is the max-stable model of Reich and Shaby (2012) which is computationally tractable for higher spatial dimensions; see also Bopp, Shaby and Huser (2021). However, this max-stable model has been criticized for its lack of flexibility in a variety of applications. Analogously, de Fondeville and Davison (2018) show how Pareto processes can be efficiently fitted to high-dimensional peaks-over-thresholds using proper scoring rules, but their approach is limited to a few thousand sites and cannot easily be accommodated to the Bayesian setting. Alternatively, Morris et al. (2017) propose a Bayesian model for high threshold exceedances based on the spatial skew-t process which lacks the strong asymptotic characterization of max-stable and Pareto processes but benefits from exceptionally faster computation. Instead of considering only block maxima or peaks-over-thresholds, as in the above-mentioned approaches, Hazra et al. (2018) consider a DPM of spatial skew-t processes where the extremes are selected probabilistically through the Dirichlet process prior; hence, this modeling approach does not require any arbitrary high thresholding but assumes that identically distributed temporal replicates are available. More recently, Bopp et al. (2020) have proposed a Bayesian analogue model where the intensity of flood-inducing precipitation is modeled using a mixture of Student's t processes. However, these approaches are applicable only for relatively low spatial dimensions, and they are not directly applicable for large spatiotemporal datasets with trend and seasonality, such as our Red Sea SST dataset.

In this paper, we propose a low-rank semiparametric Bayesian spatial mixed-effects linear model, which extends the spatial model of Hazra et al. (2018), to handle large, highly nonstationary spatiotemporal datasets. Following Fix et al. (2018), we assume the mean SST profile to be comprised of a spatially-varying trend that is linearly related to the mean annual sea surface temperature projections for the Red Sea under the Representative Concentration Pathway (RCP) 4.5 and 8.5, adopted by the Intergovernmental Panel on Climate Change (IPCC), where RCP 8.5 corresponds to the pathway with high greenhouse gas emissions, while RCP 4.5 describes a moderate mitigation pathway. Additionally, we also consider a nonlinear seasonality term (modeled using B-splines), and, for computational tractability with high spatial resolution, we consider a low-rank approximation of the spatially-varying coefficients involved within the trend and seasonality components. While the use of splines implies some degree of regularity, it is justified in our case since SST data vary smoothly over space. Moreover, unlike the max-stable model of Reich and Shaby (2012) whose dependence structure is constructed from smooth kernels, here we only consider the mean structure to be comprised of splines, while the dependence structure of residuals has a more flexible stochastic representation. Specifically, we model the residual variability using a DPM of low-rank spatial Student's t processes (LTPs), abbreviated by LTP-DPM in short. The DPM is constructed using a (truncated) stick-breaking prior (Sethuraman (1994)). The LTP mixture components are constructed by a scalar product of multivariate normal random effects with orthonormal spatial basis functions and then multiplied by an inverse-gamma random scaling. The random effects are assumed to be independent and identically distributed (i.i.d.) across time points. The proposed model has nonstationary mean, covariance, and tail dependence structure, and, under a suitable asymptotic regime, its covariance structure spans the class of all covariance structures from squared-integrable stochastic processes with continuous covariance functions. We draw posterior inference about the model parameters through an efficient Gibbs sampler.

Beyond modeling the SST data, our ultimate goal is to identify regions "at risk," where the SST level might exceed a high threshold at some future time. Similar problems arise in a wide range of scientific disciplines, for example, environmental health monitoring (Bolin and Lindgren (2015)), brain imaging (Mejia et al. (2020)), astrophysics (Beaky, Scherrer and Villumsen (1992)) and climatology (Furrer et al. (2007), French and Sain (2013)). The easiest and most naive approach for estimating exceedance regions (i.e., hotspots) is to perform site-specific exceedance tests at each grid cell separately (see, e.g., Eklundh and Olsson (2003)). However, if such an approach does not adequately account for multiple testing nor spatial dependence, it will fail to accurately represent the co-occurrence of extreme events. A better approach for estimating hotspots is to set the joint probability of exceeding a high threshold over the *whole* region being equal to some predefined value. In this spirit a variety of more advanced methods have been proposed for identifying hotspots (see, e.g., Cressie, Zhang and Craigmile (2005), Craigmile et al. (2005), French and Sain (2013), Bolin and Lindgren (2015)). In particular, French and Sain (2013) provide a sampling-based method for constructing confidence regions for Gaussian processes that contain the true exceedance regions with some predefined probability. Here, we develop a similar approach to estimate extreme hotspots by extending the Gaussian-based method of French and Sain (2013) and French and Hoeting (2016) to the more general framework of our semiparametric LTP-DPM model, which is better suited for capturing the joint tail behavior of complex spatiotemporal processes.

The paper is organized as follows. In Section 2, we present the Red Sea SST dataset and some exploratory analysis. Our proposed LTP-DPM model and its properties are discussed in Section 3. In Section 4, we discuss Bayesian computational details and the hotspot estimation technique. In Section 5, we apply the proposed methodology to the Red Sea SST dataset and discuss the results. We conclude with some discussion and perspectives on future research in Section 6.
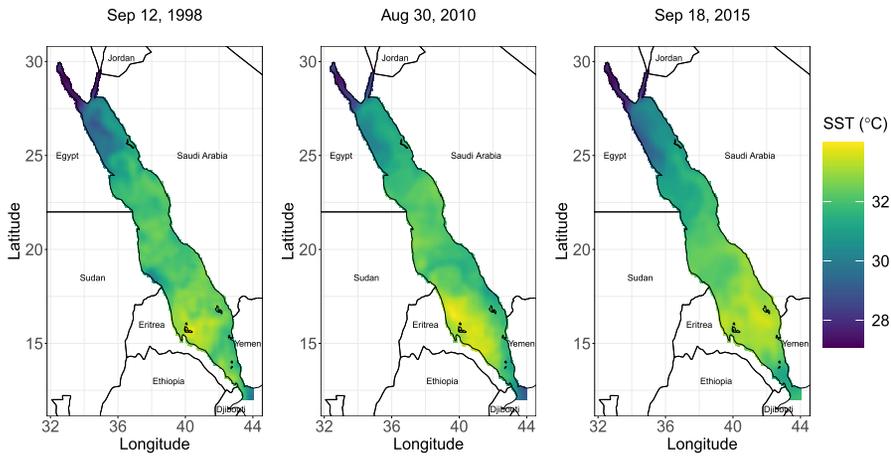
FIG. 1.   *Observed SST profiles across the Red Sea for three extremely hot days: September 12, 1998 (left), August 30, 2010 (middle) and September 18, 2015 (right). All subfigures are on the same scale.*

2. The Red Sea SST dataset and exploratory analysis.

The OSTIA project generates satellite-derived daily SST estimates (free of diurnal variability) at an output grid resolution of $0.05° \times 0.05°$ (about six km). This yields 16,703 grid cells for the whole Red Sea. The data can be freely obtained from the website http://ghrsst-pp.metoffice.com/ostia-website/, and were also (partly) analyzed for the data competition organized for the 11th International Conference on Extreme-Value Analysis, 2019, though with a different objective; see Huser (2021). Figure 1 shows spatial maps of observed SST profiles for three different days with high spatially-averaged SST. For all three days, the SST values are lowest near the Gulfs of Aqaba and Suez in the north and highest in the southern Red Sea near the coast of Eritrea and the southwest of Saudi Arabia.

Some exploratory analysis (not shown) reveals that daily SST data are highly autocorrelated in time and that the temporal dependence strength varies strongly over space. It would be extremely statistically and computationally challenging—if possible at all—to flexibly account for spatially-varying autocorrelation in a single fully Bayesian model for a dataset of this size (16,703 grid cells and 11,315 days). Therefore, for simplicity, we here analyze temporally-thinned data, keeping only one day per week at each grid cell, thus greatly reducing the temporal autocorrelation. Hence, we obtain seven (correlated) subdatasets, each comprising 1612 spatial fields (i.e., one per week), that we treat as independent time replicates. The results are observed to be consistent across the sub-datasets. To concisely summarize the results for all subdatasets, and to reduce the overall uncertainty in our final estimates, we obtain the results separately for each subdataset, check that they are indeed mutually consistent and then report the averages. As our main goal is to draw *spatial* inference, and estimate *spatial* (rather than spatiotemporal) hotspots, this approach is reasonable.

Considering the spatially-varying nature of SST across the Red Sea and the long data collection period of 31 years from 1985 to 2015, a period that faced global warming, we first conduct a preliminary site-by-site average and trend analysis of the annual mean SST as well as the quarterly mean SST by fitting simple linear regression models estimated by least squares. The mean SST across the Red Sea is lowest during the first quarter and highest during the third quarter. The trend profiles (decadal rate of change) of the first-quarterly, third-quarterly, and annual mean SST are displayed in Figure 2. For the first-quarterly and annual mean SST, the slopes are the highest near the latitude 22°N and are the lowest near the southern end of the Red Sea between Eritrea and Yemen. However, for the third-quarterly mean SST, the highest slopes are observed throughout the coast of Egypt in the northwest, and
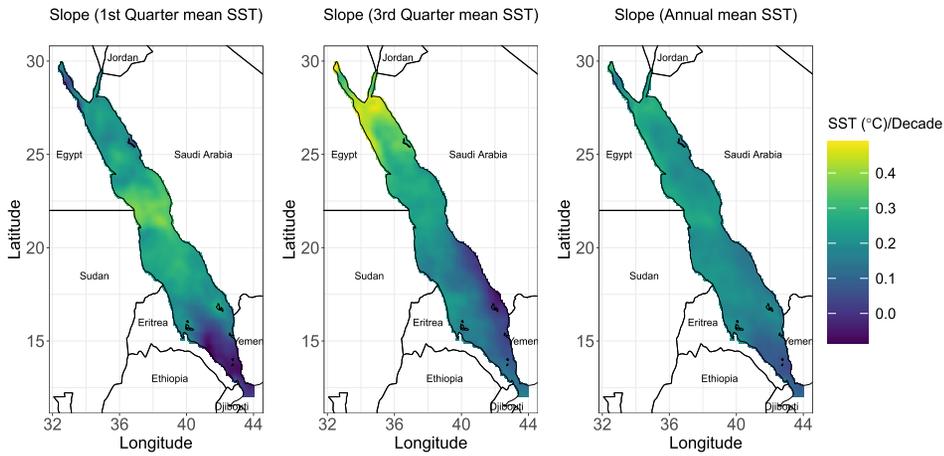
FIG. 2.   *Decadal rate of change of the quarterly average SST for the first (left) and the third (middle) quarters and also of the annual average SST (right) over the Red Sea. All subfigures are on the same scale.*

the lowest values are observed near the southwest of Saudi Arabia. Thus, Figure 2 explains the need for spatially-, as well as seasonally-, or weekly-varying coefficients for modeling the marginal SST distributions.

We then study the seasonality profile (averaged across years) at various grid cells. The results are reported in the Supplementary Material (Hazra and Huser (2021a)). Significantly different patterns are observed throughout the Red Sea. The hottest weeks (maximizing the annual weekly average SST) vary mainly between weeks 32 and 35 (above the latitude 20°N) and weeks 37 and 42 (below 20°N). The observed nonstationarity of SST across weeks explains the need for a flexible modeling of seasonality, for example, through some linear combination of local basis functions with spatially-varying coefficients.

While the seasonal pattern and the spatial variability of the mean SST profile may be estimated from observed data, future SST projections require the use of an additional covariate that accurately represents the long-term trend by incorporating climate physics and anthropogenic influence through greenhouse gas emission scenarios. For this purpose, we consider the output from the Geophysical Fluid Dynamics Laboratory (GFDL) model GFDL-CM3 simulation (Shaltout (2019)) based on current century Coupled Model Intercomparison Project Phase 5 (CMIP5) scenarios RCP 4.5 and RCP 8.5. The GFDL-CM3 simulation outputs can be downloaded from the website https://cds.climate.copernicus.eu. While the GFDL-CM3 simulations are available over a $1° \times 1°$ grid (39 grid cells for the Red Sea), we here average the annual simulated SST for the entire Red Sea region which yields only one value per year. Figure 3 displays the SST projections based on RCP 4.5 and 8.5. We performed two separate analyses based on these two RCP scenarios. The two RCPs are the same during the historic period (1985–2005) and start to diverge during the future period (2006–2100). Here, we mainly focus on our findings based on the RCP 8.5 SST trajectory which is sometimes referred to as a "business-as-usual" scenario (without mitigation measures) and sometimes as a worst-case scenario. We thus consider it as a relatively pessimistic upper bound for SST projections, while RCP 4.5 is a more optimistic, moderate case scenario. For completeness, we also report the results for RCP 4.5 in the Supplementary Material (Hazra and Huser (2021a)).

We then compute the sitewise standard deviations (SDs) of the detrended SST data (with trend obtained by spline smoothing as discussed in Section 3.2). The results are presented in the left panel of Figure 4. Again, there is a highly nonstationary pattern with high SDs near the northeast region and low SDs near the southeast region. The middle panel of Figure 4 shows the spatial correlation structure with respect to the central grid cell (38.48°E, 20.62°N).
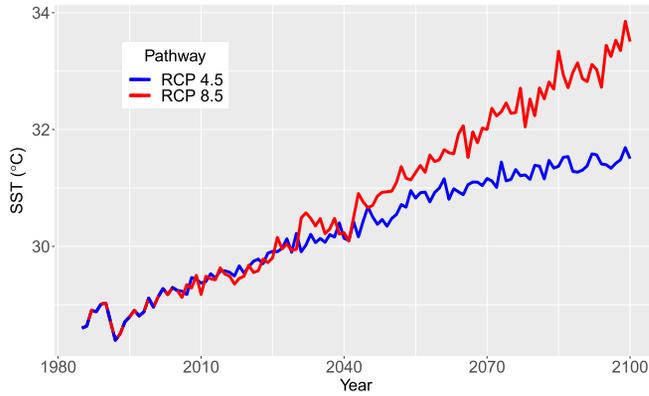
FIG. 3. *Annual mean SST projections for the Red Sea, based on RCP* 4.5 (*blue*) *and RCP* 8.5 (*red*) *scenarios. The two RCPs are the same during the historic period* (1985–2005) *and different during the future period* (2006–2100).

