Conditional Modelling of Spatio-Temporal Extremes for Red Sea Surface Temperatures

Emma S. Simpson*, Jennifer L. Wadsworth

Department of Mathematics and Statistics, Lancaster University, LA1 4YF, U.K.

Abstract

Recent extreme value theory literature has seen significant emphasis on the modelling of spatial extremes, with comparatively little consideration of spatio-temporal extensions. This neglects an important feature of extreme events: their evolution over time. Many existing models for the spatial case are limited by the number of locations they can handle; this impedes extension to space-time settings, where models for higher dimensions are required. Moreover, the spatio-temporal models that do exist are restrictive in terms of the range of extremal dependence types they can capture. Recently, conditional approaches for studying multivariate and spatial extremes have been proposed, which enjoy benefits in terms of computational efficiency and an ability to capture both asymptotic dependence and asymptotic independence. We extend this class of models to a spatio-temporal setting, conditioning on the occurrence of an extreme value at a single space-time location. We adopt a composite likelihood approach for inference, which combines information from full likelihoods across multiple space-time conditioning locations. We apply our model to Red Sea surface temperatures, show that it fits well using a range of diagnostic plots, and demonstrate how it can be used to assess the risk of coral bleaching attributed to high water temperatures over consecutive days.

Keywords: conditional extremes, environmental extremes, extremal dependence modelling, spatio-temporal modelling

1. Introduction

There are many situations where modelling extreme events may be of interest; understanding such events can allow us to prepare for, and potentially mitigate, their effect. For environmental extremes, such as high temperatures or intense rainfall, the spatial extent of an extreme event is of particular concern, with possible focus on determining regions that may be affected by a given

*Corresponding author

Email addresses: e.simpson3@lancaster.ac.uk (Emma S. Simpson), j.wadsworth@lancaster.ac.uk (Jennifer L. Wadsworth)

phenomenon. Moreover, situations may be exacerbated if the extreme events persist for a period of time, indicating that temporal modelling also requires attention. To understand and prepare for such spatio-temporal extreme events, we require modelling techniques motivated by extreme value theory. In this paper, we focus on modelling surface temperature extremes in the Red Sea by extending the so-called conditional approaches of Heffernan and Tawn (2004), Heffernan and Resnick (2007) and Wadsworth and Tawn (2019) to a spatio-temporal setting. We condition on observations at a single space-time location being above some high threshold, and construct a model for other locations across a spatio-temporal domain.

The tail dependence properties of variables are an important consideration in multivariate extreme value modelling. Many models are limited in terms of whether they can capture asymptotic dependence, when variables can take their extreme values simultaneously, or asymptotic independence, when the largest extremes occur separately. More formally, consider random variables $X(s,t)$ and $X(s+h_{\mathrm{S}}, t+h_{\mathrm{T}})$ at two space-time locations separated by spatial displacement h_{S} and temporal lag h_{T}, with $X(\cdot,\cdot) \sim F_{\cdot,\cdot}$. To assess the extremal dependence between these variables, one option is to consider the conditional survivor function

$$\chi\left(u; s, t, h_{\mathrm{S}}, h_{\mathrm{T}}\right) = \Pr\left[F_{s+h_{\mathrm{S}},t+h_{\mathrm{T}}}\{X(s+h_{\mathrm{S}}, t+h_{\mathrm{T}})\} > u \mid F_{s,t}\{X(s,t)\} > u\right], \tag{1}$$

for some value of u close to 1. If the process is isotropic in space and stationary in space and time, then the value of $\chi(u; s, t, h_{\mathrm{S}}, h_{\mathrm{T}})$ is common across all pairs of locations separated by the same spatio-temporal lag $(\|h_{\mathrm{S}}\|, h_{\mathrm{T}})$. In this case, we simplify the notation to $\chi\left(u; s, t, h_{\mathrm{S}}, h_{\mathrm{T}}\right) = \chi\left(u; \|h_{\mathrm{S}}\|, h_{\mathrm{T}}\right)$. A popular measure of asymptotic dependence is obtained by the relation $\chi(\|h_{\mathrm{S}}\|, h_{\mathrm{T}}) = \lim_{u\to 1} \chi(u; \|h_{\mathrm{S}}\|, h_{\mathrm{T}})$, with $\chi(\|h_{\mathrm{S}}\|, h_{\mathrm{T}}) = 0$ corresponding to asymptotic independence, and $\chi(\|h_{\mathrm{S}}\|, h_{\mathrm{T}}) \in (0,1]$ revealing asymptotic dependence between $X(s,t)$ and $X(s+h_{\mathrm{S}}, t+h_{\mathrm{T}})$. In practice, we expect that $\chi(u; \|h_{\mathrm{S}}\|, h_{\mathrm{T}})$ and $\chi(\|h_{\mathrm{S}}\|, h_{\mathrm{T}})$ decrease as $\|h_{\mathrm{S}}\|$ and h_{T} increase.

Existing methods for modelling spatial extremes include max-stable processes (see Davison et al. (2012)), with spatio-temporal extensions given by Davis et al. (2013) and Huser and Davison (2014). These processes arise as asymptotically-justified extensions of the generalised extreme value distribution for modelling univariate extremes (von Mises, 1936; Jenkinson, 1955) when pointwise maxima are taken over spatio-temporal domains. A drawback of max-stable processes is that they are computationally intensive to fit, limiting the number of space-time locations one can feasibly handle, though the recently proposed semiparametric approach of Buhl et al. (2019) aims to address this point. A further potential drawback of this class of models is their ability to capture only asymptotic dependence. If max-stable models are fitted to asymptotically independent data, results will be too conservative as we extrapolate into the tail. They therefore lend themselves to modelling extremes only in small spatio-temporal domains, where asymptotic dependence is more likely to arise.

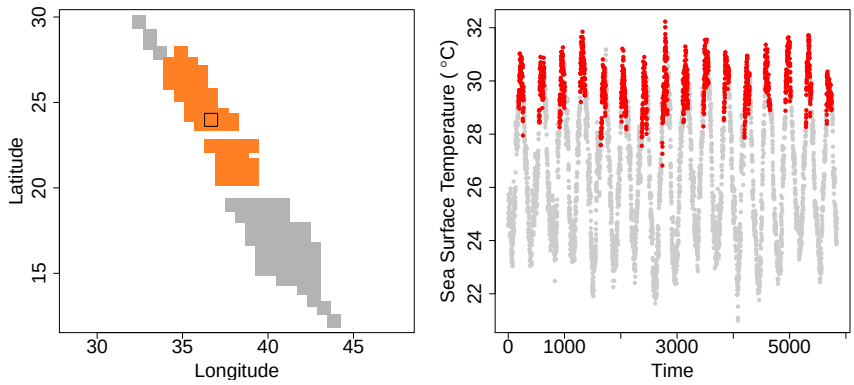

Figure 1: Locations of observations, with the northerly region we study in orange (left), and sea surface temperature data for the location outlined in black (right). The red points correspond to the months July to September.

It is also unclear which data applications would be of interest for temporally dependent maxima, since the process of taking block maxima usually leads to approximately independent observations, meaning space-time max-stable process models may not be so useful in practice. To address some of these issues, Morris et al. (2017) propose a modelling technique that allows for different tail dependence features across space. In particular, they suggest partitioning the spatial locations to allow for asymptotic dependence for nearby sites, and asymptotic independence between those that are further apart, modelled via a skew-t process. The temporal dependence in their model is captured by an AR(1) process, but in many cases it may be desirable to have models that are not restricted to one type of temporal dependence, allowing for both asymptotic dependence and asymptotic independence in time.

Recently, there has been increasing interest in using conditional approaches, first introduced by Heffernan and Tawn (2004), to model extreme events. In the multivariate setting, this involves conditioning on a single variable being above some high threshold, and modelling the behaviour of the remaining variables. In the spatial case (Wadsworth and Tawn, 2019), additional structural properties are exploited to construct models conditioning on extreme values at a single spatial location. Conditional extremes models have the advantage of being able to capture a range of tail dependence behaviours, while being computationally efficient compared to other methods. These appealing features lead us to extend the spatial conditional extremes methodology to a spatio-temporal setting, which is particularly helpful for handling the additional burden that comes from expanding the domain from space to space-time. Important new considerations for this spatio-temporal extension include: whether or not the model should exhibit separability in space and time; how to separate the data into individual events for inference; and the construction of model diagnostics that allow us to assess the fit across both domains.

