

High-Resolution Modeling and Estimation of Extreme Red Sea Surface Temperature Hotspots

Arnab Hazra and Raphaël Huser

King Abdullah University of Science and Technology (KAUST),
Thuwal, Saudi Arabia

raphael.huser@kaust.edu.sa

March 12, 2021





Credit: Vinicius M. Lube

Extreme Statistics (extSTAT) Research Group at KAUST



I am currently recruiting **2 postdocs** in the following areas:

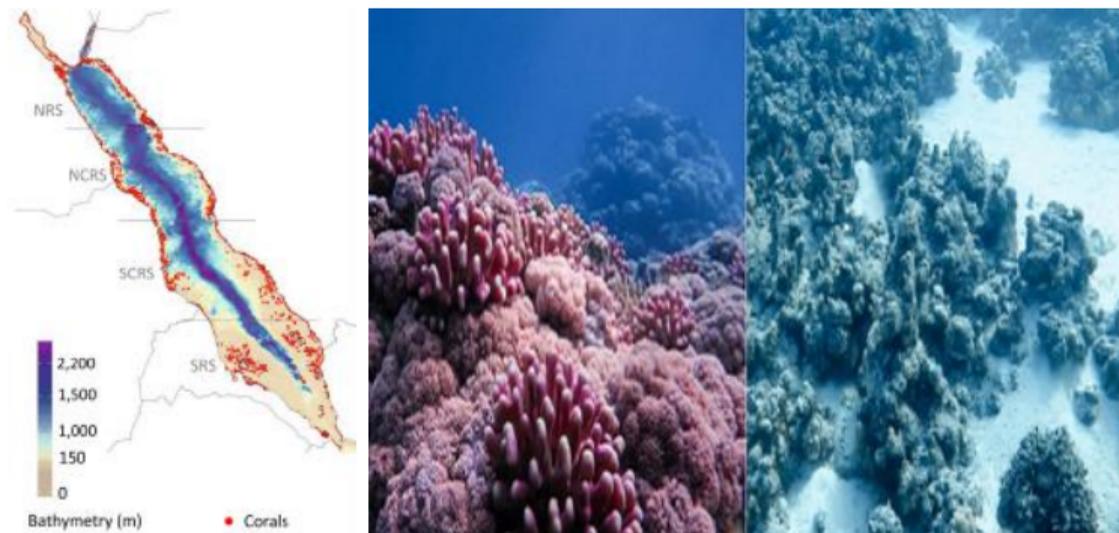
- 🍌 **Extreme-Value Theory and Statistics**, with experience in at least one of the following areas: (1) spatio-temporal statistics; (2) SPDE models; (3) computational statistics (low-rank methods, INLA, ML, etc.); (4) graphical models. Strong interest in environmental applications is desired.
- 🍌 **Spatial Statistics for Point Patterns with Applications in Geomorphology (Landslide Science)**, with experience in at least one of the following areas: (1) Spatial statistics; (2) Bayesian computational methods (INLA, MCMC, etc.); (3) Statistics of extremes; (4) Applications in geosciences. Strong interest in the development of spatial predictive models for earthquake-induced landslides is required.

Agenda

- Motivation
 - Sea surface temperature (SST) data and exploratory analysis
 - Classical models for spatial extremes, and their limitations
 - Proposed model and its properties
 - Bayesian inference and estimation of extreme hotspots
 - Data application
 - Final remarks
- Hazra, A. and Huser, R. (2020+), *Estimating high-resolution Red Sea surface temperature hotspots, using a low-rank semiparametric spatial model*, *Annals of Applied Statistics*, to appear.

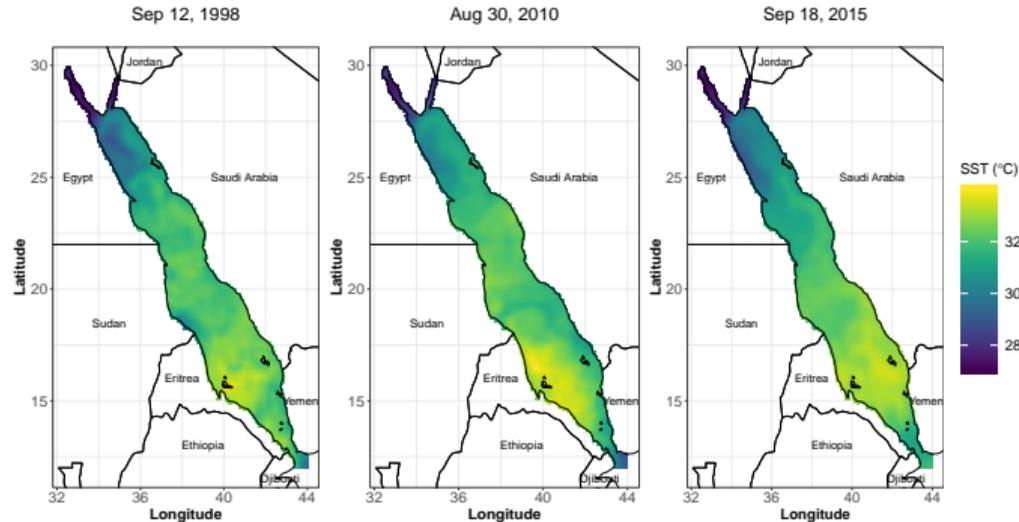
Motivation

Motivation



- 50% corals have died in last 30 years and the increasing ocean temperature is a primary cause.
- The coral reefs of the Red Sea are less studied than those of the Indian Ocean and the Pacific Ocean (Berumen et al., 2013).
- Our aim is to model, identify and predict extreme SST exceedance regions (i.e., hotspots) within the Red Sea in future years.

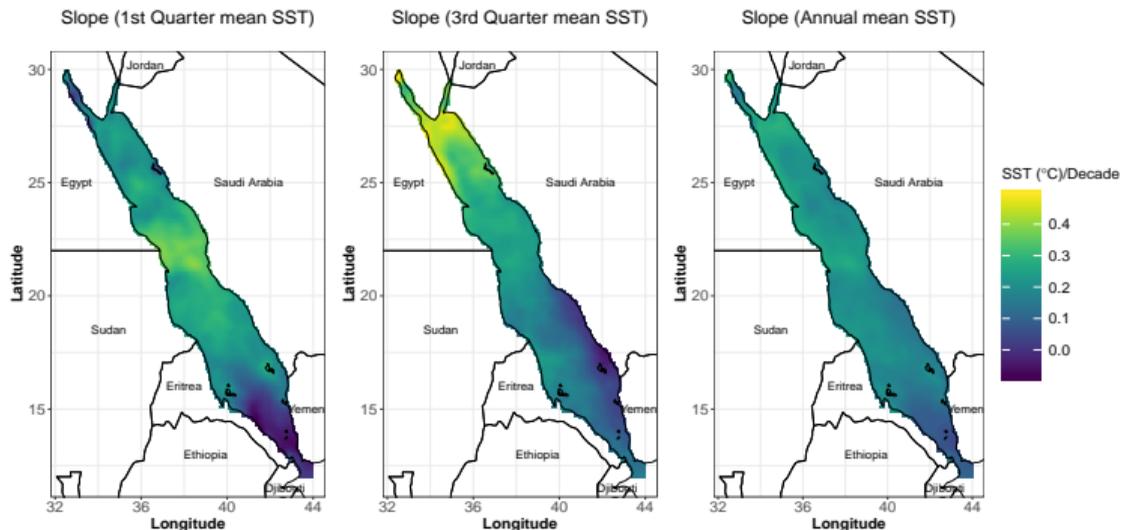
Sea Surface Temperature (SST) Data and Exploratory Analysis



- The project OSTIA (Donlon et al., 2012) generates daily SST estimates (free of diurnal variability) at a resolution $0.05^\circ \times 0.05^\circ$ (about 6 Km).
- Considering the period 1985–2015, total #grid cells = 16703 and total #weeks = 1612.
- We analyze temporally-thinned data keeping one day per week and assuming independence in time.

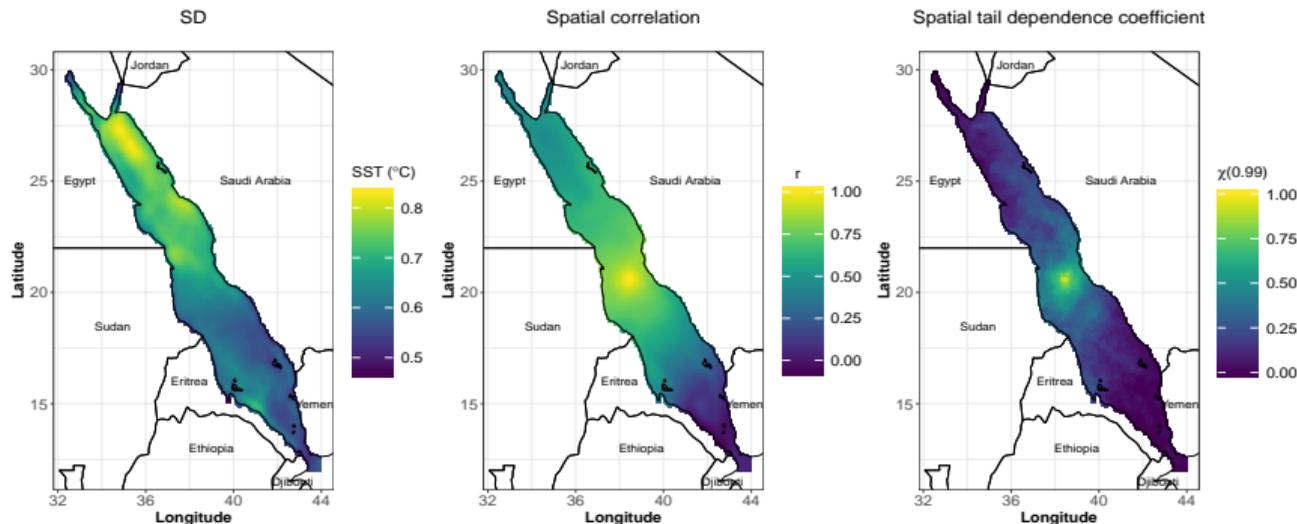
Exploratory analysis (mean structure)

- We fit a simple linear regression model of the form $Y_t = \beta_1 + \beta_2 t, t = 1, \dots, 31$, to the quarterly and annual means at each grid cell. Hence, the estimated decadal rate of change is $10\hat{\beta}_2$ (in $^{\circ}\text{C}$).