The correlation values in the northern region are significantly higher than the values in the southern region despite being at the same distance. This suggests that the spatial correlation structure is also highly nonstationary.

To investigate the extremal dependence structure, whose specification and estimation is crucial for hotspot detection, we then compute the empirical tail dependence coefficient with respect to the same center grid cell (38.48°E, 20.62°N). The tail dependence coefficient between two random variables Y_1 and Y_2 is defined as $\chi = \lim_{u \to 1} \chi_u$ (provided the limit exists), where

$$(1) \qquad \chi_u = \Pr\{Y_1 > F_1^{-1}(u) \mid Y_2 > F_2^{-1}(u)\}$$

and F_1 and F_2 are the marginal distribution functions of Y_1 and Y_2, respectively. A nonzero value of χ indicates asymptotic dependence while $\chi = 0$ indicates asymptotic independence. Here, we estimate χ with the empirical conditional probability $\hat{\chi}_u$ with $u = 0.99$. The values are reported on the right panel of Figure 4. The observed values are nonzero throughout a major portion of the spatial domain indicating the necessity for a model that can capture nonzero extremal dependence (unlike GPs) at large distances and high thresholds.

Finally, we investigate the bias in estimating high quantiles when fitting normal and Student's t distributions. High biases observed at many grid cells indicate the need for a more
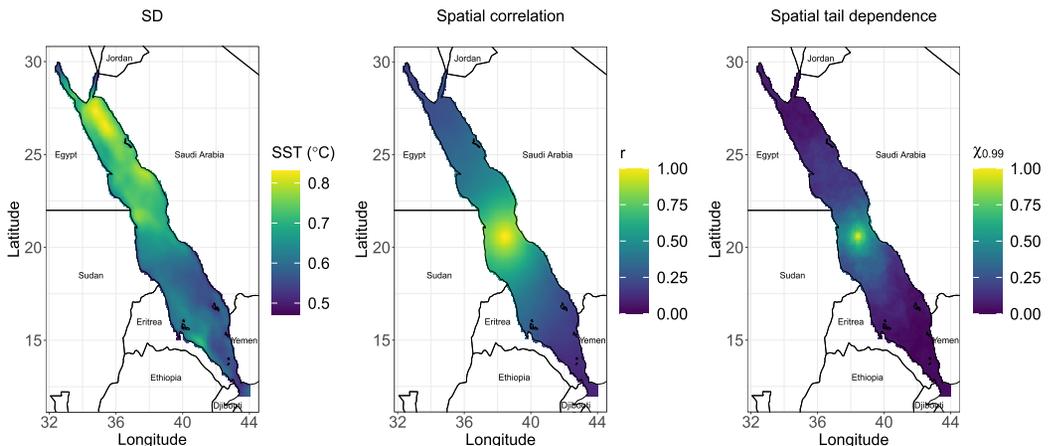

FIG. 4. *Grid cellwise SD* (*left*), *spatial correlation* (r, *middle*) *and extremal dependence* ($\chi_{0.99}$, *right*) *profiles corresponding to the grid cell* (38.48°E, 20.62°N), *respectively.*

flexible model than the usual parametric alternatives. As mixture models fit low through high quantiles of the distributions more flexibly, a semiparametric model is warranted.

More details on the exploratory analysis are provided in the Supplementary Material (Hazra and Huser (2021a)). To summarize, we need a model that accounts for spatially-varying trend and seasonality components within the mean structure, while the residual variability needs to be modeled through a mixture of spatial processes that allows for extremal dependence. Considering the high dimension, a low-rank approach is necessary to ensure that inference is practically feasible.

3. Modeling.

3.1. *General framework.* Let $\mathcal{D}$ denote the spatial domain of the Red Sea. We model the Red Sea SST data as

$$(2) \qquad\qquad Y_t(s) = \mu_t(s) + \varepsilon_t(s),$$

where $Y_t(s)$ denotes the observed SST at location $s \in \mathcal{D}$ and, at time $t \in \{1, 2, \ldots, T = 1612\}$, $\mu_t(s)$ is the mean SST profile and $\varepsilon_t(s)$ is the corresponding error component.

In Section 3.2 we first discuss the modeling of the mean term $\mu_t(s)$, and in Section 3.3 we then discuss the modeling of the residual process $\varepsilon_t(s)$ for $s \in \mathcal{D}$ and $t \in \{1, 2, \ldots, T\}$. In Section 3.4 we specify the overall Bayesian model, and in Section 3.5, we describe the model properties.

3.2. *Mean modeling.* By an abuse of notation, we write $\mu_t(s) = \mu(t_1, t_2, s)$ where $t_1 = \lceil t/52 \rceil$ denotes the year corresponding to time t and $t_2 = t - 52(t_1 - 1)$ denotes the corresponding week within the t_1th year. Here, $t_1 \in \{1, \ldots, T_1 = 31\}$ for 31 years of data and $t_2 \in \{1, \ldots, T_2 = 52\}$ for the 52 weeks within each year. Henceforth, this one-to-one relation between t and (t_1, t_2) holds for the rest of the paper. In spite of the data being observed at discrete time points, we model the mean SST profile as a continuous function of t_1 and t_2.

Let $x_{t_1}^*$ be the mean simulated SST based on a RCP scenario for the t_1th year. We assume that $\mu(t_1, t_2, s)$ is linear in $x_{t_1}^*$, and we define $\mu(t_1, t_2, s) = \beta_1(t_2, s) x_{t_1,1}^{(0)} + \beta_2(t_2, s) x_{t_1,2}^{(0)}$, where $x_{t_1,1}^{(0)}$ is a standardized version of the intercept, $x_{t_1,2}^{(0)}$ is a standardized version of $x_{t_1}^*$, and $\beta_1(t_2, s), \beta_2(t_2, s)$ denote the regression coefficients (intercept and slope with respect to $x_{t_1}^*$, respectively) that vary over space as well as for each week of a specific year. Furthermore, we write the regression coefficients $\beta_{p_0}(t_2, s)$, $p_0 = 1, 2$, as $\beta_{p_0}(t_2, s) = \sum_{p_1=1}^{P_T} \beta_{p_0, p_1}(s) x_{t_2, p_1}^{(1)}$, where $x_{t_2, p_1}^{(1)}$, $p_1 = 1, \ldots, P_T$, are cubic B-splines defined over the continuous interval $[1, T_2]$ with equidistant knots and evaluated at t_2 and thus the seasonal variation is modeled as a smooth function of week within each year. Considering one B-spline per month, we have $P_T = 12$ that helps capturing the monthly-varying features. Finally, let the grid cells within $\mathcal{D}$ be denoted by $s_1, \ldots, s_N$. In order to reduce the computational burden due to the high spatial dimension $N = 16{,}703$, we consider a low-rank approximation of the spatially-varying coefficients by specifying $\beta_{p_0, p_1}(s_n) = \sum_{p_2=1}^{P_S} \beta_{p_0, p_1, p_2} x_{n, p_2}^{(2)}$, where $x_{n, p_2}^{(2)}$ are suitable spatial basis functions. While other choices are also possible, we consider tensor products of cubic B-splines defined over a rectangular surface covering $\mathcal{D}$. A large number of basis functions P_S provides a lot of local detail, but it also increases the computational burden. On the contrary, a small P_S is computationally appealing but might over-smooth the spatial mean surfaces. In practice, this trade-off depends on the computational resources available and the required modeling accuracy.

To be more precise, we now discuss further technical details regarding the design matrices. Let X_0 denote the $(T_1 \times 2)$-dimensional matrix with (t_1, p_0)th entry, denoted by $x_{t_1, p_0}^{(0)}$. We

standardize covariates such that $x_{t_1,1}^{(0)} = T_1^{-1/2}$ and $x_{t_1,2}^{(0)} = (x_{t_1}^* - \overline{x}^*)/\{\sum_{t_1^*=1}^{T_1}(x_{t_1}^* - \overline{x}^*)^2\}^{1/2}$, with $\overline{x}^* = T_1^{-1}\sum_{t_1^*=1}^{T_1} x_{t_1}^*$. This ensures that X_0 is an orthonormal matrix which is important from a computational perspective. For the low-rank structure of $\beta_{p_0}(t_2, s)$, we consider the same B-spline basis functions for $p_0 = 1, 2$ for convenience, though other choices are also possible. We denote the $(T_2 \times P_T)$-dimensional design matrix corresponding to the seasonal effects by X_1, with (t_2, p_1)th entry $x_{t_2,p_1}^{(1)}$. For choosing suitable basis functions for the $\beta_{p_0,p_1}(s_n)$'s, taking the elongated geometry of the Red Sea into account, we place 30 equidistant B-splines along the northwest–southeast direction and 10 equidistant B-splines along the southwest–northeast direction. Out of these 300 B-splines, we only keep the $P_S = 189$ of them that represent more than 99% of the total weight over $s_1, \ldots, s_N$. The knots are well spread across all the Red Sea, and, hence, the splines are able to capture local characteristics quite well. We denote the corresponding $(N \times P_S)$-dimensional design matrix by X_2, with (n, p_2)th entry $x_{n,p_2}^{(2)}$.

In summary, we model the mean SST at spatial location s_n and time point t as

$$(3) \qquad \mu_t(\mathbf{s}_n) = \sum_{p_0=1}^{2} \sum_{p_1=1}^{P_T} \sum_{p_2=1}^{P_S} \beta_{p_0,p_1,p_2} x_{t_1,p_0}^{(0)} x_{t_2,p_1}^{(1)} x_{n,p_2}^{(2)}.$$

Let the spatial mean vector at time t be $\boldsymbol{\mu}_t = [\mu_t(\mathbf{s}_1), \ldots, \mu_t(\mathbf{s}_N)]'$, and, combining all time points, let $\boldsymbol{\mu} = [\boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_T']'$. Grouping the regression coefficients, let $\boldsymbol{\beta}_{p_0,p_1} = [\beta_{p_0,p_1,1}, \ldots, \beta_{p_0,p_1,P_S}]'$ and $\boldsymbol{\beta}_{p_0} = [\boldsymbol{\beta}_{p_0,1}', \ldots, \boldsymbol{\beta}_{p_0,P_T}']'$. Denoting the two columns of X_0 by $X_{0;1}$ and $X_{0;2}$, respectively, we can write $\boldsymbol{\mu}$ in vectorial form as $\boldsymbol{\mu} = [X_{0;1} \otimes X_1 \otimes X_2]\boldsymbol{\beta}_1 + [X_{0;2} \otimes X_1 \otimes X_2]\boldsymbol{\beta}_2$, where $\otimes$ denotes the Kronecker product between two matrices.

3.3. *Spatial dependence modeling.*

We now discuss the modeling of the residual process $\varepsilon_t(\mathbf{s})$. We here assume that the $\varepsilon_t(\mathbf{s})$'s are i.i.d. across time with zero mean, and we write $\varepsilon(\mathbf{s})$ for a generic copy of $\varepsilon_t(\mathbf{s})$. Our semiparametric Bayesian model for the spatial residual processes $\varepsilon_t(\cdot)$ is based on a Dirichlet process mixture (DPM) of parametric low-rank Student's t processes (LTPs). We first describe the construction of LTPs and then discuss the modeling based on mixtures.

3.3.1. *Low-rank Student's t process (LTP).*

LTPs are richer than LGPs, as they have heavier marginal and joint tails, and they can capture spatial extremal dependence contrary to Gaussian processes. At a spatial location $\mathbf{s}$, we model a realization from an LTP as

$$(4) \qquad \varepsilon(\mathbf{s}) = \sigma\{\mathbf{h}'(\mathbf{s})\mathbf{Z} + \eta(\mathbf{s})\}, \quad \mathbf{s} \in \mathcal{D},$$

where $\mathbf{h}(\mathbf{s})$ denotes the vector of length L comprised of some spatial basis functions evaluated at $\mathbf{s}$. The random effects are specified as $\mathbf{Z} \sim \text{Normal}_L(\mathbf{0}, \boldsymbol{\Phi})$ for some positive definite matrix $\boldsymbol{\Phi}$, while $\sigma^2 \sim \text{Inverse-Gamma}(\frac{a}{2}, \frac{a}{2} - 1)$, $a > 2$, and $\eta(\cdot)$ denote a spatial white noise process (i.e., nugget effect) such that $\eta(\mathbf{s}) \overset{\text{i.i.d.}}{\sim} \text{Normal}(0, \tau^2)$, $\tau > 0$.