We demonstrate our new spatio-temporal model through an application to Red Sea surface temperatures. This is particularly of interest as continual high sea temperatures can lead to problems for marine life, with coral bleaching being a notable concern; see McClanahan et al. (2007) and Huser (2020). While corals can recover from high sea temperatures that occur over a short period of time, prolonged extreme events can lead to irreparable damage and coral mortality, highlighting the importance of studying the temporal aspect of sea surface temperature extremes. A dataset relating to the Red Sea was the focus of the recent EVA conference data challenge (Huser, 2020). The challenge involved modelling temperature anomalies rather than the raw surface temperatures we study here, with the aim being to develop methods for missing data, which is also not our focus. Approaches developed as a result of this challenge include Castro-Camilo et al. (2020) and Ivek and Vlah (2020).

To study surface temperature extremes in the Red Sea, we have access to 16 years of daily sea surface temperature data, for the 108 locations shown in the left panel of Figure 1. These locations are a subset of $0.05^\circ \times 0.05^\circ$ grid cells covering the whole of the Red Sea. The data are obtained from a combination of satellite measurements and in situ readings; see Donlon et al. (2012). Our initial investigations indicated differing behaviour in the north and south of the Red Sea, as well as in the most north-westerly locations, corresponding to the Gulf of Suez; this issue is also identified by Huser (2020). Since the model we propose assumes spatio-temporal stationarity, we focus on the northerly region highlighted in orange, containing 54 locations, and comment further on approaches to deal with spatial non-stationarity in Section 5. To remove the need to account for seasonality in the data, we focus only on the warmest months of July to September. Example data for one location are shown in the right panel of Figure 1, with other locations in the northerly region exhibiting similar seasonal behaviour. Focussing on this time period is consistent with the claims of Maynard et al. (2008), that coral bleaching is most likely during summer months.

The remainder of the paper is constructed as follows. We detail our proposed model in Section 2, and discuss inference and simulation in Section 3. In Section 4, we present a variety of diagnostic tools for assessing model fit, as well as results on Red Sea surface temperatures in the context of coral bleaching. Section 5 concludes with a discussion of potential extensions.

2. Modelling

2.1. Model assumption

Let $w = (s, t) \in \mathcal{S} \times \mathcal{T} \subset \mathbb{R}^2 \times \mathbb{R}$ denote a space-time location, and suppose we are interested in the stationary spatio-temporal process $\{X(w) : w \in \mathcal{S} \times \mathcal{T}\}$. As with all conditional extremes approaches, we require that each marginal distribution of the process has a standard exponential upper tail, i.e., $\Pr\{X(w) > x\} \sim e^{-x}$, as $x \to \infty$, for each $w \in \mathcal{S} \times \mathcal{T}$. In practice, this can be achieved via a transformation. We focus on Laplace margins, as suggested by Keef et al.

(2013), which we obtain by applying the probability integral transform via a rank transformation to each spatial location.

Conditioning on the process at a single space-time location, $w_0 \in \mathcal{S} \times \mathcal{T}$, being above a high threshold u, we assume that for a finite number of space-time locations $w_1, \ldots, w_{dm}$, corresponding to d spatial locations at m time-points, there exist functions $a_{w-w_0}(\cdot)$ and $b_{w-w_0}(\cdot)$ such that

$$\left(\left[\frac{X(w_i) - a_{w_i - w_0}\{X(w_0)\}}{b_{w_i - w_0}\{X(w_0)\}} \right]_{i=1,\ldots,dm}, X(w_0) - u \right) \Bigg| X(w_0) > u \xrightarrow{d} \left(\left\{ Z^0(w_i) \right\}_{i=1,\ldots,dm}, E \right), \tag{2}$$

as $u \to \infty$. That is, after normalisation, the process $\{X(w) : w \in \mathcal{S} \times \mathcal{T}\}$ converges to $\{Z^0(w) : w \in \mathcal{S} \times \mathcal{T}\}$ in the sense of finite dimensional distributions. The variable E is independent of $\{Z^0(w)\}$ in the limit and follows a standard exponential distribution. The function $a_{w-w_0}(x)$ should take values in $[0, x]$, with $a_0(x) = x$, and be non-increasing with respect to $\|s - s_0\|$ and $\|t - t_0\|$. This implies that the process $\{Z^0(w)\}$, subsequently referred to as the residual process, must satisfy $Z^0(w_0) = 0$. We note that although we are working with finite dimensional distributions here, these are given useful structure by having the spatio-temporal index domain.

To model $\{X(w)\}$ using assumption (2), choices need to be made about the form of the normalising functions and the residual process; we discuss these in Sections 2.2 and 2.3, respectively. Wadsworth and Tawn (2019) demonstrate that (2) holds for a wide variety of underlying spatial dependence structures, detailing specific forms of $a_{s-s_0}(\cdot)$, $b_{s-s_0}(\cdot)$ and $\{Z^0(s) : s \in \mathcal{S}\}$. It will therefore hold for analogous dependence structures with spatio-temporal indexing. We check the validity of assumption (2) for our Red Sea data in Section 4.

2.2. The normalising functions $a_{w-w_0}(\cdot)$ and $b_{w-w_0}(\cdot)$

In order to exploit assumption (2) for modelling purposes, we impose parametric forms on the normalising functions $a_{w-w_0}(\cdot)$ and $b_{w-w_0}(\cdot)$. In the spatial case, under the assumption of isotropy, Wadsworth and Tawn (2019) suggest taking $a_{s-s_0}(x) = x\alpha_{\mathrm{S}}(s - s_0)$, with

$$\alpha_{\mathrm{S}}(s - s_0) = \begin{cases} 1, & \|s - s_0\| \leq \Delta_{\mathrm{S}}, \\ \exp\left[-\left\{ \left(\|s - s_0\| - \Delta_{\mathrm{S}} \right) / \lambda_{\mathrm{S}} \right\}^{\kappa_{\mathrm{S}}} \right], & \|s - s_0\| > \Delta_{\mathrm{S}}, \end{cases} \tag{3}$$

for $\|s - s_0\|$ denoting the distance between s and s_0, $\lambda_{\mathrm{S}} > 0$, $\kappa_{\mathrm{S}} \in [0, 2]$ and $\Delta_{\mathrm{S}} \geq 0$. If $\Delta_{\mathrm{S}} > 0$, this yields $a_{s-s_0}(x) = x$ for $\|s - s_0\| \leq \Delta_{\mathrm{S}}$, allowing for asymptotic dependence up to distances of Δ_{S} from the conditioning site, and asymptotic independence beyond that. Although (3) assumes isotropy, anisotropy is straightforward to handle, as discussed in Section 2.4. In the spatio-temporal setting, a possible extension of this is to take

$$a_{w-w_0}(x) = x\alpha_{\mathrm{S}}(s - s_0)\alpha_{\mathrm{T}}(t - t_0), \tag{4}$$

with $\alpha_{\mathrm{T}}(t-t_0)$ defined analogously to (3), where we recall that the conditioning site is denoted by $w_0 = (s_0, t_0)$, and more generally we set $w = (s, t)$. This allows for asymptotic dependence within some space-time neighbourhood of the conditioning site, controlled by the parameters $(\Delta_{\mathrm{S}}, \Delta_{\mathrm{T}})$, and asymptotic independence outside this neighbourhood. In modelling the Red Sea surface temperatures, we found the inclusion of these parameters not to have a significant effect; we therefore fix $\Delta_{\mathrm{S}}, \Delta_{\mathrm{T}} = 0$, corresponding to an asymptotically independent model.