- The slope (and intercept) profiles vary spatially as well as seasonally.
- Overall, the mean SST varies across space, weeks and years. The effects are not additive.

Exploratory analysis (dependence structure)



- The spatial covariance structure is nonstationary.
- $\chi(u) = \Pr \{ Y_1 > F_1^{-1}(u) \mid Y_2 > F_2^{-1}(u) \}$ and $\chi = \lim_{u \rightarrow 1} \chi(u)$. The spatial extremal dependence is nonzero over a large region.

Objectives

We want a model that reflects the following features and facts:

- The spatiotemporal mean structure is nonstationary and the effects of space, week and year are not additive [$\mu(lon, lat, year, week) \neq \mu_1(lon) + \mu_2(lat) + \mu_3(year) + \mu_4(week)$].

Objectives

We want a model that reflects the following features and facts:

- The spatiotemporal mean structure is nonstationary and the effects of space, week and year are not additive [$\mu(lon, lat, year, week) \neq \mu_1(lon) + \mu_2(lat) + \mu_3(year) + \mu_4(week)$].
- Since our main interest is in hotspot estimation by the end of 21st century, our model should comprise a trend component that involves climate model-based projections (RCP 4.5/8.5).

Objectives

We want a model that reflects the following features and facts:

- 1 The spatiotemporal mean structure is nonstationary and the effects of space, week and year are not additive [$\mu(lon, lat, year, week) \neq \mu_1(lon) + \mu_2(lat) + \mu_3(year) + \mu_4(week)$].
- 2 Since our main interest is in hotspot estimation by the end of 21st century, our model should comprise a trend component that involves climate model-based projections (RCP 4.5/8.5).
- 3 The spatial covariance is nonstationary and spatial extremal dependence is nonzero.

Objectives

We want a model that reflects the following features and facts:

- 1 The spatiotemporal mean structure is nonstationary and the effects of space, week and year are not additive [$\mu(lon, lat, year, week) \neq \mu_1(lon) + \mu_2(lat) + \mu_3(year) + \mu_4(week)$].
- 2 Since our main interest is in hotspot estimation by the end of 21st century, our model should comprise a trend component that involves climate model-based projections (RCP 4.5/8.5).
- 3 The spatial covariance is nonstationary and spatial extremal dependence is nonzero.
- 4 Our model should be semiparametric (allows for high flexibility over large spatio-temporal domains, and can capture various forms of joint PDFs).

Objectives

We want a model that reflects the following features and facts:

- ❶ The spatiotemporal mean structure is nonstationary and the effects of space, week and year are not additive [$\mu(lon, lat, year, week) \neq \mu_1(lon) + \mu_2(lat) + \mu_3(year) + \mu_4(week)$].
- ❷ Since our main interest is in hotspot estimation by the end of 21st century, our model should comprise a trend component that involves climate model-based projections (RCP 4.5/8.5).
- ❸ The spatial covariance is nonstationary and spatial extremal dependence is nonzero.
- ❹ Our model should be semiparametric (allows for high flexibility over large spatio-temporal domains, and can capture various forms of joint PDFs).
- ❺ The spatial dimension is large and hence requires specifying a dependence structure that leads to feasible computation and inference.

Objectives

We want a model that reflects the following features and facts:

- 1 The spatiotemporal mean structure is nonstationary and the effects of space, week and year are not additive [$\mu(lon, lat, year, week) \neq \mu_1(lon) + \mu_2(lat) + \mu_3(year) + \mu_4(week)$].
- 2 Since our main interest is in hotspot estimation by the end of 21st century, our model should comprise a trend component that involves climate model-based projections (RCP 4.5/8.5).
- 3 The spatial covariance is nonstationary and spatial extremal dependence is nonzero.
- 4 Our model should be semiparametric (allows for high flexibility over large spatio-temporal domains, and can capture various forms of joint PDFs).
- 5 The spatial dimension is large and hence requires specifying a dependence structure that leads to feasible computation and inference.
- 6 Closed-form posterior distributions should be available for as many parameters as possible.

Classical Models for Spatial Extremes, and Their Limitations

- 🟡 **High-dimensional geostatistics:** Vecchia (1988), Wikle and Cressie (1999), Stein (1999), Higdon (2002), Rue and Held (2005), Furrer et al. (2006), Banerjee et al. (2008), Heaton et al. (2019).

Pros and **Cons:** Apt for modeling high-dimensional spatial data but usually assume the process to be Gaussian, which implies asymptotic extremal independence.

- 🟡 **Semiparametric geostatistics:** Gelfand et al. (2005), Duan et al. (2007), Reich et al. (2013).

Pros and **Cons:** More flexible than Gaussian process, Bayesian computation is easy in low-dimension but leads to asymptotic extremal independence.

- **Max-stable processes:** Padoan et al. (2010), Davison and Huser (2015), Castruccio et al. (2016), Davison et al. (2019), Huser et al. (2019)
- Max-stable processes have been widely used for modeling spatial extremes defined as *spatially-indexed block maxima*, because of their **appealing asymptotic characterization**:

$$\frac{\max\{Y_1(\mathbf{s}), \dots, Y_n(\mathbf{s})\} - b_n(\mathbf{s})}{a_n(\mathbf{s})} \xrightarrow{D} Z(\mathbf{s}) \sim \text{Max-stable process},$$

for suitably chosen sequences of functions $a_n(\mathbf{s}) > 0$ and $b_n(\mathbf{s})$.

- **Max-stable theory is well understood.**
- Max-stable processes admit a **spectral representation**:

$$Z(\mathbf{s}) = \sup_{i=1,2,\dots} \xi_i W_i(\mathbf{s}),$$

where $\{\xi_i\} \sim \text{PPP}(\xi^{-2}d\xi) \perp\!\!\!\perp \{W_i(\mathbf{s})\}$ i.i.d. copies of a process $W(\mathbf{s}) \geq 0$ with $E\{W(\mathbf{s})\} = 1$.

- Max-stable processes **capture asymptotic extremal dependence**, i.e.,
 $\chi = \lim_{u \rightarrow 1} \Pr\{Z(\mathbf{s}_1) > F_1^{-1}(u) \mid Z(\mathbf{s}_2) > F_2^{-1}(u)\} > 0$.

However:

- Often **too rigid in finite samples** due to the strong restriction imposed by the max-stability property.
- Spatial block maxima may not correspond to real observations** (as maxima at different sites may occur on different days)...
- Max-stable processes are **computationally extremely intensive to fit** in moderate (or even small) dimensions, because the joint density (i.e., likelihood for a single replicate) has the form

$$g(z_1, \dots, z_D) = \exp\{-V(z_1, \dots, z_D)\} \sum_{\pi \in \mathcal{P}_D} \prod_{i=1}^{|\pi|} \{-V_{\tau_i}(z_1, \dots, z_D)\},$$

where D is the number of sites, \mathcal{P}_D is the set of all partitions of $\{1, \dots, D\}$ and V is the exponent function.

Spatial extremes literature— r -Pareto processes

- r -**Pareto processes**: Ferreira and de Haan (2014), Thibaud and Opitz (2015), de Fondeville and Davison (2018)
- r -Pareto processes are analogous to max-stable processes, in the sense that they have an **appealing asymptotic characterization** for spatial threshold exceedances. On a standard Pareto scale:

$$\frac{Y(\mathbf{s})}{u} \mid r\{Y(\mathbf{s})\} > u \xrightarrow{D} Z(\mathbf{s}) \sim r\text{-Pareto process}, \quad \text{as } u \rightarrow \infty$$

where $r(\cdot)$ is some homogeneous *risk functional*.

- r -Pareto processes are **meant to model the original spatial extreme events** that effectively occurred.
- r -Pareto processes also **capture asymptotic extremal dependence**, and also admit a **convenient stochastic representation**.
- r -Pareto have **simpler (censored) likelihood functions** of the form

$$\prod_{i=1}^n \left[-\frac{V_{i_i} \{\max(\tilde{\mathbf{y}}_i, u)\}}{K_r(u)} \right]$$

However:

- Still **limited to moderate dimensions** ($D \sim 50$) because of multivariate censoring, unless some less efficient scoring rule approaches are used (and how to use this approach in a Bayesian context is unclear)
- Practical issue: **how to choose the threshold u** , especially in a non-stationary context?

- 🕒 **Semiparametric spatial extremes:** Hazra et al. (2018), Bopp et al. (2020).
- 🕒 **Idea:** developing a **semiparametric spatial mixture model** that is **flexible both in the bulk and the tails**, so that it provides a good fit to the whole dataset (from low to high quantiles)
- 🕒 Mixtures can probabilistically (and “automatically”) cluster observations from the tail and the bulk, thus **bypassing threshold selection**.
- 🕒 Depending on the mixture components, the spatial mixture **may or may not capture asymptotic extremal dependence**.
- 🕒 Bayesian **computations are quite simple** if the model is based on (potentially skewed) Gaussian or Student’s t mixture components.
- 🕒 However, the existing semiparametric spatial extremes literature is **still limited to relatively low spatial dimensions**.

Proposed Model and its Properties

The proposed model

- We model the Red Sea SST data as $Y_t(\mathbf{s}_n) = \mu_t(\mathbf{s}_n) + \epsilon_t(\mathbf{s}_n)$, $t = 1, \dots, T = 1612$.

The proposed model

- We model the Red Sea SST data as $Y_t(\mathbf{s}_n) = \mu_t(\mathbf{s}_n) + \epsilon_t(\mathbf{s}_n)$, $t = 1, \dots, T = 1612$.
- We write $\mu_t(\mathbf{s}_n) = \mu(t_1, t_2, \mathbf{s}_n)$ where $t_1 = \lceil t/52 \rceil$ (i.e., year), and $t_2 = t - 52(t_1 - 1)$ (i.e., week).

