

# Patterns in Spatio-Temporal Extremes

Marco Oesting<sup>1</sup> and Raphaël Huser<sup>2</sup>

December 22, 2022

## Abstract

In environmental science applications, extreme events frequently exhibit a complex spatio-temporal structure, which is difficult to describe flexibly and estimate in a computationally efficient way using state-of-art parametric extreme-value models. In this paper, we propose a computationally-cheap non-parametric approach to investigate the probability distribution of temporal clusters of spatial extremes, and study within-cluster patterns with respect to various characteristics. These include risk functionals describing the overall event magnitude, spatial risk measures such as the size of the affected area, and measures representing the location of the extreme event. Under the framework of functional regular variation, we verify the existence of the corresponding limit distributions as the considered events become increasingly extreme. Furthermore, we develop non-parametric estimators for the limiting expressions of interest and show their asymptotic normality under appropriate mixing conditions. Uncertainty is assessed using a multiplier block bootstrap. The finite-sample behavior of our estimators and the bootstrap scheme is demonstrated in a spatio-temporal simulated example. Our methodology is then applied to study the spatio-temporal dependence structure of high-dimensional sea surface temperature data for the southern Red Sea. Our analysis reveals new insights into the temporal persistence, and the complex hydrodynamic patterns of extreme sea temperature events in this region.

**Keywords:** Cluster of extremes; Extreme event; Functional time series; Ordinal pattern; Regular variation; Sea surface temperature data; Spatio-temporal extremes

---

<sup>1</sup>Stuttgart Center for Simulation Science (SC SimTech) & Institute for Stochastics and Applications, University of Stuttgart, 70569 Stuttgart, Germany, E-mail: marco.oesting@mathematik.uni-stuttgart.de

<sup>2</sup>Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. E-mail: raphael.huser@kaust.edu.sa

# 1 Introduction

Many environmental extreme events, such as heavy precipitation [Castro-Camilo and Huser, 2020], heat waves [Zhong et al., 2022], or storms [Oesting et al., 2017], have a complex spatio-temporal structure that has a direct influence on their statistical properties. With appropriate statistical methods, this structure can be revealed by the use of spatio-temporal data, which have become available in increasingly higher resolutions due to ongoing technical progress. At the same time, the relevant events of interest are by definition rare, so observations usually contain only a few (if any) individual extreme events, which makes traditional empirical approaches inefficient (if useful at all). An alternative more reliable approach for tail extrapolation is to use specialized parametric models backed up by extreme-value theory. Such models include max-stable processes [de Haan, 1984, Davison et al., 2012, Davison and Huser, 2015, Davison et al., 2019],  $r$ -Pareto processes [Ferreira and de Haan, 2014, Dombry and Ribatet, 2015, Thibaud and Opitz, 2015, de Fondeville and Davison, 2018], or more recent processes that possess higher flexibility at sub-asymptotic levels [Wadsworth and Tawn, 2012, Huser and Wadsworth, 2019, Wadsworth and Tawn, 2022]; see Huser and Wadsworth [2022] for a recent review of these parametric modeling approaches. However, one faces two major challenges with such parametric approaches: it is both difficult to develop (i) models that are relatively parsimonious and yet flexible enough to capture complex tail dependence structures, and (ii) fast inference methods to fit these models efficiently with large datasets.

With regard to challenge (i), the extreme-value statistics community has so far primarily focused on developing either purely temporal or purely spatial models. Only in recent years, truly spatio-temporal models have been emerging [see, e.g., Davis et al., 2013, Huser and Davison, 2014, de Fondeville and Davison, 2022, Simpson et al., 2022], but they are often restricted to be stationary, spatially isotropic, space-time separable, or subject to unrealistic dependence assumptions [see, e.g., Huser and Wadsworth, 2022, Hazra et al., 2021]. In most

cases, the resulting models mainly aim at representing the marginal distributions and the pairwise dependencies with maximum accuracy, often disregarding higher-order dependence interactions. Certain crucial characteristic features of the events caused by the underlying physical processes, such as a storm growing, moving over an area, and weakening, are indeed very difficult to describe flexibly using parametric models.

With regard to challenge (ii), such spatio-temporal parametric extreme-value models are often computationally intensive to fit in high dimensions, especially when they are based on max-stable processes [Padoan et al., 2010, Castruccio et al., 2016, Huser et al., 2019]. Recent progress has been made to fit  $r$ -Pareto processes efficiently using a gradient-scoring approach [de Fondeville and Davison, 2018], and to make fast inference for the conditional extremes model using R-INLA [Simpson et al., 2022]. Amortized inference approaches based on neural Bayes estimators have also been recently advocated to fit max-stable and other extreme-value processes efficiently [Sainsbury-Dale et al., 2022]. Nevertheless, developing truly flexible spatio-temporal models that can reliably capture a variety of joint tail events of interest and within-cluster behaviors, and that can be fitted in truly high dimensions (e.g., of the order of  $5000 \times 11000$  dependent spatio-temporal locations as in our data application in Section 7) remains currently inaccessible.

In this paper, instead of relying on parametric models, we propose an alternative and computationally-cheap non-parametric approach that relies on mild standard assumptions about the regularity of the joint tail (important for reliable tail extrapolation), and which allows us to efficiently answer a wide variety of key scientific questions, such as

(Q1) How long does an extreme event last?

(Q2) Does the magnitude of the event increase/decrease within its course?

(Q3) Does the spatial area affected by the event shrink/grow?

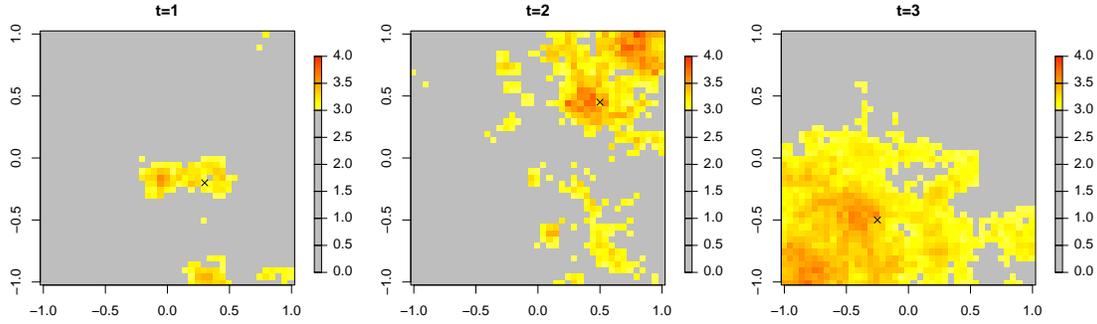


Figure 1: An extreme event as a realization of a spatio-temporal random field at three instances of time ( $t = 1$ ,  $t = 2$  and  $t = 3$ ; from left to right). The exceedances of the critical threshold  $u$  (here  $u = 3$ ) are displayed in colors. The black crosses mark the componentwise median of the area affected by the extreme event.

(Q4) Does the “center” of the extreme event move in a certain way?

While these questions are quite basic, the answers may be not straightforward, even through appealing visual plots. To illustrate this, Figure 1 displays exceedances of a critical high threshold in three consecutive time points, based on some simulated data. Thus, one might say that this is an extreme event of duration 3 or more, which provides a partial answer to (Q1). Furthermore, it can be seen that the size of the area affected by the extreme event is progressively growing in the displayed period (Q3), while, as the black crosses indicate, the “center” of this event moves first from the center to the northeast and further in the opposite direction to the southwest (Q4). The answer to question (Q2) requires a closer look at the actual values of the magnitude, e.g., the spatial maximum, indicated by the darkest red color in the figure, which increases from  $t = 1$  to  $t = 2$  and decreases thereafter.

More careful probabilistic answers to any of the questions raised above requires a more precise definition of extreme events. Here, we employ the idea of clusters of extremes, for which various definitions already exist in the context of univariate time series [see e.g., Ferro and Segers, 2003, Markovich, 2014], and we extend them to the functional time series setting. As the above discussion also demonstrates, the zig-zag behavior of certain characteristics of extremes are of interest to address questions such as (Q2), (Q3) and (Q4). We shall describe

such a zig-zag behavior by using ordinal patterns, a popular concept in time series analysis which has found applications in various areas [see Bandt and Pompe, 2002, Keller et al., 2007, Sinn et al., 2013, for instance]. Here, we generalize this tool to describe certain features of spatio-temporal extreme events. Our proposed methodology extends the recent approach of Oesting and Schnurr [2020] from univariate time series to process-valued time series.

As hinted above, we need assumptions on the space-time process that allow us to reliably extrapolate beyond the range of the observed data. Our main assumption, which is quite typical in extreme-value statistics, is that the spatio-temporal process of interest,  $X$ , is regularly varying as a time series. As we do not only consider the process itself, but also functionals applied to the time series, we need to introduce a refined and more flexible notion of regular variation that can be applied in a more general setting. This notion is introduced in Section 2. In Section 3, this concept is then used to define extremes and clusters of extremes in a functional setting and to deduce limit expressions for the cluster size distribution. The within-cluster behavior, in particular with respect to risk functionals, spatial risk measures, and location measures, is further analyzed in Section 4. In Section 5, we then introduce specialized ratio estimators for the corresponding limit distributions, and show that they are asymptotically normal under appropriate mixing and anti-clustering conditions. In Section 6, the finite-sample behavior of our proposed ratio estimators is illustrated in a spatio-temporal simulated example. The estimators are then used in Section 7 to analyze the complex extremal behavior of sea surface temperature in the southern Red Sea. We finally conclude with some discussion and perspective on future research in Section 8.

## 2 Regular variation of functional time series

In the following, we consider a nonnegative stochastic process  $\{X(\mathbf{s}, t)\}_{\mathbf{s} \in S, t \in \mathbb{Z}}$  on a spatio-temporal domain  $S \times \mathbb{Z}$  where the spatial domain  $S \subset \mathbb{R}^d$  is compact and the temporal

domain  $\mathbb{Z}$  is discrete. We assume that, for each time  $t \in \mathbb{Z}$ , the process  $\{X(\mathbf{s}, t)\}_{\mathbf{s} \in S}$  possesses continuous sample paths, denoted by  $\{X(\mathbf{s}, t)\}_{\mathbf{s} \in S} \in C_+(S)$ , i.e., the time series is  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  is  $C_+(S)$ -valued. For simplicity, we will assume that the time series is stationary, i.e., the space-time process  $X$  is stationary in time (but not necessarily in space).

Our proposed extreme-value methodology relies on the more general concept of regular variation for functional time series. Here, we follow Hult and Lindskog [2006] and Dombry et al. [2018] in considering regular variation in a rather broad setting for stationary time series on a metric space. More precisely, we shall study stationary time series  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  that take values in some complete separable metric space  $E$ , but for simplicity, we will restrict ourselves to the following cases:

- $E = [0, \infty)$ , i.e., we have a time series of random variables,
- $E = [0, \infty)^d$ , i.e., we have a time series of random vectors,
- $E = C_+(S)$ , i.e., we have a time series of nonnegative sample-continuous stochastic processes on some spatial domain such as mentioned above,

equipped with the metric  $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_\infty$  for all  $\mathbf{x}, \mathbf{y} \in E$ . As these spaces satisfy all the conditions listed in Dombry et al. [2018], regular variation of an  $E$ -valued time series can be defined in the following way.

**Definition 2.1.** An  $E$ -valued stationary time series  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  is *regularly varying*, if, for any finite set  $T \subset \mathbb{Z}$ , the random vector  $(\mathbf{X}_t)_{t \in T}$  is regularly varying on  $E^{|T|}$ , i.e., if there exists a nondecreasing sequence  $\{a_n\}_{n \in \mathbb{N}}$  tending to  $\infty$  and a non-zero Radon measure  $\mu_T$  on  $E^T \setminus \{\mathbf{0}\}$ , the *exponent measure*, such that

$$n \mathbb{P}((a_n^{-1} \mathbf{X}_t)_{t \in T} \in A) \xrightarrow{n \rightarrow \infty} \mu_T(A)$$

for any measurable set  $A \subset E^T \setminus \{\mathbf{0}\}$  bounded away from  $\mathbf{0}$  such that  $\mu_T(\partial A) = 0$ .