Normalising function (4) exhibits space-time separability, but one could instead consider a non-separable form for $a_{w-w_0}(\cdot)$. A natural choice is to exploit the class of known, non-separable covariance functions, such as those introduced by Cressie and Huang (1999) and studied further by Gneiting (2002). One possibility, inspired by the class of Gneiting (2002), is to take

$$a_{w-w_0}(x) = x\left(\lambda_{\mathrm{T}}\|t-t_0\|^{2\kappa_{\mathrm{T}}}+1\right)^{-1}\exp\left\{-\frac{\lambda_{\mathrm{S}}\|s-s_0\|^{2\kappa_{\mathrm{S}}}}{(\lambda_{\mathrm{T}}\|t-t_0\|^{2\kappa_{\mathrm{T}}}+1)^{\eta\kappa_{\mathrm{S}}}}\right\}, \tag{5}$$

for $\lambda_{\mathrm{S}}, \lambda_{\mathrm{T}} > 0$, $\kappa_{\mathrm{S}}, \kappa_{\mathrm{T}} \in (0, 1]$ and $\eta \in [0, 1]$. The larger the parameter η, the stronger the interaction between space and time, with $\eta = 0$ corresponding to separability. We note that while certain covariance functions provide a convenient way to introduce non-separability into the function $a_{w-w_0}(x)$, we are not restricted to this class of models since covariance functions must induce positive definiteness in their covariance matrices, but there is no such constraint on the function $a_{w-w_0}(x)$. Exploiting the covariance function in (5) also allows us to satisfy the condition that $a_{w-w_0}(x) \in [0, x]$, and leads to the intuitive model property that $a_{w-w_0}(x)$ is non-increasing with respect to $\|s-s_0\|$ and $\|t-t_0\|$. Asymptotic dependence could be incorporated into (5) via an approach analogous to the use of $(\Delta_{\mathrm{S}}, \Delta_{\mathrm{T}})$ in the separable case.

Three theoretically-motivated forms for the normalising function $b_{s-s_0}(\cdot)$ are proposed by Wadsworth and Tawn (2019). We focus on one such example in the spatio-temporal extension, taking

$$b_{w-w_0}(x) = 1 + \{a_{w-w_0}(x)\}^{\beta}, \tag{6}$$

for some $\beta \in [0, 1]$. This can be used for either of our proposed forms of $a_{w-w_0}(x)$. The advantage of using (6) as the model for $b_{w-w_0}(x)$ is that the function can vary with distance, while being controlled by only a single extra parameter, but other options could also be used.

2.3. The residual process $\{Z^0(w)\}$

We construct a model for the residual process by first considering the stationary space-time Gaussian process $\{Z(w) : w \in \mathcal{S} \times \mathcal{T}\}$, with mean μ and standard deviation $\sigma > 0$. We propose using a separable covariance function with powered exponential components in space and time, i.e., the covariance corresponding to sites $w_i = (s_i, t_i)$ and $w_j = (s_j, t_j)$ is

$$\sigma^2 \cdot \exp\left\{-\left(\|s_i - s_j\|/\phi_{\mathrm{S}}\right)^{p_{\mathrm{S}}}\right\} \cdot \exp\left\{-\left(\|t_i - t_j\|/\phi_{\mathrm{T}}\right)^{p_{\mathrm{T}}}\right\}. \tag{7}$$

To ensure the condition that $Z^0(w_0) = 0$ is satisfied, the residual process is taken as the Gaussian process with distribution equal to that of $\{Z(w)\}|Z(w_0) = 0$. We note that taking this Gaussian form for the residual process does not induce Gaussianity in the process $\{X(w)\}$ due to the appearance of the normalising functions.

Recall that we are interested in modelling observations at space-time locations $w_1, \ldots, w_{dm}$. The corresponding mean vector of the Gaussian process $\{Z(w) : w \in \mathcal{S} \times \mathcal{T}\}$ is $\boldsymbol{\mu} = \mu \cdot \mathbf{1}_{dm} \in \mathbb{R}^{dm}$, with $\mathbf{1}_{dm}$ denoting a vector of 1s, and we denote the covariance matrix, constructed via function (7), as $\Sigma \in \mathbb{R}^{dm \times dm}$. These two components can be partitioned to represent the space-time locations to be modelled, and the conditioning site, i.e., $\boldsymbol{\mu} = \mu \cdot (\mathbf{1}_{dm-1}, 1)$, and

$$\Sigma = \begin{bmatrix} \Sigma^* & \Sigma_0^* \\ \Sigma_0^{*T} & \sigma^2 \end{bmatrix}, \text{ for } \Sigma^* \in \mathbb{R}^{(dm-1)\times(dm-1)} \text{ and } \Sigma_0^* \in \mathbb{R}^{dm-1}.$$

Then the conditional Gaussian process $\{Z(w)\}|Z(w_0) = 0$ follows a multivariate Gaussian distribution with mean $\boldsymbol{\mu}_{|0} \in \mathbb{R}^{dm-1}$ and covariance matrix $\Sigma_{|0} \in \mathbb{R}^{(dm-1)\times(dm-1)}$ of the form

$$\boldsymbol{\mu}_{|0} = \mu \cdot \left(\mathbf{1}_{dm-1} - \sigma^{-2}\Sigma_0^*\right); \quad \Sigma_{|0} = \Sigma^* - \sigma^{-2}\Sigma_0^*\Sigma_0^{*T}. \tag{8}$$

In place of (7), we could also consider a non-separable covariance function. However, most of the structure in our model is captured by the normalising functions $a_{w-w_0}(x)$ and $b_{w-w_0}(x)$, and separability in the covariance function of $\{Z(w)\}$ actually induces non-separability in $\{Z^0(w)\}$. We therefore choose to focus only on covariance function (7).

Wadsworth and Tawn (2019) propose what they term a delta-Laplace distribution for the marginal form of their residual process $\{Z^0(s) : s \in \mathcal{S}\}$, which has univariate Gaussian and Laplace distributions as special cases. We expect independence between observations at large spatial lags, and the delta-Laplace distribution allows for the recovery of the original Laplace margins in such cases. It is possible to include such a marginal distribution in our spatio-temporal extension, however, we found that at the distances studied in our Red Sea example, this is not necessary, and retaining the Gaussian margins of the conditional Gaussian process leads to improvements in terms of computational efficiency.

We note that our choices of normalising functions and residual process lead to a model that is symmetric with respect to time, i.e., the model assumes that the process has the same behaviour going forwards or backwards from time t_0. In Section 4.3, we demonstrate that this reasonable for our data example, but in other cases the assumption of temporal symmetry may not be valid. If such an issue arises, it may be better to consider the forwards and backwards cases separately. In such cases, models of the form we have proposed could still be used, but different parameter estimates may be obtained in each situation. Depending on the data set and relevant issues that need to be studied, it may also be possible to restrict the analysis to focus on just one of the cases.

2.4. Spatial anisotropy

We account for spatial anisotropy through a transformation of the spatial coordinates, setting

$$s^* = \begin{pmatrix} 1 & 0 \\ 0 & 1/L \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} s, \tag{9}$$

where the parameter $\theta \in [-\pi/2, 0]$ controls rotation, and $L > 0$ relates to the amount by which the coordinates are stretched, with $L = 1$ recovering the isotropic model. Such an approach is often termed *geometric anisotropy*. These parameters are estimated, along with the other model parameters, using the methods discussed in Section 3.2.

3. Inference and simulation

3.1. Extreme events for inference

We propose carrying out inference using a likelihood approach. To reduce the computational burden of these calculations, we first aim to separate the data into shorter space-time blocks that allow us to capture the evolution of single extreme events.

Suppose that an extreme value at a particular location is defined as any observation above the high threshold u. One option, proposed by Huser and Wadsworth (2019), is to consider clusters of days with extreme values at any site, separated by k days with no extreme values, to belong to a single extreme event, for some value k. This provides an intuitive way to separate extreme events, but for our data, we found this often produced long clusters that were not computationally feasible to handle. A simpler alternative is to divide the data into blocks of m days, for some choice of m, taking care not to have blocks that span multiple years. If the chosen block-length, m, does not exactly divide the length of each summer period, we remove an appropriate number of observations at the end of each one. The value of m should be large enough to provide sufficient information about temporal changes in the extremal dependence, but small enough to ensure computational feasibility.