For the N spatial grid cells $\mathbf{s}_1, \ldots, \mathbf{s}_N$, let $\boldsymbol{\varepsilon} = [\varepsilon(\mathbf{s}_1), \ldots, \varepsilon(\mathbf{s}_N)]'$ be the vector of observed values from the process $\varepsilon(\cdot)$. Moreover, let $\mathbf{H}$ be the $(N \times L)$-dimensional matrix, whose columns are the different spatial basis functions evaluated at $\mathbf{s}_1, \ldots, \mathbf{s}_N$. After marginalization over the random effects $\mathbf{Z}$ and σ^2, the joint distribution of $\boldsymbol{\varepsilon}$ is

$$(5) \qquad \boldsymbol{\varepsilon} \sim T_a\left(\mathbf{0}_N, \frac{a-2}{a}(\mathbf{H}\boldsymbol{\Phi}\mathbf{H}' + \tau^2\mathbf{I}_N)\right),$$

where $T_{\tilde{a}}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$ denotes the multivariate Student's t distribution with location vector $\tilde{\boldsymbol{\mu}}$, dispersion matrix $\tilde{\boldsymbol{\Sigma}}$ and degrees of freedom $\tilde{a}$, $\mathbf{I}_N$ is the N-by-N identity matrix and $\mathbf{0}_N$ is the

zero vector of length N. In case (temporal) replications are available from the spatial process $\varepsilon(\cdot)$, a rough estimate of the spatial covariance matrix of $\boldsymbol{\varepsilon}$ (which exists since $a > 2$) can be obtained upfront using the sample covariance matrix $\widehat{\boldsymbol{\Sigma}} = \widehat{\mathrm{Cov}}(\boldsymbol{\varepsilon})$, with ε obtained from a preliminary least-squares fit of model (2) based on the specification (3) after subtracting the fitted mean from the data. Here, we consider spatial basis functions to be the $L \ll N$ eigenvectors of $\widehat{\boldsymbol{\Sigma}}$ with the largest corresponding eigenvalues. In other words, the matrix $\boldsymbol{H}$ is comprised of empirical orthogonal functions (EOFs). Specifically, let $\boldsymbol{\Delta}$ be the diagonal matrix with the L largest eigenvalues of $\widehat{\boldsymbol{\Sigma}}$ and $\boldsymbol{H}$ be the matrix with column vectors equal to the corresponding eigenvectors. Then, we have the low-rank approximation $\widehat{\boldsymbol{\Sigma}} \approx \boldsymbol{H}\boldsymbol{\Delta}\boldsymbol{H}'$, where the right-hand side has rank L. Other choices of basis functions $\boldsymbol{h}(\boldsymbol{s})$ are also possible (even in case replicates are unavailable); a detailed discussion about the choice of $\boldsymbol{h}(\boldsymbol{s})$ is provided in Wikle (2010). For illustration, the six most important EOFs for the Red Sea data are displayed in the Supplementary Material (Hazra and Huser (2021a)). They reveal interesting geographic patterns which indicate that our EOF-based approach makes sense from a practical perspective. While obtaining the $(N \times N)$-dimensional matrix $\widehat{\boldsymbol{\Sigma}}$ is computationally intensive, we here need to compute it only once upfront (before running the MCMC algorithm), and the function `cova` from the R package `Rfast` (Papadakis et al. (2017)) can be exploited to speed up the computation. Furthermore, obtaining all the eigenvalues and the corresponding eigenvectors would also be time-consuming, but the function `eigs_sym` from the R package `rARPACK` (Qiu and Mei (2016)) can be used to very efficiently compute a small number of highest eigenvalues and the corresponding eigenvectors. While $\boldsymbol{\Phi} = \boldsymbol{\Delta}$ could be assumed to be fixed, we consider instead $\boldsymbol{\Phi}$ in (5) to be unknown and use an informative prior on $\boldsymbol{\Phi}$ with prior mean equal to $\boldsymbol{\Delta}$. The nugget component $\boldsymbol{\eta} = [\eta(\boldsymbol{s}_1), \ldots, \eta(\boldsymbol{s}_N)]'$ is important for $\mathrm{Cov}(\boldsymbol{\varepsilon})$ to be full-rank, though τ^2 is expected to be small. The specific parametrization of σ^2 ensures that $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{H}' + \tau^2 \boldsymbol{I}_N$, so, plugging the prior mean of $\boldsymbol{\Phi}$, we get the approximation $\mathrm{Cov}(\boldsymbol{\varepsilon}) \approx \boldsymbol{H}\boldsymbol{\Delta}\boldsymbol{H}' + \tau^2 \boldsymbol{I}_N \approx \widehat{\boldsymbol{\Sigma}}$. Hence, overall, we construct a zero-mean LTP, where the covariance structure resembles the sample covariance estimated in a preliminary step.

3.3.2. *Dirichlet process mixture* (*DPM*) *of LTPs.* In case we do not have any (temporal) replicates of the process $\varepsilon(\cdot)$, parametric assumptions are required (though $\widehat{\boldsymbol{\Sigma}}$—and hence $\boldsymbol{H}$—are not available and another choice of basis functions is required). However, for the Red Sea data, independent temporal replicates of the anomalies, $\varepsilon_t(\boldsymbol{s})$, $t = 1, \ldots, T$, are available, and we can thus estimate the underlying spatial process semiparametrically.

Considering that our focus mainly lies in inferences from the tail, we here extend the construction (4) by modeling the residuals using a DPM in the same spirit as Hazra et al. (2018), where the characteristics of the bulk and the tail of the anomaly process are described by different mixture components with component-specific parameters. Thus, our approach can automatically and probabilistically cluster observations into weeks characterized by "normal conditions" or weeks characterized by "abnormal (extreme) conditions" without any artificial and subjective thresholding. Thus, tail inference is expected to be minimally influenced by observations from the bulk while providing a reasonable fit on the entire probability range.

Now, we assume that, for all t, $\boldsymbol{\varepsilon}_t = [\varepsilon_t(\boldsymbol{s}_1), \ldots, \varepsilon_t(\boldsymbol{s}_N)]'$ are *i.i.d.*, N-dimensional realizations from a LTP-DPM model with K mixture components for some natural number K. The corresponding multivariate density function is

$$(6) \qquad f_{\mathrm{DPM}}(\boldsymbol{\varepsilon}) = \sum_{k=1}^{K} \pi_k f_T(\boldsymbol{\varepsilon}; \boldsymbol{\Theta}_k),$$

where $\pi_k > 0$ are the mixture probabilities with $\sum_{k=1}^{K} \pi_k = 1$, $\boldsymbol{\Theta}_k$ denotes the set of parameters of the kth LTP component and $f_T(\cdot)$ denotes the density of an N-dimensional realization from the LTP in (5). When $K = \infty$, the model becomes fully nonparametric.

The main advantage of the LTP-DPM model lies in its hierarchical Bayesian model representation. The model can be rewritten as a clustering model, where, conditional on the random cluster label g_t with probability mass function $\Pr(g_t = k) = \pi_k$, $k \in \{1, \ldots, K\}$, we have $\boldsymbol{\varepsilon}_t \sim f_T(\cdot \mid \boldsymbol{\Theta}_{g_t})$. Thus, our LTP-DPM model assumes that weeks with similar residuals can be clustered together, their distribution being described by the same LTP. Treating the cluster labels g_t as unknown, the model accounts for uncertainty in cluster allocation.

We assign a truncated stick-breaking prior distribution (Sethuraman (1994)) on the mixture probabilities, π_k, $k \in \{1, \ldots, K\}$. Precisely, we have the following constructive representation: $\pi_1 = V_1 \sim \text{Beta}(1, \delta)$ for some Dirichlet process concentration parameter $\delta > 0$ and, subsequently, $\pi_k = (1 - \sum_{i=1}^{k-1} \pi_i) V_k$ for $k = 1, \ldots, K - 1$ with $V_k \overset{\text{i.i.d.}}{\sim} \text{Beta}(1, \delta)$. As we consider K to be finite, we set $V_K = 1$ so that $\sum_{k=1}^{K} \pi_k = 1$. In our MCMC implementation, we exploit the one-to-one correspondence between the π_k's and the V_k's by iteratively updating the latter to estimate the former. We write $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_K]' \sim \text{Stick-Breaking}(\delta)$.

3.4. Overall model.

Considering two types of spatial basis functions within $\mu_t(\boldsymbol{s})$ and $\varepsilon_t(\boldsymbol{s})$ leads to collinearity issues between fixed and random effects. To resolve this issue, we divide the matrix $\boldsymbol{X}_2$ from Section 3.2 into two parts based on its projection onto the column space of the matrix $\boldsymbol{H}$ from Section 3.3. The projection matrix is $\boldsymbol{P}_H = \boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}' = \boldsymbol{H}\boldsymbol{H}'$, as $\boldsymbol{H}$ is an orthonormal matrix. We have $\boldsymbol{X}_2 = \boldsymbol{X}_{2;1} + \boldsymbol{X}_{2;2}$ where $\boldsymbol{X}_{2;1} = \boldsymbol{P}_H \boldsymbol{X}_2$ and $\boldsymbol{X}_{2;2} = (\boldsymbol{I}_N - \boldsymbol{P}_H)\boldsymbol{X}_2$, and then rewrite $\boldsymbol{\mu} = \sum_{i=1}^{2} \sum_{j=1}^{2} (\boldsymbol{X}_{0;i} \otimes \boldsymbol{X}_1 \otimes \boldsymbol{X}_{2;j}) \boldsymbol{\beta}_{i;j}$ with two separate vectors of coefficients corresponding to $\boldsymbol{X}_{2;1}$ and $\boldsymbol{X}_{2;2}$ for each of the intercept-related and slope-related terms. For a specific grid cell $\boldsymbol{s}_n$ and time point t, $\mu_t(\boldsymbol{s}_n) = \sum_{i=1}^{2} \sum_{j=1}^{2} x_{t_1}^{(0;i)}(\boldsymbol{x}_{t_2}^{(1)} \otimes \boldsymbol{x}_n^{(2;j)}) \boldsymbol{\beta}_{i;j}$, where $x_{t_1}^{(0;i)}$ is the t_1th entry of $\boldsymbol{X}_{0;i}$, $\boldsymbol{x}_{t_2}^{(1)}$ is the t_2th row of $\boldsymbol{X}_1$, and $\boldsymbol{x}_n^{(2;j)}$ is the nth row of $\boldsymbol{X}_{2;j}$.

Overall, given the cluster label $g_t = k$, our hierarchical model is defined as follows:

$$Y_t(\boldsymbol{s}) = \mu_t(\boldsymbol{s}) + \sigma_t\{\boldsymbol{h}'(\boldsymbol{s})\boldsymbol{Z}_t + \eta_t(\boldsymbol{s})\},$$

$$(7) \qquad \boldsymbol{Z}_t \sim \text{Normal}_L(\boldsymbol{0}, \boldsymbol{\Phi}_k), \qquad \eta_t(\boldsymbol{s}) \overset{\text{i.i.d.}}{\sim} \text{Normal}(0, \tau_k^2),$$

$$\sigma_t^2 \sim \text{Inverse-Gamma}\left(\frac{a_k}{2}, \frac{a_k}{2} - 1\right),$$

where $\mu_t(\boldsymbol{s})$ is the mean SST profile, $\boldsymbol{h}(\boldsymbol{s})$ is the vector of spatial basis functions as described in Section 3.3, and $\boldsymbol{Z}_t$ and σ_t denote independent copies of the corresponding random effects. The set of parameters of the LTP corresponding to time t is $\boldsymbol{\Theta}_k = \{\boldsymbol{\Phi}_k, \tau_k^2, a_k\}$. We treat the cluster-specific parameters $\boldsymbol{\Theta}_k$ as unknown and put hyperpriors on them. We assume that $\boldsymbol{\Theta}_k \overset{\text{i.i.d.}}{\sim} G_{\boldsymbol{\Theta}}$, where the components of $\boldsymbol{\Theta}_k$, that is, $\boldsymbol{\Phi}_k$, τ_k^2 and a_k, are treated as independent of each other. Our choice of hyperpriors is discussed in Section 4.1.

3.5. Model properties.

For model (7) the conditional mean and covariance structure of $Y_t(\boldsymbol{s}_n)$, given the coefficients $\boldsymbol{\beta}_{i;j}$, and the cluster-specific parameters $\boldsymbol{\Theta}_k$ are

$$E\{Y_t(\boldsymbol{s}_n) \mid \boldsymbol{\beta}_{i;j}, i, j = 1, 2\} = \mu_t(\boldsymbol{s}_n) = \sum_{i=1}^{2} \sum_{j=1}^{2} x_{t_1}^{(0;i)}(\boldsymbol{x}_{t_2}^{(1)} \otimes \boldsymbol{x}_n^{(2;j)}) \boldsymbol{\beta}_{i;j},$$

$$(8)$$

$$\text{Cov}\{Y_t(\boldsymbol{s}_{n_1}), Y_t(\boldsymbol{s}_{n_2}) \mid \boldsymbol{\Theta}_k; k = 1, \ldots, K\} = \sum_{k=1}^{K} \pi_k(\boldsymbol{h}_{n_1} \boldsymbol{\Phi}_k \boldsymbol{h}'_{n_2} + \tau_k^2 \mathbb{I}_{\{n_1=n_2\}}),$$

where $\boldsymbol{h}_n$ denotes the nth row of $\boldsymbol{H}$ and $\mathbb{I}_{\{n_1=n_2\}} = 1$ if $n_1 = n_2$ and zero otherwise.

The mean structure is nonstationary both in space and time, nonlinear, and includes interaction terms between the spatial and temporal effects. Thus, the model can capture spatially-varying seasonality and trend as well as seasonally-varying trends at each spatial grid cell. Because we specify high-resolution B-spline knot locations both over space and time, our model can capture the local features in the mean behavior reasonably well.

The covariance between the observations at grid cells s_{n_1} and s_{n_2} is dependent on both s_{n_1} and s_{n_2} and cannot be reduced to a function of the spatial lag $s_{n_1} - s_{n_2}$. Hence, the covariance structure is also nonstationary. For the Dirichlet process atoms $\boldsymbol{\Phi}_k$, $k = 1, \ldots, K$, we consider priors so that $E(\boldsymbol{\Phi}_k) = \boldsymbol{\Delta}$; recall Section 3.3. Marginalizing over $\boldsymbol{\Phi}_k$, we get $\mathrm{Cov}\{Y_t(s_{n_1}), Y_t(s_{n_2})\} = \boldsymbol{h}_{n_1} \boldsymbol{\Delta} \boldsymbol{h}'_{n_2} + \tau^2 \mathbb{I}_{\{n_1 = n_2\}}$, where $\tau^2 = \sum_{k=1}^{K} \pi_k \tau_k^2$. Considering τ^2 to be small, we get the approximation $\mathrm{Cov}(\boldsymbol{Y}_t) \approx \widehat{\boldsymbol{\Sigma}}$, where $\boldsymbol{Y}_t = [Y_t(s_1), \ldots, Y_t(s_N)]'$ and $\widehat{\boldsymbol{\Sigma}}$ is as discussed in Section 3.3. Thus, the LTP-DPM model is centered around a low-rank Student's t process, constructed as in Section 3.3 with mean structure discussed in Section 3.2.