It can be shown that the sequence  $\{a_n\}_{n \in \mathbb{N}}$  is necessarily regularly varying with some index  $1/\alpha > 0$ , i.e.,  $a_{[nt]}/a_n \rightarrow t^{1/\alpha}$  for all  $t > 0$  as  $n \rightarrow \infty$ , and that the measure  $\mu_T$  is positively homogeneous of order  $-\alpha$  where  $\alpha$  is the same for all finite  $T \subset \mathbb{Z}$ . Thus,  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  is also called regularly varying with index  $\alpha$ . Note that also the sequence  $\{a_n\}_{n \in \mathbb{N}}$  can be chosen independently of the index set  $T \subset \mathbb{Z}$ , for instance, by the relation

$$n \mathbb{P}(a_n^{-1} \|\mathbf{X}_0\|_\infty > 1) \xrightarrow{n \rightarrow \infty} 1.$$

Here, we repeat the characterization of regularly varying time series via the so-called tail process and spectral tail process, which are essential concepts needed to derive the asymptotic properties of our estimators defined below. For  $E = \mathbb{R}^d$ , the result has been stated in Basrak and Segers [2009], Theorem 2.1, Theorem 3.1 and Corollary 3.2 and, in a very general setting, in Dombry et al. [2018], Lemma 3.5.

**Theorem 2.2.** *Let  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  be an  $E$ -valued stationary time series. Then, the following statements are equivalent (where  $\mathcal{L}(\cdot)$  denotes “the probability law of”):*

(i)  *$X$  is regularly varying with index  $\alpha$ .*

(ii) *There exists an  $E$ -valued time series  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  with  $\mathbb{P}(\|\mathbf{Y}_0\|_\infty > y) = y^{-\alpha}$  for  $y \geq 1$ , called the tail process, such that, for every finite  $T \subset \mathbb{Z}$ ,*

$$\mathcal{L}(\{x^{-1} \mathbf{X}_t\}_{t \in T} \mid \|\mathbf{X}_0\|_\infty > x) \xrightarrow{x \rightarrow \infty} \mathcal{L}(\{\mathbf{Y}_t\}_{t \in T}) \quad \text{weakly in } E^T.$$

(iii) *There exists an  $E$ -valued time series  $\{\Theta_t\}_{t \in \mathbb{Z}}$  with  $\|\Theta\|_\infty = 1$  a.s., called the spectral tail process, such that, for every finite  $T \subset \mathbb{Z}$ ,*

$$\mathcal{L}(\{\|\mathbf{X}_0\|_\infty^{-1} \mathbf{X}_t\}_{t \in T} \mid \|\mathbf{X}_0\|_\infty > x) \xrightarrow{x \rightarrow \infty} \mathcal{L}(\{\Theta_t\}_{t \in T}) \quad \text{weakly in } E^T.$$

*The tail process and the spectral tail process are related by  $\mathcal{L}(\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}) = \mathcal{L}(\{P\Theta_t\}_{t \in \mathbb{Z}})$ , where  $P$  is an  $\alpha$ -Pareto variable independent of the process  $\{\Theta_t\}_{t \in \mathbb{Z}}$ .*

Note that, even if  $\mathbf{X}_t$  is not univariate, but is a regularly varying multivariate or functional time series, we can transform it to a univariate time series by taking norms. Then, by definition,  $\{\|\mathbf{X}_t\|_\infty\}_{t \in \mathbb{Z}}$  is a regularly varying univariate time series with tail process  $\{\|\mathbf{Y}_t\|_\infty\}_{t \in \mathbb{Z}}$  and spectral tail process  $\{\|\Theta_t\|_\infty\}_{t \in \mathbb{Z}}$ . Thus, applying the norm allows for the application of classical tools for the analysis of extremes in regularly varying time series. In practice, however, extremes are not only defined via the norm, but in terms of more general *risk functionals*, see also Dombry and Ribatet [2015], Thibaud and Opitz [2015] and de Fondeville and Davison [2018]. Thus, in the following, we analyze the extremal behavior of the time series with respect to such a functional.

### 3 Clusters of extremes with respect to a homogeneous risk functional

#### 3.1 Risk functionals and regular variation

Building upon the background theory laid out in Section 2, we study the functional time series  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  and consider extremes defined in terms of threshold exceedances of a so-called risk functional, i.e., a functional  $r : C_+(S) \rightarrow [0, \infty)$  that is continuous and positively 1-homogeneous, i.e.,

$$r(cf) = cr(f), \quad c \geq 0, \quad f \in C_+(S).$$

As a first result, we generalize the fact that  $\{\|\mathbf{X}_t\|_\infty\}_{t \in \mathbb{Z}}$  is a regularly varying univariate time series and state that the same holds true for any time series of the type  $\{r(\mathbf{X}_t)\}_{t \in \mathbb{Z}}$ . This statement is closely related to results in Dombry and Ribatet [2015] and Dombry et al. [2018]. For completeness, we provide a short proof in the Supplementary Material.

**Proposition 3.1.** *Let  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  be a stationary  $C_+(S)$ -valued time series that is regularly varying with index  $\alpha > 0$  and spectral tail process  $\{\Theta_t\}_{t \in \mathbb{Z}}$ , and let  $r : C_+(S) \rightarrow [0, \infty)$  be a risk functional as defined above. Then, the univariate time series  $\{r(\mathbf{X}_t)\}_{t \in \mathbb{Z}}$  is regularly*

varying with index  $\alpha$  and spectral tail process  $\{\Theta_t^r\}_{t \in \mathbb{Z}}$  whose law is given by

$$\mathbb{P}(\{\Theta_t^r\}_{t \in \mathbb{Z}} \in A) = \frac{1}{\mathbb{E}\{r(\Theta_0)^{\alpha}\}} \int_{C_+(S)^{\mathbb{Z}}} \mathbf{1} \left\{ \left\{ \frac{r(\theta_t)}{r(\theta_0)} \right\}_{t \in \mathbb{Z}} \in A \right\} r(\theta_0)^{\alpha} \mathbb{P}_{\Theta}(\mathrm{d}\theta)$$

for any measurable set  $A \subset C_+(S)^{\mathbb{Z}}$ . In particular,  $r(\Theta_0^r) = 1$  a.s..

### 3.2 Clusters of extremes

Denoting a  $C_+(S)$ -valued random object  $\mathbf{X}_t$  at some fixed time  $t \in \mathbb{Z}$  as an extreme observation if and only if  $r(\mathbf{X}_t) > u$  for some large threshold  $u$ , we can then follow the notation of Oesting and Schnurr [2020] and call an  $\ell$ -dimensional vector of processes  $(\mathbf{X}_i)_{i=t}^{t+\ell-1}$  a  $u$ -exceedance cluster of size  $\ell \in \mathbb{N}$  with respect to  $r$  if and only if

$$r(\mathbf{X}_{t-1}) \leq u, r(\mathbf{X}_t) > u, \dots, r(\mathbf{X}_{t+\ell-1}) > u \text{ and } r(\mathbf{X}_{t+\ell}) \leq u.$$

A natural interpretation of these clusters is that all the data points within a cluster form an extreme event (such as a severe precipitation event spanning several time points, or a high sea temperature persisting over consecutive days), while different extreme events are separated by at least one non-exceedance of the prescribed threshold  $u$ . This definition naturally gives rise to questions that may be potentially interesting from a scientific perspective, such as ‘‘How long does an extreme event last?’’, or in other words, ‘‘What is the size of a cluster of  $u$ -exceedances with respect to  $r$ ?’’. To answer such questions, we need to characterize the distribution of the size  $C_u^r$  of a randomly selected  $u$ -exceedance cluster from the time series  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  with respect to  $r$ . The cluster size distribution may be defined as

$$\begin{aligned} \mathbb{P}(C_u^r = \ell) &= \frac{\mathbb{P}(\text{there is a cluster of } u\text{-exceedances of size } \ell \text{ starting at } t = 0)}{\mathbb{P}(\text{there is a cluster of } u\text{-exceedances starting at } t = 0)} \\ &= \frac{\mathbb{P}(r(\mathbf{X}_{-1}) \leq u, r(\mathbf{X}_0) > u, \dots, r(\mathbf{X}_{\ell-1}) > u, r(\mathbf{X}_{\ell}) \leq u)}{\mathbb{P}(r(\mathbf{X}_{-1}) \leq u, r(\mathbf{X}_0) > u)}, \quad \ell \in \mathbb{N}. \end{aligned}$$

Analogously to Oesting and Schnurr [2020], regular variation of  $\{r(\mathbf{X}_t)\}_{t \in \mathbb{Z}}$  stated in Prop. 3.1 implies the existence of a limit distribution as  $u \rightarrow \infty$ , namely,

$$\lim_{u \rightarrow \infty} \mathbb{P}(C_u^r = \ell) = \frac{\mathbb{P}(P \cdot r(\Theta_{-1}^r) \leq 1, P \cdot r(\Theta_0^r) > 1, \dots, P \cdot r(\Theta_{\ell-1}^r) > 1, P \cdot r(\Theta_{\ell}^r) \leq 1)}{\mathbb{P}(P \cdot r(\Theta_{-1}^r) \leq 1, P \cdot r(\Theta_0^r) > 1)}$$

$$\begin{aligned}
&= \frac{\mathbb{P}(\max\{1/r(\Theta_0^r), \dots, 1/r(\Theta_{\ell-1}^r)\} < P \leq \min\{1/r(\Theta_{-1}^r), 1/r(\Theta_\ell^r)\})}{\mathbb{P}(1/r(\Theta_0^r) < P \leq 1/r(\Theta_{-1}^r))} \\
&= \frac{\mathbb{E} [\min\{r(\Theta_0^r)^\alpha, \dots, r(\Theta_{\ell-1}^r)^\alpha\} - \max\{r(\Theta_{-1}^r)^\alpha, r(\Theta_\ell^r)^\alpha\}]_+}{\mathbb{E} [r(\Theta_0^r) - r(\Theta_{-1}^r)^\alpha]_+} \\
&= \frac{\mathbb{E} [\min\{1, r(\Theta_1^r)^\alpha, \dots, r(\Theta_{\ell-1}^r)^\alpha\} - \max\{r(\Theta_{-1}^r)^\alpha, r(\Theta_\ell^r)^\alpha\}]_+}{\mathbb{E} [1 - r(\Theta_{-1}^r)^\alpha]_+}, \tag{1}
\end{aligned}$$

for  $l \in \mathbb{N}$ , where  $P$  is an  $\alpha$ -Pareto random variable independent of the spectral tail process  $\{\Theta_t^r\}_{t \in \mathbb{Z}}$ . Note that, for the last equality, we used that  $r(\Theta_0^r) = 1$  a.s..

### 3.3 Extension to the non-heavy-tailed case

Remind that, for the definition of clusters, we required the  $C_+(S)$ -valued time series  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  to be regularly varying. Segers et al. [2017] showed that this holds if and only if it is regularly varying as a random element of the metric space  $C_+(S)^\mathbb{Z}$ . Thus, by results such as Thm- 2.4 in Lin and de Haan [2001] or Lemma 2.2 in Davis and Mikosch [2008],  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  is regularly varying with index  $\alpha$  if and only if  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  is in the max-domain of attraction of a max-stable process  $Z = \{\mathbf{Z}_t\}_{t \in \mathbb{Z}}$  with  $\alpha$ -Fréchet margins, i.e.,  $a_n^{-1} \max_{i=1, \dots, n} X^{(i)} \rightarrow_d Z$  as processes in  $C_+(S)^\mathbb{Z}$ , where  $X^{(1)}, X^{(2)}, \dots$  are independent copies of  $X$  and  $a_n$  are constants as in Def. 2.1. Thus, we can extend our results to stationary  $C(S)$ -valued time series  $\{\mathbf{W}_t\}_{t \in \mathbb{Z}}$  that can be monotonically transformed to a regularly varying  $C_+(S)$ -valued time series  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ , i.e.,

$$X(\mathbf{s}, t) = \phi(W(\mathbf{s}, t)), \quad \mathbf{s} \in S, t \in \mathbb{Z}, \tag{2}$$

where  $\phi : (-\infty, w^*) \rightarrow [0, \infty)$  is some nondecreasing bijective marginal transformation and  $w^* \in \mathbb{R} \cup \{\infty\}$  is the (joint) upper endpoint of the marginal distributions of  $W^*$ .