To determine an appropriate value of m, we investigate the spatio-temporal dependence present in our data, using the measure defined in (1). Our model assumes stationarity, and we have spatial isotropy under transformation (9), so we proceed with the notation $\chi(u; \|h_{\mathrm{S}}\|, h_{\mathrm{T}})$. For a finite number of realisations of $X(s,t)$ and $X(s + h_{\mathrm{S}}, t + h_{\mathrm{T}})$, it is not possible to evaluate $\chi(\|h_{\mathrm{S}}\|, h_{\mathrm{T}}) = \lim_{u\to 1} \chi(u; \|h_{\mathrm{S}}\|, h_{\mathrm{T}})$ itself, so we instead consider an empirical estimate of (1), denoted by $\widehat{\chi}(u; \|h_{\mathrm{S}}\|, h_{\mathrm{T}})$. Suppose we observe the process at times $t = 1, \ldots, n$ at two spatial locations, s and $s + h_{\mathrm{S}}$. We obtain realisations for these two locations where the temporal lag is h_{T}, by considering the pairs of observations

$\{x(s,i), x(s+h_{\mathrm{S}}, i+h_{\mathrm{T}})\}$, for $i = 1, \ldots, n - h_{\mathrm{T}}$. Then,

$$\widehat{\chi}(u; \|h_{\mathrm{S}}\|, h_{\mathrm{T}}) = \frac{\sum_{i=1}^{n-h_{\mathrm{T}}} \mathbb{1}\{x(s+h_{\mathrm{S}}, i+h_{\mathrm{T}}) > q_u, x(s,i) > q_u\}}{\sum_{i=1}^{n-h_{\mathrm{T}}} \mathbb{1}\{x(s,i) > q_u\}},$$

with q_u denoting the quantile of the Laplace distribution corresponding to threshold u. We evaluate $\widehat{\chi}(u; \|h_{\mathrm{S}}\|, h_{\mathrm{T}})$ for observations from each pair of spatial locations, and at increasing time lags. The results are shown by the grey points in Figures 2 and 3 for $u = 0.95$ and $u = 0.975$, respectively; these plots will later be used to assess model fit. As expected, the strongest extremal dependence occurs for pairs of nearby locations and at short time lags. From these plots, it appears that the pairwise spatial dependence becomes stable for time lags greater than three or four. We therefore expect that taking $m \geq 4$ should enable us to sufficiently capture the changing spatio-temporal dependence in our process. For the remainder of the paper we take $m = 5$, corresponding to a maximum time lag of four, but results for other values of m are presented in Section A of the Supplementary Material.

3.2. Inference

To estimate the parameters of our model, we adopt a composite likelihood approach similar to that of Wadsworth and Tawn (2019), where the idea is to allow different space-time locations to act as the conditioning site, and pool information from each one. In our case, these conditioning sites correspond to all combinations of the spatial locations in Figure 1, and the time points in our chosen blocks of length m. With $w^* = (s^*, t)$ denoting locations under transformation (9), the likelihood contribution of conditioning site w_i is

$$\mathcal{L}_i(\boldsymbol{\theta}) = \prod_{\ell=1}^{n_i} \left(f_i^{Z^0} \left[\left\{ \frac{x_j^\ell - a_{w_j^* - w_i^*}(x_i^\ell)}{b_{w_j^* - w_i^*}(x_i^\ell)} \right\}_{j \in \{1,\ldots,dm\} \setminus \{i\}} \right] \cdot \prod_{j \in \{1,\ldots,dm\} \setminus \{i\}} b_{w_j^* - w_i^*}(x_i^\ell)^{-1} \right),$$

for the n_i observations where $x(w_i^*) > u$, where x_j^ℓ denotes the value of the ℓth such observation at space-time location w_j^*, and $f_i^{Z^0}$ represents the density of the residual process $\{Z^0(w^*)\}$ at all locations except w_i^*. Here, $\boldsymbol{\theta}$ represents all model parameters relating to geometric anisotropy, the residual process Z^0 and normalising functions a_{w-w_0} and b_{w-w_0}, and is discussed more explicitly for the models we use in Section 4.2.

The overall likelihood is given by $\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{dm} \mathcal{L}_i(\boldsymbol{\theta})$. This is a composite likelihood since the same observation can be included in the likelihood calculation multiple times, depending on how many space-time locations take values above the threshold u. The parameters $\boldsymbol{\theta}$ can be estimated using standard maximum likelihood techniques, and we take u to be the 0.95 quantile of the marginal Laplace distributions, which results in an average of 73 threshold exceedances per spatial location, or approximately 14 per space-time location. We use `optim`

in R for inference, which works well in general, but can require repeated iterations to ensure convergence is reached. We subsequently drop the s^* and w^* notation, but note that geometric anisotropy is included for the remainder of the paper.

3.3. Simulation given extremes at a single location

Many of the diagnostics we introduce in Section 4 and in the Supplementary Material rely on the simulation of events from our fitted models. To simulate a single spatio-temporal event $\{x(w) : w \in \mathcal{S} \times \mathcal{T}\}$, conditioning on $X(w_0) > v$, for $v \geq u$, i.e., at a level at least as extreme as where we fit the model, the steps are:

1. simulate $e \sim \text{Exp}(1)$, and let $x(w_0) = v + e$;
2. simulate $\{z^0(w) : w \in \mathcal{S} \times \mathcal{T}\}$ from the model for the residual process defined in Section 2.3;
3. set $\{x(w) : w \in \mathcal{S} \times \mathcal{T}\} = a_{w-w_0}\left[x(w_0)\right] + b_{w-w_0}\left[x(w_0)\right] \cdot \{z^0(w)\}$.

We note that the resulting simulations are on the Laplace scale, which is sufficient in our setting as we are particularly interested in the dependence structure of our variables, and in Section 4.6 we will use quantiles to approximate critical levels associated with coral bleaching. One could marginally invert the rank transform to obtain simulations on the original scale.

3.4. Importance sampling

Rather than conditioning on extreme values at a single space-time location, it may be more useful to condition on situations where extremes occur anywhere in some spatio-temporal domain. Wadsworth and Tawn (2019) introduce an importance sampling algorithm to handle this case, which is straightforward to adapt to our spatio-temporal setting. Suppose we are interested in conditioning on $\max_{w \in D} X(w) > v$, for some $v \geq u$, where D denotes a subset of locations in the spatio-temporal domain $\mathcal{S} \times \mathcal{T}$, which may or may not be identical to the locations $\{w_1, \ldots, w_{dm}\}$ used for inference. For some function $g(\cdot)$, the aim is to estimate a quantity of the form

$$\mathbb{E}\left[g\left\{X(w) : w \in D\right\} \,\middle|\, \max_{w \in D} X(w) > v\right]. \tag{10}$$

This is achieved by taking the following steps:

1. sample a location $\tilde{w}$ from the set D, each with probability $|D|^{-1}$;
2. simulate a single observation using the method in Section 3.3, with $\tilde{w}$ as the conditioning site;
3. repeat steps 1 and 2 n times, generating samples $\boldsymbol{X}_{D,1}, \ldots, \boldsymbol{X}_{D,n}$;
4. obtain an estimate of (10) via

$$\frac{\sum_{j=1}^{n}\left[g(\boldsymbol{X}_{D,j}) \cdot \left\{\sum_{x \in \boldsymbol{X}_{D,j}} \mathbb{1}(x > v)\right\}^{-1}\right]}{\sum_{j=1}^{n}\left\{\sum_{x \in \boldsymbol{X}_{D,j}} \mathbb{1}(x > v)\right\}^{-1}}.$$

We demonstrate results using this approach in Section 4.5.

4. Model diagnostics and results

4.1. Pairwise $\chi(u; \|h_{\mathrm{S}}\|, h_{\mathrm{T}})$ in space and time

As a first assessment of how well our model captures the spatio-temporal dependence present in the data, we again consider the measure $\chi(u; \|h_{\mathrm{S}}\|, h_{\mathrm{T}})$ discussed in Section 3.1. In Figures 2 and 3, we compare estimates of $\chi(u; \|h_{\mathrm{S}}\|, h_{\mathrm{T}})$ calculated from the data, to estimates obtained from 10,000 simulations from our fitted model, with both the separable model (4) and non-separable model (5) for $a_{w-w_0}(x)$, and $u = 0.95, 0.975$. The spatial distances are calculated using the Euclidean distance on locations under transformation (9). We demonstrate this transformation in Section B of the Supplementary Material.

We see from Figures 2 and 3 that the two models perform similarly well overall. The main difference is the performance at short spatial distances as the time lag increases, where the separable model overestimates the extremal dependence, and the non-separable model performs better. Due to this advantage, we use the non-separable model for the remainder of the paper.