Considering $\varepsilon_t(\cdot)$ as an infinite-dimensional spatial process and ignoring the nugget term, for $L = \infty$, we can write $\varepsilon_t(s) = \sum_{l=1}^{\infty} Z_{lt} h_l(s)$. For any pair l_1 and l_2 with $l_1 \neq l_2$, the vector $[Z_{l_1 t}, Z_{l_2 t}]'$ follows a DPM of bivariate zero-mean Student's t distributions which spans any bivariate zero-mean distribution when $K = \infty$. Integrating with respect to the priors of $\boldsymbol{\Phi}_k$, $\mathrm{Cov}(Z_{l_1 t}, Z_{l_2 t}) = 0$. Thus, for $L, K = \infty$, the covariance of the proposed model satisfies the criteria of the Karhunen–Loève Theorem (Alexanderian (2015)) and spans the covariance structure of all squared-integrable stochastic processes with continuous covariance functions. Under suitable regularity conditions, posterior consistency of the proposed model holds (Wu and Ghosal (2010), Ghosal and van der Vaart (2017)).

The spatial tail dependence coefficient between two different grid cells s_{n_1} and s_{n_2}, defined in (1), is nonstationary for the proposed LTP-DPM model, and it is easy to show that it is determined by the heaviest-tailed mixture component in (6). Using the tail properties of Student's t copulas (Demarta and McNeil (2005)), its expression may be written explicitly as $\chi(s_{n_1}, s_{n_2}) = 2\overline{F}_T(\sqrt{(a_m + 1)\{1 - r_m(s_{n_1}, s_{n_2})\}/\{1 + r_m(s_{n_1}, s_{n_2})\}}; 0, 1, a_m + 1)$, where $m = \arg\min_k\{a_k\}$ and $r_m(s_{n_1}, s_{n_2}) = \boldsymbol{h}_{n_1} \boldsymbol{\Phi}_m \boldsymbol{h}'_{n_2} / \sqrt{\prod_{\tilde{n} = n_1, n_2}(\boldsymbol{h}_{\tilde{n}} \boldsymbol{\Phi}_m \boldsymbol{h}'_{\tilde{n}} + \tau_m^2)}$ denotes the underlying spatial correlation of the Gaussian term characterizing the mth mixture component, and $\overline{F}_T(\cdot; 0, 1, a) = 1 - F_T(\cdot; 0, 1, a)$ is the survival function for a standard (univariate) Student's t distribution with a degrees of freedom; see, also Hazra et al. (2018). Here, $\chi(s_{n_1}, s_{n_2}) > 0$ for any pair of sites, so our model can capture asymptotic dependence, unlike Gaussian processes, and is an increasing function of $r_m(s_{n_1}, s_{n_2})$. Asymptotic dependence holds, however, even when $r_m(s_{n_1}, s_{n_2}) = 0$ (unless $a_m \to \infty$). Therefore, the model cannot capture full independence, and dependence persists at large distances, even in the tails. This is a downside of the proposed model in case the spatial domain is large and the process of interest is rough across the domain (e.g., with precipitation or wind speed data), but this should not be a big limitation for our Red Sea SST data which remain strongly dependent at large distances. From equation (8) and the expression for $\chi(s_{n_1}, s_{n_2})$ above, we can see that the dependence structure (both covariance, and tail dependence) is nonstationary and is not impacted by the smooth spatiotemporal mean structure.

4. Bayesian inference and identification of hotspots.

4.1. *Hyperpriors and an efficient Gibbs sampler.* We draw posterior inference about the model parameters in our model using Markov chain Monte Carlo (MCMC) sampling. Except for the degrees of freedom parameters, a_k, $k = 1, \ldots, K$, of the LTP components of the LTP-DPM model, conjugate priors exist for all other model parameters which allows Gibbs sampling. While Metropolis–Hastings sampling would be possible for the a_k's, we prefer to consider discrete uniform priors on a fine grid of values which allows drawing samples from

the full conditional distributions in a fast and easy manner. Due to its numerical stability, this strategy is often considered in the literature; for example, Gelfand, Kottas and MacEachern (2005) use it for posterior sampling from the range parameter of the spatial Matérn covariance; see also Morris et al. (2017) and Hazra, Reich and Staicu (2020).

For the vectors of fixed effects involved within the mean terms, $\boldsymbol{\beta}_{i;j}$, $i = 1, 2$, $j = 1, 2$, we consider the priors $\boldsymbol{\beta}_{i;j} \sim \text{Normal}_P(\mu_{i;j} \mathbf{1}_P, \sigma_{i;j}^2 \boldsymbol{I}_P)$, where $\mathbf{1}_P$ is the P-dimensional vector of ones and $\boldsymbol{I}_P$ is the P-by-P identity matrix with $P = P_{\mathcal{T}} P_{\mathcal{S}} = 2268$. While the posterior distribution of $\boldsymbol{\beta}_{i;j}$ is also a P-variate normal distribution, we exploit the Kronecker product structure of the high-dimensional posterior covariance matrix to avoid its computationally challenging inversion. For the hyperparameters $\mu_{i;j}$, we consider the relatively noninformative priors $\mu_{i;j} \sim \text{Normal}(0, s_{i;j}^2)$ with $s_{1;j} = 10^2$ and $s_{2;j} = 10$ for $j = 1, 2$. The parameter vectors $\boldsymbol{\beta}_{1;j}$, $j = 1, 2$, correspond to the intercept term, while $\boldsymbol{\beta}_{2;j}$, $j = 1, 2$, correspond to the slopes with respect to the simulated RCP-based SST estimates. The absolute values of the intercept-related terms are likely to be large, while the slope terms are likely to be small, and, thus, we consider flatter prior for the $\mu_{1;j}$'s. For the hyperparameters $\sigma_{i;j}^2$, we consider the priors $\sigma_{i;j}^2 \sim \text{Inverse-Gamma}(a_{i;j}, b_{i;j})$, $i = 1, 2$, $j = 1, 2$. We fix the hyperparameters to $a_{1;j} = b_{1;j} = 0.01$ and $a_{2;j} = b_{2;j} = 0.1$ for $j = 1, 2$. While both priors are quite noninformative, we choose the hyperpriors differently, following a similar logic as the one used when considering the priors for the $\mu_{i;j}$'s.

The parameters involved in the distribution of the error terms $\varepsilon_t(\boldsymbol{s}_n)$ are the component-specific parameters and hyperparameters of the DPM model described in Section 3.3. For the purpose of computation, we fix the number of components in the stick-breaking prior by setting $V_K = 1$ for some finite integer K. The choice of K is problem-specific and leads to a bias–variance trade-off. Large K is desirable to increase the model flexibility (i.e., decrease the bias), but considering K to be very large may lead to spurious estimates, as the sampling from the parameters of an LTP component depends on the observations from that specific cluster and a large K may lead to very few observations within some clusters (thus increasing the variance). In our application, we fit different models with $K = 1, 5, 10$ and compare them by cross-validation (see Section 5.1). The prior choices for the DPM model parameters are $\boldsymbol{\Phi}_k \overset{\text{i.i.d.}}{\sim} \text{Inverse-Wishart}(L + 2, \boldsymbol{\Delta})$, where $\boldsymbol{\Delta}$ is the diagonal containing the L largest eigenvalues of $\widehat{\boldsymbol{\Sigma}}$ (recall Section 3.3), $\tau_k^2 \overset{\text{i.i.d.}}{\sim} \text{Inverse-Gamma}(1, 1)$, and $a_k \overset{\text{i.i.d.}}{\sim} \text{Discrete-Uniform}(2.1, 2.2, \ldots, 40.0)$. The prior for $\boldsymbol{\Phi}_k$ is conjugate, and it ensures that $E(\boldsymbol{\Phi}_k) = \boldsymbol{\Delta}$, although the variance of its elements is infinite. We choose hyperparameters for the prior of τ_k^2 so that the mass is mainly distributed near zero. The existence of a continuous conjugate prior for a_k is unknown, and the discrete uniform prior avoids the need of Metropolis–Hastings sampling, as mentioned above. The support of the distribution considered here covers a wide range of degrees of freedom. The smallest values within the support of a_k correspond to a heavy-tailed and strongly dependent process, while the largest values practically correspond to a near-Gaussian behavior. For the mixing probabilities π_k, we consider a stick-breaking prior as discussed in Section 3.3. The distribution of $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_K]'$ involves the unknown concentration hyperparameter δ. We here consider a fairly noninformative conjugate hyperprior, $\delta \sim \text{Gamma}(0.1, 0.1)$.

The full conditional distributions required for model fitting and prediction are provided in the Supplementary Material (Hazra and Huser (2021a)). In our data application, we implement the MCMC algorithm in R (http://www.r-project.org), and our commented R code is also provided in the Supplementary Material (Hazra and Huser (2021b)). We generate 60,000 posterior samples and discard the first 10,000 iterations as burn-in period. Subsequently, we thin the Markov chains by keeping one out of five consecutive samples, and thus, we finally obtain $B = 10,000$ samples for drawing posterior inference. Convergence and mixing are monitored through trace plots.

4.2. *Hotspot estimation.* Our main goal is to exploit the observed Red Sea SST data to identify extreme hotspots, that is, to construct a "confidence region" that contains joint threshold exceedances of some (very) high threshold u at some future time t_0 with a pre-defined probability. Our proposed approach developed below generalizes French and Sain (2013) (who focus on Gaussian processes only, in the frequentist setting) and French and Hoeting (2016) (extended to the Bayesian setting) to the more general and flexible case of LTP-DPMs which better capture the joint tail behavior of complex spatiotemporal processes. While other approaches based on alternative definitions of hotspots are also possible (see, e.g., Bolin and Lindgren (2015)), the pragmatic sampling-based method of French and Sain (2013) that we adopt here is easy to implement and computationally convenient, as it does not rely on the solution of a complex optimization problem as in Bolin and Lindgren (2015).

Let $\mathcal{D}' = \{s_1, \ldots, s_N\} \subset \mathcal{D}$ be our discretized domain of interest. We define the exceedance region above a fixed threshold u at time t_0 as $E^0_{u+} = \{s \in \mathcal{D}' : Y_{t_0}(s) \geq u\}$. Note that, because $Y_{t_0}(\cdot)$ is a random process, E^0_{u+} is a random set. Our goal is to find a region $\mathcal{D}^0_{u+}$ that contains the "true" exceedance region E^0_{u+} with some predefined probability $1 - \alpha$, that is, $\Pr(E^0_{u+} \subseteq \mathcal{D}^0_{u+}) = 1 - \alpha$. While the target region $\mathcal{D}^0_{u+}$ is generally not unique, a valid solution to this problem may be obtained by viewing it through the lenses of multiple hypothesis testing. The proposed approach works essentially as follows. For each fixed grid cell $s_n \in \mathcal{D}'$, we first individually test the null hypothesis $H_0 : Y_{t_0}(s_n) = u$ against the alternative $H_1 : Y_{t_0}(s_n) < u$ based on some test statistic denoted by $\tilde{Y}_{t_0}(s_n)$. An obvious choice for $\tilde{Y}_{t_0}(s_n)$ is to exploit (a rescaled version of) $\widehat{Y}_{t_0}(s_n)$, a predictor of $Y_{t_0}(s_n)$. Then, to find the exceedance region, we combine these single-cell tests together by collecting all grid cells $s_n \in \mathcal{D}'$ where we fail to reject H_0. However, if each test is performed individually at the confidence level $1 - \alpha$, this approach accurately identifies marginal threshold exceedances, while it fails to correctly represent joint exceedances at the required *overall* confidence level. Thus, we finally account for multiple testing by appropriately adjusting the critical value of the tests in order to reach an overall familywise error rate α. Notice that, because of spatial proximity, all the individual tests are dependent, and a sampling-based approach is here used to precisely set the critical value.

We now describe how we design the test statistic and how we set the critical value. First, consider the marginal predictive distributions based on our proposed LTP-DPM model (7). We have $Y_{t_0}(s_n) = \mu_{t_0}(s_n) + \varepsilon_{t_0}(s_n)$. Let $t_{01} = \lceil t_0/52 \rceil$ and $t_{02} = t_0 - 52(t_{01} - 1)$ be the year and week, respectively, corresponding to the time point of interest, t_0. From Section 3.2 and Section 3.4, the mean of $Y_{t_0}(s_n)$ has the form

$$(9) \qquad \mu_{t_0}(s_n) = \sum_{i=1}^{2} \sum_{j=1}^{2} x_{t_{01}}^{(0;i)} (x_{t_{02}}^{(1)} \otimes x_n^{(2;j)}) \beta_{i;j},$$

where $x_{t_{01}}^{(0;1)} = T_1^{-1/2}$, $x_{t_{01}}^{(0;2)} = (x_{t_{01}}^* - \overline{x}^*)/\{\sum_{t_1^*=1}^{T_1}(x_{t_1^*}^* - \overline{x}^*)^2\}^{1/2}$ (with $\overline{x}^*$ as in Section 3.2) and $x_{t_{02}}^{(1)}$ is the t_{02}th row of X_1. The density of the corresponding error term $\varepsilon_{t_0}(s_n)$ is

$$(10) \qquad f(\varepsilon) = \sum_{k=1}^{K} \pi_k f_T \left\{ \varepsilon; 0, \frac{a_k - 2}{a_k}(h_n \Phi_k h_n' + \tau_k^2), a_k \right\},$$

with $f_T(\cdot; 0, \tilde{\sigma}^2, \tilde{a})$ the univariate Student's t density function with location 0, scale $\tilde{\sigma}$ and degrees of freedom $\tilde{a}$. Now, let $\widehat{Y}_{t_0}(s_n)$ be the posterior mean of $Y_{t_0}(s_n)$, estimated by averaging B roughly independent posterior predictive samples of $Y_{t_0}(s_n)$ based on (9) and (10). Here, we have $B = 10^4$, and we simulate posterior samples for the entire spatial process at once to account for spatial dependence among the tests.