The proposed model

- We model the Red Sea SST data as $Y_t(\mathbf{s}_n) = \mu_t(\mathbf{s}_n) + \epsilon_t(\mathbf{s}_n)$, $t = 1, \dots, T = 1612$.
- We write $\mu_t(\mathbf{s}_n) = \mu(t_1, t_2, \mathbf{s}_n)$ where $t_1 = \lceil t/52 \rceil$ (i.e., year), and $t_2 = t - 52(t_1 - 1)$ (i.e., week).
- We assume $\mu(t_1, t_2, \mathbf{s}) = \beta_1(t_2, \mathbf{s})x_{t_1,1}^{(0)} + \beta_2(t_2, \mathbf{s})x_{t_1,2}^{(0)}$.

Here, $\beta_{p_0}(t_2, \mathbf{s}) = \sum_{p_1=1}^{P_T=12} \beta_{p_0,p_1}(\mathbf{s})x_{t_2,p_1}^{(1)}$ and $\beta_{p_0,p_1}(\mathbf{s}_n) = \sum_{p_2=1}^{P_S=189} \beta_{p_0,p_1,p_2}x_{n,p_2}^{(2)}$, where

- $x_{t_1,1}^{(0)} = 1/\sqrt{31}$ (intercept),
- $x_{t_1,2}^{(0)}$ = rescaled annual mean SST projection based on RCP 4.5/8.5 scenarios,
- $x_{t_2,p_1}^{(1)}$ = splines across months evaluated at t_2 (within a year),
- $x_{n,p_2}^{(2)}$ = splines across space evaluated at \mathbf{s}_n .

The proposed model

- We model the Red Sea SST data as $Y_t(\mathbf{s}_n) = \mu_t(\mathbf{s}_n) + \epsilon_t(\mathbf{s}_n)$, $t = 1, \dots, T = 1612$.
- We write $\mu_t(\mathbf{s}_n) = \mu(t_1, t_2, \mathbf{s}_n)$ where $t_1 = \lceil t/52 \rceil$ (i.e., year), and $t_2 = t - 52(t_1 - 1)$ (i.e., week).
- We assume $\mu(t_1, t_2, \mathbf{s}) = \beta_1(t_2, \mathbf{s})x_{t_1,1}^{(0)} + \beta_2(t_2, \mathbf{s})x_{t_1,2}^{(0)}$.
Here, $\beta_{p_0}(t_2, \mathbf{s}) = \sum_{p_1=1}^{P_T=12} \beta_{p_0,p_1}(\mathbf{s})x_{t_2,p_1}^{(1)}$ and $\beta_{p_0,p_1}(\mathbf{s}_n) = \sum_{p_2=1}^{P_S=189} \beta_{p_0,p_1,p_2}x_{n,p_2}^{(2)}$, where
 - $x_{t_1,1}^{(0)} = 1/\sqrt{31}$ (intercept),
 - $x_{t_1,2}^{(0)}$ = rescaled annual mean SST projection based on RCP 4.5/8.5 scenarios,
 - $x_{t_2,p_1}^{(1)}$ = splines across months evaluated at t_2 (within a year),
 - $x_{n,p_2}^{(2)}$ = splines across space evaluated at \mathbf{s}_n .
- Vectorized form: $\boldsymbol{\mu} = [\mathbf{X}_{0;1} \otimes \mathbf{X}_1 \otimes \mathbf{X}_2] \boldsymbol{\beta}_1 + [\mathbf{X}_{0;2} \otimes \mathbf{X}_1 \otimes \mathbf{X}_2] \boldsymbol{\beta}_2$.

The proposed model

- We model the Red Sea SST data as $Y_t(\mathbf{s}_n) = \mu_t(\mathbf{s}_n) + \epsilon_t(\mathbf{s}_n)$, $t = 1, \dots, T = 1612$.
- We write $\mu_t(\mathbf{s}_n) = \mu(t_1, t_2, \mathbf{s}_n)$ where $t_1 = \lceil t/52 \rceil$ (i.e., year), and $t_2 = t - 52(t_1 - 1)$ (i.e., week).
- We assume $\mu(t_1, t_2, \mathbf{s}) = \beta_1(t_2, \mathbf{s})x_{t_1,1}^{(0)} + \beta_2(t_2, \mathbf{s})x_{t_1,2}^{(0)}$.
Here, $\beta_{p_0}(t_2, \mathbf{s}) = \sum_{p_1=1}^{P_T=12} \beta_{p_0,p_1}(\mathbf{s})x_{t_2,p_1}^{(1)}$ and $\beta_{p_0,p_1}(\mathbf{s}_n) = \sum_{p_2=1}^{P_S=189} \beta_{p_0,p_1,p_2}x_{n,p_2}^{(2)}$, where
 - $x_{t_1,1}^{(0)} = 1/\sqrt{31}$ (intercept),
 - $x_{t_1,2}^{(0)}$ = rescaled annual mean SST projection based on RCP 4.5/8.5 scenarios,
 - $x_{t_2,p_1}^{(1)}$ = splines across months evaluated at t_2 (within a year),
 - $x_{n,p_2}^{(2)}$ = splines across space evaluated at \mathbf{s}_n .
- Vectorized form: $\boldsymbol{\mu} = [\mathbf{X}_{0;1} \otimes \mathbf{X}_1 \otimes \mathbf{X}_2] \boldsymbol{\beta}_1 + [\mathbf{X}_{0;2} \otimes \mathbf{X}_1 \otimes \mathbf{X}_2] \boldsymbol{\beta}_2$.
- We assume that the density of ϵ_t is $f_{\text{DPM}}(\boldsymbol{\epsilon}) = \sum_{k=1}^K \pi_k f_T(\boldsymbol{\epsilon} \mid \Theta_k = \{\boldsymbol{\Phi}_k, \tau_k^2, \mathbf{a}_k\})$, $\mathbf{a}_k > 2 \forall k$.

The proposed model

- We model the Red Sea SST data as $Y_t(\mathbf{s}_n) = \mu_t(\mathbf{s}_n) + \epsilon_t(\mathbf{s}_n)$, $t = 1, \dots, T = 1612$.
- We write $\mu_t(\mathbf{s}_n) = \mu(t_1, t_2, \mathbf{s}_n)$ where $t_1 = \lceil t/52 \rceil$ (i.e., year), and $t_2 = t - 52(t_1 - 1)$ (i.e., week).
- We assume $\mu(t_1, t_2, \mathbf{s}) = \beta_1(t_2, \mathbf{s})x_{t_1,1}^{(0)} + \beta_2(t_2, \mathbf{s})x_{t_1,2}^{(0)}$.
Here, $\beta_{p_0}(t_2, \mathbf{s}) = \sum_{p_1=1}^{P_T=12} \beta_{p_0,p_1}(\mathbf{s})x_{t_2,p_1}^{(1)}$ and $\beta_{p_0,p_1}(\mathbf{s}_n) = \sum_{p_2=1}^{P_S=189} \beta_{p_0,p_1,p_2}x_{n,p_2}^{(2)}$, where
 - $x_{t_1,1}^{(0)} = 1/\sqrt{31}$ (intercept),
 - $x_{t_1,2}^{(0)}$ = rescaled annual mean SST projection based on RCP 4.5/8.5 scenarios,
 - $x_{t_2,p_1}^{(1)}$ = splines across months evaluated at t_2 (within a year),
 - $x_{n,p_2}^{(2)}$ = splines across space evaluated at \mathbf{s}_n .
- Vectorized form: $\boldsymbol{\mu} = [\mathbf{X}_{0;1} \otimes \mathbf{X}_1 \otimes \mathbf{X}_2] \boldsymbol{\beta}_1 + [\mathbf{X}_{0;2} \otimes \mathbf{X}_1 \otimes \mathbf{X}_2] \boldsymbol{\beta}_2$.
- We assume that the density of ϵ_t is $f_{\text{DPM}}(\boldsymbol{\epsilon}) = \sum_{k=1}^K \pi_k f_T(\boldsymbol{\epsilon} \mid \Theta_k = \{\boldsymbol{\Phi}_k, \tau_k^2, a_k\})$, $a_k > 2 \forall k$.
- $f_T(\boldsymbol{\epsilon} \mid \Theta = \{\boldsymbol{\Phi}, \tau^2, a\})$ is the multivariate t density of $T_a(\mathbf{0}_N, \frac{a-2}{a} [\mathbf{H}\boldsymbol{\Phi}\mathbf{H}' + \tau^2 \mathbf{I}_N])$.

$\epsilon(\mathbf{s})$: Low-rank dependence structure based on empirical orthogonal functions (EOFs)

- Estimate $\mu_t(\mathbf{s}_n)$ empirically based on the data (e.g., LS estimates).

$\epsilon(\mathbf{s})$: Low-rank dependence structure based on empirical orthogonal functions (EOFs)

- Estimate $\mu_t(\mathbf{s}_n)$ empirically based on the data (e.g., LS estimates).
- Consider $\hat{\epsilon}_t(\mathbf{s}_n) = Y_t(\mathbf{s}_n) - \hat{\mu}_t(\mathbf{s}_n)$.

$\epsilon(\mathbf{s})$: Low-rank dependence structure based on empirical orthogonal functions (EOFs)

- Estimate $\mu_t(\mathbf{s}_n)$ empirically based on the data (e.g., LS estimates).
- Consider $\hat{\epsilon}_t(\mathbf{s}_n) = Y_t(\mathbf{s}_n) - \hat{\mu}_t(\mathbf{s}_n)$.
- Form the sample spatial covariance matrix $\hat{\Sigma}$ of $\hat{\epsilon}_t, t = 1, \dots, T$.

$\epsilon(\mathbf{s})$: Low-rank dependence structure based on empirical orthogonal functions (EOFs)

- Estimate $\mu_t(\mathbf{s}_n)$ empirically based on the data (e.g., LS estimates).
- Consider $\hat{\epsilon}_t(\mathbf{s}_n) = Y_t(\mathbf{s}_n) - \hat{\mu}_t(\mathbf{s}_n)$.
- Form the sample spatial covariance matrix $\hat{\Sigma}$ of $\hat{\epsilon}_t, t = 1, \dots, T$.
- Perform eigen-decomposition of $\hat{\Sigma}$ and let the highest L eigenvalues be $\lambda_1, \dots, \lambda_L$ and $\mathbf{h}_1, \dots, \mathbf{h}_L$ be the corresponding eigenvectors.