4.2. Diagnostic based on $\{Z^0(w)\}$

After anisotropy parameters, the remaining parameters in our model can be separated into those related to the normalising functions $a_{w-w_0}(\cdot)$ and $b_{w-w_0}(\cdot)$, and those from the model of the residual process $\{Z^0(w)\}$. Under the non-separable model, the former is the set $\boldsymbol{\theta}_{\mathrm{norm}} = (\lambda_{\mathrm{S}}, \kappa_{\mathrm{S}}, \lambda_{\mathrm{T}}, \kappa_{\mathrm{T}}, \eta, \beta)$ and the latter is $\boldsymbol{\theta}_{\mathrm{res}} = (\mu, \sigma, \phi_{\mathrm{S}}, p_{\mathrm{S}}, \phi_{\mathrm{T}}, p_{\mathrm{T}})$. We obtain estimates of these parameters through maximising the composite likelihood discussed in Section 3.2, and denote these by $\widehat{\boldsymbol{\theta}}_{\mathrm{norm}}$ and $\widehat{\boldsymbol{\theta}}_{\mathrm{res}}$, respectively.

Empirical realisations of the residual process can be obtained by normalising the observed data, transformed to Laplace margins, using functions (5) and (6) with parameters $\widehat{\boldsymbol{\theta}}_{\mathrm{norm}}$, and conditioning on the observation at site w_0 being above the modelling threshold u. For this same choice of conditioning site, we can simulate directly from the multivariate Gaussian distribution with mean vector and covariance matrix in (8) determined by parameter values $\widehat{\boldsymbol{\theta}}_{\mathrm{res}}$, in order to obtain realisations of our fitted residual process. If the model fits well, the empirical and fitted residuals should exhibit similar behaviour.

For the model fitted using block-length $m = 5$, Figure 4 demonstrates the behaviour of the empirical and fitted residuals across a diagonal spatial transect, for three different conditioning sites. This transect was chosen as it includes some of the longest spatial ranges present in our data. We take u to be the 0.95 quantile of the observations in Laplace margins. The number of simulations from the fitted residual process was taken to be the same as the number of empirical observations. The diagnostic plots demonstrate reasonable agreement between the empirical and fitted residuals.

Figure 4 represents a purely spatial diagnostic. In Section C of the Supplementary Material, we present similar results to assess the temporal aspect of our model. These diagnostic plots also indicate a successful model fit.

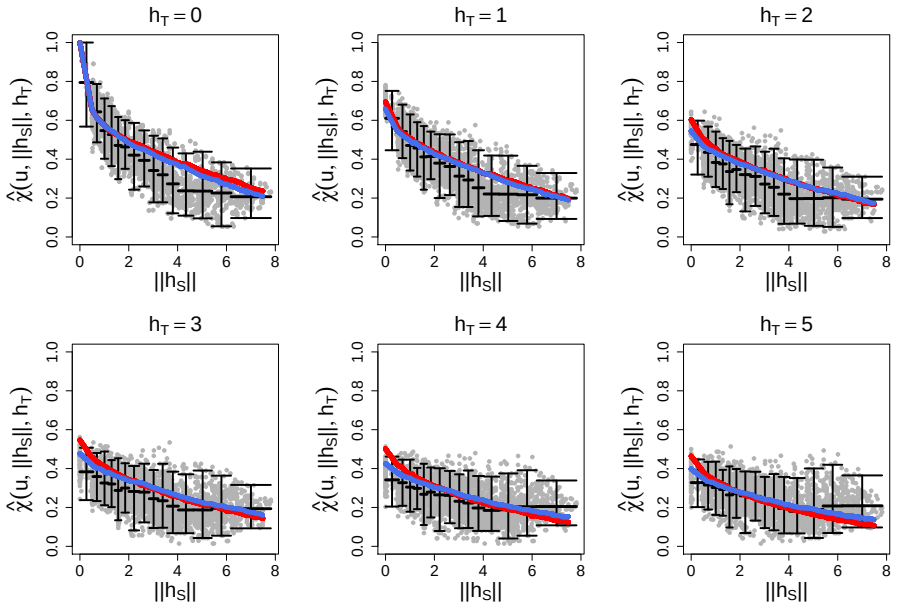

Figure 2: Pairwise estimates of $\chi(0.95; \|h_{\mathrm{S}}\|, h_{\mathrm{T}})$ for the Red Sea data (grey), with the mean and 0.025 and 0.975 quantiles for sections of the distances (black). The solid lines show $\chi(0.95; \|h_{\mathrm{S}}\|, h_{\mathrm{T}})$ estimates for our fitted models: separable (red); non-separable (blue). The distances $\|h_{\mathrm{S}}\|$ are measured in the transformed space.

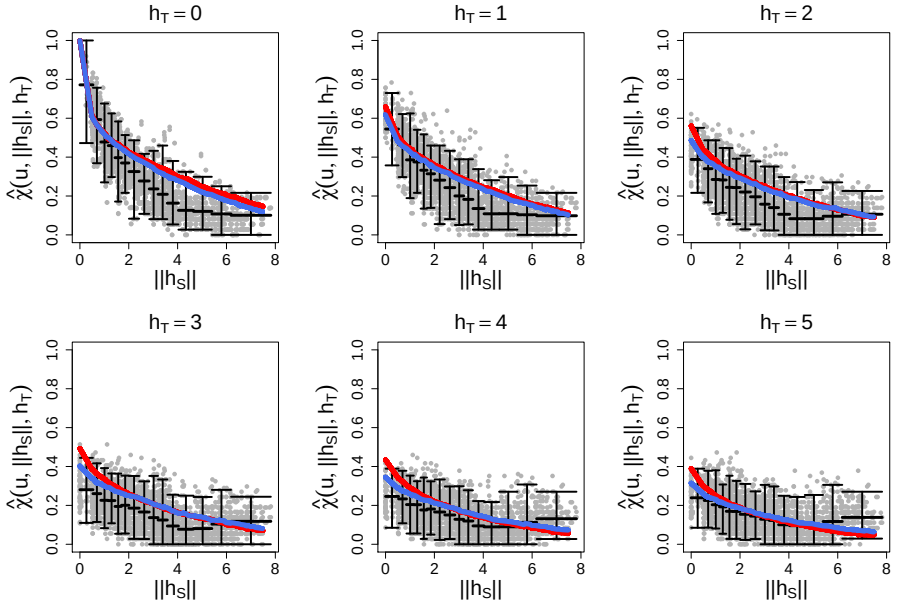

Figure 3: Pairwise estimates of $\chi(0.975; \|h_{\mathrm{S}}\|, h_{\mathrm{T}})$ for the Red Sea data (grey), with the mean and 0.025 and 0.975 quantiles for sections of the distances (black). The solid lines show $\chi(0.975; \|h_{\mathrm{S}}\|, h_{\mathrm{T}})$ estimates for our fitted models: separable (red); non-separable (blue). The distances $\|h_{\mathrm{S}}\|$ are measured in the transformed space.

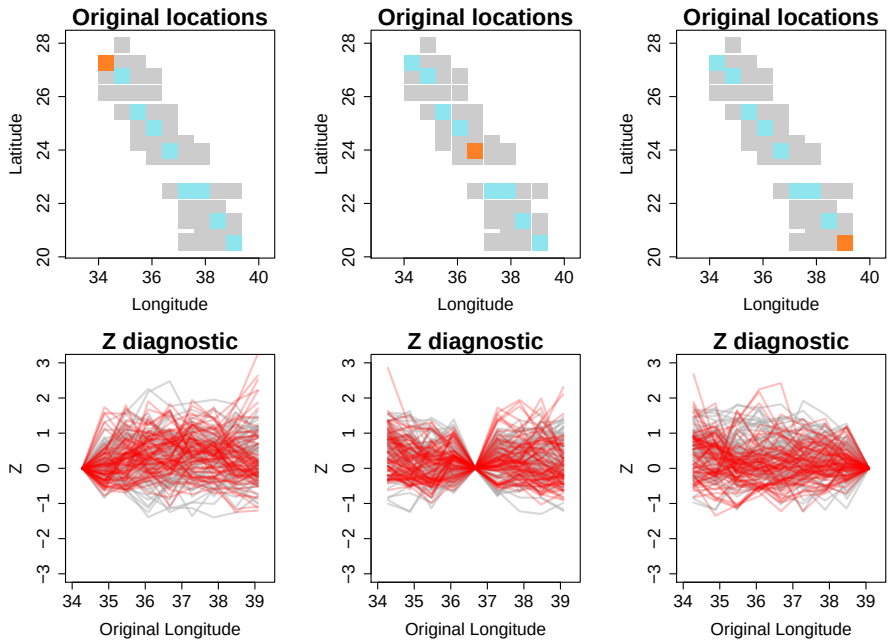

Figure 4: Top row: locations of the conditioning sites (orange) and spatial transect (blue); bottom row: comparison of residuals simulated from the fitted model (grey) and empirical residuals (red) across space.