From the Bayesian central limit theorem, we then have the large-sample approximation $\widehat{Y}_{t_0}(s_n) \mid Y_{t_0}(s_n) \overset{\cdot}{\sim} \text{Normal}(Y_{t_0}(s_n), \tilde{\sigma}_{t_0}^2(s_n)/B)$, where $\tilde{\sigma}_{t_0}(s_n)$ denotes the posterior standard deviation of $Y_{t_0}(s_n)$, which we estimate from the B posterior samples. This leads us to consider the test statistic

$$\text{(11)} \qquad \tilde{Y}_{t_0}(s_n) = \sqrt{B} \frac{\widehat{Y}_{t_0}(s_n) - u}{\tilde{\sigma}_{t_0}(s_n)}$$

and to define a rejection region for $H_0 : Y_{t_0}(s_n) = u$ (against $H_1 : Y_{t_0}(s_n) < u$) of the form $\{\tilde{Y}_{t_0}(s_n) < C_\alpha\}$ for some critical level C_α. Under H_0, $\tilde{Y}_{t_0}(s_n)$ has an approximate standard normal distribution. We further note that this test is equivalent to testing $H_0 : Y_{t_0}(s_n) > u$ against $H_1 : Y_{t_0}(s_n) < u$.

We finally need to adjust the critical value C_α, so that the familywise Type I error rate is α. This ensures a confidence level of $1 - \alpha$ for the confidence region $\mathcal{D}_{u+}^0$. A Type I error can only occur at locations s_n within the true exceedance region, and, hence, to control the overall Type I error, we need to consider the test statistic values only within E_{u+}^0. Therefore, the critical value C_α should be chosen such that $\Pr(\min_{s_n \in E_{u+}^0} \{\tilde{Y}_{t_0}(s_n)\} < C_\alpha) = \alpha$. However, in practice, we ignore the true exceedance region E_{u+}^0, and, as a result, the distribution of $\min_{s_n \in E_{u+}^0} \{\tilde{Y}_{t_0}(s_n)\}$ is also unknown. To circumvent this issue, we can again exploit the B posterior samples to approximate the region E_{u+}^0 as well as the required probability. The critical level C_α is obtained as the empirical α-quantile based on posterior samples of $\min_{s_n \in E_{u+}^0} \{\tilde{Y}_{t_0}(s_n)\}$. The steps for sampling from the posterior predictive distribution, along with estimating the critical value C_α empirically, and for obtaining the estimated exceedance set $\mathcal{D}_{u+}^0$ are provided in Algorithm 1. The proof that the familywise error rate is indeed α, as required for the described algorithm, is provided in the Supplementary Material (Hazra and Huser (2021a)).

Algorithm 1 Hotspot estimation at a future time t_0

- Fix an overall Type-I error α and a threshold level u.
- Fit the model (7) to the data, and obtain B posterior samples from the parameters and hyperparameters.
- For each posterior sample $b = 1, \ldots, B$:
    - calculate $\mu_{t_0}(s_n)$ according to (9), and let $\mu_{t_0}^{(b)}(s_n)$ denote the the bth sample.
    - plug in $\pi_k = \pi_k^{(b)}$, and simulate $g_{t_0}^{(b)}$ from $\Pr(g_{t_0} = k) = \pi_k^{(b)}$. Assume that $g_{t_0}^{(b)} = k_b$.
    - plug in $a_k = a_k^{(b)}$, and simulate $\sigma_{t_0}^{2(b)}$ from $\sigma_{t_0}^2 \sim \text{Inverse-Gamma}(\frac{a_{k_b}^{(b)}}{2}, \frac{a_{k_b}^{(b)}}{2} - 1)$.
    - plug in $\Phi_k = \Phi_k^{(b)}$, and simulate $Z_{t_0}^{(b)}$ from $Z_{t_0} \sim \text{Normal}_L(\mathbf{0}, \Phi_{k_b}^{(b)})$.
    - plug in $\tau_k^2 = \tau_k^{2(b)}$, and simulate $\eta_{t_0}^{(b)}(s_n), n = 1, \ldots, N$, from $\eta_{t_0}(s_n) \overset{\text{i.i.d.}}{\sim} \text{Normal}(0, \tau_{k_b}^{2(b)})$.
    - calculate $\varepsilon_{t_0}^{(b)}(s_n) = \sigma_{t_0}^{(b)}\{h_n' Z_{t_0}^{(b)} + \eta_{t_0}^{(b)}(s_n)\}$, the bth sample from $\varepsilon_{t_0}(s_n)$.
    - calculate $Y_{t_0}^{(b)}(s_n) = \mu_{t_0}^{(b)}(s_n) + \epsilon_{t_0}^{(b)}(s_n)$, the bth posterior predictive sample from $Y_{t_0}(s_n)$.
- Based on $Y_{t_0}^{(b)}(s_n), b = 1, \ldots, B$:
    - calculate $\tilde{Y}_{t_0}(s_n)$ according to (11), at each s_n.
    - identify the exceedance region $E_{u+}^b = \{s_n \in \mathcal{D}' : Y_{t_0}^{(b)}(s_n) \geq u\}$, for each b.
    - calculate $\min_{s_n \in E_{u+}^b} \{\tilde{Y}_{t_0}(s_n)\}$, for each b.
    - estimate C_α by $\widehat{C}_\alpha$, the empirical α-quantile of $\{\min_{s_n \in E_{u+}^b} \{\tilde{Y}_{t_0}(s_n)\}; b = 1, \ldots, B\}$.
- Return $\mathcal{D}_{u+}^0 = \{s_n \in \mathcal{D}' : \tilde{Y}_{t_0}(s_n) \geq \widehat{C}_\alpha\}$.

5. Data application.

5.1. *Model fitting and cross-validation study.*   Here, we present the results based on RCP 8.5, considering it as a relatively pessimistic greenhouse gas emission scenario. The results based on RCP 4.5 (moderate mitigation pathway) are provided in the Supplementary Material (Hazra and Huser (2021a)). To fit our model, we first need to fix the number of mixture components, K, and the number of spatial basis functions (EOFs), L. Both K and L affect the bias-variance trade-off and the computational burden. We specify $K = 5, 10$ and $L = \arg\max_{l \in \{1,\dots,N\}} \{\lambda_l \geq q\lambda_1\}$ with $q = 0.005, 0.01$, where $\lambda_1 > \lambda_2 > \cdots, \lambda_N$ are the ordered eigenvalues of $\widehat{\Sigma}$ (recall Section 3.3). For $q = 0.01$ and $q = 0.005$, we denote the values of L by L_1 and L_2, respectively. For the Red Sea SST data, we obtain $L_1 = 15$ and $L_2 = 24$ for all the seven weekly subdatasets. We then compare the different choices of K and L by cross-validation. Moreover, we also compare our proposed LTP-DPM model to some simpler parametric and semiparametric submodels, namely, LGP ($K = 1, a_k = \infty$), LTP ($K = 1$) and LGP-DPM ($a_k = \infty$).

For model comparison we divide each weekly Red Sea SST subdataset into two parts, using the years 1985–2010 (1352 weeks) for training and keeping the years 2011–2015 (260 weeks) for testing. For each spatiotemporal observation in the test set, we estimate the posterior predictive distribution, based on the posterior samples, and confront the estimated distribution with the test observation. Because our primary goal is to predict high threshold exceedances, we use the Brier score (BS) and the tail-weighted continuous rank probability score (TWCRPS), proposed by Gneiting and Raftery (2007) and Gneiting and Ranjan (2011), respectively. For a single test sample y, the BS at a given level u is defined as $\mathrm{BS}_u(y, F) = \{\mathbb{I}_{\{y>u\}} - \overline{F}(u)\}^2$, where $\overline{F}(\cdot) = 1 - F(\cdot)$ is the survival function corresponding to the posterior predictive distribution F. For a single test sample y, the TWCRPS is defined as $\mathrm{TWCRPS}_w(y, F) = \int_{-\infty}^{\infty} w(x)\{F(x) - \mathbb{I}_{\{y \leq x\}}\}^2 \, dx$, where $w(\cdot)$ is a nonnegative weight function. To focus on the upper tail, we use $w(x) = \mathbb{I}_{\{x>u\}}$ so that $\mathrm{TWCRPS}_w(y, F) = \int_u^{\infty} \mathrm{BS}_x(y, F) \, dx$. Then, we define the Brier skill score (BSS) and tail-weighted continuous rank probability skill score (TWCRPSS) for a model M as

$$\mathrm{BSS}_M = \frac{\mathrm{BS}_{\mathrm{LGP}} - \mathrm{BS}_M}{\mathrm{BS}_{\mathrm{LGP}}} \times 100\%,$$

(12)

$$\mathrm{TWCRPSS}_M = \frac{\mathrm{TWCRPS}_{\mathrm{LGP}} - \mathrm{TWCRPS}_M}{\mathrm{TWCRPS}_{\mathrm{LGP}}} \times 100\%,$$

using the LGP with $L = L_1$ as the benchmark, where BS_M and TWCRPS_M are the shorthand notation for the BS and TWCRPS for a model M, respectively. Higher values of BSS or TWCRPSS indicate better prediction performance (compared to the benchmark). We report the results by averaging values over the test set. We chose L_1 basis functions for the LGP benchmark because it corresponds to the simplest (i.e., most parsimonious) model fitted here. Figure 5 reports the skill scores plotted as a function of the threshold u which ranges between the 95% and the 99.9%-quantiles of the SST data. As all scores are positive, the LGP benchmark is consistently the worst, while our proposed LTP-DPM model (blue curves) performs better than the others in most cases. Comparing across the choices of K and L, the predictive performances of the LTP-DPM model for the cases $K = 10, L = L_1$ and $K = 5, L = L_2$ are comparable, while the skill scores are noticeably lower for $K = 5, L = L_1$ (most parsimonious LTP-DPM) and slightly lower for $K = 10, L = L_2$ (most complex LTP-DPM). Among the two intermediate cases with best performances ($K = 10, L = L_1$ and $K = 5, L = L_2$), the model with $K = 10, L = L_1$ is the most parsimonious. As there is no visible difference in their performances, we, therefore, proceed with $K = 10, L = L_1$ to draw inference.
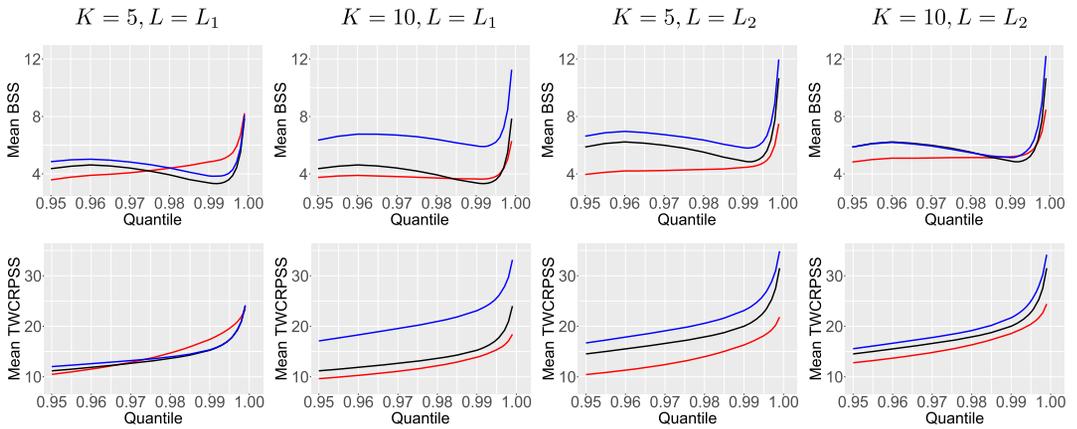
FIG. 5. *Averaged BSS (top) and TWCRPSS (bottom) for the models LTP (black), LGP-DPM (red) and LTP-DPM (blue) considering the LGP as the reference model, plotted as a function of the threshold u ranging from the 95%-quantile to the 99.9%-quantile of the observed data. Higher values of BSS and TWCRPSS indicate better prediction performance.*

Detailed goodness-of-fit diagnostics for the LTP-DPM model are provided in the Supplementary Material (Hazra and Huser (2021a)). To summarize, the mean and standard deviation profiles are very well estimated overall, as demonstrated by very small root mean squared differences between their corresponding empirical and fitted model-based estimates. Moreover, while the difference between empirical and fitted pairwise correlations varies between -0.1 and 0.1, the empirical and fitted χ-coefficients appear to be more variable. There is indeed a higher degree of uncertainty involved in the empirical χ-coefficients, due to the sparsity of extreme events, which results in larger differences with their model-based counterparts. This, however, does not necessarily indicate a lack of fit, as further explained in the Supplementary Material (Hazra and Huser (2021a)).