$\epsilon(\mathbf{s})$: Low-rank dependence structure based on empirical orthogonal functions (EOFs)

- Estimate $\mu_t(\mathbf{s}_n)$ empirically based on the data (e.g., LS estimates).
- Consider $\hat{\epsilon}_t(\mathbf{s}_n) = Y_t(\mathbf{s}_n) - \hat{\mu}_t(\mathbf{s}_n)$.
- Form the sample spatial covariance matrix $\hat{\Sigma}$ of $\hat{\epsilon}_t, t = 1, \dots, T$.
- Perform eigen-decomposition of $\hat{\Sigma}$ and let the highest L eigenvalues be $\lambda_1, \dots, \lambda_L$ and $\mathbf{h}_1, \dots, \mathbf{h}_L$ be the corresponding eigenvectors.
- Choose $\mathbf{H} = [\mathbf{h}_1 | \dots | \mathbf{h}_L]$ and let $\mathbf{\Delta} = \text{diag}(\lambda_1, \dots, \lambda_L)$.

$\epsilon(\mathbf{s})$: Low-rank dependence structure based on empirical orthogonal functions (EOFs)

- Estimate $\mu_t(\mathbf{s}_n)$ empirically based on the data (e.g., LS estimates).
- Consider $\hat{\epsilon}_t(\mathbf{s}_n) = Y_t(\mathbf{s}_n) - \hat{\mu}_t(\mathbf{s}_n)$.
- Form the sample spatial covariance matrix $\hat{\Sigma}$ of $\hat{\epsilon}_t, t = 1, \dots, T$.
- Perform eigen-decomposition of $\hat{\Sigma}$ and let the highest L eigenvalues be $\lambda_1, \dots, \lambda_L$ and $\mathbf{h}_1, \dots, \mathbf{h}_L$ be the corresponding eigenvectors.
- Choose $\mathbf{H} = [\mathbf{h}_1 | \dots | \mathbf{h}_L]$ and let $\mathbf{\Delta} = \text{diag}(\lambda_1, \dots, \lambda_L)$.
- $\hat{\Sigma} \approx \mathbf{H}\mathbf{\Delta}\mathbf{H}' + \psi^2 \mathbf{I}_N$ for some small ψ^2 .

$\epsilon(\mathbf{s})$: Low-rank dependence structure based on empirical orthogonal functions (EOFs)

- Estimate $\mu_t(\mathbf{s}_n)$ empirically based on the data (e.g., LS estimates).
- Consider $\hat{\epsilon}_t(\mathbf{s}_n) = Y_t(\mathbf{s}_n) - \hat{\mu}_t(\mathbf{s}_n)$.
- Form the sample spatial covariance matrix $\hat{\Sigma}$ of $\hat{\epsilon}_t, t = 1, \dots, T$.
- Perform eigen-decomposition of $\hat{\Sigma}$ and let the highest L eigenvalues be $\lambda_1, \dots, \lambda_L$ and $\mathbf{h}_1, \dots, \mathbf{h}_L$ be the corresponding eigenvectors.
- Choose $\mathbf{H} = [\mathbf{h}_1 | \dots | \mathbf{h}_L]$ and let $\mathbf{\Delta} = \text{diag}(\lambda_1, \dots, \lambda_L)$.
- $\hat{\Sigma} \approx \mathbf{H}\mathbf{\Delta}\mathbf{H}' + \psi^2 \mathbf{I}_N$ for some small ψ^2 .
- For the proposed model, $\text{Cov}(\epsilon_t) = \mathbf{H}(\sum_{k=1}^K \pi_k \Phi_k) \mathbf{H}' + (\sum_{k=1}^K \pi_k \tau_k^2) \mathbf{I}_N$.

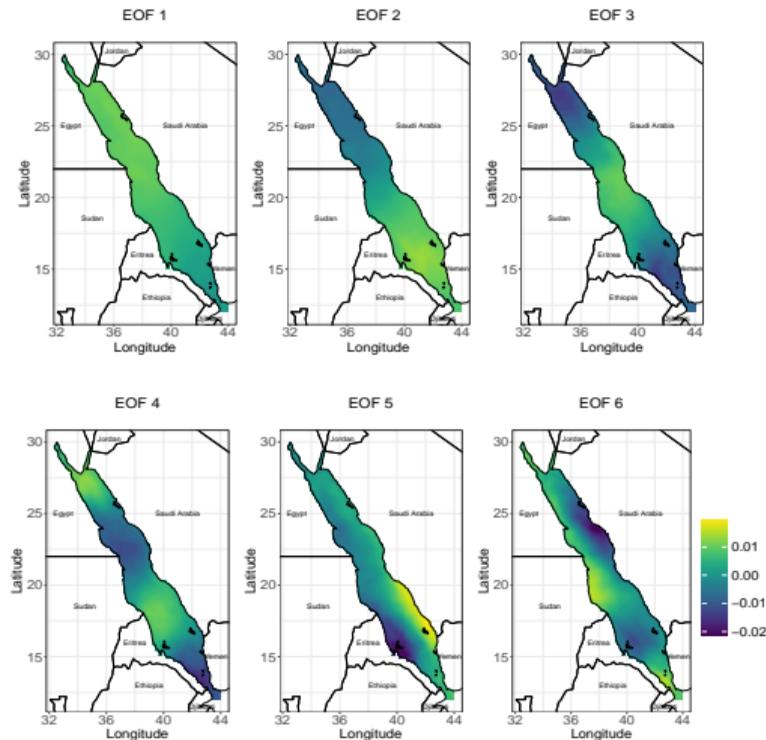


Figure: The proportions of the total variance of SST explained by EOFs 1–6 are 49.40%, 17.78%, 6.14%, 2.87%, 2.26%, and 1.82%, respectively.

Overall model (so-called LTP-DPM)

- Different weeks may be grouped in different “clusters” with different spatial characteristics.
- Hierarchically, given the cluster index $g_t = k$, the process can be written as

$$\begin{aligned} Y_t(\mathbf{s}) &= \mu_t(\mathbf{s}) + \sigma_t [\mathbf{h}'(\mathbf{s})\mathbf{Z}_t + \eta_t(\mathbf{s})], \\ \mathbf{Z}_t &\sim \text{Normal}_L(\mathbf{0}, \Phi_k), \eta_t(\mathbf{s}) \stackrel{iid}{\sim} \text{Normal}(0, \tau_k^2), \\ \sigma_t^2 &\sim \text{Inverse-Gamma}\left(\frac{a_k}{2}, \frac{a_k}{2} - 1\right). \end{aligned}$$

- For computational reasons, we split the mean vector as $\mu = \sum_{i=1}^2 \sum_{j=1}^2 [\mathbf{X}_{0;i} \otimes \mathbf{X}_1 \otimes \mathbf{X}_{2;j}] \beta_{i;j}$, where $\mathbf{X}_{2;1} = \mathbf{P}_H \mathbf{X}_2$, and $\mathbf{X}_{2;2} = (\mathbf{I}_N - \mathbf{P}_H) \mathbf{X}_2$ with $\mathbf{P}_H = \mathbf{H}\mathbf{H}'$.
- We consider the priors:
 - $\Phi_k \sim \text{Inverse-Wishart}(L + 2, \Delta)$, so that $E(\Phi_k) = \Delta$ and $\text{Cov}(\epsilon_t) \approx \hat{\Sigma}$.
 - $\pi_k \sim \text{Stick-Breaking}(\delta)$.

- Bulk properties are similar to DPM models, i.e., nonstationary mean and covariance structure:

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

$$\text{Cov}\{Y_t(\mathbf{s}_{n_1}), Y_t(\mathbf{s}_{n_2}) \mid \Theta_k; k = 1, \dots, K\} = \sum_{k=1}^K \pi_k \left(\mathbf{h}_{n_1} \Phi_k \mathbf{h}'_{n_2} + \tau_k^2 \mathbb{I}_{\{n_1=n_2\}} \right).$$

- Regarding tail properties, we get asymptotic extremal dependence with

$$\chi(\mathbf{s}_{n_1}, \mathbf{s}_{n_2}) = 2\bar{F}_T \left(\sqrt{(a_m + 1) \frac{1 - r_m(\mathbf{s}_{n_1}, \mathbf{s}_{n_2})}{1 + r_m(\mathbf{s}_{n_1}, \mathbf{s}_{n_2})}}; 0, 1, a_m + 1 \right) > 0, \quad m = \arg \min_k \{a_k\}.$$

- Asymptotic extremal independence if $a_k = \infty$ for all k .

Bayesian Inference and Estimation of Extreme Hotspots

- We exploit the separable structure of the design matrices for fast updating of $P = P_{\mathcal{T}}P_{\mathcal{S}} = 2268$ dimensional regression coefficients.
- For the DPM models, we fix K within the MCMC and choose the best K using cross-validation.
- Except the df parameters a_k , $k = 1, \dots, K$, the other parameters have conjugate priors.
- We consider the priors $a_k \stackrel{iid}{\sim} \text{Discrete-Uniform}(2.1, 2.2, \dots, 40.0)$ and sample straightforwardly.
- Computation time (for 60,000 MCMC samples) is ≈ 10 hours which is approximately 2.2 times of fitting a LGP (≈ 4.5 hours).