4.3. Diagnostic based on $\{X(w)\}$

Another useful diagnostic is to consider simulations from our fitted model, obtained via the method outlined in Section 3.3, conditioning on extreme values at some conditioning site $w_0 = (s_0, 0)$. We can then consider the propagation of these events forwards and backwards in time, and compare the simulated events to observations from the data on Laplace margins. In Figure 5, we demonstrate this technique for the conditioning site and spatial transect in the top-central panel of Figure 4. In these plots, we also see reasonably similar behaviour between the simulated events and observed data. We note that the model appears to have captured the behaviour of the process both forwards and backwards in time from the conditioning site, indicating that our choice to use a temporally-symmetric model for this data is reasonable.

4.4. Evaluating parameter uncertainty

Due to our use of a composite likelihood, it is necessary to evaluate the uncertainty in our parameter estimates, for which we use the bootstrap (Efron, 1979). For a time series of length n, Künsch (1989) proposes creating blocks of length b containing consecutive observations, i.e., starting at observation $1, 2, \ldots, n + 1 - b$. One can then randomly sample n/b of these blocks, with

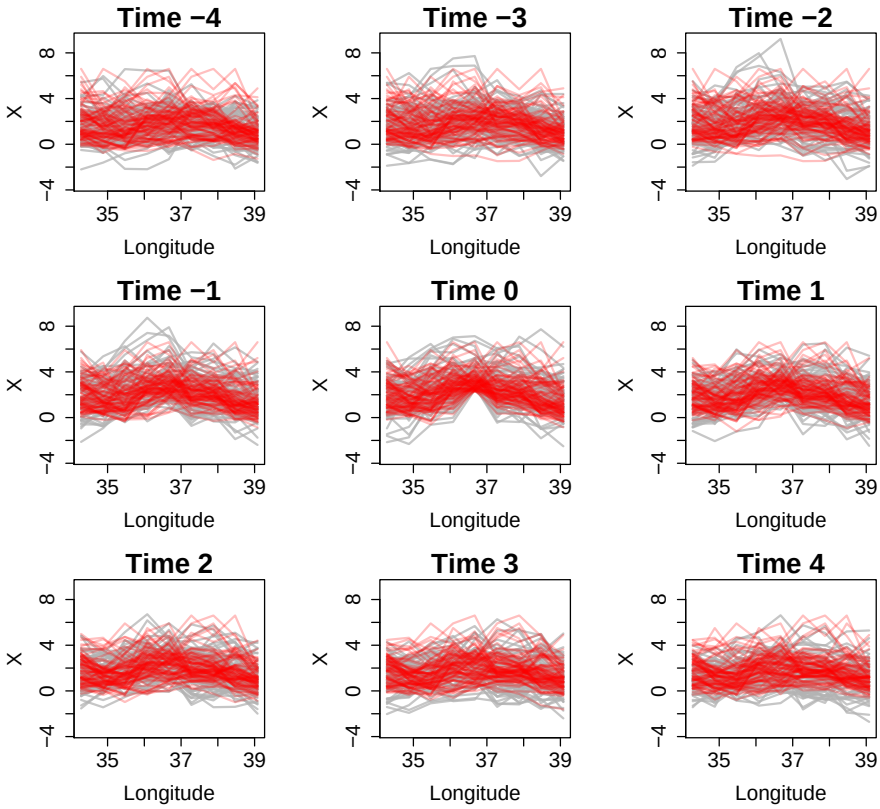

Figure 5: Comparison of simulations from the fitted model (grey) and observed data (red) on Laplace scale at different time lags. The conditioning site and spatial transect correspond to those in the top-central panel of Figure 4

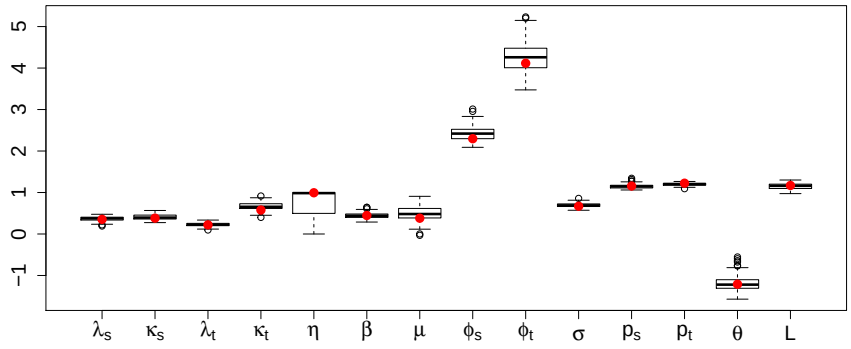

Figure 6: Boxplots of bootstrapped parameter estimates for 100 bootstrapped samples, with the estimated parameters values for the original data shown in red.

replacement, to obtain each bootstrapped sample. We take this approach, but remove blocks that span more than one summer to ensure each block contains consecutive observations. Since we separate the data into blocks of length m to carry out our likelihood inference, it is possible that not all the temporal dependence in the data has been captured in our model. We can account for this issue by taking a block-size $b > m$ for the bootstrapped samples.

A drawback of the block bootstrap approach is that it can lead to non-stationarity in the bootstrapped samples, even if the original observations exhibit stationarity. Politis and Romano (1994) propose a stationary bootstrapping approach to overcome this issue, where the first observation in each block is still sampled uniformly across all observations, but the lengths of the blocks are randomly sampled from a $Geom(p)$ distribution. The use of random block-lengths does not guarantee that the blocks, of fixed length m, that we subsequently create for our inferential approach contain consecutive observations, so we continue with the block bootstrap approach.

As in the preceding investigations, we fix the block-length used to carry out our inference as $m = 5$. Then each summer contains 90 observations, and we have $n = 90 \times 16 = 1440$ observations overall. We take a block-length of $b = 20$ for the bootstrap, with 72 blocks making up each of our bootstrapped samples. We fit the non-separable version of our model to 100 such samples, with boxplots of the resulting parameter estimates shown in Figure 6. The majority of the estimated values of the parameter η in equation (5) are equal or very close to one, indicating strong interaction between space and time, and that the non-separable model is a reasonable choice.

4.5. Conditioning on an extreme at any location

We now consider quantities of the form (10), estimated via the importance sampling algorithm discussed in Section 3.4. This method can be applied for any set of space-time locations within the domain $\mathcal{S} \times \mathcal{T}$. In this section, we focus on the 54 locations where we have observations, at five time-points, corresponding to our chosen block-length m, i.e., a total of 270 locations $w_1, \ldots, w_{270}$.

In Figure 7, we consider results for two different functions g. The left panel corresponds to the expected number of space-time locations that exceed the threshold v, given that $\max_{i=1,\ldots,270} X(w_i) > v$. The right panel shows results based on the average temperature over all sites, on the Laplace scale, under the same conditioning event. We simulate $n = 10,000$ importance samples for each threshold. To demonstrate uncertainty in the estimates, we present results obtained using the bootstrapped parameter estimates in Figure 6, and compare these to empirical estimates obtained using the data.