Now, assume that the risk functional  $r$  satisfies

$$\phi(r(f)) = r(\phi(f)) \tag{3}$$

for any continuous function  $f : S \rightarrow (-\infty, y^*)$ . For instance, irrespective of the choice of  $\phi$ , Condition (3) holds true for the following specific examples:

- the spatial minimum/maximum :  $r(f) = \min_{\mathbf{s} \in S} f(\mathbf{s})$  and  $r(f) = \max_{\mathbf{s} \in S} f(\mathbf{s})$ ,
- spatial quantiles:

$$r(f) = \inf \left\{ q \in \mathbb{R} : |S|^{-1} \int_S \mathbf{1}\{f(\mathbf{s}) \geq q\} d\mathbf{s} > p \right\}, \quad p \in (0, 1), \quad |S| = \int_S 1 d\mathbf{s}.$$

Under Assumptions (2) and (3) it is reasonable to consider  $\phi^{-1}(u)$ -exceedance clusters in  $\{\mathbf{W}_t\}_{t \in \mathbb{Z}}$  with respect to  $r$  for  $u \rightarrow \infty$ , that is, as  $\phi^{-1}(u) \rightarrow w^*$ . The probability that a randomly selected cluster of exceedances is of size  $\ell$  then equals

$$\begin{aligned} & \frac{\mathbb{P}(r(\mathbf{W}_{-1}) \leq \phi^{-1}(u), r(\mathbf{W}_0) > \phi^{-1}(u), \dots, r(\mathbf{W}_{\ell-1}) > \phi^{-1}(u), r(\mathbf{W}_\ell) \leq \phi^{-1}(u))}{\mathbb{P}(r(\mathbf{W}_{-1}) \leq \phi^{-1}(u), r(\mathbf{W}_0) > \phi^{-1}(u))} \\ &= \frac{\mathbb{P}(r(\mathbf{X}_{-1}) \leq u, r(\mathbf{X}_0) > u, \dots, r(\mathbf{X}_{\ell-1}) > u, r(\mathbf{X}_\ell) \leq u)}{\mathbb{P}(r(\mathbf{X}_{-1}) \leq u, r(\mathbf{X}_0) > u)}. \end{aligned}$$

and, by (1), the limit for the right-hand side exists as  $X = \phi(W)$  is regularly varying.

Examples of such processes  $\{\mathbf{W}_t\}_{t \in \mathbb{Z}} = \{W(\mathbf{s}, t)\}_{\mathbf{s} \in S, t \in \mathbb{Z}}$  comprise all processes with tail-equivalent marginal distributions that are in the max-domain of attraction of an arbitrary max-stable process  $Z$ , denoted by  $W \in \text{MDA}(Z)$ . This is due to the fact that  $Y^* \in \text{MDA}(Z)$  if and only if the following two conditions hold [see Aulbach et al., 2015, Corollary 2.2]:

- $W(\mathbf{s}, t) \in \text{MDA}(Z(\mathbf{s}, t))$  for all fixed  $\mathbf{s} \in S, t \in \mathbb{Z}$ ,
- and the copula process of  $W$  is in the max-domain of attraction of a marginally transformed version of  $Z$  with reversed Weibull margins.

In other words,  $\phi(W)$  is thus still in the max-domain of attraction of some max-stable process as long as the marginal distributions are attracted by a generalized extreme value distribution. In particular, if the marginal distributions of  $X = \phi(W)$  are in the max-domain of attraction of an  $\alpha$ -Fréchet distribution,  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  is regularly varying as discussed above.

## 4 Within-cluster behavior of extremes

### 4.1 General setting

In the following sections, we consider a regularly varying time series  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ , and take a closer look at the behavior of the time series within a cluster of exceedances. More precisely, we shall study the behavior of the vector of normalized processes  $(u^{-1}\mathbf{X}_i)_{i=t}^{t+\ell-1}$  provided that  $(\mathbf{X}_i)_{i=t}^{t+\ell-1}$  is a cluster of  $u$ -exceedances with respect to a risk functional  $r$ , i.e., we condition on the event  $r(\mathbf{X}_{t-1}) \leq u, r(\mathbf{X}_t) > u, \dots, r(\mathbf{X}_{t+\ell-1}) > u$  and  $r(\mathbf{X}_{t+\ell}) \leq u$ .

While regular variation guarantees the existence of a limit expression for

$$\mathbb{P}((u^{-1}\mathbf{X}_t, \dots, u^{-1}\mathbf{X}_{t+\ell-1}) \in A \mid r(\mathbf{X}_{t-1}) \leq u, r(\mathbf{X}_t) > u, \dots, r(\mathbf{X}_{t+\ell-1}) > u, r(\mathbf{X}_{t+\ell}) \leq u)$$

as  $u \rightarrow \infty$  whenever  $\mathbb{P}((P\Theta_0, \dots, P\Theta_{\ell-1}) \in \partial A) = 0$  for a broad range of Borel sets  $A \subset (C_+(S))^\ell$ , we shall henceforth study the limiting within-cluster behavior of certain functionals applied separately to each of the spatial fields  $\mathbf{X}_t, \dots, \mathbf{X}_{t+\ell-1}$ . More precisely, we here consider three types of functionals, which have a practical interest:

- the risk functional  $r$  itself (Section 4.2), which provides insights into the (evolving) severity of an extreme event during a cluster of exceedances.
- spatial risk measures (Section 4.3), which provide insights into the (evolving) spatial extent of an extreme event during a cluster of exceedances.
- multivariate location measures (Section 4.4), which provide insights into the (evolving) position of an extreme event during a cluster of exceedances.

### 4.2 Limit expressions for risk functionals

As the risk functional  $r$  is used to define extremes, it is natural to further analyze its behavior inside an extreme event, i.e., inside a cluster of  $u$ -exceedances with respect to  $r$ . As the time

series  $\{r(\mathbf{X}_t)\}_{t \in \mathbb{Z}}$  is regularly varying by Prop. 3.1, we can exploit the results from Oesting and Schnurr [2020] to derive the limiting distribution of the corresponding extremal ordinal patterns and other within-cluster characteristics.

Note that there are various definitions of  $\ell$ -ordinal patterns being related either on the rank [e.g., in Bandt, 2020] or the order [e.g., in Keller et al., 2007] of data. We here make use of the rank-based definition because we consider it more straightforward to interpret, thus deviating from the notion used in Oesting and Schnurr [2020], for instance, who propose to use ordinal patterns to describe stylized facts of extreme events. More precisely, we define the  $\ell$ -ordinal pattern as the mapping  $\Pi$  that maps a vector  $(X_i)_{i=1}^\ell \in \mathbb{R}^\ell$  (without ties) to the unique permutation  $\pi$  of  $\{1, \dots, \ell\}$  such that

$$X_i < X_j \iff \pi(i) < \pi(j),$$

i.e.,  $\Pi(X)$  is the vector consisting of the ranks of the components of  $X$ . For instance, whenever  $X$  is itself some permutation of  $\{1, \dots, \ell\}$ , the ordinal pattern  $\Pi(X)$  equals  $X$ . Note that, in most of the applications we have in mind, ties occur very rarely. Thus, we do not consider this case further, but shall report any pattern including ties separately.

Denoting the probability that the  $\ell$ -ordinal pattern of  $(r(\mathbf{X}_i))_{i=0}^{\ell-1}$  for a cluster of  $u$ -exceedances  $(\mathbf{X}_i)_{i=0}^{\ell-1}$  equals a specific permutation  $\pi$  by  $\mathbb{P}_{u,\ell}(\pi)$ , i.e.,

$$\mathbb{P}_{u,\ell}(\pi) = \mathbb{P}(\Pi((r(\mathbf{X}_i))_{i=0}^{\ell-1}) = \pi \mid r(\mathbf{X}_{-1}) \leq u, r(\mathbf{X}_0) > u, \dots, r(\mathbf{X}_{\ell-1}) > u, r(\mathbf{X}_\ell) \leq u),$$

we obtain the limit distribution

$$\begin{aligned} & \lim_{u \rightarrow \infty} \mathbb{P}_{u,\ell}(\pi) \\ = & \frac{\mathbb{P}(\Pi((r(\Theta_i^r))_{i=0}^{\ell-1}) = \pi, P \cdot r(\Theta_{-1}^r) \leq 1, P \cdot r(\Theta_0^r) > 1, \dots, P \cdot r(\Theta_{\ell-1}^r) > 1, P \cdot r(\Theta_\ell^r) \leq 1)}{\mathbb{P}(P \cdot r(\Theta_{-1}^r) \leq 1, P \cdot r(\Theta_0^r) > 1, \dots, P \cdot r(\Theta_{\ell-1}^r) > 1, P \cdot r(\Theta_\ell^r) \leq 1)} \\ = & \frac{\mathbb{E}[(r(\Theta_0^r)^\alpha \wedge \dots \wedge r(\Theta_{\ell-1}^r)^\alpha - r(\Theta_{-1}^r)^\alpha \vee (r(\Theta_\ell^r)^\alpha)_+ \mathbf{1}\{\Pi((r(\Theta_i^r))_{i=0}^{\ell-1}) = \pi\}]}{\mathbb{E}(r(\Theta_0^r)^\alpha \wedge \dots \wedge r(\Theta_{\ell-1}^r)^\alpha - r(\Theta_{-1}^r)^\alpha \vee (r(\Theta_\ell^r)^\alpha)_+}. \end{aligned} \quad (4)$$

**Remark 4.1.** As the ordinal pattern is invariant under marginal transformations, by the same arguments as given in Section 3.3, the results above also hold true more generally for marginally-transformed processes  $W$  defined such that  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}} = \{\phi(\mathbf{W}_t)\}_{t \in \mathbb{Z}}$  is regularly varying as in (2), provided that (3) holds.

Even more generally, it follows directly from Prop. 3.1 that

$$\begin{aligned} & \lim_{u \rightarrow \infty} \mathbb{P}((u^{-1}r(\mathbf{X}_i))_{i=0}^{\ell-1} \in A \mid r(\mathbf{X}_{-1}) \leq u, r(\mathbf{X}_0) > u, \dots, r(\mathbf{X}_{\ell-1}) > u, r(\mathbf{X}_\ell) \leq u) \\ &= \frac{\mathbb{P}((P \cdot r(\Theta_i^r))_{i=0}^{\ell-1} \in A, P \cdot r(\Theta_{-1}^r) \leq 1, P \cdot r(\Theta_0^r) > 1, \dots, P \cdot r(\Theta_{\ell-1}^r) > 1, P \cdot r(\Theta_\ell^r) \leq 1)}{\mathbb{P}(P \cdot r(\Theta_{-1}^r) \leq 1, P \cdot r(\Theta_0^r) > 1, \dots, P \cdot r(\Theta_{\ell-1}^r) > 1, P \cdot r(\Theta_\ell^r) \leq 1)} \end{aligned}$$

for all measurable  $A \in (0, \infty)^\ell$  satisfying  $\mathbb{P}((P \cdot r(\Theta_i^r))_{i=0}^{\ell-1} \in \partial A) = 0$ . However, in this work, we mainly focus on ordinal patterns.

### 4.3 Limit expressions for spatial risk measures

Besides the behavior of the risk functional itself inside a cluster, there are also other quantities of practical interest, such as the spatial extent of an extreme event for instance. To this end, Koch [2017] proposed to consider the normalized spatially aggregated “loss” due to a (random) function  $f \in C_+(S)$  defined by

$$L_u(f) = \frac{1}{|S|} \int_S E(\mathbf{s}) \mathbf{1}\{f(\mathbf{s}) > u\} \, d\mathbf{s},$$

where  $E$  is a positive *exposure function*. Here, it is important to note that, due to the indicator function,  $L_u$  is not homogeneous and thus not a valid risk functional.