- ⦿ We extend the approach of French and Sain (2013) (developed for Gaussian processes) to the context of our LTP-DPM model.
- ⦿ We define an exceedance region as $E_{u^+}^0 = \{\mathbf{s}_n \in \mathcal{D} : Y_{t_0}(\mathbf{s}_n) \geq u\}$ for threshold u at time t_0 .
- ⦿ We want to find a “confidence region” $\mathcal{D}_{u^+}^0$ so that $\Pr(E_{u^+}^0 \subseteq \mathcal{D}_{u^+}^0) = 1 - \alpha$ for some predefined probability α .
- ⦿ For a future time t_0 , we test $H_0 : Y_{t_0}(\mathbf{s}_n) = u$ versus $H_1 : Y_{t_0}(\mathbf{s}_n) < u$ for each $\mathbf{s}_n \in \mathcal{D}$ on the basis of some test statistic $\tilde{Y}_{t_0}(\mathbf{s}_n)$ and collect all $\mathbf{s}_n \in \mathcal{D}$ where we fail to reject H_0 .
- ⦿ An obvious choice for $\tilde{Y}_{t_0}(\mathbf{s}_n)$ is exploiting $\hat{Y}_{t_0}(\mathbf{s}_n)$, a predictor of $Y_{t_0}(\mathbf{s}_n)$.
- ⦿ We need to adjust the critical value of the tests to achieve an overall family-wise error rate of α , while accounting for spatial dependence.
- ⦿ We rely on posterior predictive sampling.

Hotspot estimation

- Generate $B = 10^4$ posterior samples from $\mathbf{Y}_{t_0} = [Y_{t_0}(\mathbf{s}_1), \dots, Y_{t_0}(\mathbf{s}_N)]'$.

Hotspot estimation

- Generate $B = 10^4$ posterior samples from $\mathbf{Y}_{t_0} = [Y_{t_0}(\mathbf{s}_1), \dots, Y_{t_0}(\mathbf{s}_N)]'$.
- By Bayesian CLT, posterior average $\hat{Y}_{t_0}(\mathbf{s}_n) | Y_{t_0}(\mathbf{s}_n) \sim \text{Asymptotic-Normal}(Y_{t_0}(\mathbf{s}_n), \tilde{\sigma}_{t_0}^2(\mathbf{s}_n))$.

Hotspot estimation

- Generate $B = 10^4$ posterior samples from $\mathbf{Y}_{t_0} = [Y_{t_0}(\mathbf{s}_1), \dots, Y_{t_0}(\mathbf{s}_N)]'$.
- By Bayesian CLT, posterior average $\hat{Y}_{t_0}(\mathbf{s}_n) | Y_{t_0}(\mathbf{s}_n) \sim \text{Asymptotic-Normal}(Y_{t_0}(\mathbf{s}_n), \tilde{\sigma}_{t_0}^2(\mathbf{s}_n))$.
- We consider $\tilde{Y}_{t_0}(\mathbf{s}_n) = \frac{\hat{Y}_{t_0}(\mathbf{s}_n) - u}{\tilde{\sigma}_{t_0}(\mathbf{s}_n)} \sim \text{Normal}(0, 1)$ under H_0 .

Hotspot estimation

- Generate $B = 10^4$ posterior samples from $\mathbf{Y}_{t_0} = [Y_{t_0}(\mathbf{s}_1), \dots, Y_{t_0}(\mathbf{s}_N)]'$.
- By Bayesian CLT, posterior average $\hat{Y}_{t_0}(\mathbf{s}_n) \mid Y_{t_0}(\mathbf{s}_n) \sim \text{Asymptotic-Normal}(Y_{t_0}(\mathbf{s}_n), \tilde{\sigma}_{t_0}^2(\mathbf{s}_n))$.
- We consider $\tilde{Y}_{t_0}(\mathbf{s}_n) = \frac{\hat{Y}_{t_0}(\mathbf{s}_n) - u}{\tilde{\sigma}_{t_0}(\mathbf{s}_n)} \sim \text{Normal}(0, 1)$ under H_0 .
- The critical value C_α is chosen such that

$$\Pr \left(\min_{\mathbf{s}_n \in E_{u^+}^0} \left\{ \tilde{Y}_{t_0}(\mathbf{s}_n) \right\} < C_\alpha \right) = \alpha.$$

Hotspot estimation

- Generate $B = 10^4$ posterior samples from $\mathbf{Y}_{t_0} = [Y_{t_0}(\mathbf{s}_1), \dots, Y_{t_0}(\mathbf{s}_N)]'$.
- By Bayesian CLT, posterior average $\hat{Y}_{t_0}(\mathbf{s}_n) \mid Y_{t_0}(\mathbf{s}_n) \sim \text{Asymptotic-Normal}(Y_{t_0}(\mathbf{s}_n), \tilde{\sigma}_{t_0}^2(\mathbf{s}_n))$.
- We consider $\tilde{Y}_{t_0}(\mathbf{s}_n) = \frac{\hat{Y}_{t_0}(\mathbf{s}_n) - u}{\tilde{\sigma}_{t_0}(\mathbf{s}_n)} \sim \text{Normal}(0, 1)$ under H_0 .
- The critical value C_α is chosen such that

$$\Pr \left(\min_{\mathbf{s}_n \in E_{u^+}^0} \{ \tilde{Y}_{t_0}(\mathbf{s}_n) \} < C_\alpha \right) = \alpha.$$

- We identify $E_{u^+}^b = \{ \mathbf{s}_n \in \mathcal{D} : Y_{t_0}^{(b)}(\mathbf{s}_n) \geq u \}$ for each b , calculate $\tilde{Y}_{t_0}(\mathbf{s}_n)$, and $\min_{\mathbf{s}_n \in E_{u^+}^b} \{ \tilde{Y}_{t_0}(\mathbf{s}_n) \}$.

Hotspot estimation

- Generate $B = 10^4$ posterior samples from $\mathbf{Y}_{t_0} = [Y_{t_0}(\mathbf{s}_1), \dots, Y_{t_0}(\mathbf{s}_N)]'$.
- By Bayesian CLT, posterior average $\hat{Y}_{t_0}(\mathbf{s}_n) \mid Y_{t_0}(\mathbf{s}_n) \sim \text{Asymptotic-Normal}(Y_{t_0}(\mathbf{s}_n), \tilde{\sigma}_{t_0}^2(\mathbf{s}_n))$.
- We consider $\tilde{Y}_{t_0}(\mathbf{s}_n) = \frac{\hat{Y}_{t_0}(\mathbf{s}_n) - u}{\tilde{\sigma}_{t_0}(\mathbf{s}_n)} \sim \text{Normal}(0, 1)$ under H_0 .
- The critical value C_α is chosen such that

$$\Pr \left(\min_{\mathbf{s}_n \in E_{u^+}^0} \{ \tilde{Y}_{t_0}(\mathbf{s}_n) \} < C_\alpha \right) = \alpha.$$

- We identify $E_{u^+}^b = \{ \mathbf{s}_n \in \mathcal{D} : Y_{t_0}^{(b)}(\mathbf{s}_n) \geq u \}$ for each b , calculate $\tilde{Y}_{t_0}(\mathbf{s}_n)$, and $\min_{\mathbf{s}_n \in E_{u^+}^b} \{ \tilde{Y}_{t_0}(\mathbf{s}_n) \}$.
- We repeat this procedure for each b and estimate C_α by \hat{C}_α , the sample α -th quantile of $\{ \min_{\mathbf{s}_n \in E_{u^+}^b} \{ \tilde{Y}_{t_0}(\mathbf{s}_n) \}; b = 1, \dots, B \}$.

Hotspot estimation

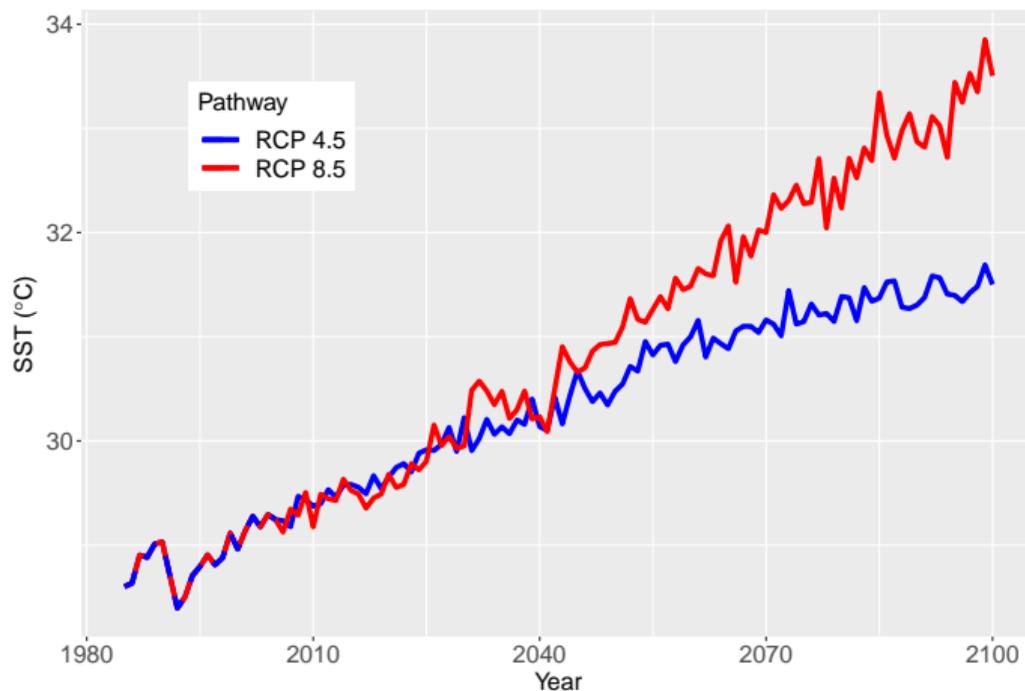
- Generate $B = 10^4$ posterior samples from $\mathbf{Y}_{t_0} = [Y_{t_0}(\mathbf{s}_1), \dots, Y_{t_0}(\mathbf{s}_N)]'$.
- By Bayesian CLT, posterior average $\hat{Y}_{t_0}(\mathbf{s}_n) \mid Y_{t_0}(\mathbf{s}_n) \sim \text{Asymptotic-Normal}(Y_{t_0}(\mathbf{s}_n), \tilde{\sigma}_{t_0}^2(\mathbf{s}_n))$.
- We consider $\tilde{Y}_{t_0}(\mathbf{s}_n) = \frac{\hat{Y}_{t_0}(\mathbf{s}_n) - u}{\tilde{\sigma}_{t_0}(\mathbf{s}_n)} \sim \text{Normal}(0, 1)$ under H_0 .
- The critical value C_α is chosen such that

$$\Pr\left(\min_{\mathbf{s}_n \in E_{u^+}^0} \{\tilde{Y}_{t_0}(\mathbf{s}_n)\} < C_\alpha\right) = \alpha.$$

- We identify $E_{u^+}^b = \{\mathbf{s}_n \in \mathcal{D} : Y_{t_0}^{(b)}(\mathbf{s}_n) \geq u\}$ for each b , calculate $\tilde{Y}_{t_0}(\mathbf{s}_n)$, and $\min_{\mathbf{s}_n \in E_{u^+}^b} \{\tilde{Y}_{t_0}(\mathbf{s}_n)\}$.
- We repeat this procedure for each b and estimate C_α by \hat{C}_α , the sample α -th quantile of $\{\min_{\mathbf{s}_n \in E_{u^+}^b} \{\tilde{Y}_{t_0}(\mathbf{s}_n)\}; b = 1, \dots, B\}$.
- Finally, $\mathcal{D}_{u^+}^0 = \{\mathbf{s}_n \in \mathcal{D} : \tilde{Y}_{t_0}(\mathbf{s}_n) \geq \hat{C}_\alpha\}$.