The results in Figure 7 reveal that our model overestimates the number of simultaneously extreme locations, particularly at lower thresholds, as well as the average temperature across all locations. These discrepancies may be explained by the results in Figures 2 and 3; for short time-lags, we slightly overestimate the pairwise extremal dependence, the effect of which could have been exacerbated when considering results for all locations simultaneously. However, the empirical

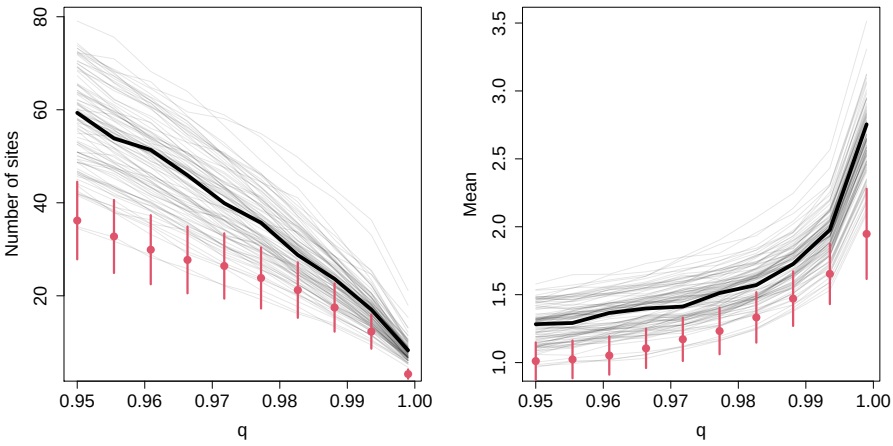

Figure 7: Importance sampling results for the number of sites exceeding a threshold (left) and the average value of the temperatures on the Laplace scale (right), given an exceedance anywhere in the domain. The threshold v is given in terms of quantiles of the Laplace distribution. The solid black lines show results using our parameter estimates, and the grey lines show the bootstrapped results. The red circles provide empirical estimates from the data with 95% confidence intervals.

estimates are within the range of our bootstrapped calculations in both cases, so taking uncertainty into account, our model appears to perform adequately.

4.6. Investigation into the risk of coral bleaching

Sea surface temperatures can be used as an indicator of potentially damaging conditions for coral life. The average sea surface temperature at different coral reefs varies depending on geographic location, with different varieties of coral adapting to their different habitats. Therefore, there is no single, high temperature threshold to which coral bleaching can be attributed. While Jokiel and Brown (2004) collate information about temperature thresholds associated with this phenomenon for a range of locations, none of these are particularly close to the northern Red Sea. Rather than using one fixed threshold to explain coral bleaching at different locations, Genevier et al. (2019) propose using high quantiles that vary with space (an approach adopted by Hazra and Huser (2020)), and time. Goreau and Hayes (2005) suggest that coral bleaching may occur if temperatures persistently exceed a level corresponding to one degree above the average temperature in the warmest month, which for our data is August. In Section D of the Supplementary Material, we show that this is well approximated by taking the 0.961 quantile of summer sea surface temperatures for each location we study, so this is the level on which we focus our study.

Coral reefs are located around much of the coast of the northern Red Sea, with those in the north-west generating particularly high revenue from tourism (Fine et al., 2019). Although coral bleaching is currently less of an issue in

the northern Red Sea than the south, its potential impact on tourism means studying extreme sea surface temperatures in this region may still be of interest, and we use this example to demonstrate how our model could be employed to investigate the probability of high sea surface temperature events for other locations that may be more at risk of coral bleaching.

When modelling extreme events through time at a single location, it is usual to decluster the observations into sets of days that correspond to the same event. In practice, we achieve this declustering using the runs method of Smith and Weissman (1994), where clusters of observations containing exceedances above a high threshold v are separated by r consecutive days of non-exceedances. We note that the method of Ferro and Segers (2003) offers an alternative to the approach of Smith and Weissman (1994), with the desirable feature that the declustering is automated, and this could also be used here. Of interest may be the average length of such clusters, but in studying the effect of high temperatures this may not be the most important factor. In their heatwave application, one alternative proposed by Winter and Tawn (2016) is to consider the maximum number of consecutive exceedances within a cluster, since these events can have a large impact in practice. We follow this approach, aiming to estimate the number of clusters per year containing a maximum of at least n consecutive exceedances. We first consider events at any single location in the northern Red Sea, and then extend the approach to consider simultaneous exceedances at a set of spatial locations.

Since corals particularly grow in shallow water, we focus on locations where the water depth is less than 150m (Genevier et al., 2019), shown in the left panel of Figure 8 using bathymetry data from the GEBCO Compilation Group (2019). Winter and Tawn (2016) discuss the importance of *within-cluster* and *over-cluster* results; in our single location setting, the former is the probability that the maximum number of consecutive exceedances in a cluster is at least n, while the latter is the expected number of clusters in a single year. Assuming independence of clusters, we can multiply together these two values to obtain our estimate.

To study the expected number of clusters per year, we apply the runs method of Smith and Weissman (1994), with parameter $r = 10$, at each of the 54 locations, and take empirical values. Due to our assumption of spatial stationarity, we take the average over all these locations to obtain our estimate. We apply this same approach for other values of r in Section E of the Supplementary Material; $r = 10$ corresponds to the value above which the estimated number of clusters stabilises, suggesting independence between clusters. We estimate the distribution of the maximum number of consecutive exceedances in a cluster via simulation, based on our fitted model. To simulate single events, Smith et al. (1997) suggest conditioning on the maximum value within a cluster being above the threshold v, and simulating forwards and backwards in time. This can be achieved by applying the simulation method in Section 3.3 within a rejection sampling routine, with rejection if the maximum value does not occur at the conditioning site; see Winter and Tawn (2016). We obtain 250,000 observations using this technique, and take empirical estimates for the required distribution,

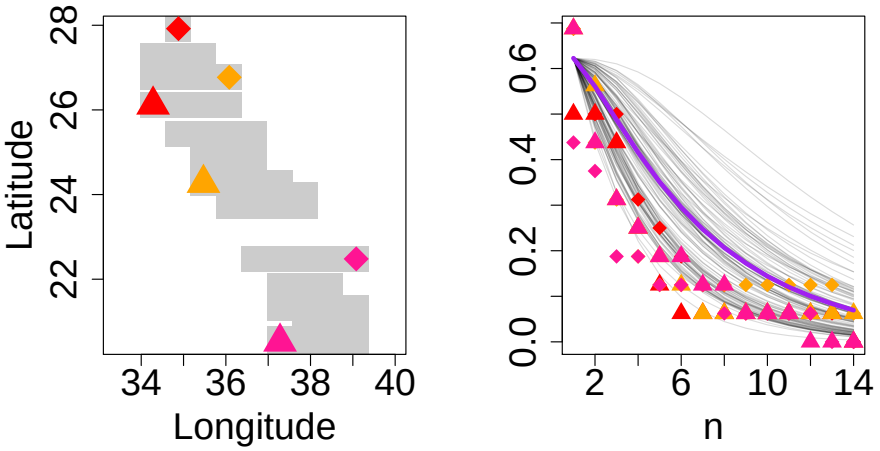

Figure 8: Left: six locations where the water depth is less than 150m and coral has the potential to grow. Right: expected number of clusters per year where the maximum number of consecutive exceedances of the 0.961 quantile at a single location is at least n, estimated using the fitted model parameters (purple) and bootstrapped parameter estimates (grey); empirical results corresponding to the highlighted locations in the left panel are also shown.

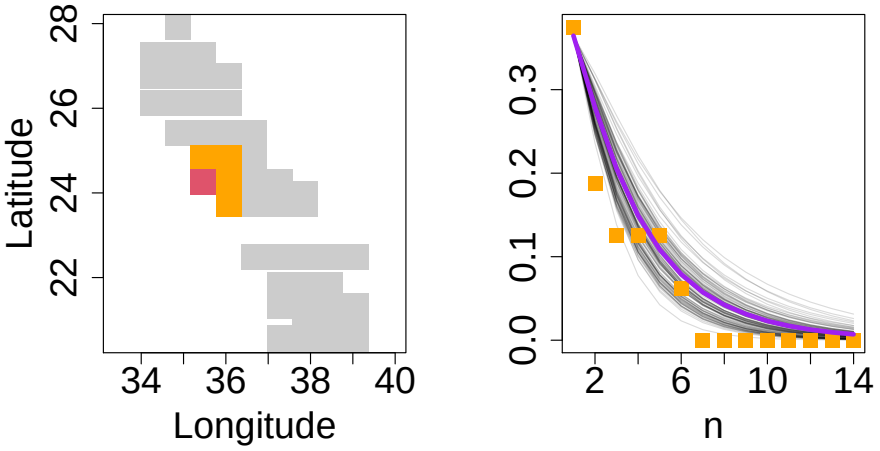

Figure 9: Left: one of the shallow-water locations (red) and its neighbours (orange). Right: expected number of clusters per year where the maximum number of consecutive days with exceedances of the 0.961 quantile at all of these locations is at least n, estimated using the fitted model parameters (purple) and bootstrapped parameter estimates (grey); empirical estimates are shown in orange.

ensuring that the conditioning site is within the contributing cluster, and that each cluster begins and ends with $r = 10$ non-exceedances of the threshold. This is repeated using each set of bootstrapped parameter estimates. The results are shown in the right panel of Figure 8. Here, we also present empirical estimates for the six shallow-water locations, and our model appears to give reasonable results. Further, at four of these six locations, the data does not contain a block of consecutive exceedances longer than 13 days, i.e., our model enables us to consider the expected occurrence of unobserved events.