In the special case where  $E \equiv 1$ ,  $L_u$  corresponds to the percentage of the area  $S$  that is affected by an extreme event defined as a pointwise exceedance of the threshold  $u$ . We are interested in the joint limiting behavior of the spatially aggregated “losses” of  $\mathbf{X}_t, \dots, \mathbf{X}_{t+\ell-1}$  provided that  $(\mathbf{X}_i)_{i=t}^{t+\ell-1}$  forms a cluster of extremes. More generally, we can consider the functional  $m : \text{Bin}(S) \rightarrow [0, \infty]$  where  $\text{Bin}(S) = \{f : S \rightarrow \{0, 1\}, f \text{ measurable}\}$  is the space

of measurable binary functions on  $S$ . For instance, choosing

$$m(f) = |S|^{-1} \int_S E(\mathbf{s}) \cdot f(\mathbf{s}) \, d\mathbf{s},$$

we can rewrite  $L_u(\mathbf{X}_t) = m(\mathbf{1}\{u^{-1}\mathbf{X}_t > 1\})$ .

In the following, with a slight abuse of notation, we write  $m(A)$  for  $m(\mathbf{1}\{A\})$ . Furthermore, we assume that  $m(\mathbf{0}_S) = 0$  where  $\mathbf{0}_S$  denotes the zero function on  $S$ . Under these assumption, the following result which is proven in the Supplementary Material, on the limit distribution of spatial risk measures can be stated.

**Proposition 4.2.** *Let  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  be a stationary  $C_+(S)$ -valued time series that is regularly varying with index  $\alpha > 0$  and spectral tail process  $\{\Theta_t\}_{t \in \mathbb{Z}}$ , and let  $m : \text{Bin}(S) \rightarrow [0, \infty)$  be such that  $m(\mathbf{0}_S) = 0$ . If the joint distribution of  $(m(\{\Theta_0 > \eta\}), \dots, m(\{\Theta_{\ell-1} > \eta\}))$  depends continuously on  $\eta \in [0, 1]$  for each  $\ell > 0$ , then, for any measurable set  $A \subset (0, \infty)^\ell$  with  $\mathbb{P}((m(\{\Theta_0 > \eta\}), \dots, m(\{\Theta_{\ell-1} > \eta\})) \in \partial A) = 0$  for a.e.  $\eta \in [0, 1]$ , we have that*

$$\begin{aligned} & \mathbb{P}((m(\{\mathbf{X}_0 > u\}), \dots, m(\{\mathbf{X}_{\ell-1} > u\})) \in A \mid \\ & \quad r(\mathbf{X}_{-1}) \leq u, r(\mathbf{X}_0) > u, \dots, r(\mathbf{X}_{\ell-1}) > u, r(\mathbf{X}_\ell) \leq u) \\ & \xrightarrow{u \rightarrow \infty} (\mathbb{E}[\min\{1, r(\Theta_1)^\alpha, \dots, r(\Theta_{\ell-1})^\alpha\} - \max\{r(\Theta_{-1})^\alpha, r(\Theta_\ell)^\alpha\}]_+)^{-1} \\ & \quad \cdot \int_0^1 \mathbb{P}((m(\{\Theta_0^\alpha > \eta\}), \dots, m(\{\Theta_{\ell-1}^\alpha > \eta\})) \in A, \\ & \quad \quad r(\Theta_{-1})^\alpha \leq \eta, r(\Theta_0)^\alpha > \eta, \dots, r(\Theta_{\ell-1})^\alpha > \eta, r(\Theta_\ell)^\alpha \leq \eta) \, d\eta. \end{aligned}$$

As an example, similarly to Section 3.1, we can consider ordinal patterns, but now applied to the spatial risk measure  $m$  rather than the risk functional  $r$ . Specifically, by a simple application of the results above, we can find the asymptotic distribution of

$$\Pi((m(\{\mathbf{X}_i > u\}))_{i=0}^{\ell-1} \mid r(\mathbf{X}_{-1}) \leq u, r(\mathbf{X}_0) > u, \dots, r(\mathbf{X}_{\ell-1}) > u, r(\mathbf{X}_\ell) \leq u),$$

as  $u \rightarrow \infty$ . These ordinal patterns provide important information about the evolution of the spatial extent (i.e., spatial risk) of an extreme event of a certain duration.

**Remark 4.3.** For a process  $W$  related to  $X$  via  $\mathbf{X}_t = \phi(\mathbf{W}_t)$  for all  $t \in \mathbb{Z}$  as in (2), we can directly rewrite  $m(\{\mathbf{X}_t > u\}) = m(\{\mathbf{W}_t > \phi^{-1}(u)\})$ . Thus, if (3) holds, we have that

$$\begin{aligned} & \mathbb{P}((m(\{\mathbf{Y}_0 > \phi^{-1}(u)\}), \dots, m(\{\mathbf{Y}_{\ell-1} > \phi^{-1}(u)\})) \in A \mid \\ & \quad r(\mathbf{Y}_{-1}) \leq \phi^{-1}(u), r(\mathbf{Y}_0) > \phi^{-1}(u), \dots, r(\mathbf{Y}_{\ell-1}) > \phi^{-1}(u), r(\mathbf{Y}_\ell) \leq \phi^{-1}(u)) \\ &= \mathbb{P}((m(\{\mathbf{X}_0 > u\}), \dots, m(\{\mathbf{X}_{\ell-1} > u\})) \in A \mid \\ & \quad r(\mathbf{X}_{-1}) \leq u, r(\mathbf{X}_0) > u, \dots, r(\mathbf{X}_{\ell-1}) > u, r(\mathbf{X}_\ell) \leq u) \end{aligned}$$

and Prop. 4.2 applies if  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  is regularly varying.

#### 4.4 Limit expressions for location measures

Beyond the intensity and the spatial extent, we may be interested to infer whether an extreme event moves over space, i.e., whether (and how) its position changes during a cluster of exceedances. To tackle this question, we consider general functionals  $c : C_+(S) \rightarrow S$  describing the location of an extreme event. Having in mind the interpretation as the “center” of the area affected by the extreme event, we want to allow for cases where this location depends on the area where the threshold (here, set to 1 for simplicity) is exceeded. Furthermore, once this area is fixed, we want the location to be invariant under scalar multiplication, i.e., we assume the existence of some functional  $\tilde{c} : C_+(S) \times \text{Bin}(S) \rightarrow S$  such that

$$c(f) = \tilde{c}\left(\frac{f}{r(f)}, \mathbb{1}\{f > 1\}\right) =: \tilde{c}\left(\frac{f}{r(f)}, \{f > 1\}\right),$$

and certain continuity conditions specified below are satisfied. Typical examples of such functionals include:

- the peak location,  $c(f) = \arg \max_{\mathbf{s} \in S} f(\mathbf{s})$ ;
- the centroid of the region exceeding the threshold,

$$c(f) = \frac{1}{\int_{\{\mathbf{s} \in S: f(\mathbf{s}) > 1\}} \mathbb{1} \, d\mathbf{s}} \int_{\{\mathbf{s} \in S: f(\mathbf{s}) > 1\}} \mathbf{s} \, d\mathbf{s};$$

where the second integral is to be interpreted componentwise;

- the centroid weighted by the intensity,  $f(\mathbf{s})$ , of the event itself,

$$c(f) = \frac{1}{\int_S f(\mathbf{s}) \, d\mathbf{s}} \int_S \mathbf{s} f(\mathbf{s}) \, d\mathbf{s} = \begin{pmatrix} \frac{\int_S s_1 f(\mathbf{s}) \, d\mathbf{s}}{\int_S f(\mathbf{s}) \, d\mathbf{s}} \\ \vdots \\ \frac{\int_S s_d f(\mathbf{s}) \, d\mathbf{s}}{\int_S f(\mathbf{s}) \, d\mathbf{s}} \end{pmatrix};$$

- the location defined by the componentwise median of the coordinates of all points exceeding the threshold,

$$c(f) = \text{median}\{\mathbf{s} \in S : f(\mathbf{s}) > 1\}.$$

We are interested in the behavior of  $c(u^{-1}\mathbf{X}_t)$  if  $\mathbf{X}_t$  is extreme in the sense that  $r(\mathbf{X}_t) > u$  for some large threshold  $u$  over consecutive time points. The following proposition, proven in the Supplementary Material, characterizes the limiting distribution of such an event.

**Proposition 4.4.** *Let  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  be a stationary  $C_+(S)$ -valued time series that is regularly varying with index  $\alpha > 0$  and spectral tail process  $\{\Theta_t\}_{t \in \mathbb{Z}}$ . Let  $r : C_+(S) \rightarrow [0, \infty)$  be a continuous positively 1-homogeneous risk functional and let  $c : C_+(S) \rightarrow S$  be of the form*

$$c(f) = \tilde{c} \left( \frac{f}{r(f)}, \mathbf{1}\{f > 1\} \right),$$

for some functional  $\tilde{c} : C_+(S) \times \text{Bin}(S) \rightarrow S$  where we formally set  $\tilde{c}(\cdot, \mathbf{0}_S) = \infty$ . Moreover, assume that, for  $\ell > 0$ , the joint distribution of  $(c(\eta^{-1}\Theta_0), \dots, c(\eta^{-1}\Theta_{\ell-1}))$  depends continuously on  $\eta \in (0, 1]$ . Then, for any  $\ell \in \mathbb{N}$  and any measurable set  $A \subset S^\ell$  satisfying

$$\mathbb{P}((c(\Theta_0), \dots, c(\Theta_{\ell-1})) \in \partial A) = 0,$$

we have that

$$\begin{aligned} & \mathbb{P}((c(u^{-1}\mathbf{X}_0), \dots, c(u^{-1}\mathbf{X}_{\ell-1})) \in A \mid r(\mathbf{X}_{-1}) \leq u, r(\mathbf{X}_0) > u, \dots, r(\mathbf{X}_{\ell-1}) > u, r(\mathbf{X}_\ell) \leq u) \\ & \xrightarrow{u \rightarrow \infty} [\mathbb{E}(\min\{1, r(\Theta_1)^\alpha, \dots, r(\Theta_{\ell-1})^\alpha\} - \max\{r(\Theta_{-1})^\alpha, r(\Theta_\ell)^\alpha\})_+]^{-1} \end{aligned}$$

$$\cdot \int_0^1 \mathbb{P}((\tilde{c}(\Theta_0, \{\Theta_0^\alpha > \eta\}), \dots, \tilde{c}(\Theta_{\ell-1}, \{\Theta_{\ell-1}^\alpha > \eta\})) \in A, \\ r(\Theta_{-1})^\alpha \leq \eta, r(\Theta_0)^\alpha > \eta, \dots, r(\Theta_{\ell-1})^\alpha > \eta, r(\Theta_\ell)^\alpha \leq \eta) d\eta.$$

**Remark 4.5.** Note that, by the same arguments as given in Section 3.3, the results above also hold true for marginally-transformed processes  $W$  such that  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}} = \{\phi(\mathbf{W}_t)\}_{t \in \mathbb{Z}}$  is regularly varying as in (2), provided that  $c$  is invariant under the transformation  $\phi$ , i.e.,  $c(\phi(f)) = c(f)$  for any continuous function  $f : S \rightarrow (-\infty, y^*)$ . This is true for the peak location functional, the centroid of the region exceeding the threshold, and for the componentwise median location of threshold exceedances, but not for the weighted centroid.