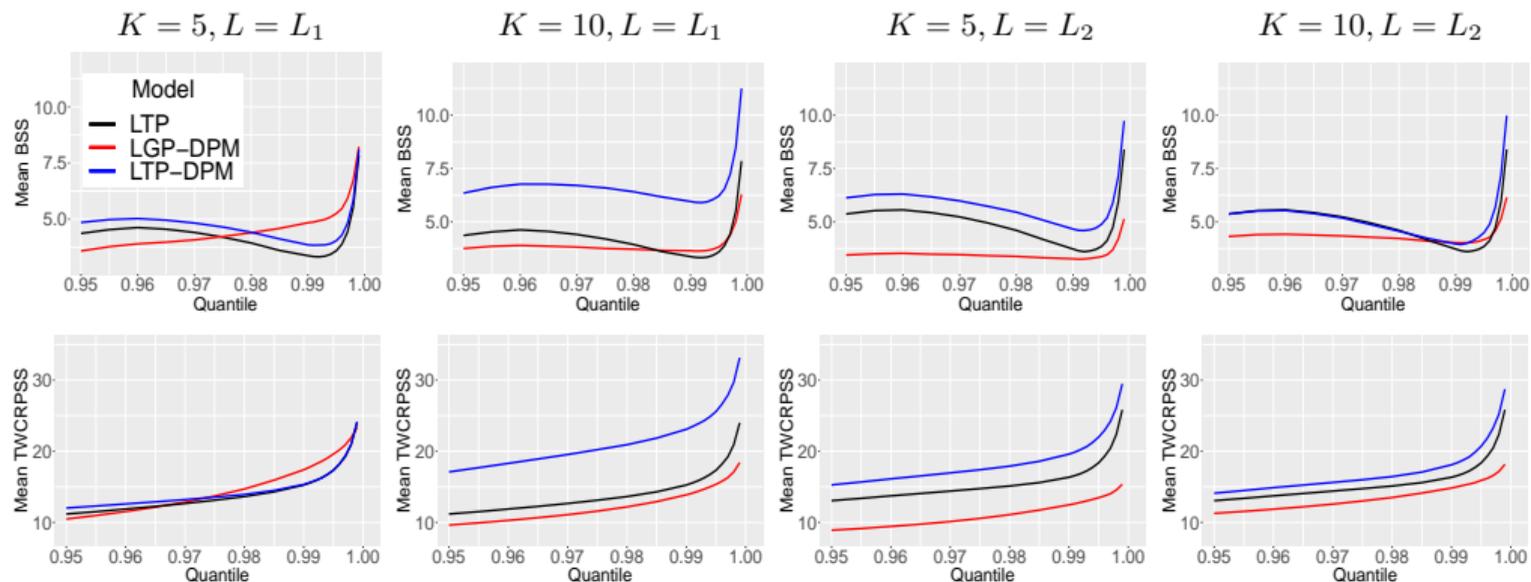
Data Application

Representative concentration pathway (RCP) scenarios



Cross-validation results

- Training set = 1985–2010 (26 years) and Test set = 2011–2015 (5 years).
- We compare models LGP ($K = 1, a_k = \infty$), LTP ($K = 1$), LGP-DPM ($a_k = \infty$) and LTP-DPM across K and L based on BSS and TWCRPSS.



LGP-DPM versus LTP-DPM

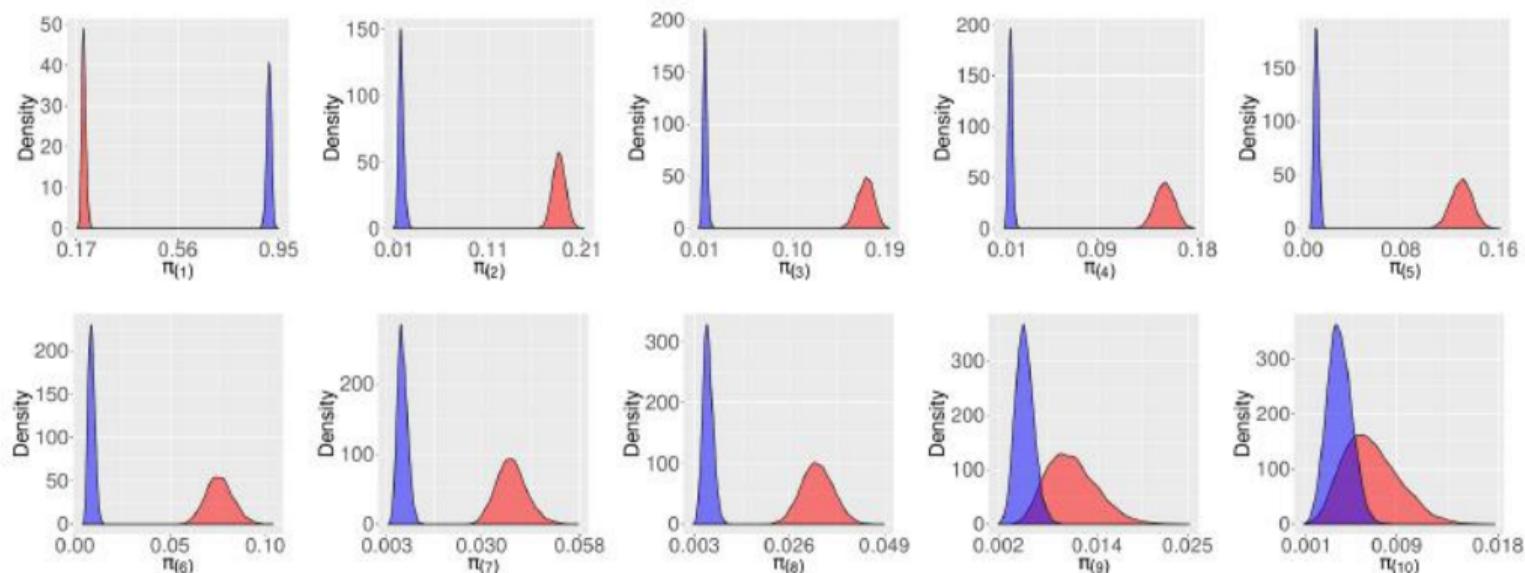
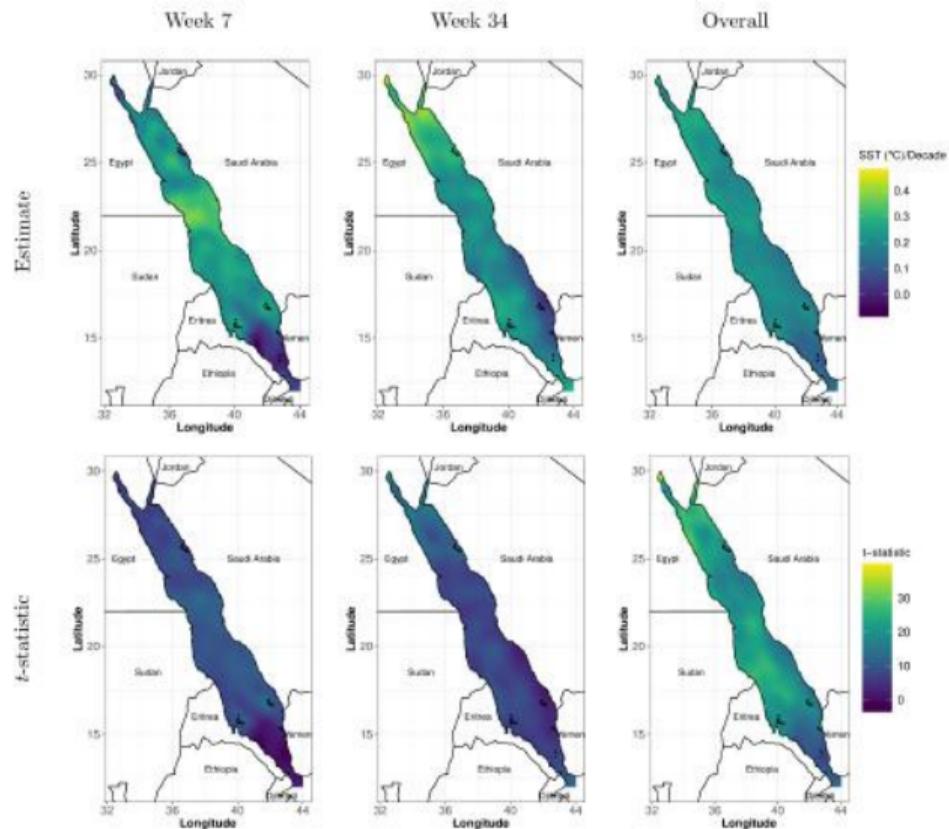
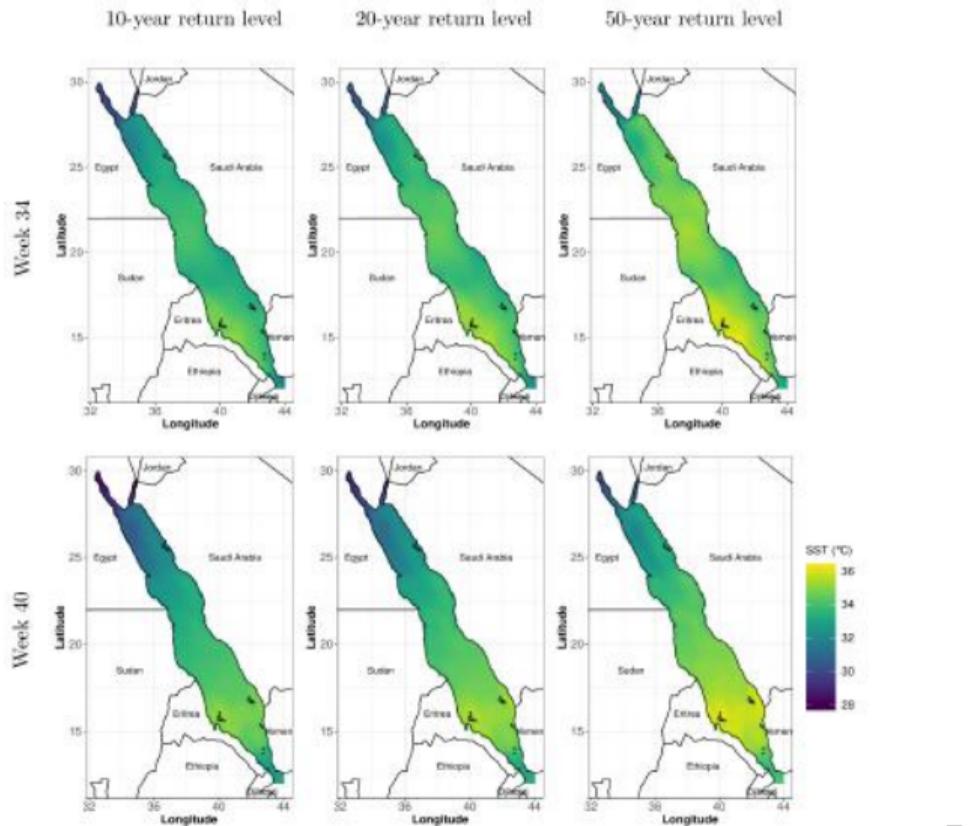