Genevier et al. (2019) consider events where the sea surface temperature is above some critical level for seven consecutive days. For any single location in the northern Red Sea, we estimate the expected number of clusters with a maximum of at least seven consecutive exceedances to be 0.247 per year. This corresponds to a return period of 4.05 years with a 95% confidence interval of $(2.28, 9.10)$ years.

Since extreme surface temperatures occurring over larger spatial domains could have more impact on marine life, for one of the locations in Figure 8 taken as s_0, we consider simultaneous extremes at this and its neighbouring sites, which correspond to locations s where $\|s - s_0\| < 1$ in the transformed coordinates; these sites are shown in the left panel of Figure 9. We now define a cluster as a series of days where all these sites exceed their 0.961 quantile simultaneously, separated by $r = 10$ days where this is not the case, and are interested in the maximum number of consecutive, simultaneous exceedances in a cluster. The estimate of the number of clusters is again obtained empirically, but we now only average over those locations with four, five or six neighbours within a distance of 1, to give a reasonable approximation for the location of interest.

To estimate the distribution of the maximum number of consecutive exceedances within a cluster, we note that our definition of a cluster means we must condition on there being an exceedance at all locations within the cluster, rather than at a single location as in the previous case. We employ a similar rejection sampling technique as previously, simulating forwards and backwards conditioning on $X(w_0) > v$, with the criteria for rejection being that at least one site is below the threshold v, and the maximum within the cluster does not occur at the conditioning time.

Results are shown in the right panel of Figure 9; we now have no observations of a maximum number of consecutive exceedances greater than six, but are able to extrapolate beyond this using our model. The expected number of clusters per year where all locations exceed their 0.961 quantiles for a maximum of at least seven consecutive days is estimated to be 0.057. This corresponds to a return period of 17.4 years with a 95% confidence interval of $(9.5, 41.8)$ years.

5. Discussion

In this paper, we presented an approach to modelling extreme events over space and time, by extending the conditional spatial extremes model of Wadsworth

and Tawn (2019) to a spatio-temporal setting. The model is constructed by conditioning on exceedances above a threshold at a single location, with inference carried out via a composite likelihood approach, allowing for contributions from different conditioning sites. We used this approach to model sea surface temperature extremes in the north of the Red Sea, and proposed a range of diagnostic techniques, showing that the data were well described by the model. The resulting model fit was used to demonstrate how one could assess the risk of coral bleaching, by estimating the return period of clusters where sea surface temperatures exceed a high threshold over consecutive days, across one or several spatial locations.

One issue that could be considered further is the differing behaviour observed in the north and south of the Red Sea. We chose to concentrate on modelling surface temperatures only in the north to simplify our approach, but it would be useful to have techniques available to model the full set of locations simultaneously. One way to deal with the non-stationarity in the data may be to allow the model parameters to depend on the spatial conditioning site in some way. We considered allowing the parameters λ_{S} or κ_{S} to depend on the spatial coordinates of the conditioning site, and separately tried to include water depth as a covariate in the parameter λ_{S}. Although these attempts did not sufficiently improve our model fit, it is possible that a different covariate could have explained some of the spatial non-stationarity present in the data, and covariate modelling as a technique may be successful in other applications.

A final aspect that could benefit from further attention in the future is that of threshold selection. This is a topic of notorious difficultly within extreme value statistics, with the univariate case seeing a particularly large amount of research; see for instance Scarrott and MacDonald (2012) or Northrop et al. (2017). The problem lies in choosing a threshold low enough that there is sufficient data to carry out reliable inference, but high enough that the asymptotic assumptions of the model hold. A common technique is to use plots to assess the stability of parameter estimates, or some other measure related to the data, but these plots can be difficult to interpret, and the resulting threshold choice subjective. Our choice to use a fixed quantile in the present paper has the benefit of simplicity, but we acknowledge that, as with all extremal modelling, there may be some sensitivity to this choice.

Acknowledgements

This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-CRG2017-3434. We thank the two anonymous referees for their helpful comments.

References

Buhl, S., Davis, R. A., Klüppelberg, C., and Steinkohl, C. (2019). Semiparametric estimation for isotropic max-stable space-time processes. *Bernoulli*,

25(4A):2508–2537.

Castro-Camilo, D., Mhalla, L., and Opitz, T. (2020). Bayesian space-time gap filling for inference on extreme hot-spots: an application to Red Sea surface temperatures. *Extremes (to appear)*.

Cressie, N. and Huang, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94:1330–1340.

Davis, R. A., Klüppelberg, C., and Steinkohl, C. (2013). Statistical inference for max-stable processes in space and time. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 75(5):791–819.

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, 27(2):161–186.

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.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Ferro, C. A. T. and Segers, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65(2):545–556.

Fine, M., Cinar, M., Voolstra, C. R., Safa, A., Rinkevich, B., Laffoley, D., Hilmi, N., and Allemand, D. (2019). Coral reefs of the Red Sea - challenges and potential solutions. *Regional Studies in Marine Science*, 25:100498.

GEBCO Compilation Group (2019). GEBCO 2019 Grid. `doi:10.5285/836f016a-33be-6ddc-e053-6c86abc0788e`.

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. *Global Change Biology*, 25(7):2338–2351.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97:590–600.

Goreau, T. J. F. and Hayes, R. L. (2005). Global coral reef bleaching and sea surface temperature trends from satellite-derived hotspot analysis. *World Resource Review*, 17(2):254–293.

Hazra, A. and Huser, R. (2020). Estimating high-resolution Red Sea surface temperature hotspots, using a low-rank semiparametric spatial model. *arXiv:1912.05657*.

Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17(2):537–571.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66(3):497–546.

Huser, R. (2020). Editorial: EVA 2019 data competition on spatio-temporal prediction of Red Sea surface temperature extremes. *Extremes (to appear)*.

Huser, R. and Davison, A. C. (2014). Space-time modelling of extreme events. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 76(2):439–461.

Huser, R. and Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, 114:434–444.

Ivek, T. and Vlah, D. (2020). BlackBox: Generalizable reconstruction of extremal values from incomplete spatio-temporal data. *arXiv:2005.02140*.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81:158–171.

Jokiel, P. L. and Brown, E. K. (2004). Global warming, regional trends and inshore environmental conditions influence coral bleaching in Hawaii. *Global Change Biology*, 10:1627–1641.

Keef, C., Tawn, J. A., and Lamb, R. (2013). Estimating the probability of widespread flood events. *Environmetrics*, 24(1):13–21.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241.

Maynard, J. A., Turner, P. J., Anthony, K. R. N., Baird, A. H., Berkelmans, R., Eakin, C. M., Johnson, J., Marshall, P. A., Packer, G. R., Rea, A., and Willis, B. L. (2008). ReefTemp: An interactive monitoring system for coral bleaching using high-resolution SST and improved stress predictors. *Geophysical Research Letters*, 35(5):1–5.

McClanahan, T. R., Ateweberhan, M., Muhando, C. A., Maina, J., and Mohammed, M. S. (2007). Effects of climate and seawater temperature variation on coral bleaching and mortality. *Ecological Monographs*, 77(4):503–525.

Morris, S. A., Reich, B. J., Thibaud, E., and Cooley, D. (2017). A space-time skew-t model for threshold exceedances. *Biometrics*, 73(3):749–758.

Northrop, P. J., Attalides, N., and Jonathan, P. (2017). Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(1):93–120.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89:1303–1313.

Scarrott, C. J. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT - Statistical Journal*, 10(1):33–60.

Smith, R. L., Tawn, J. A., and Coles, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika*, 84(2):249–268.

Smith, R. L. and Weissman, I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(3):515–528.

von Mises, R. (1936). La distribution de la plus grande de n valeurs. *Rev. Math. Union Interbalcanique*, 1:141–160.

Wadsworth, J. L. and Tawn, J. A. (2019). Higher-dimensional spatial extremes via single-site conditioning. *arXiv:1912.06560*.

Winter, H. C. and Tawn, J. A. (2016). Modelling heatwaves in central France: a case-study in extremal dependence. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(3):345–365.