## 5 Inference for limit distributions

Even though Equation (1), Equation (4), Prop. 4.2 and Prop. 4.4 provide explicit expressions for some specific cases in terms of the spectral tail process  $\{\Theta_t\}_{t \in \mathbb{Z}}$ , all the limit expressions considered in Section 3 and Section 4 are of the form

$$\lim_{u \rightarrow \infty} \frac{\mathbb{P}((u^{-1} \mathbf{X}_i)_{i=-1}^\ell \in A)}{\mathbb{P}((u^{-1} \mathbf{X}_i)_{i=-1}^\ell \in A_0)} = \frac{\mathbb{P}((P \cdot \Theta_i)_{i=-1}^\ell \in A)}{\mathbb{P}((P \cdot \Theta_i)_{i=-1}^\ell \in A_0)} = \frac{\mathbb{P}((\mathbf{Y}_i)_{i=-1}^\ell \in A)}{\mathbb{P}((\mathbf{Y}_i)_{i=-1}^\ell \in A_0)} = \frac{\mu_{\{-1, \dots, \ell\}}(A)}{\mu_{\{-1, \dots, \ell\}}(A_0)},$$

where  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  and  $\mu_T$ ,  $T \subset \mathbb{Z}$ , denote the tail process and the exponent measure, given in Def. 2.1 and Thm. 2.2, respectively. In all of the cases considered so far, the sets  $A_0$  and  $A$  include the condition that an extreme event occurs at time  $t = 0$ , i.e.,

$$A_0, A \subset \{(\mathbf{X}_{-1}, \dots, \mathbf{X}_\ell) \in (C_+(S))^{\ell+2} : r(\mathbf{X}_0) > 1\}.$$

Given observations  $\mathbf{X}_{-1}, \dots, \mathbf{X}_n \in C_+(S)$ , a natural estimator for the desired limiting expression  $\mu_{\{-1, \dots, \ell\}}(A)/\mu_{\{-1, \dots, \ell\}}(A_0)$  is the ratio estimator

$$\widehat{R}_{n,u}(A, A_0) = \frac{\sum_{k=0}^{n-\ell} \mathbf{1}\{(u^{-1} \mathbf{X}_i)_{i=k-1}^{k+\ell} \in A\}}{\sum_{k=0}^{n-\ell} \mathbf{1}\{(u^{-1} \mathbf{X}_i)_{i=k-1}^{k+\ell} \in A_0\}} =: \frac{\widehat{\mathbb{P}}_{n,u}(A)}{\widehat{\mathbb{P}}_{n,u}(A_0)}. \quad (5)$$

We shall impose certain conditions on the time series  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  to ensure asymptotic normality of the estimator. More precisely, we follow Oesting and Schnurr [2020] who adapt conditions

used in Davis and Mikosch [2009] for the estimation of the extremogram, in order to show asymptotic normality for ratio estimators of the same type as in (5) in the context of a regularly varying univariate time series. We here extend these results to the functional time series setting. We impose weak (and eventually vanishing) long-range temporal dependence. Specifically, consider the  $\alpha$ -mixing coefficient

$$\alpha_h = \sup_{A, B \in \mathcal{B}(C_+(S)^\mathbb{N})} |\mathbb{P}(\{\mathbf{X}_t\}_{t \leq 0} \in A, \{\mathbf{X}_t\}_{t \geq h} \in B) - \mathbb{P}(\{\mathbf{X}_t\}_{t \leq 0} \in A) \mathbb{P}(\{\mathbf{X}_t\}_{t \geq h} \in B)|, \quad (6)$$

where  $h \in \mathbb{N}_0$ . We assume the following mixing and anti-clustering condition:

**Condition (M).** There exist a sequence  $\{u_n\}_{n \in \mathbb{N}} \subset \mathbb{R}$  of thresholds and an intermediate sequence  $\{r_n\}_{n \in \mathbb{N}} \subset \mathbb{N}$  with  $\lim_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} r_n = \infty$ ,  $\lim_{n \rightarrow \infty} n \mathbb{P}(\|\mathbf{X}_0\|_\infty > u_n) = \infty$ ,  $\lim_{n \rightarrow \infty} r_n \mathbb{P}(\|\mathbf{X}_0\|_\infty > u_n) = 0$  such that

$$\lim_{n \rightarrow \infty} \frac{1}{\mathbb{P}(\|\mathbf{X}_0\|_\infty > u_n)} \sum_{h=r_n}^{\infty} \alpha_h = 0 \quad (7)$$

and, for all  $\varepsilon > 0$ ,

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \sum_{h=k}^{r_n} \mathbb{P}(\|\mathbf{X}_h\|_\infty > \varepsilon u_n \mid \|\mathbf{X}_0\|_\infty > \varepsilon u_n) = 0. \quad (8)$$

We can then show the following result that guarantees asymptotic normality of our estimator, based on a proof analogous to the results of Oesting and Schnurr [2020]:

**Proposition 5.1.** *Let  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  be a stationary  $C_+(S)$ -valued time series that is regularly varying with index  $\alpha > 0$  and tail process  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  whose finite-dimensional distributions are denoted by  $(\mu_I)_{I \subset \mathbb{Z}}$ . Moreover, let  $A, A_0 \subset C_+(S)^{\ell+1}$  be continuity sets with respect to  $\mu_{\{-1, \dots, \ell\}}$  bounded away from zero. We further assume that Condition (M) holds and that the  $\alpha$ -mixing coefficients satisfy  $\alpha_h \in \mathcal{O}(h^{-\delta})$  for some  $\delta > 2$ . If the sequence of thresholds  $\{u_n\}_{n \in \mathbb{N}}$  additionally satisfies*

$$\lim_{n \rightarrow \infty} n^{\delta/(4+\delta)} \mathbb{P}(\|\mathbf{X}_0\|_\infty > u_n) = \infty, \quad (9)$$

and

$$\lim_{n \rightarrow \infty} \sqrt{n \mathbb{P}(\|\mathbf{X}_0\|_\infty > u_n)} \left[ \frac{\mathbb{P}(\{\mathbf{X}_i\}_{i=-1}^\ell \in u_n A)}{\mathbb{P}(\{\mathbf{X}_i\}_{i=-1}^\ell \in u_n A_0)} - \frac{\mu_{\{-1, \dots, \ell\}}(A)}{\mu_{\{-1, \dots, \ell\}}(A_0)} \right] = 0,$$

then

$$\sqrt{n \mathbb{P}(\|\mathbf{X}_0\|_\infty > u_n)} \left( \widehat{R}_{n, u_n}(A, A_0) - \frac{\mu_{\{-1, \dots, \ell\}}(A)}{\mu_{\{-1, \dots, \ell\}}(A_0)} \right)$$

converges in distribution to a centered Gaussian random variable as  $n \rightarrow \infty$ .

The asymptotic variance in Prop. 5.1 can be estimated via bootstrap techniques. Here, we make use of a multiplier block bootstrap which has proven to be appropriate for estimators of a similar type [Drees, 2015]. More precisely, we divide the time series data  $\mathbf{X}_{-1}, \dots, \mathbf{X}_n$  of length  $n + 1$  into  $b_n$  data sets of length  $\ell_n + 2$ , i.e., we consider blocks

$$(\mathbf{X}_t)_{t=-1}^{\ell_n}, (\mathbf{X}_t)_{t=\ell_n+1}^{2\ell_n+3}, \dots, (\mathbf{X}_t)_{t=n-\ell_n-1}^n$$

and denote the counterparts of  $\widehat{\mathbb{P}}_{n, u}(A)$  and  $\widehat{\mathbb{P}}_{n, u}(A_0)$  based on each of these  $b_n$  blocks by

$$\widehat{\mathbb{P}}_{\ell_n, u_n}^{(1)}(A), \dots, \widehat{\mathbb{P}}_{\ell_n, u_n}^{(b_n)}(A) \quad \text{and} \quad \widehat{\mathbb{P}}_{\ell_n, u_n}^{(1)}(A_0), \dots, \widehat{\mathbb{P}}_{\ell_n, u_n}^{(b_n)}(A_0),$$

respectively. Then, a bootstrap sample of  $\widehat{R}_{n, u_n}(A, A_0)$  can be obtained by repeatedly sampling i.i.d. random variables  $\xi_1, \dots, \xi_{b_n}$  with  $\mathbb{E}(\xi_j) = 0$  and  $\text{Var}(\xi_j) = 1$  and considering

$$\frac{\sum_{j=1}^{b_n} (1 + \xi_j) \widehat{\mathbb{P}}_{\ell_n, u_n}^{(j)}(A)}{\sum_{j=1}^{b_n} (1 + \xi_j) \widehat{\mathbb{P}}_{\ell_n, u_n}^{(j)}(A_0)}.$$

We analyze the performance of this multiplier block bootstrap scheme numerically in our simulation study in Section 6.

## 6 Simulation study

In order to demonstrate the behavior of our estimators for finite sample sizes, we apply them to a simulated realization of a regularly varying space-time stochastic process. Here, we choose a stationary space-time Brown–Resnick process [Brown and Resnick, 1977, Kabluchko

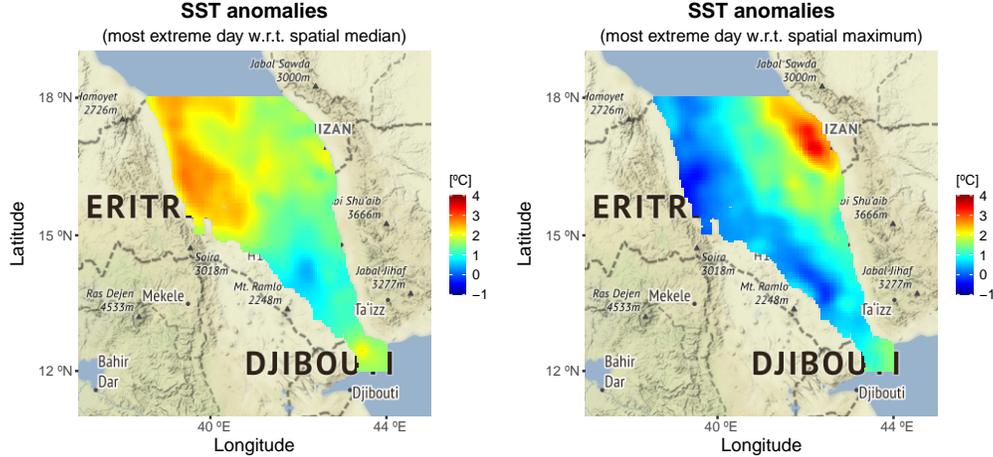


Figure 2: The study region in the Southern Red Sea consisting of 5562 grid cells. The plots show the sea surface temperature anomalies (i.e., the detrended SST data) for the most extreme day with respect to the spatial median (left) and maximum (right), respectively.

et al., 2009, Davis et al., 2013, Huser and Davison, 2014, Engelke et al., 2015] with unit Fréchet margins, simulated on a grid  $S$  consisting of  $|S| = 1\,367$  grid points in an irregular domain delimited by a box with coordinates in  $40.0\text{--}42.5^\circ\text{E}$  and  $16.5\text{--}18.0^\circ\text{N}$ . That is, the spatial domain is a subset the southern Red Sea region mapped in Figure 2 and considered in the application in Section 7, while the length of the time series is 12 000, also similar to the application. The properties of the Brown–Resnick process are determined by the variogram of the underlying space-time Gaussian process, which we choose here to be separable with

$$\gamma(\mathbf{h}, t) = \left\| \begin{pmatrix} 2.6h_1 \\ 2.4h_2 \end{pmatrix} \right\|^{1.9} + |t|^{1.1}, \quad \mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}, \mathbf{s} \in S, t = 1, \dots, 12\,000.$$

The variogram parameters were chosen so that the dependence structure of the resulting Brown–Resnick process resembles that of the sea temperature data in the data in Section 7.