Figure: Posterior densities of the ten ordered stick-breaking probabilities $\pi_{(1)} > \dots > \pi_{(K)}$, based on fitting the LGP-DPM (red) and LTP-DPM (blue) models with $K = 10$ mixture components and $L = L_1$ spatial basis functions.

Estimated decadal rate of change



Estimated return levels



Marginal/joint exceedance probabilities

- Type-I: Considering equal bleaching threshold across the Red Sea (Jokiel and Brown, 2004), we estimate

$$\Pr(\cup_{\mathbf{s}_n \in \mathcal{D}_0} \{Y_{t_0}(\mathbf{s}_n) > u\}) \quad \text{and} \quad \Pr(\cap_{\mathbf{s}_n \in \mathcal{D}_0} \{Y_{t_0}(\mathbf{s}_n) > u\})$$

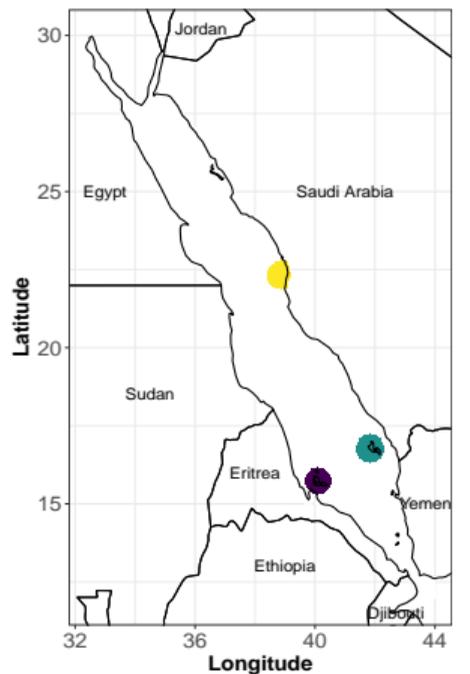
for a range of high temperature values u .

- Type-II: Considering spatially-varying bleaching threshold (Genevier et al., 2019) as well as the adaptive nature of coral reefs to moderate climate change (Logan et al., 2014), we estimate

$$\Pr(\cup_{\mathbf{s}_n \in \mathcal{D}_0} \{Y_{t_0}(\mathbf{s}_n) > Q_{t_0}^{(n)}(p)\}) \quad \text{and} \quad \Pr(\cap_{\mathbf{s}_n \in \mathcal{D}_0} \{Y_{t_0}(\mathbf{s}_n) > Q_{t_0}^{(n)}(p)\})$$

where $Q_{t_0}^{(n)}(p)$ denotes the p -th quantile function of the distribution of $Y_{t_0}(\mathbf{s}_n)$.

Three important regions

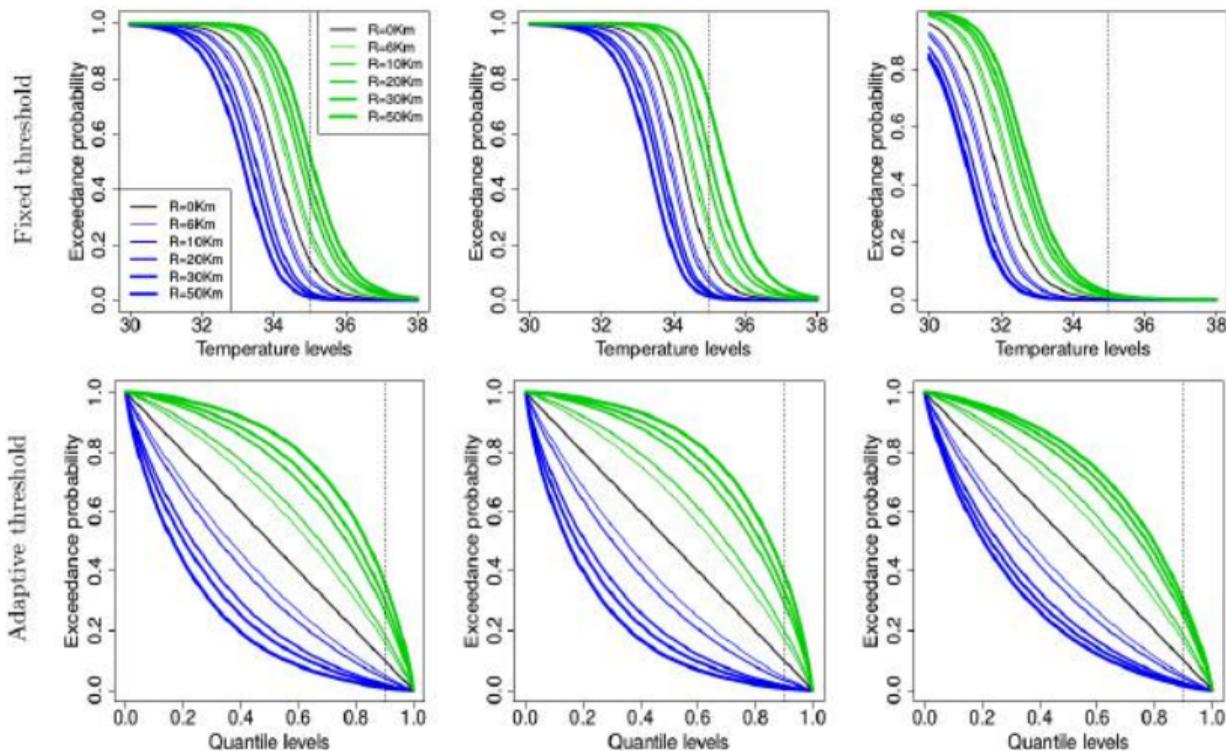


Estimated exceedance probabilities (Week 40, Year 2099, RCP 8.5)