To explore the DPM models further, we investigate the posterior distribution of stick-breaking probabilities π_k. Because the π_k's are not identifiable due to label-switching within the MCMC algorithm, Figure 6 displays the estimated densities of the ordered probabilities, $\pi_{(k)}$, $k = 1, \ldots, K$, for $K = 10$ and $L = L_1$. For the LTP-DPM model, one single cluster fits a large proportion of the temporal replicates, while the remaining nine clusters capture the behavior of "abnormal weeks," that is, extreme events. This property is desirable, as it allows
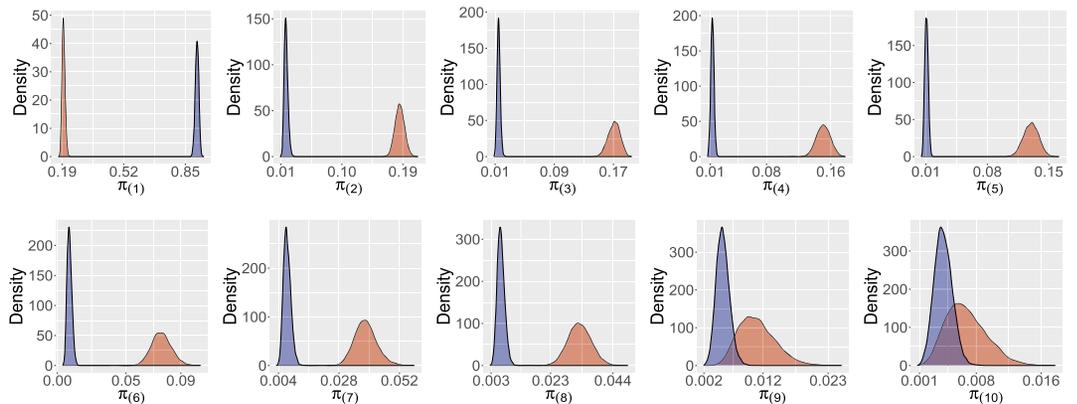

FIG. 6. *Posterior densities of the 10 ordered stick-breaking probabilities $\pi_{(1)} > \cdots > \pi_{(K)}$, based on fitting the LGP-DPM (red) and LTP-DPM (blue) models with $K = 10$ mixture components and $L = L_1$ spatial EOFs.*
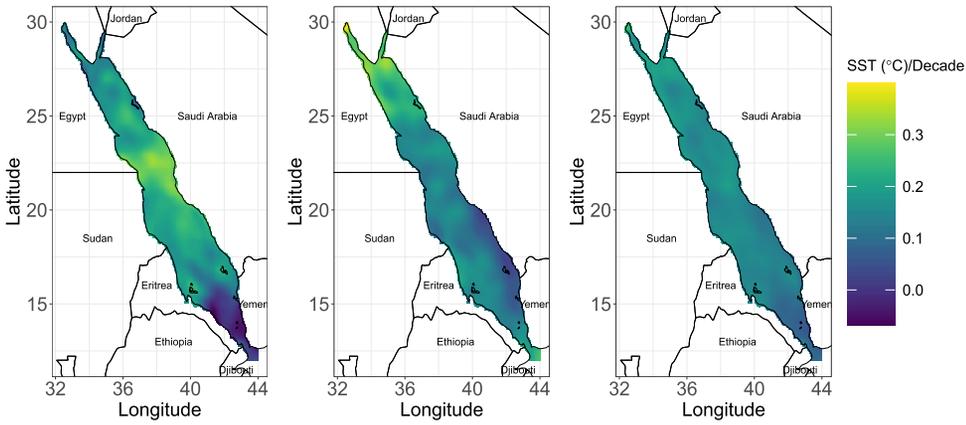
FIG. 7.   *Estimated decadal rate of change in mean SST profile for the coolest (seventh, left) and the warmest (34th, middle) weeks along with the estimated overall rate obtained by averaging across the 52 weeks (right). The subfigures are on the same scale.*

us to have separate control over extreme events, being described by different mixture-specific parameters. On the other hand, the LGP-DPM model allocates large probabilities to many clusters. The thin-tailed LGP cluster components fail to allocate the bulk of the distribution into a single cluster for our heavy-tailed SST data. This makes the identification of the bulk and the tail more challenging and requires significantly more clusters for model fitting.

5.2. *Estimated time trend and return levels.*   We now discuss the estimated spatial maps of the decadal rate of change (DRC) in mean SST and the return levels (adjusted for nonstationarity) based on our best LTP-DPM model.

As the mean SST incorporates the simulated RCP-based SST projections as covariate, we define the DRC for a specific week t_2 at a spatial location s as $\mathrm{DRC}_{t_2}(s) = 10\{\mu(T_1, t_2, s) - \mu(1, t_2, s)\}/(T_1 - 1)$, where the notation is as described in Section 3.2. We present in Figure 7 the spatial maps of the estimated weekly-varying DRC in mean SST across the Red Sea for Weeks 7 and 34 (the coolest and warmest weeks based on the averaged observed SST, respectively) along with the overall DRC obtained by averaging across the 52 weeks. The spatial patterns of DRC for the two weeks are significantly different and broadly consistent with the exploratory analysis (Figure 2). For Week 7 (winter), the highest DRC values are near the latitude 22°N while the lowest values are observed near the southern end of the Red Sea. For Week 34 (summer), the highest DRC values are near the northern tip of the Gulf of Suez and over a large coastal region of Egypt as well as the coastal region of northwest Saudi Arabia while the lowest values are observed near the coastal region of the southwest Saudi Arabia. The spatial map of the overall DRC is smoother than the weekly profiles, with higher values being observed near the coastal region of Egypt. In addition, we also calculate the corresponding t-statistic for each spatial location; see the Supplementary Material (Hazra and Huser (2021a)). A value of $|t| > 2$ indicates a (sitewise) significant change at the 95% confidence level in mean SST over the years 1985–2015. While negative estimates of DRC are obtained near the coast of Eritrea for Week 7 (for a total 896 grid cells), all such DRC values are not significant except for 35 grid cells. Considering the overall DRC, the t-statistics vary between 5.92 and 39.18, indicating that the overall DRC is positive and highly significant at all the grid cells.

A T_0-year return level of a stationary weekly time series is (approximately) the $(1 - 1/[52T_0])$th quantile of its marginal distribution, given that there are about 52 weeks per year. The LTP-DPM model, however, has a nonstationary mean and involves both trend and
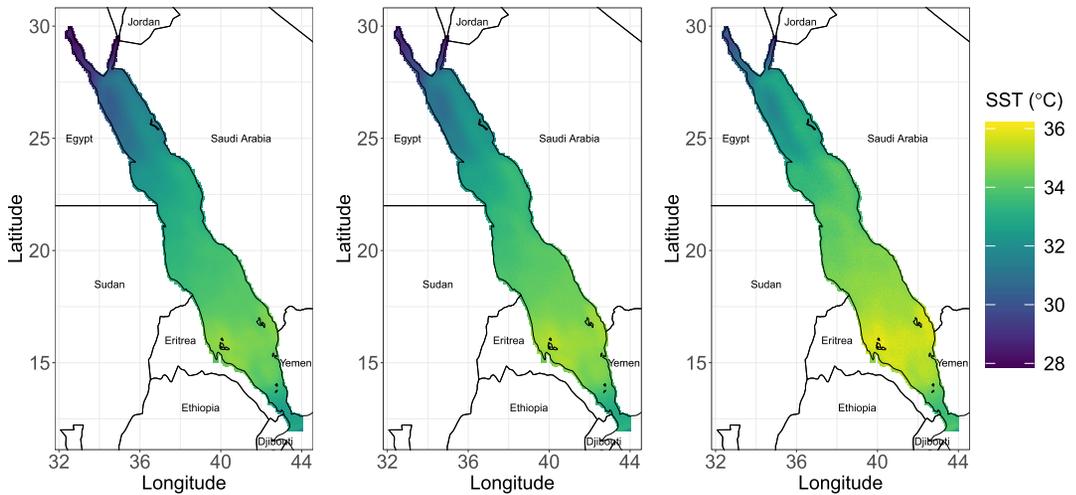
FIG. 8. *Estimated* 10 *(left)*, 20 *(middle) and* 50 *(right)-year return levels for Week* 40, *taking* 2020 *as the reference year and following* Cheng et al. (2014) *to account for nonstationarity. All subfigures share the same scale.*

seasonality. Thus, quantiles—hence return levels—change over time. As the model involves a trend that is linearly related to the RCP 8.5-based SST estimates for each week, we follow Cheng et al. (2014), consider the average RCP 8.5-based SST estimate within the return period and calculate the mean SST and return level accordingly. Here, we concentrate specifically on some of the warmest weeks. While the SST spatial average is maximum for Week 34, Week 40 is the hottest week for 2860 grid cells spread between Eritrea and southwest Saudi Arabia which covers some major coral reefs. The results for Weeks 33 and 34 are discussed in the Supplementary Material (Hazra and Huser (2021a)). For the Week 40, Figure 8 provides the spatial maps of 10-year, 20-year and 50-year estimated return levels considering 2020 as the reference year. For all three return periods, estimates are lower near the Gulf of Suez and the Gulf of Aqaba and higher across a large region between the coast of Eritrea and the southwest of Saudi Arabia. This suggests that very extreme sea temperatures might occur over the next century, potentially causing severe ecological damages to endemic species (e.g., certain types of corals) and may have also economic consequences for neighboring countries.

5.3. *Estimated joint exceedance probabilities.* We now investigate model-based estimates of joint exceedance probabilities in more detail and consider three specific regions: the Dahlak Islands of Eritrea (40.13°E, 15.77°N), the Farasan Islands of Saudi Arabia (41.88°E, 16.82°N) and the region of Thuwal in Saudi Arabia (38.88°E, 22.37°N), where large coral reefs are present. Considering the RCP 8.5-based Red Sea SST projections until 2100, the highest projection (33.85°C) corresponds to the year 2099; see Figure 3. For this particular year, we estimate two types of exceedance probabilities, namely, $\Pr(\bigcup_{s_n \in \mathcal{D}_0} \{Y_{t_0}(s_n) > u\})$ and $\Pr(\bigcap_{s_n \in \mathcal{D}_0} \{Y_{t_0}(s_n) > u\})$, for a range of high temperature values u and different neighborhoods $\mathcal{D}_0$, that is, sets of grid cells within a certain distance from the three specific locations. For $\mathcal{D}_0$, we consider distances 0 km (for which both types of probabilities coincide with the marginal exceedance probability), 6 km (including the first order neighbors), 10 km, 20 km, 30 km and 50 km. A large value of $\Pr(\bigcup_{s_n \in \mathcal{D}_0} \{Y_{t_0}(s_n) > u\})$ indicates a high probability that *at least one* grid cell within $\mathcal{D}_0$ exceeds the threshold u. A large value of $\Pr(\bigcap_{s_n \in \mathcal{D}_0} \{Y_{t_0}(s_n) > u\})$ indicates a high probability that *all* grid cells within $\mathcal{D}_0$ exceed the threshold u. These probabilities are different across weeks. The results for Weeks 33 and 34 are provided in the Supplementary Material (Hazra and Huser (2021a)). For Week 40 the values are presented in the top panel of Figure 9. For Dahlak Islands of Eritrea and
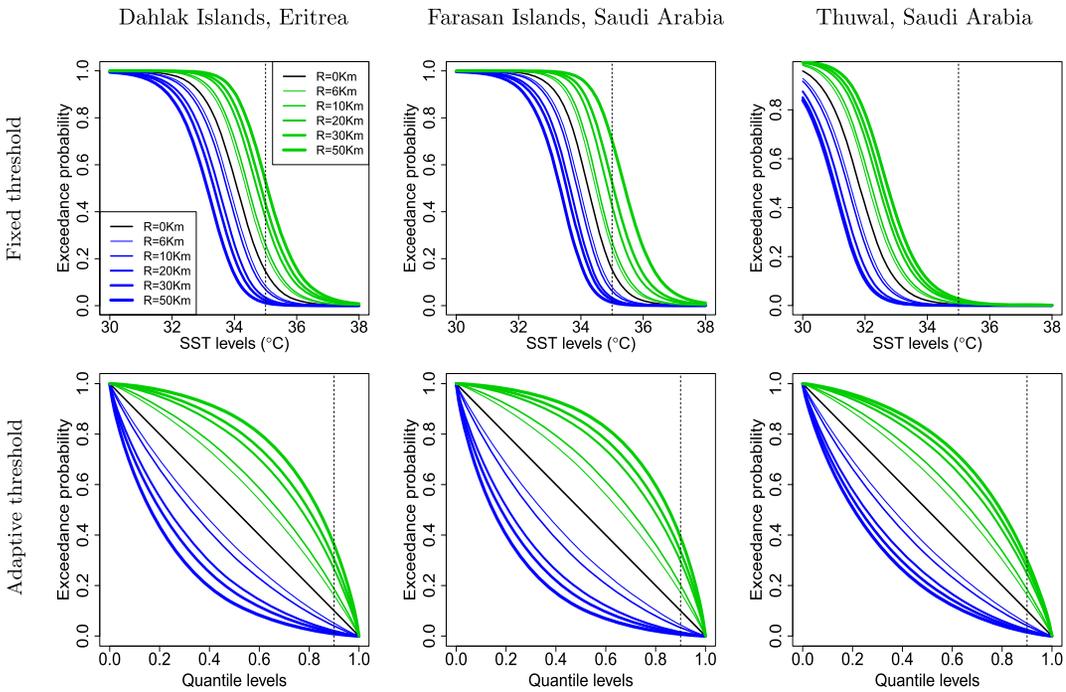
FIG. 9.   *Exceedance probabilities of SST across all the grid cells (blue) or at least one grid cell (green) within 0 km (the marginal case), 6 km (including the first order neighbors), 10 km, 20 km, 30 km and 50 km distances from (40.13°E, 15.77°N) near Dahlak Islands of Eritrea, (41.88°E, 16.82°N) near Farasan Islands of Saudi Arabia and (38.88°E, 22.37°N) near Thuwal of Saudi Arabia at different temperature levels for Week 40 of the reference year 2099 (corresponds to the highest RCP 8.5-based SST projection until 2100), considering fixed (top) or spatially-varying (bottom) temperature thresholds. Vertical dashed lines show reference thresholds corresponding to $u = 35°C$ (top) and the $p = 0.9$-quantile (bottom).*

Farasan Islands of Saudi Arabia, the exceedance probabilities are high even for high thresholds. As Week 40 is a postsummer week in the northern Red Sea, the exceedance probabilities are comparatively lower for Thuwal, Saudi Arabia. At a threshold of 35°C, the estimated probability values (combining both types of exceedance probabilities) range within $(0.0122, 0.5277)$, $(0.0137, 0.7031)$ and $(0.0004, 0.0264)$ for the three regions, respectively.