The simulation makes use of the extremal functions algorithm of Dombry et al. [2016]. Here, note that an exact simulation of the max-stable process would on average require the simulation of  $|S| \times \{1, \dots, 12\,000\} = 1\,367 \times 12\,000$  Gaussian processes on the whole spatio-temporal domain  $S \times \{1, \dots, 12\,000\}$  and, thus, would be extremely time-consuming. Therefore, we use an approximate version of the exact algorithm of Dombry et al. [2016],

which ensures exactness only on a subgrid consisting of every second point, similarly to Oesting and Storkorb [2022]. Furthermore, we assume that an extremal function cannot have an influence over a temporal lag exceeding 18, that is, the extremal function at the space-time point  $(\mathbf{s}, t)$  is simulated on a smaller domain  $\mathcal{S} \times \{t - 18, \dots, t - 1, t, t + 1, \dots, t + 18\}$  only. Thus, the computing time for the simulation is reduced significantly, while the accuracy is hardly affected, as preliminary experiments on smaller domains indicate.

As the simulated process is regularly varying with (heavy-tailed) Fréchet margins, the theory developed above applies and allows us to use any homogeneous functional  $r$  for defining extremes. Here, we illustrate our methodology with the spatial average

$$r(f) = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{s} \in \mathcal{S}} f(\mathbf{s}),$$

as the corresponding tail process for space-time Brown–Resnick processes turns out to be easy to simulate from [see Dombry et al., 2016, for instance].

We apply our proposed ratio estimators to estimate the following two quantities:

- the cluster size distribution, i.e.,  $\mathbb{P}(C_u^r = \ell)$ ,  $\ell = 1, 2, \dots$  (recall Section 3.2),
- the distribution of ordinal patterns with respect to the actual risk functional  $r$  (recall Section 4.2); here, we focus on the patterns of the first  $\ell$  time points within a cluster, i.e., we focus on all clusters of size greater than or equal to  $\ell$ , with  $\ell = 2, 3$ ,

based on the choice of the threshold  $u$  as the 95th percentile of  $r(\mathbf{X}_t)$ ,  $t = 1, \dots, 12\,000$ . For each estimate, the corresponding 95% confidence interval is obtained using the multiplier block bootstrap described at the end of Section 5 with a block length of 1 000. The results are displayed in Figure 3 and Figure 4, respectively.

Our results based on estimators of the form  $\widehat{R}_{n, u_n}(A, A_0)$  can then be compared to the theoretical limit  $\mu_{\{-1, \dots, \ell\}}(A) / \mu_{\{-1, \dots, \ell\}}(A_0)$  which can be accurately approximated via direct Monte Carlo simulations from the tail process. Similarly, the asymptotic variance in Prop. 5.1

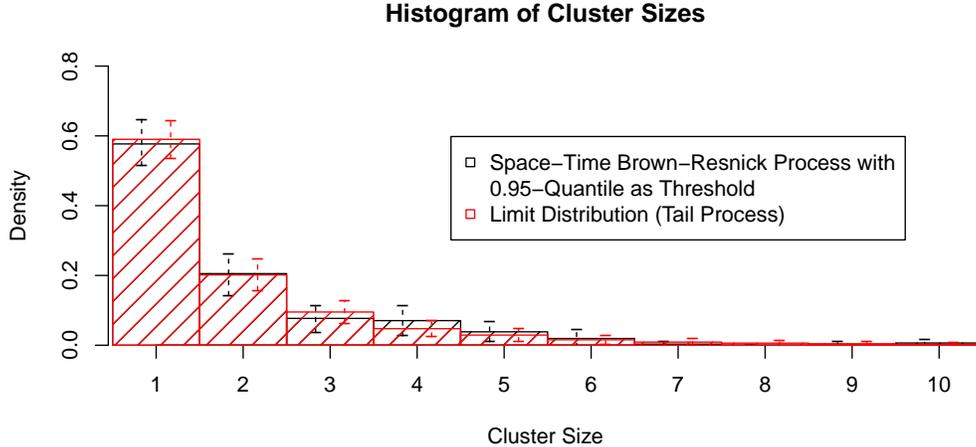


Figure 3: Cluster size distribution for the simulated example. The black bars correspond to the estimated probabilities with the 95% confidence intervals obtained by the multiplier block bootstrap, while the red bars are the result of Monte Carlo simulations from the tail process with the corresponding confidence intervals being calculated according to Prop. 5.1.

can also be calculated precisely. Assuming asymptotic normality, theoretical confidence intervals can thus be derived, as well.

From Figures 3 and 4, it can be seen that the estimated distribution of the cluster size and the ordinal patterns in clusters of size at least 2 are very close to the theoretical values. For patterns of size at least 3, there are somewhat larger deviations, which can be explained by the fact that even in a time series of length 12 000, the number of long clusters is still very small. As can be seen, the bootstrap-based confidence intervals are fairly wide, but still cover the theoretical values, which indicates that our estimator performs well overall. Comparing the width of the asymptotic confidence intervals and the bootstrap-based ones, there are slight differences, but no systematic ones. This suggests that the bootstrap procedure also works satisfactorily.

## 7 Application to Red Sea surface temperature data

Sea surface temperature (SST) is known to be a major factor influencing the survival of marine life and the sustainability of fragile ecosystems, such as coral reefs [Reaser et al., 2000,

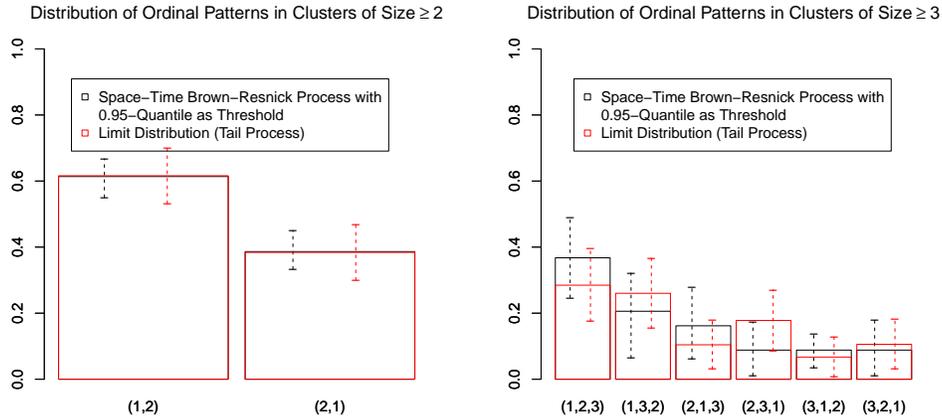


Figure 4: Distribution of the ordinal pattern for the beginning of clusters of size  $\ell \geq 2$  (left) and  $\ell \geq 3$  (right), for the simulated examples. Analogously to Figure 3, the black bars correspond to the estimated probabilities with 95% confidence intervals obtained via bootstrap and the red bars are the result of Monte Carlo simulations from the tail process with confidence intervals according to Prop. 5.1.

Berumen et al., 2013, Lewandowska et al., 2014]. Therefore, the spatiotemporal analysis of extreme SST data is important, and an entire special issue of the journal *Extremes* was devoted to the spatiotemporal prediction of extremes with application to SST data over the whole Red Sea; see <https://link.springer.com/journal/10687/volumes-and-issues/24-1> and the Editorial by Huser [2021]. Apart from spatiotemporal prediction, previous research has also focused on identifying time trends in SST data due to climate change and estimating extreme SST hotspots [Hazra and Huser, 2021], as well as on the modeling of spatial and spatiotemporal extremal dependence in SST data [Simpson et al., 2022, Huser et al., 2022]. However, the characterization of complex spatiotemporal patterns and the within-cluster behavior in functional time series of SST extremes has never been investigated. We here re-analyze the Red Sea data used in the above papers to illustrate our proposed methodology. Our analysis of spatiotemporal patterns complements previous studies by providing new insights into the complex hydrodynamic behavior of the SST data across the study domain, and shedding light on the spatial extent and persistence of extreme SST events.

The dataset we use comes from the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) project, producing satellite-derived daily SST data at  $0.05^\circ \times 0.05^\circ$  resolution [Donlon et al., 2012]. Over the whole Red Sea, daily SST data are available at 16703 grid cells between 1985–2015 (i.e., 31 years worth of daily data) and we here focus on the southern Red Sea, below latitude  $18^\circ\text{N}$ , resulting in 5562 high-resolution grid cells in total. Figure 2 shows the study region and SST data for two days that are considered as extreme with respect to different functionals.

Our methodological framework assumes temporal stationarity, so we first need to appropriately detrend the data. Let  $\{\mathcal{Y}(\mathbf{s}, t)\}_{\mathbf{s} \in S, t \in T}$  denote the sea surface temperature process over the study domain  $S$  (comprising 5562 pixels) for the time period  $T = \{1, \dots, 11315\}$ . We assume that the data can be described, at each location  $\mathbf{s}$ , by the linear regression model

$$\mathcal{Y}(\mathbf{s}, t) = \beta_0(\mathbf{s}) + \beta_1(\mathbf{s})t + \sum_{j=1}^{12} \beta_{1+j}(\mathbf{s})\psi_j(t) + X(\mathbf{s}, t), \quad t = 1, \dots, 11315,$$

where  $\{\psi_j(t)\}$  denote 12 cyclic cubic splines (i.e., one for each month) capturing the seasonal cycle in a flexible way,  $t$  is a linear time trend capturing the effect of climate change,  $\{\beta_j(\mathbf{s})\}_{j=0, \dots, 13}$  are site-specific regression coefficients to estimate from the data, and the process  $\{X(\mathbf{s}, t)\}_{\mathbf{s} \in S, t \in T}$  is a zero-mean stochastic residual component, assumed to be stationary in time, but not necessarily in space. We estimate regression parameters by least squares at each site separately, pooling information from neighboring grid cells within a disk of radius 30 km. This local inference approach gives high flexibility to estimate complex spatiotemporal trends, while the pooling of information allows one to smooth estimates spatially and borrow strength across nearby sites to reduce the overall uncertainty.

We then apply our new spatiotemporal pattern estimation approach to  $\{\hat{x}(\mathbf{s}, t)\}_{\mathbf{s} \in S, t \in T}$ , pseudo-observations from the residual component  $\{X(\mathbf{s}, t)\}_{\mathbf{s} \in S, t \in T}$  that are obtained by subtracting the estimated SST mean from the realizations  $\{y(\mathbf{s}, t)\}_{\mathbf{s} \in S, t \in T}$  of  $\{\mathcal{Y}(\mathbf{s}, t)\}_{\mathbf{s} \in S, t \in T}$  (after replacing the  $\beta_j(\mathbf{s})$  coefficients with their least square estimates). Note that the resid-

uals  $\hat{x}(\mathbf{s}, t)$  can be interpreted as the SST anomaly, and extremes from  $\hat{x}(\mathbf{s}, t)$  represent abnormal sea temperature excesses with respect to the usual SST distribution at the specific space-time point  $(\mathbf{s}, t)$ . We also considered rescaling  $\hat{x}(\mathbf{s}, t)$  by its standard deviation (which varies moderately over space and time), but this makes the interpretation of the results more difficult and we thus proceed with  $\hat{x}(\mathbf{s}, t)$  (detrended, but not rescaled).

Remind that, even though  $X$  may not be regularly varying itself because SST data often have a bounded upper tail rather than a heavy tail, our methodology can still be applied for a variety of functionals, as long as some marginal transformation of  $X$  is regularly varying. This condition is equivalent to  $X$  being in the max-domain of attraction of some max-stable process, as explained throughout Sections 3 and 4. In some applications, this may be considered quite restrictive as it implies asymptotic dependence. However, extremes in sea surface processes are usually strongly spatially dependent, even at large distances, and they also tend to persist over consecutive days, if not weeks. Therefore, asymptotic dependence is a natural assumption here, and statistical models with this property have in fact already been applied to this dataset [see, e.g., Hazra and Huser, 2021, Huser et al., 2022].

We here choose to consider two different risk functionals, namely

- the spatial maximum,  $r_{\max}(f) = \max_{\mathbf{s} \in S} f(\mathbf{s})$ , and
- the spatial median,  $r_{\text{med}}(f) = \text{median}_{\mathbf{s} \in S} f(\mathbf{s})$ .