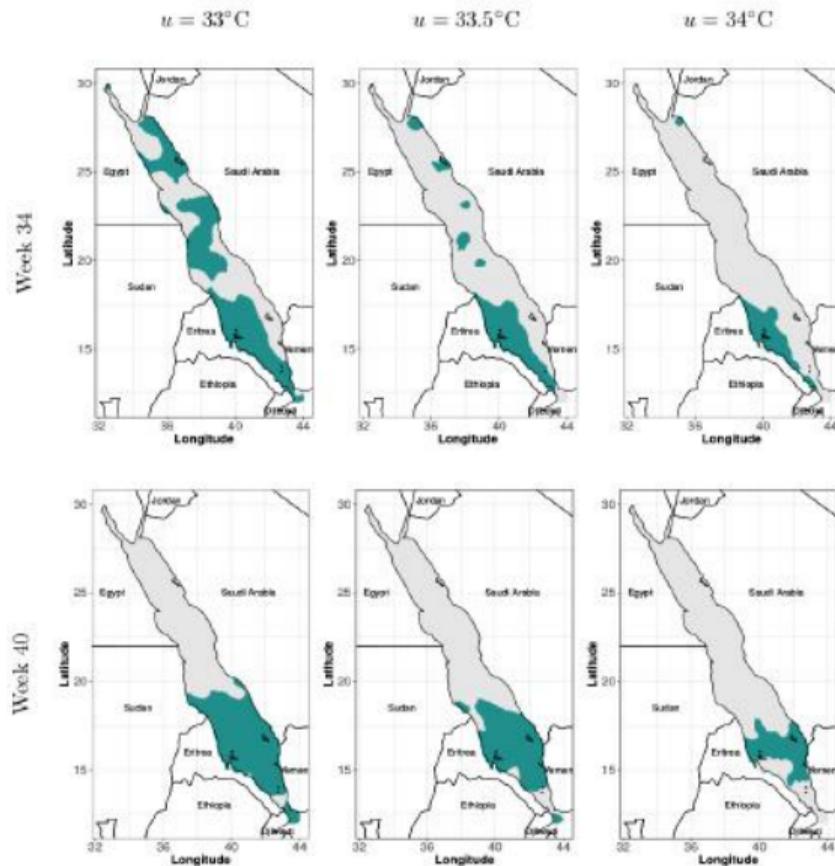
Dahlak Islands, Eritrea

Farasan Islands, Saudi Arabia

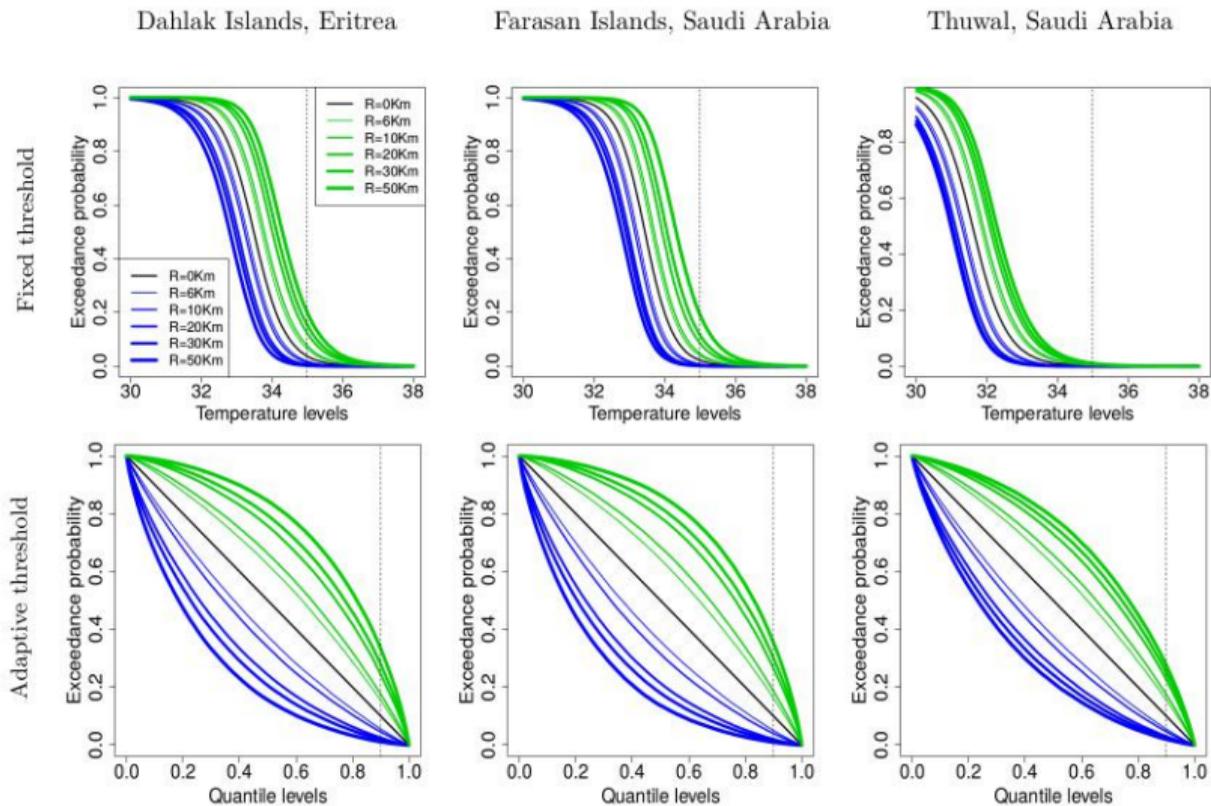
Thuwal, Saudi Arabia



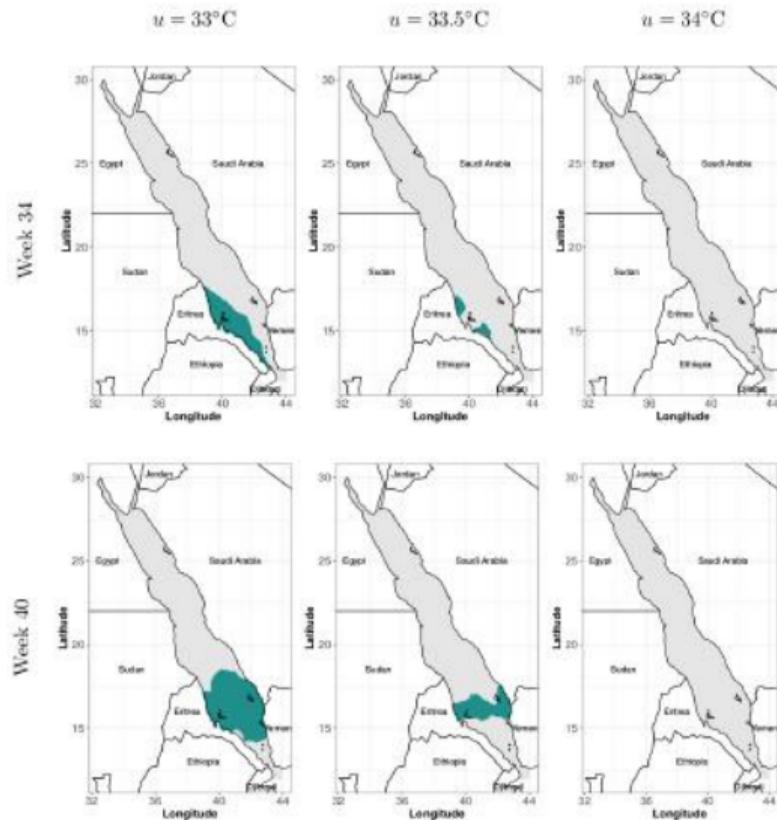
Estimated hotspots (Year 2099, RCP 8.5)



Estimated exceedance probabilities (Week 40, Year 2009, RCP 4.5)



Estimated hotspots (Year 2099, RCP 4.5)



Final Remarks

Conclusions

- We have proposed a low-rank semiparametric Bayesian model for large spatial data where independent replications are available.
- The model has nonstationary mean and covariance structures, and can capture asymptotic extremal dependence.
- Inference can be drawn using an efficient Gibbs sampler, and spatial hotspots can be estimated efficiently using posterior predictive samples.
- Our results suggest that large areas of the Red Sea (that include major coral reefs) are likely to experience very extreme SSTs simultaneously around the end of the 21st century, which may have dramatic consequences in terms of coral bleaching and mortality if mitigation measures are not implemented.
- In future research, it would be interesting to generalize our approach to estimate spatio-temporal dependence, and assess whether spatial hotspots tend to persist over consecutive days, which plays a key role in coral bleaching and mortality.

Thank you!

Hazra, A. and Huser, R. (2020+), *Estimating high-resolution Red Sea surface temperature hotspots, using a low-rank semiparametric spatial model*, *Annals of Applied Statistics*, to appear.

Preprint available at <https://arxiv.org/abs/1912.05657>

References I

- Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848.
- Berumen, M. L., Hoey, A. S., Bass, W. H., Bouwmeester, J., Catania, D., Cochran, J. E., Khalil, M. T., Miyake, S., Mughal, M., Spaet, J. L., et al. (2013). The status of coral reef ecology research in the Red Sea. *Coral Reefs*, 32(3):737–748.
- Bopp, G. P., Shaby, B. A., Forest, C. E., and Mejía, A. (2020). Projecting flood-inducing precipitation with a Bayesian analogue model. *Journal of Agricultural, Biological, and Environmental Statistics*, 25:229–249.
- Castruccio, S., Huser, R., and Genton, M. G. (2016). High-order composite likelihood inference for max-stable distributions and processes. *Journal of Computational and Graphical Statistics*, 25:1212–1229.
- Davison, A. C. and Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and its Application*, 2:203–235.
- Davison, A. C., Huser, R., and Thibaud, E. (2019). Spatial Extremes. In Gelfand, A. E., Fuentes, M., Hoeting, J. A., and Smith, R. L., editors, *Handbook of Environmental and Ecological Statistics*, pages 711–744. CRC Press.
- de Fondeville, R. and Davison, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika*, 105(3):575–592.
- 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 Sensing of Environment*, 116:140–158.

References II

- Duan, J. A., Guindani, M., and Gelfand, A. E. (2007). Generalized spatial Dirichlet process models. *Biometrika*, 94(4):809–825.
- Ferreira, A. and de Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli*. To appear.
- French, J. P. and Sain, S. R. (2013). Spatio-temporal exceedance locations and confidence regions. *Annals of Applied Statistics*, 7(3):1421–1449.
- Furrer, R., Genton, M. G., and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523.
- Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association*, 100(471):1021–1035.
- Genevier, L. G., Jamil, T., Raitzos, D. E., Krokos, G., and Hoteit, I. (2019). Marine heatwaves reveal coral reef zones susceptible to bleaching in the red sea. *Global Change Biology*, 25:2338–2351.
- Hazra, A., Reich, B. J., Shaby, B. A., and Staicu, A.-M. (2018). A semiparametric Bayesian model for spatiotemporal extremes. *arXiv preprint arXiv:1812.11699*.
- Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D. W., Sun, F., and Zammit-Mangion, A. (2019). A case study competition among methods for analysing large spatial data. *Journal of Agricultural, Biological, and Environmental Statistics*, 24(3):398–425.

References III

- Higdon, D. (2002). Space and space-time modeling using process convolutions. In *Quantitative methods for current environmental issues*, pages 37–56. Springer.
- Huser, R., Dombry, C., Ribatet, M., and Genton, M. G. (2019). Full likelihood inference for max-stable data. *Stat*, 8:e218.
- Jokiel, P. L. and Brown, E. K. (2004). Global warming, regional trends and inshore environmental conditions influence coral bleaching in hawaii. *Global Change Biology*, 10(10):1627–1641.
- Logan, C. A., Dunne, J. P., Eakin, C. M., and Donner, S. D. (2014). Incorporating adaptive responses into future projections of coral bleaching. *Global Change Biology*, 20(1):125–139.
- Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277.
- Reich, B. J., Bandyopadhyay, D., and Bondell, H. D. (2013). A nonparametric spatial model for periodontal data with nonrandom missingness. *Journal of the American Statistical Association*, 108(503):820–831.
- Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC.
- Stein, M. (1999). *Statistical Interpolation of Spatial Data*. Springer, New York.
- Thibaud, E. and Opitz, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, 102(4):855–870.
- Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 50(2):297–312.
- Wikle, C. K. and Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, 86(4):815–829.