We also estimate the same exceedance probabilities, but, instead of calculating them over a range of spatially-constant temperature levels, we use an adaptive site-specific threshold. Fixed or adaptive thresholds might be useful in different contexts: while a fixed threshold is directly interpretable on the temperature scale, a spatially-varying threshold is more in line with coral bleaching theory (see, e.g., Genevier et al. (2019)). For any $p \in (0, 1)$, let $Q_{t_0}^{(n)}(p)$ be the pth quantile function of the distribution of $Y_{t_0}(s_n)$ at the nth grid cell, s_n, for the reference time t_0. We estimate the probabilities $\Pr(\bigcup_{s_n \in \mathcal{D}_0}\{Y_{t_0}(s_n) > Q_{t_0}^{(n)}(p)\})$ and $\Pr(\bigcap_{s_n \in \mathcal{D}_0}\{Y_{t_0}(s_n) > Q_{t_0}^{(n)}(p)\})$ for the same choices of $\mathcal{D}_0$ as above. Let $Q^{(n)}(p)$ be the pth quantile function of the (temporally invariant) distribution of $\varepsilon_{t_0}(s_n)$. Following (2), we have $Q_{t_0}^{(n)}(p) = \mu_{t_0}(s_n) + Q^{(n)}(p)$, so $\Pr(\bigcup_{s_n \in \mathcal{D}_0}\{Y_{t_0}(s_n) > Q_{t_0}^{(n)}(p)\}) = \Pr(\bigcup_{s_n \in \mathcal{D}_0}\{\varepsilon_{t_0}(s_n) > Q^{(n)}(p)\})$ which does not change over time. Similarly, we have that $\Pr(\bigcap_{s_n \in \mathcal{D}_0}\{Y_{t_0}(s_n) > Q_{t_0}^{(n)}(p)\}) = \Pr(\bigcap_{s_n \in \mathcal{D}_0}\{\varepsilon_{t_0}(s_n) > Q^{(n)}(p)\})$. From low to high values of p, we present the two types of exceedance probabilities in the bottom panel of Figure 9. At a marginal quantile level of 0.9, the estimated probabilities range within $(0.0088, 0.3605)$, $(0.0070, 0.3933)$ and $(0.0155, 0.3014)$ for the three regions, respectively, indicating high risk of simultaneously large temperatures in these regions.
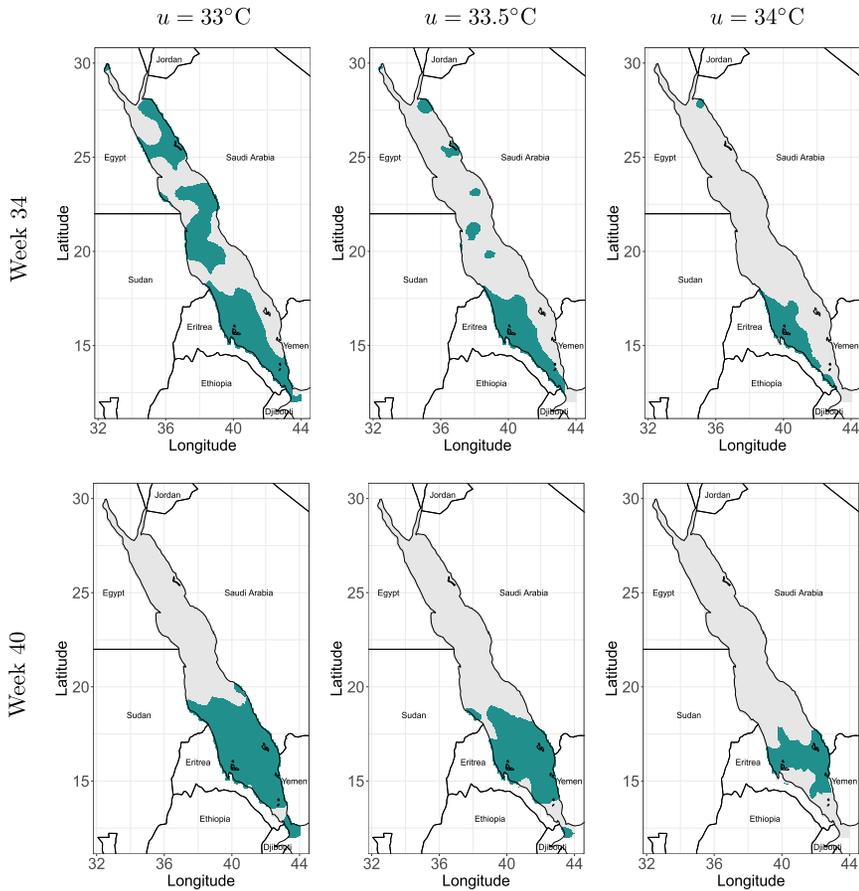
FIG. 10.   *The* 95% *confidence regions* $\mathcal{D}_{u+}^0$ *of the Red Sea SST profile, projected to year* 2099 (*corresponds to the highest RCP* 8.5-*based SST projection until* 2100), *for exceedance levels u* $= 33°C$, $u = 33.5°C$ *and* $u = 34°C$ *for Weeks* 34 *and* 40.

5.4. *Estimated hotspots.*   After exploring the joint threshold exceedance probabilities for fixed regions, we now focus on identifying the regions at risk themselves, that is, hotspots, for the year 2099. We estimate the 95% confidence regions $\mathcal{D}_{u+}^0$ for joint exceedance levels $u = 33°C$, $u = 33.5°C$ and $u = 34°C$. The maps of $\mathcal{D}_{u+}^0$ for Weeks 34 and 40 are presented in Figure 10. The results for other summer weeks are provided in the Supplementary Material (Hazra and Huser (2021a)).

For Week 40, $\mathcal{D}_{u+}^0$ stretches over a major portion of the southern Red Sea for each temperature threshold considered. When $u = 33°C$, the estimated $\mathcal{D}_{u+}^0$ covers almost entirely the area within the latitudes 13.5°N and 19°N and also a small region near coastal Djibouti. When $u = 33.5°C$, the estimated $\mathcal{D}_{u+}^0$ covers a large region between the latitudes 14°N and 18°N. Finally, when $u = 34°C$, the estimated $\mathcal{D}_{u+}^0$ covers a narrower region stretched between the latitudes 15.5°N and 17.5°N, including, in particular, the Dahlak Islands of Eritrea and Farasan Islands of Saudi Arabia which host many endemic species, including corals. The number of grid cells within $\mathcal{D}_{u+}^0$ are 7136, 5487, and 2819 for the three different thresholds, respectively. For Week 34, the estimated hotspots include 8905, 4186, and 2321 grid cells for the three different thresholds, respectively, and mostly affect the southwestern part of the Red Sea for high SST thresholds.

While the estimated hotspots based on RCP 8.5 presented here are quite alarming, the results for RCP 4.5 (see the Supplementary Material (Hazra and Huser (2021a))) are slightly

less concerning. Nevertheless, from our analysis, it is likely that large parts of the southern Red Sea, including major coral reefs, will experience joint extreme SST events within the current century which may have important consequences in terms of coral bleaching and mortality. As corals are believed to be somewhat resilient to their local environment, it would also be interesting from an applied perspective to estimate hotspots for spatially-varying thresholds. Moreover, for a better risk assessment of coral bleaching, a future research direction could be to extend our model for capturing spatiotemporal dependence, in order to measure the persistence of high SST values over time and, eventually, to estimate spatiotemporal hotspots.

6. Discussion and perspectives.   In this paper, we have proposed a low-rank semiparametric Bayesian spatial model for high-dimensional spatiotemporal data with spatial tail-dependence, where the observations are assumed to be independent across time. The proposed model has a flexible mean structure that captures trend, seasonality and accounts for spatial variability in the mean component. Using B-splines for modeling seasonality and spatial variability helps to identify local spatiotemporal features. Relaxing the parametric Gaussian process (GP) assumption that is generally used in the analysis of high-dimensional spatial data, we propose here a flexible semiparametric model that better captures the marginal and dependence structures. While a finite Dirichlet process mixture of GPs relaxes the parametric GP assumption, it is not apt for modeling spatial data with strong extremal dependence, and we circumvent this issue through a mixture of low-rank Student's t processes. Furthermore, the covariance structure of the proposed model is constructed from empirical orthogonal functions (EOFs), and it provides a reasonable sparse approximation to the highly nonstationary sample spatial covariance while allowing for inference in high dimensions. We have also developed a hotspot estimation method tailored for our proposed model which allows us to identify regions at risk of joint extreme events.

Our statistical analysis revealed several important features of the Red Sea SST data. The decadal rate of change in mean SST varies spatially as well as seasonally; during winter weeks the highest values are near the latitude 22°N, while the lowest values are observed near the southern end of the Red Sea. During summer weeks, while the mean SST is generally lower in the northern Red Sea (particularly, near the Gulf of Aqaba and Suez) compared to other regions, its increasing rate is higher in those regions compared to the southern part. Considering the overall decadal rate of change (averaged across weeks), the estimates are positive and significant at every grid cell; higher values are observed all along the coast of the northern Red Sea. Moreover, when including RCP 8.5-based SST as a covariate, the estimated 50-year return levels turn out to be very high ($\approx$ 35–36°C) during the warmest weeks in the southern Red Sea across a large region stretching from Eritrea to Southwestern Saudi Arabia. Furthermore, joint exceedance probabilities for regions near the Dahlak Islands and the Farasan Islands are estimated to be alarmingly high. Both areas host major coral reefs, and our study indicates a high chance that the sea surface temperature will jointly exceed high thresholds which has a direct impact on the risk of coral bleaching (even though several other factors than SST are also important). Our estimation and prediction of hotspots are useful to identify regions "at risk" from an environmental and ecological perspective. Considering the very high (fixed) threshold of 34°C, the estimated hotspots based on the RCP 8.5 scenario indeed cover the Dahlak Islands and the Farasan Islands. Climate change mitigation measures are necessary to safeguard coral reefs and ecosystems which are not only important for ecological reasons but also because they support tourism in these regions.

While the proposed model is able to capture several features of the data very flexibly, it has a few downsides. A first drawback is that the model assumes that the temporal replicates are independent in time which may not be realistic with daily or subdaily data. The

spatiotemporal model of Hazra et al. (2018), which assumes a copula structure in time, is a possible solution to address this issue, but it is limited to small spatiotemporal datasets. No closed-form expression of the full posterior exists for the random effects involved in that model, and, hence, the analysis becomes computationally challenging for large temporal dimensions. The MCMC algorithm based on deterministic transformations proposed by Dutta and Bhattacharya (2014) might be a computationally feasible solution. Additionally, the model of Hazra et al. (2018) assumes that the temporal dependence structure is spatially invariant which may not be a realistic assumption for large geographic regions like the Red Sea. Another drawback is that our proposed model does not capture independence, as the distance between sites increases and its spatial extremal dependence remains nonzero throughout the entire spatial domain. Although this does not seem to be an issue for our Red Sea data, which are strongly dependent, it could be problematic in other data applications with shorter-range spatial dependence. The approach of Morris et al. (2017), which breaks down long-range dependence by introducing random partitions of the spatial domain, could be a possible solution to address this issue.

SUPPLEMENTARY MATERIAL

Supplementary material for "Estimating high-resolution Red Sea surface temperature hotspots, using a low-rank semiparametric spatial model" (DOI: 10.1214/20-AOAS1418SUPPA; .pdf). We first provide further results for the exploratory analysis, in particular regarding the spatial variation of the SST seasonality profile across the Red Sea, the estimated EOFs, and the comparison between marginal data quantiles and quantiles from fitted univariate normal and Student's t distributions. Then we discuss the MCMC steps for model fitting and future prediction, and provide diagnostics to demonstrate the convergence and mixing of the MCMC chains. We also prove that our hotspot estimation approach attains the desired overall familywise error rate. We finally provide further results and diagnostics for the fitted LTP-DPM model, in particular t-statistics for the estimated SST decadal rate of change, return level maps, joint exceedance probabilities, and estimated hotspots for a few summer weeks for the RCP 8.5-based analysis. We also discuss some results of the RCP 4.5-based analysis.

R **code for "Estimating high-resolution Red Sea surface temperature hotspots, using a low-rank semiparametric spatial model"** (DOI: 10.1214/20-AOAS1418SUPPB; .zip). We provide the R code for fitting the model, as well as for drawing inferences from the MCMC chains. We also provide a detailed simulation example (similar to the Red Sea surface temperature dataset) to illustrate the hotspot estimation method step by step.

REFERENCES

ALEXANDERIAN, A. (2015). A brief note on the Karhunen–Loève expansion. Preprint. Available at arXiv:1509.07526.

ANDERES, E., HUSER, R., NYCHKA, D. and CORAM, M. (2013). Nonstationary positive definite tapering on the plane. *J. Comput. Graph. Statist.* **22** 848–865. MR3173746 https://doi.org/10.1080/10618600.2012.729982

BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 825–848. MR2523906 https://doi.org/10.1111/j.1467-9868.2008.00663.x

BEAKY, M. M., SCHERRER, R. J. and VILLUMSEN, J. V. (1992). Topology of large-scale structure in seeded hot dark matter models. *Astrophys*. *J*. **387** 443–448.