At first, we estimate the length,  $\ell$ , of extremal clusters defined as consecutive spatial fields  $f_0, \dots, f_{\ell-1}$ , such that  $r_{\max}(f_{-1}) \leq u, r_{\max}(f_i) > u, r_{\max}(f_\ell) \leq u$ , for all  $i = 0, \dots, \ell - 1$ , or  $r_{\text{med}}(f_{-1}) \leq u, r_{\text{med}}(f_i) > u, r_{\text{med}}(f_\ell) \leq u$ , for all  $i = 0, \dots, \ell - 1$ , where  $u$  is a high threshold, here taken as the empirical 95% quantile of the data  $\{\hat{x}(\mathbf{s}, t)\}_{\mathbf{s} \in S, t \in T}$ . Figure 5 reports the estimated cluster size distribution. The results for the risk functional defined as the spatial maximum and those for the spatial median are very similar to each other: about 35%–40%

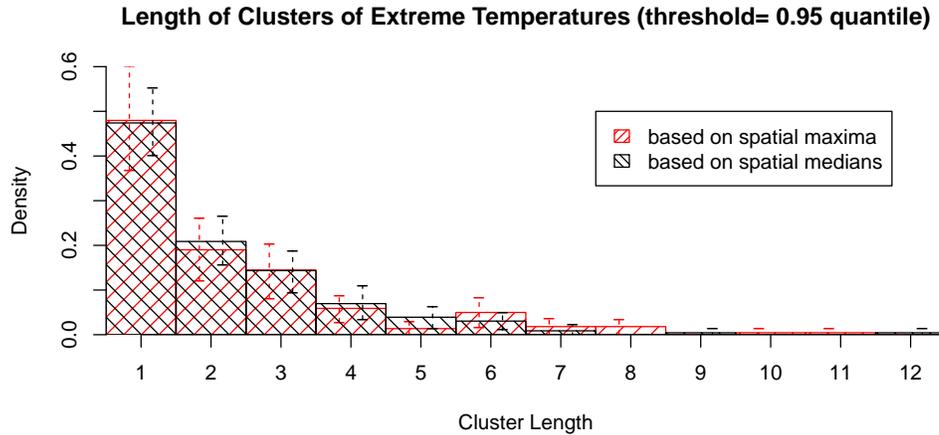


Figure 5: Cluster size distribution for extreme SST defined as high spatial maxima (red) and spatial medians (black) exceeding the 95% empirical quantile. Vertical dashed segments display 95% confidence intervals, obtained via our proposed multiplier block bootstrap.

of clusters comprise individual events, 20%–25% are of size two, 15%–20% of size three, and they are very rarely longer than a week (7 consecutive days).

We then zoom into the clusters themselves and estimate ordinal patterns defined in terms of the respective risk functionals, giving an indication of how the intensity of the extreme event evolves through time. Figure 6 displays, for each risk functional, the distribution of these ordinal patterns for the beginning of clusters of size at least two, and at least three. While we could restrict ourselves to clusters of size *exactly* two or three, the resulting sample size would be too low for accurate estimation. Again, there is strong similarity between results for the spatial maximum and spatial median. The ordinal pattern (1, 2) appears about 60%–70% of the time, indicating that the value of the risk functional at the start of a cluster of size at least two tends to increase. In other words, spatial extreme events that persist over consecutive days tend to strengthen at the beginning, which makes sense from a physical perspective. A similar story holds for clusters of size at least three, with the increasing ordinal pattern (1, 2, 3) occurring more than 30% of the time. Interestingly, the ordinal patterns (2, 1, 3) and (3, 1, 2), which describe a “V-shape” (i.e., with a decrease

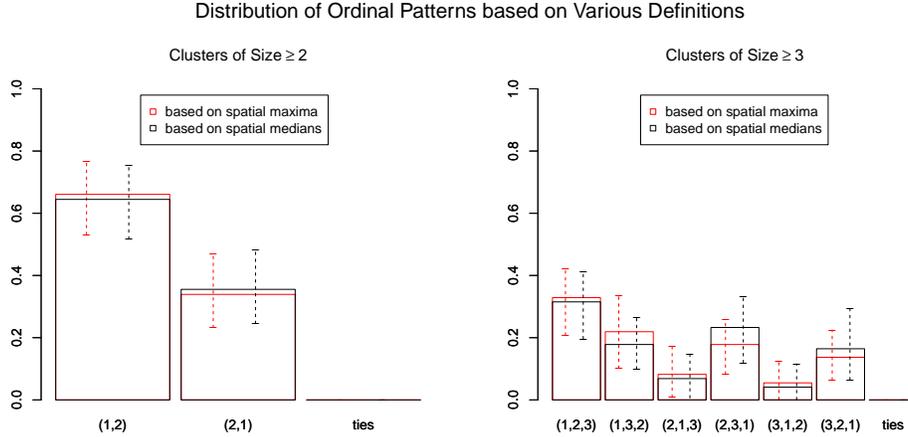


Figure 6: Bar charts displaying the distribution of ordinal patterns with respect to the intensity of extreme SST defined through the spatial maximum (red) and spatial median (black), for the beginning of clusters of size at least 2 (left) and 3 (right). Vertical dashed segments display 95% confidence intervals, obtained via our proposed multiplier block bootstrap.

followed by an increase) are the least likely. Conversely, the patterns (1, 3, 2) and (2, 3, 1) describing an “inverse V-shape” characteristic of relatively short clusters are the most likely combined. While these results make sense, our methodology is helpful to precisely quantify the frequency of these complex within-cluster behaviors.

To investigate further within-cluster characteristics, we then estimate the distribution of patterns for the relative area affected by extremes, i.e., the number of grid cells within a spatial field exceeding the chosen threshold divided by the total number of grid cells. We here only consider clusters defined by the spatial maximum. Note that the estimated area affected by extremes cannot be zero, given that we consider clusters of exceedances (i.e., the maximum exceeds the threshold). Figure 7 reports the results. The results are similar to those discussed above, with the increasing pattern (1, 2) being the most frequent one. This indicates that the spatial extent of extreme events also tends to increase at the beginning of a cluster of size at least two. Similar results hold for clusters of size at least three.

Finally, to investigate the movement of high sea temperatures across the study region

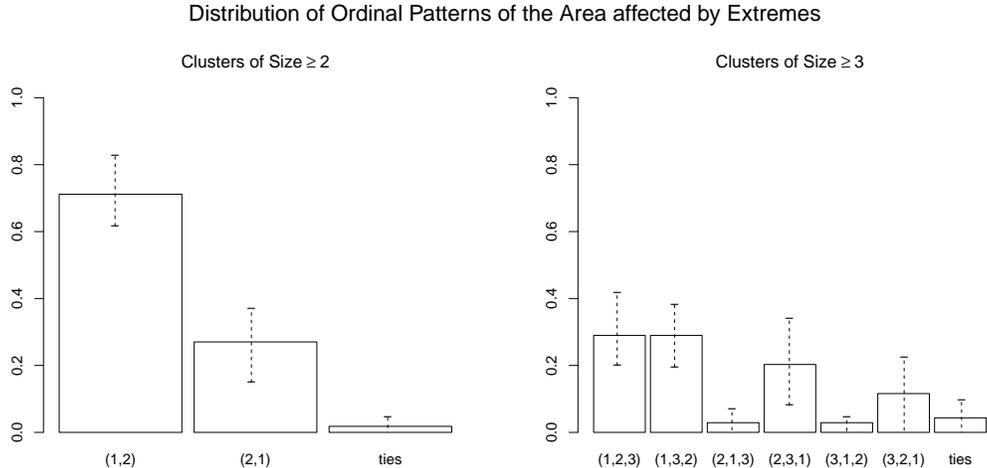


Figure 7: Bar charts displaying the distribution of ordinal patterns with respect to the relative area affected by extreme SST, based on exceedances of the spatial maximum for the beginning of clusters of size at least 2 (left) and 3 (right). Vertical dashed segments display 95% confidence intervals, obtained via our proposed multiplier block bootstrap.

during a spatial extreme episode, we estimate the distribution of ordinal patterns for the latitude and longitude (Figure 8) of the centroid of the area affected by extremes (as defined in Section 4.4). The results indicate that when a spatiotemporal extreme event occurs, the area of highest SST intensity tends to move from South to North, and from East to West, most of the time. These results seem to be broadly aligned with the southern Red Sea circulation literature and might reflect the formation of eddies [Raitsos et al., 2013, Zhan et al., 2014], though further in-depth analyses would be required to confirm this. In future research it would be interesting to exploit our methodology to study these hydrodynamic properties in more detail using alternative specifically customized functionals.

## 8 Conclusion

In order to investigate the probabilistic behavior of temporal clusters formed by spatiotemporal extreme events, and to understand the complex within-cluster behavior of the process under study with respect to various spatial summaries of interest, we have developed a novel non-parametric method that allows accurately estimating the limiting distributions

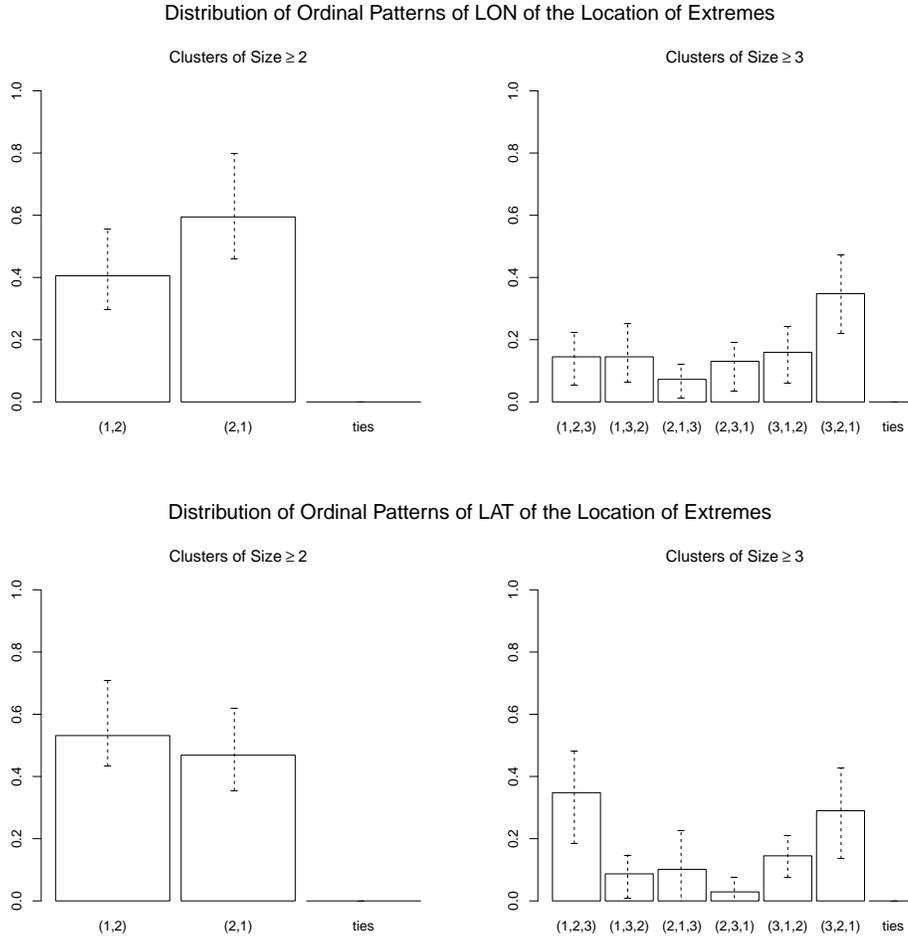


Figure 8: Bar charts displaying the distribution of ordinal patterns with respect to the longitude (top) and latitude (bottom) of the centroid of the region of the most extreme temperatures exceeding the overall 95% quantile for the beginning of clusters of extreme SST of size at least 2 (left) and 3 (right). Vertical dashed segments display 95% confidence intervals, obtained via our proposed multiplier block bootstrap.

of such quantities in a flexible and computationally-efficient way.

Our proposed methodology relies on the theoretical framework of functional regular variation for reliable joint tail extrapolation beyond the range of the data, and it is based on the notion of risk functionals to account for various definitions of extreme events. While our work has mostly focused on describing the within-cluster behavior of extreme events (i.e., their temporal evolution) in terms of the ordinal patterns of various functional characteristics, our theoretical results apply more generally. The characteristics that our framework can track

over time are very diverse, including the magnitude of the extreme event (characterized by the risk functional itself), spatial risk measures (e.g., the size of the affected area), and the location of the extreme event (characterized by various types of location measures). Our framework is, however, not limited to these characteristics. In future research, it would be interesting to extend our method to estimate the within-cluster behavior of an extreme event with respect to *spatio-temporal* functionals of the process capturing, for example, its local rotational circulation properties. This could prove helpful for identifying the fingerprints of well-known physical phenomena, such as circular currents of water (i.e., eddies) in data analyses similar to our sea surface temperature application.

In contrast to state-of-the-art parametric extreme-value models, the great benefit of our non-parametric approach is that it is not subject to stringent assumptions (e.g., spatial stationarity and isotropy, autoregressive temporal dependence structure, etc.) and can thus be used to flexibly estimate complex spatio-temporal characteristics in a wide range of real data applications. Moreover, since our proposed estimators have a very simple form, they can be computed efficiently, even in ultra-high dimensions. For example, in our Red Sea surface temperature application, the full dataset comprises about 63 million dependent spatio-temporal variables, and our results can be obtained in just a few seconds. Furthermore, our proposed approach is not only feasible in such high dimensions, but it also strongly benefits from massive datasets, as our non-parametric estimators then become more accurate.

Although regular variation is a key assumption here, our results still hold under monotone marginal transformations of the process under study (e.g., when it is expressed on a different marginal scale) for a wide range of functionals (i.e., specifically, those that are invariant under monotone transformations). This implies that our methodology is not strictly restricted to marginally heavy-tailed processes, but it can also be applied more generally to spatio-temporal processes exhibiting asymptotic dependence with light or bounded marginal tails.

The asymptotic dependence assumption, however, is a crucial condition that we cannot easily bypass. In our sea surface temperature data application, this does not seem to be a limitation, but empirical evidence has shown that some other environmental processes tend to support asymptotic independence instead. In future research, it would therefore be interesting to extend our methodology to the framework of hidden regular variation, in order to accommodate both asymptotic dependence and independence in a flexible way.

## Supplementary Materials

The supplementary materials include the proofs of Propositions 3.1, 4.2 and 4.4.

## Funding

MO gratefully acknowledges partial support by the project “Climate Change and Extreme Events - ClimXtreme Module B - Statistics (subproject B3.1)” funded by the German Federal Ministry of Education and Research (BMBF) with the grant number 01LP1902I. RH was partially supported by funding from King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-CRG2020-4394.

## References

- S. Aulbach, M. Falk, M. Hofmann, and M. Zott. Max-stable processes and the functional  $D$ -norm revisited. *Extremes*, 18(2):191–212, 2015.
- C. Bandt. Order patterns, their variation and change points in financial time series and Brownian motion. *Statistical Papers*, 61(4):1565–1588, 2020.
- C. Bandt and B. Pompe. Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.*, 88(17):174102, 2002.
- B. Basrak and J. Segers. Regularly varying multivariate time series. *Stochastic Process. Appl.*, 119(4):1055–1080, 2009.
- M. L. Berumen, A. S. Hoey, W. H. Bass, J. Bouwmeester, D. Catania, J. E. Cochran, M. T. Khalil, S. Miyake, M. Mughal, J. L. Spaet, et al. The status of coral reef ecology research in the Red Sea. *Coral Reefs*, 32(3):737–748, 2013.

- B. M. Brown and S. I. Resnick. Extreme values of independent stochastic processes. *J. Appl. Probab.*, 14(4):732–739, 1977.
- D. Castro-Camilo and R. Huser. Local likelihood estimation of complex tail dependence structures, applied to U.S. precipitation extremes. *J. Americ. Stat. Assoc.*, 115:1037–1054, 2020.
- S. Castruccio, R. Huser, and M. G. Genton. High-order composite likelihood inference for max-stable distributions and processes. *J Comput Graph Stat*, 25:1212–1229, 2016.
- R. A. Davis and T. Mikosch. Extreme value theory for space–time processes with heavy-tailed distributions. *Stochastic Process. Appl.*, 118(4):560–584, 2008.
- R. A. Davis and T. Mikosch. The extremogram: A correlogram for extreme events. *Bernoulli*, 15(4):977–1009, 2009.
- R. A. Davis, C. Klüppelberg, and C. Steinkohl. Max-stable processes for modeling extremes observed in space and time. *J. Korean Stat. Soc.*, 42(3):399–414, 2013.
- A. C. Davison and R. Huser. Statistics of extremes. *Annu. Rev. Stat. Appl.*, 2:203–235, 2015.
- A. C. Davison, S. Padoan, and M. Ribatet. Statistical modelling of spatial extremes (with Discussion). *Stat. Sci.*, 27(2):161–186, 2012.
- A. C. Davison, R. Huser, and E. Thibaud. Spatial extremes. In A. E. Gelfand, M. Fuentes, J. A. Hoeting, and R. L. Smith, editors, *Handbook of Environmental and Ecological Statistics*, pages 711–744. CRC Press, 2019.
- R. de Fondeville and A. C. Davison. High-dimensional peaks-over-threshold inference. *Biometrika*, 105(3):575–592, 2018.
- R. de Fondeville and A. C. Davison. Functional peaks-over-threshold analysis. *J. Royal Stat. Soc. B*, 84(4):1392–1422, 2022.
- L. de Haan. A spectral representation for max-stable processes. *Ann. Probab.*, 12(4):1194–1204, 1984.
- C. Dombry and M. Ribatet. Functional regular variations, Pareto processes and peaks over threshold. *Stat. Interface*, 8(1):9–17, 2015.
- C. Dombry, S. Engelke, and M. Oesting. Exact simulation of max-stable processes. *Biometrika*, 103(2):303–317, 2016.
- C. Dombry, E. Hashorva, and P. Soulier. Tail measure and tail spectral process of regularly varying time series. *Ann. Appl. Probab.*, 28(6):3884–3921, 2018.
- C. J. Donlon, M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer. The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sensing of Environment*, 116:140–158, 2012.

- H. Drees. Bootstrapping empirical processes of cluster functionals with application to extremograms. *arXiv preprint arXiv:1511.00420*, 2015.
- S. Engelke, A. Malinowski, Z. Kabluchko, and M. Schlather. Estimation of Huesler–Reiss distributions and Brown–Resnick processes. *J. Royal Stat. Soc. B*, 77(1):239–265, 2015.
- A. Ferreira and L. de Haan. The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20(4):1717–1737, 2014.
- C. A. Ferro and J. Segers. Inference for clusters of extreme values. *J. Royal Stat. Soc. B*, 65(2):545–556, 2003.
- A. Hazra and R. Huser. Estimating high-resolution Red Sea surface temperature hotspots, using a low-rank semiparametric spatial model. *Ann. Appl. Stat.*, 15:572–596, 2021.
- A. Hazra, R. Huser, and D. Bolin. Realistic and fast modeling of spatial extremes over large geographical domains. *arXiv preprint arXiv:2112.10248*, 2021.
- H. Hult and F. Lindskog. Regular variation for measures on metric spaces. *Publications de l’Institut Mathématique*, 80(94):121–140, 2006.
- R. Huser. Editorial: EVA 2019 data competition on spatio-temporal prediction of Red Sea surface temperature extremes. *Extremes*, 24:91–104, 2021.
- R. Huser and A. C. Davison. Space-time modelling of extreme events. *J. Royal Stat. Soc. B*, 76(2):439–461, 2014.
- R. Huser and J. L. Wadsworth. Modeling spatial processes with unknown extremal dependence class. *J. Americ. Stat. Assoc.*, 114:434–444, 2019.
- R. Huser and J. L. Wadsworth. Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews (WIREs): Computational Statistics*, 14(1):e1537, 2022.
- R. Huser, C. Dombry, M. Ribatet, and M. G. Genton. Full likelihood inference for max-stable data. *Stat*, 8:e218, 2019.
- R. Huser, M. L. Stein, and P. Zhong. Vecchia likelihood approximation for accurate and fast inference in intractable spatial extremes models. *arXiv preprint 2203.05626*, 2022.
- Z. Kabluchko, M. Schlather, and L. de Haan. Stationary max-stable fields associated to negative definite functions. *Ann. Probab.*, 37(5):2042–2065, 2009.
- K. Keller, M. Sinn, and J. Emonds. Time series from the ordinal viewpoint. *Stochastics and Dynamics*, 7(02):247–272, 2007.
- E. Koch. Spatial risk measures and applications to max-stable processes. *Extremes*, 20(3):635–670, 2017.
- A. M. Lewandowska, D. G. Boyce, M. Hofmann, B. Matthiessen, U. Sommer, and B. Worm. Effects of sea surface warming on marine plankton. *Ecology letters*, 17(5):614–623, 2014.

- T. Lin and L. de Haan. On convergence toward an extreme value distribution in  $C[0, 1]$ . *Ann. Probab.*, 29(1):467–483, 2001.
- N. M. Markovich. Modeling clusters of extreme values. *Extremes*, 17(1):97–125, 2014.
- M. Oesting and A. Schnurr. Ordinal patterns in clusters of subsequent extremes of regularly varying time series. *Extremes*, 23(4):521–545, 2020.
- M. Oesting and K. Strokorb. A comparative tour through the simulation algorithms for max-stable processes. *Stat. Sci.*, 37(1):42–63, 2022.
- M. Oesting, M. Schlather, and P. Friederichs. Statistical post-processing of forecasts for extremes using bivariate Brown–Resnick processes with an application to wind gusts. *Extremes*, 20(2):309–332, 2017.
- S. A. Padoan, M. Ribatet, and S. A. Sisson. Likelihood-based inference for max-stable processes. *J. Americ. Stat. Assoc.*, 105(489):263–277, 2010.
- D. E. Raitsos, Y. Pradhan, R. J. W. Brewin, G. Stenchikov, and I. Hoteit. Remote sensing the phytoplankton seasonal succession of the Red Sea. *PLoS One*, 8(6):e64909, 2013.
- J. K. Reaser, R. Pomerance, and P. O. Thomas. Coral bleaching and global climate change: scientific findings and policy recommendations. *Biol. Conserv.*, 14(5):1500–1511, 2000.
- M. Sainsbury-Dale, A. Zammit-Mangion, and R. Huser. Fast optimal estimation with intractable models using permutation-invariant neural networks. arXiv preprint arXiv:2208.12942, 2022.
- J. Segers, Y. Zhao, and T. Meinguet. Polar decomposition of regularly varying time series in star-shaped metric spaces. *Extremes*, 20(3):539–566, 2017.
- E. Simpson, T. Opitz, and J. Wadsworth. High-dimensional modeling of spatial and spatio-temporal conditional extremes using INLA and Gaussian Markov random fields. arXiv preprint arXiv:2011.04486, 2022.
- M. Sinn, K. Keller, and B. Chen. Segmentation and classification of time series using ordinal pattern distributions. *Eur. Phys. J. Special Topics*, 222(2):587–598, 2013.
- E. Thibaud and T. Opitz. Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, 102(4):855–870, 2015.
- J. L. Wadsworth and J. A. Tawn. Dependence modelling for spatial extremes. *Biometrika*, 99(2):253–272, 2012.
- J. L. Wadsworth and J. A. Tawn. Higher-dimensional spatial extremes via single-site conditioning. *Spatial Statistics*, 51:100677, 2022.
- P. Zhan, A. C. Subramanian, F. Yao, and I. Hoteit. Eddies in the Red Sea: a statistical and dynamical study. *Journal of Geophysical Research: Oceans*, 119(6):3909–3925, 2014.
- P. Zhong, R. Huser, and T. Opitz. Modeling nonstationary temperature maxima based on extremal dependence changing with event magnitude. *Ann. Appl. Stat.*, 16:272–299, 2022.