BERUMEN, M. L., HOEY, A. S., BASS, W. H., BOUWMEESTER, J., CATANIA, D., COCHRAN, J. E., KHALIL, M. T., MIYAKE, S., MUGHAL, M. et al. (2013). The status of coral reef ecology research in the Red Sea. *Coral Reefs* **32** 737–748.

BOLIN, D. and LINDGREN, F. (2015). Excursion and contour uncertainty regions for latent Gaussian models. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **77** 85–106. MR3299400 https://doi.org/10.1111/rssb.12055

BOPP, G., SHABY, B. A. and HUSER, R. (2021). A hierarchical max-infinitely divisible spatial model for extreme precipitation. *J. Amer. Statist*. *Assoc*. **116** 93–106. https://doi.org/10.1080/01621459.2020.1750414

BOPP, G. P., SHABY, B. A., FOREST, C. E. and MEJÍA, A. (2020). Projecting flood-inducing precipitation with a Bayesian analogue model. *J. Agric. Biol. Environ. Stat*. **25** 229–249. MR4089687 https://doi.org/10.1007/s13253-020-00391-6

CASTRO-CAMILO, D., HUSER, R. and RUE, H. (2019). A spliced gamma-generalized Pareto model for short-term extreme wind speed probabilistic forecasting. *J. Agric. Biol. Environ. Stat*. **24** 517–534. MR3996457 https://doi.org/10.1007/s13253-019-00369-z

CASTRUCCIO, S., HUSER, R. and GENTON, M. G. (2016). High-order composite likelihood inference for max-stable distributions and processes. *J. Comput. Graph. Statist*. **25** 1212–1229. MR3572037 https://doi.org/10.1080/10618600.2015.1086656

CHENG, L., AGHAKOUCHAK, A., GILLELAND, E. and KATZ, R. W. (2014). Non-stationary extreme value analysis in a changing climate. *Clim. Change* **127** 353–369.

COOLEY, D. and SAIN, S. R. (2010). Spatial hierarchical modeling of precipitation extremes from a regional climate model. *J. Agric. Biol. Environ. Stat*. **15** 381–402. MR2787265 https://doi.org/10.1007/s13253-010-0023-9

CRAIGMILE, P. F., CRESSIE, N., SANTNER, T. J. and RAO, Y. (2005). A loss function approach to identifying environmental exceedances. *Extremes* **8** 143–159. MR2275915 https://doi.org/10.1007/s10687-006-7964-y

CRESSIE, N. A. C., ZHANG, J. and CRAIGMILE, P. F. (2005). Geostatistical prediction of spatial extremes and their extent. In *Geostatistics for Environmental Applications*, *Proceedings of the Fifth Conference on Geostatistics for Environmental Applications* (P. Renard, H. Demougeot-Renard and R. Froidevaux, eds.) 27–37. Springer, Berlin.

DAVISON, A. C. and HUSER, R. (2015). Statistics of extremes. *Annu. Rev. Stat. Appl*. **2** 203–235. https://doi.org/10.1146/annurev-statistics-010814-020133

DAVISON, A. C., HUSER, R. and THIBAUD, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Math. Geosci*. **45** 511–529. MR3079649 https://doi.org/10.1007/s11004-013-9469-y

DAVISON, A., HUSER, R. and THIBAUD, E. (2019). Spatial extremes. In *Handbook of Environmental and Ecological Statistics* M. Fuentes, J. A. Hoeting and R. L. Smith, eds.). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 711–744. CRC Press, Boca Raton, FL. MR3889918

DAVISON, A. C., PADOAN, S. A. and RIBATET, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci*. **27** 161–186. MR2963980 https://doi.org/10.1214/11-STS376

DE FONDEVILLE, R. and DAVISON, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika* **105** 575–592. MR3842886 https://doi.org/10.1093/biomet/asy026

DEMARTA, S. and MCNEIL, A. J. (2005). The t copula and related copulas. *Int. Stat. Rev*. **73** 111–129.

DONLON, C. J., MARTIN, M., STARK, J., ROBERTS-JONES, J., FIEDLER, E. and WIMMER, W. (2012). The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sens. Environ*. **116** 140–158.

DUAN, J. A., GUINDANI, M. and GELFAND, A. E. (2007). Generalized spatial Dirichlet process models. *Biometrika* **94** 809–825. MR2416794 https://doi.org/10.1093/biomet/asm071

DUTTA, S. and BHATTACHARYA, S. (2014). Markov chain Monte Carlo based on deterministic transformations. *Stat. Methodol*. **16** 100–116. MR3110892 https://doi.org/10.1016/j.stamet.2013.08.006

EKLUNDH, L. and OLSSON, L. (2003). Vegetation index trends for the African Sahel 1982–1999. *Geophys. Res. Lett*. **30** 1430.

FIX, M. J., COOLEY, D., SAIN, S. R. and TEBALDI, C. (2018). A comparison of U.S. precipitation extremes under RCP8.5 and RCP4.5 with an application of pattern scaling. *Clim. Change* **146** 335–347.

FRENCH, J. P. and HOETING, J. A. (2016). Credible regions for exceedance sets of geostatistical data. *Environmetrics* **27** 4–14. MR3457233 https://doi.org/10.1002/env.2371

FRENCH, J. P. and SAIN, S. R. (2013). Spatio-temporal exceedance locations and confidence regions. *Ann. Appl. Stat*. **7** 1421–1449. MR3127953 https://doi.org/10.1214/13-AOAS631

FUENTES, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *J. Amer. Statist. Assoc*. **102** 321–331. MR2345545 https://doi.org/10.1198/016214506000000852

FURRER, R., GENTON, M. G. and NYCHKA, D. (2006). Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Statist*. **15** 502–523. MR2291261 https://doi.org/10.1198/106186006X132178

FURRER, R., KNUTTI, R., SAIN, S., NYCHKA, D. and MEEHL, G. (2007). Spatial patterns of probabilistic temperature change projections from a multivariate Bayesian analysis. *Geophys. Res. Lett.* **34** L06711.

GELFAND, A. E., KOTTAS, A. and MACEACHERN, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *J. Amer. Statist. Assoc.* **100** 1021–1035. MR2201028 https://doi.org/10.1198/016214504000002078

GELFAND, A. E. and SCHLIEP, E. M. (2016). Spatial statistics and Gaussian processes: A beautiful marriage. *Spat. Stat.* **18** 86–104. MR3573271 https://doi.org/10.1016/j.spasta.2016.03.006

GENEVIER, L. G. C., JAMIL, T., RAITSOS, D. E., KROKOS, G. and HOTEIT, I. (2019). Marine heatwaves reveal coral reef zones susceptible to bleaching in the Red Sea. *Glob. Change Biol.* **25** 2338–2351. https://doi.org/10.1111/gcb.14652

GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. *Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge Univ. Press, Cambridge. MR3587782 https://doi.org/10.1017/9781139029834

GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. MR2345548 https://doi.org/10.1198/016214506000001437

GNEITING, T. and RANJAN, R. (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *J. Bus. Econom. Statist.* **29** 411–422. MR2848512 https://doi.org/10.1198/jbes.2010.08110

HAZRA, A. and HUSER, R. (2021a). Supplement to "Estimating high-resolution Red Sea surface temperature hotspots, using a low-rank semiparametric spatial model." https://doi.org/10.1214/20-AOAS1418SUPPA.

HAZRA, A. and HUSER, R. (2021b). R code for "Estimating high-resolution Red Sea surface temperature hotspots, using a low-rank semiparametric spatial model." https://doi.org/10.1214/20-AOAS1418SUPPB.

HAZRA, A., REICH, B. J. and STAICU, A.-M. (2020). A multivariate spatial skew-*t* process for joint modeling of extreme precipitation indexes. *Environmetrics* **31** e2602, 19. MR4098481 https://doi.org/10.1002/env.2602

HAZRA, A., REICH, B. J., SHABY, B. A. and STAICU, A.-M. (2018). A semiparametric Bayesian model for spatiotemporal extremes. Preprint. Available at arXiv:1812.11699.

HEATON, M. J., DATTA, A., FINLEY, A. O., FURRER, R., GUINNESS, J., GUHANIYOGI, R., GERBER, F., GRAMACY, R. B., HAMMERLING, D. et al. (2019). A case study competition among methods for analysing large spatial data. *J. Agric. Biol. Environ. Stat.* **24** 398–425.

HIGDON, D. (2002). Space and space–time modeling using process convolutions. In *Quantitative Methods for Current Environmental Issues* 37–56. Springer, London. MR2059819

HUSER, R. (2021). Editorial: EVA 2019 data competition on spatio-temporal prediction of Red Sea surface temperature extremes. *Extremes* **24** 91–104. https://doi.org/10.1007/s10687-019-00369-9

HUSER, R., OPITZ, T. and THIBAUD, E. (2017). Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. *Spat. Stat.* **21** 166–186. MR3692183 https://doi.org/10.1016/j.spasta.2017.06.004

HUSER, R. and WADSWORTH, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *J. Amer. Statist. Assoc.* **114** 434–444. MR3941266 https://doi.org/10.1080/01621459.2017.1411813

HUSER, R. and WADSWORTH, J. L. (2020). Advances in statistical modeling of spatial extremes. *Wiley Interdiscip. Rev.: Comput. Stat.* e1537. https://doi.org/10.1002/wics.1537

HUSER, R., DOMBRY, C., RIBATET, M. and GENTON, M. G. (2019). Full likelihood inference for max-stable data. *Stat* **8** e218, 14. MR3938281 https://doi.org/10.1002/sta4.218

LEWANDOWSKA, A. M., BOYCE, D. G., HOFMANN, M., MATTHIESSEN, B., SOMMER, U. and WORM, B. (2014). Effects of sea surface warming on marine plankton. *Ecol. Lett.* **17** 614–623.

MEJIA, A. F., YUE, Y., BOLIN, D., LINDGREN, F. and LINDQUIST, M. A. (2020). A Bayesian general linear modeling approach to Cortical surface FMRI data analysis. *J. Amer. Statist. Assoc.* **115** 501–520. MR4107654 https://doi.org/10.1080/01621459.2019.1611582

MORRIS, S. A., REICH, B. J., THIBAUD, E. and COOLEY, D. (2017). A space–time skew-*t* model for threshold exceedances. *Biometrics* **73** 749–758. MR3713109 https://doi.org/10.1111/biom.12644

OPITZ, T. (2013). Extremal *t* processes: Elliptical domain of attraction and a spectral representation. *J. Multivariate Anal.* **122** 409–413. MR3189331 https://doi.org/10.1016/j.jmva.2013.08.008

OPITZ, T., HUSER, R., BAKKA, H. and RUE, H. (2018). INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes* **21** 441–462. MR3855716 https://doi.org/10.1007/s10687-018-0324-x

PADOAN, S. A., RIBATET, M. and SISSON, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.* **105** 263–277. MR2757202 https://doi.org/10.1198/jasa.2009.tm08577

PAPADAKIS, M., TSAGRIS, M., DIMITRIADIS, M., TSAMARDINOS, I., FASIOLO, M., BOR-BOUDAKIS, G. and BURKARDT, J. (2017). Rfast: Fast R functions. R Package Version 1.

QIU, Y. and MEI, J. (2016). Package 'rARPACK'.

REASER, J. K., POMERANCE, R. and THOMAS, P. O. (2000). Coral bleaching and global climate change: Scientific findings and policy recommendations. *Conserv. Biol.* **14** 1500–1511.

REICH, B. J. and SHABY, B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. *Ann. Appl. Stat.* **6** 1430–1451. MR3058670 https://doi.org/10.1214/12-AOAS591

RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields*: *Theory and Applications. Monographs on Statistics and Applied Probability* **104**. CRC Press/CRC, Boca Raton, FL. MR2130347 https://doi.org/10.1201/9780203492024

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

SANG, H. and GELFAND, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environ. Ecol. Stat.* **16** 407–426. MR2749848 https://doi.org/10.1007/s10651-007-0078-0

SANG, H. and GELFAND, A. E. (2010). Continuous spatial process models for spatial extreme values. *J. Agric. Biol. Environ. Stat.* **15** 49–65. MR2755384 https://doi.org/10.1007/s13253-009-0010-1

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4** 639–650. MR1309433

SHALTOUT, M. (2019). Recent sea surface temperature trends and future scenarios for the Red Sea. *Oceanologia* **61** 484–504.

SMITH, R. L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript.

STEIN, M. L. (1999). *Interpolation of Spatial Data*: *Some Theory for Kriging. Springer Series in Statistics*. Springer, New York. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6

STEIN, M. L., CHI, Z. and WELTY, L. J. (2004). Approximating likelihoods for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 275–296. MR2062376 https://doi.org/10.1046/j.1369-7412.2003.05512.x

THIBAUD, E. and OPITZ, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika* **102** 855–870. MR3431558 https://doi.org/10.1093/biomet/asv045

VECCHIA, A. V. (1988). Estimation and model identification for continuous spatial processes. *J. Roy. Statist. Soc. Ser. B* **50** 297–312. MR0964183

WIKLE, C. K. (2010). Low-rank representations for spatial processes. In *Handbook of Spatial Statistics* (A. E. Gelfand, P. Diggle, P. Guttorp and M. Fuentes, eds.). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 107–118. CRC Press, Boca Raton, FL. MR2730946 https://doi.org/10.1201/9781420072884-c8

WIKLE, C. K. and CRESSIE, N. (1999). A dimension-reduced approach to space–time Kalman filtering. *Biometrika* **86** 815–829. MR1741979 https://doi.org/10.1093/biomet/86.4.815

WU, Y. and GHOSAL, S. (2010). The L_1-consistency of Dirichlet mixtures in multivariate Bayesian density estimation. *J. Multivariate Anal.* **101** 2411–2419. MR2719871 https://doi.org/10.1016/j.jmva.2010.06.012